Cerebral Cortex
doi:10.1093/cercor/bhm049

Beyond Superior Temporal Cortex: Intersubject Correlations in Narrative Speech Comprehension

Stephen M. Wilson[1,2], Istvan Molnar-Szakacs[1,3,4,5] and Marco Iacoboni[1,3,4,6]

[1]Ahmanson-Lovelace Brain Mapping Center, [2]Neuroscience Interdepartmental Program, [3]Department of Psychiatry and Biobehavioral Sciences, [4]Semel Institute for Neuroscience and Human Behavior, [5]Center for the Biology of Creativity and [6]Brain Research Institute, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA

The role of superior temporal cortex in speech comprehension is well established, but the complete network of regions involved in understanding language in ecologically valid contexts is less clearly understood. In a functional magnetic resonance imaging (fMRI) study, we presented 24 subjects with auditory or audiovisual narratives, and used model-free intersubject correlational analyses to reveal brain areas that were modulated in a consistent way across subjects during the narratives. Conventional comparisons to a resting state were also performed. Both analyses showed the expected recruitment of superior temporal areas, however, the intersubject correlational analyses also revealed an extended network of areas involved in narrative speech comprehension. Two findings stand out in particular. Firstly, many areas in the "default mode" network (typically deactivated relative to rest) were systematically modulated by the time-varying properties of the auditory or audiovisual input. These areas included the anterior cingulate and adjacent medial frontal cortex, and the posterior cingulate and adjacent precuneus. Secondly, extensive bilateral inferior frontal and premotor regions were implicated in auditory as well as audiovisual language comprehension. This extended network of regions may be important for higher-level linguistic processes, and interfaces with extralinguistic cognitive, affective, and interpersonal systems.

Keywords: default mode, narrative speech comprehension

Introduction

The central role of the superior temporal cortex in speech comprehension has been known for over a century, since the pioneering work of Wernicke (1874). Wernicke proposed that the left posterior superior temporal cortex in particular was crucial for receptive language abilities. Recent studies with aphasic patients and especially neuroimaging have greatly expanded our understanding of superior temporal areas involved in language comprehension (Scott et al. 2000; Hickok and Poeppel 2000, 2004; Wise et al. 2001; Scott and Wise 2004), and have revealed that the earliest stages of speech perception are bilateral (Hickok and Poeppel 2000, 2004). Essentially, all studies of auditory language comprehension have revealed bilateral temporal activation for sentences (e.g., Scott et al. 2000; Rodd et al. 2005) and also for narratives, that is, language in an ecologically valid context (Mazoyer et al. 1993; Dehaene et al. 1997; Perani et al. 1998; Giraud et al. 2000; Papathanassiou et al. 2000; Ahmad et al. 2003; Crinion et al. 2003; Tzourio-Mazoyer et al. 2004; Crinion and Price 2005; Skipper et al. 2005; Schmithorst et al. 2006; Alho et al. 2006). In most studies of narrative comprehension, activations have also been reported in the left inferior frontal gyrus (IFG) (e.g., Mazoyer et al. 1993; Skipper et al. 2005).

However, it is clear that in everyday use, language must interface with numerous other systems such as working memory, conceptual knowledge, emotion, and social cognition. So we would expect that many brain regions beyond superior temporal cortex must be involved in narrative speech comprehension. Several neuroimaging studies of narrative comprehension have indeed suggested the involvement of a number of regions beyond classical perisylvian language areas (Fletcher et al. 1995; Xu et al. 2005; for review see Mar 2004). In particular, medial prefrontal cortex has been implicated in a number of studies, and has been interpreted as reflecting social cognitive processes such as "theory of mind" (Fletcher et al. 1995; Gallagher et al. 2000; Ferstl and von Cramon 2002; Xu et al. 2005). Several studies have also shown the involvement of more posterior midline regions in posterior cingulate cortex and/or the precuneus, which may be involved in linking incoming information with prior knowledge, or episodic memory retrieval (Fletcher et al. 1995; Maguire et al. 1999; Ferstl and von Cramon 2002; Xu et al. 2005; Schmithorst et al. 2006). Narrative-related activations have been observed in the posterior superior temporal sulcus (STS) or angular gyrus; these regions are important for a range of cognitive processes including attention, mental imagery and social cognition, all of which would plausibly be components of understanding discourse (Fletcher et al. 1995; Gallagher et al. 2000; Ferstl and von Cramon 2002; Xu et al. 2005; Schmithorst et al. 2006). Another theme is that a shift toward greater right-hemisphere involvement of numerous regions is associated with language in context (e.g., St. George et al. 1999; Robertson et al. 2000; Xu et al. 2005).

Many of the studies that have succeeded in identifying extraperisylvian regions involved in narrative comprehension have employed written materials (Fletcher et al. 1995; St. George et al. 1999; Gallagher et al. 2000; Robertson et al. 2000; Xu et al. 2005) or have relied on subtle manipulations of the extent to which sentences cohere with one another (Ferstl and von Cramon 2002) or with prior context (Maguire et al. 1999). However, in marked contrast to these findings, studies of auditory narrative comprehension, where listening to narratives has been contrasted with resting blocks or acoustic control conditions, have not consistently identified any regions besides the superior temporal cortex bilaterally and the left IFG (Mazoyer et al. 1993; Dehaene et al. 1997; Perani et al. 1998; Giraud et al. 2000; Papathanassiou et al. 2000; Ahmad et al. 2003; Crinion et al. 2003; Tzourio-Mazoyer et al. 2004; Crinion and Price 2005; Skipper et al. 2005; Schmithorst et al. 2006; Alho et al. 2006). Extraperisylvian regions identified in small subsets of these studies include the right IFG (Dehaene et al. 1997; Tzourio-Mazoyer et al. 2004), the precuneus (Perani et al. 1998; Schmithorst et al. 2006), and regions in the vicinity of the angular gyrus (Perani et al. 1998; Crinion et al. 2003; Schmithorst et al. 2006).

One possibility is that the necessity of comparing speech comprehension with some baseline obscures activity in brain areas involved in higher levels of comprehension, beyond auditory processing. Some studies of speech comprehension have used resting baselines (e.g., Mazoyer et al. 1993; Skipper et al. 2005), whereas many others have used acoustically matched control conditions such as backwards speech (e.g., Dehaene et al. 1997; Crinion et al. 2003), but in either case, higher-level cognitive processes which are difficult to constrain presumably take place during the baseline conditions. Even regions which are neither activated nor deactivated relative to a baseline might nevertheless be involved in speech comprehension, because mean signal could be statistically equivalent even though distinct processes are taking place in each condition.

In particular, a set of brain areas termed the "default mode" network (Raichle et al. 2001) have been observed to be consistently deactivated relative to rest or passive sensory processing when subjects engage in a variety of different tasks; these default mode areas include the anterior cingulate and adjacent medial frontal cortex, the posterior cingulate and adjacent precuneus, and the left and right angular gyri (Shulman et al. 1997; Binder et al. 1999; Gusnard and Raichle 2001; Mazoyer et al. 2001; Raichle et al. 2001; McKiernan et al. 2003, 2006). These areas are thought to be involved in ongoing internal processes at rest, such as semantic processing, and monitoring of internal states and the external environment. Semantic processing is an important aspect of speech comprehension, so some default mode areas may be essential components of a wider language comprehension network (Binder et al. 1999; McKiernan et al. 2003, 2006; Iacoboni et al. 2004). Furthermore, the content of perceived speech can provide information concerning the environment, or influence the listener's internal state directly, so other default mode areas may also interface with areas involved in speech perception.

To circumvent the issues which arise when comparing a condition of interest with a baseline, we presented subjects with naturalistic auditory or audiovisual narratives, and used model-free intersubject correlational analysis (Hasson et al. 2004) to identify cortical areas which are systematically modulated by the linguistic input and the processing it entails. This method of analysis requires no control condition, instead identifying as significant those voxels which tend to respond similarly across subjects over the course of a stimulus that varies in time along dimensions of interest. This implies that neural activity in these voxels must be sensitive to time-varying properties of the stimulus, such as dynamic changes in demands on phonological, syntactic, semantic, or extralinguistic processing. Our results revealed the involvement of numerous regions not typically reported in studies of auditory narrative comprehension, including much of the default mode network, and extensive bilateral inferior frontal and premotor areas. This extended set of regions may be important for higher-level linguistic processes and interfaces with conceptual and affective representations.

Materials and Methods

Participants

A total of 24 native English-speaking participants were scanned with functional magnetic resonance imaging (fMRI). Twelve subjects (3 males, mean age 24.2, range 19–33 years) listened to auditory cartoon narratives, and 12 subjects (6 males, mean age 24.7 years, range 20–31 years) viewed and listened to audiovisual cartoon narratives. All participants gave written informed consent and were compensated for their participation, and the study was approved by the UCLA Institutional Review Board.

Experimental Design

The auditory and audiovisual stimuli consisted of cartoon narrations (McNeill 1992). We showed an actor Looney Tunes cartoons from the video "Carrotblanca" (Fig. 1*a*, Warner Brothers Family Entertainment) and she was videotaped while recounting the plots of various stories to a listener behind the camera (Fig. 1*b*). The actor's hands and face were visible at all times, so language-related visual stimuli included mouth movements, head movements, and numerous beat, iconic and other gestures. The actor, who was not a professional, was given no instructions regarding the storytelling, however, she was chosen because she naturally produced prolific and expressive gestures.

In the fMRI experiment, each subject was scanned during 2 runs. In one run, the narratives "Carrotblanca" (6′32″) and "Hare Do" (6′41″) were presented, and in the other run "Dripalong Daffy" (4′40″), "The Scarlet Pumpernickel" (4′31″) and "Box Office Bunny" (2′57″) were presented. There were 16 s of rest (with blank screen) between narratives, as well as at the start and end of each run. The order of runs, and of narratives within runs, was counterbalanced across subjects. Subjects in the auditory-only and audiovisual groups heard exactly the same soundtracks, so the only difference between the groups was the presence or absence of visual information.

Subjects were instructed simply to watch and/or listen to the narratives, and were told that they would be asked questions about the plots. The soundtracks were presented through scanner-compatible headphones at a volume sufficiently loud that the speech could be readily perceived over the scanner noise. The sound volume was set individually for each subject to a comfortable level during preliminary scans. Subjects in the auditory-only condition in particular reported that it was necessary to concentrate and pay attention in order to follow the plots of the narratives over the background scanner noise. When asked questions after the scanning session, subjects in both groups had no difficulty in recalling elements of the plots.

The visual stimuli were presented through custom-made goggles (Resonance Technology Inc., Northridge, CA).

Image Acquisition

Functional images were acquired on a 3-T Siemens Allegra scanner at the Ahmanson-Lovelace Brain Mapping Center at UCLA. There were 2 functional runs (repetition time [TR] = 2000 ms; echo time [TE] = 25 ms; flip angle = 90°; 36 axial slices with interleaved acquisition; 3 × 3 × 4 mm resolution; field of view = 192 × 192 × 144 mm). The number of volumes acquired was 421 for the 2 longer narratives, or 397 for the 3 shorter narratives. In addition, 2 volumes were acquired and discarded to allow for magnetization to reach steady state.

For registration purposes, high-resolution T2-weighted images coplanar with the functional images were acquired (TR = 5000 ms; TE = 33 ms; flip angle = 90°; 36 axial slices; 1.5 × 1.5 × 4 mm resolution; field of view = 192 × 192 × 144 mm).

Image Processing

The fMRI data were preprocessed using tools from FMRIB Software Library (Smith et al., 2004): after skull stripping and motion correction, the data were smoothed with a Gaussian kernel (8mm FWHM) and mean signal intensity was normalized across subjects.

Functional images were aligned using FMRIB's Linear Image Registration Tool to high-resolution coplanar images via an affine transformation with 6 degrees of freedom. High-resolution coplanar images were then aligned to the standard Montreal Neurological Institute (MNI) average of 152 brains using an affine transformation with 12 degrees of freedom.

Standard Analysis

A standard subtraction analysis comparing auditory or audiovisual language comprehension with rest was performed with the FMRISTAT toolbox (Worsley et al. 2002) in MATLAB (Mathworks, Natick, MA). A general linear model was fit to the data from each voxel in each subject,

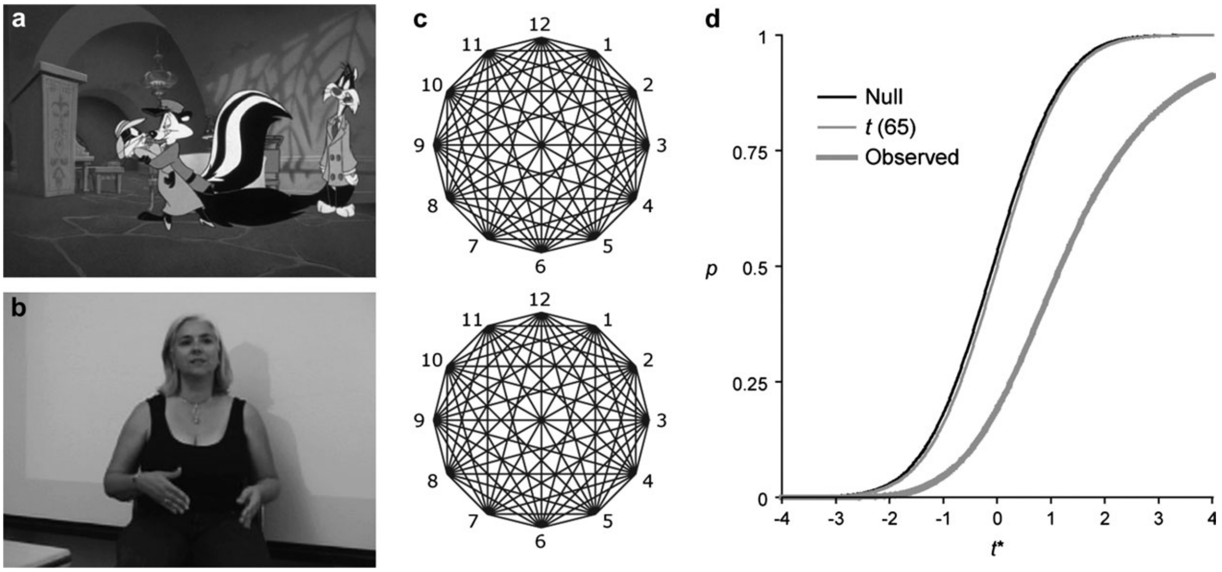

Figure 1. Materials and methods. (*a*) Frame from the movie "Carrotblanca." (*b*) Frame from stimulus video of the actor retelling the narrative. (*c*) Each group comprised 12 subjects, and 66 pairwise correlational maps were created for each group by correlating voxel timecourses for each pair of subjects. (*d*) Distribution of voxel values under null hypothesis (randomly offset time series), $t(65)$, and the observed distribution. Under the null hypothesis, the distribution of voxel values was similar to $t(65)$.

in functional space. The boxcar design matrix was convolved with a hemodynamic response function modeled as a difference of 2 gamma functions. Temporal drift was removed by adding a cubic spline in the frame times to the design matrix (one covariate per 2 min of scan time), and spatial drift was removed by adding a covariate in the whole volume average. Six motion parameters (3 each for translation and rotation) were also included as confounds of no interest. Autocorrelation parameters were estimated at each voxel and used to whiten the data and design matrix. The 2 runs within each subject were combined using a fixed effects model, then the resulting statistical images were registered to MNI space by concatenating the transformation matrices derived above.

Group analysis was performed for each of the 2 groups (auditory only, and audiovisual) with FMRISTAT, using a mixed effects linear model (Worsley et al. 2002). Standard deviations from individual subject analyses were passed up to the group level. Variance ratio images were not smoothed (i.e., a conventional group analysis was performed). The resulting t statistic images were thresholded at $t > 3.106$ (df = 11, $P < 0.005$ uncorrected) at the voxel level, with a minimum cluster size then applied so that only clusters significant at $P < 0.05$ (corrected) according to Gaussian Random Field (GRF) theory were reported.

The 2 groups were compared with one another using a mixed effects linear model implemented with FMRISTAT. In this case, t statistic images were thresholded at $t > 2.819$ (df = 22, $P < 0.005$ uncorrected), before being corrected based on GRF theory as above.

Statistical parameter maps were displayed as overlays on a high-resolution single subject T1 image (Holmes et al. 1998) using Analysis of Functional Neuroimages (Cox 1996). In the tables of regions showing significant signal increases or decreases, anatomical labels were determined manually by inspecting significant regions in relation to the anatomical data averaged across the subjects, with reference to an atlas of neuroanatomy (Duvernoy 1999). In cases of large activated areas spanning more than one region, prominent local maxima were identified and tabulated separately.

Supplementary analyses were performed including further continuously varying explanatory variables in addition to the "boxcar" variable which modeled the narratives, in order to model some of the internal structure of the narrative blocks. In the auditory-only condition, the root mean square (RMS) energy of the speech signal was included, and for the audiovisual condition, this auditory variable was included along with 2 additional variables quantifying the speed of motion of the actor's left and right hands. Each of these variables varied continuously with bins of 100 ms. The RMS energy was determined using a custom MATLAB script, and the actor's hand positions were tracked manually on the videos and the difference between positions at each 100-ms interval was calculated using another MATLAB script. Each continuous variable was convolved with the same hemodynamic response function as the boxcar variable.

Intersubject Correlational Analysis

The intersubject correlational analysis was based on the methods described by Hasson et al. (2004). Each subject's preprocessed functional data was transformed to MNI space, and split up according to narrative. Then a model was fit for each narrative consisting of temporal drift terms (a cubic spline in the frame times, one covariate per 2 min of scan time), 6 motion parameters as above, and the whole volume average, none of which were convolved with a hemodynamic response function. Removing the whole volume average is similar to factoring out what is termed the "nonspecific component" by Hasson et al. (2004). The whole volume average was highly correlated across subjects watching the same movies, and removing it reduces estimates of intersubject correlation (Hasson et al. 2004). Furthermore, the first 16 s of each narrative were excluded, so that common responses to the onset of the narrative (following on from rest) could not account for intersubject correlations. Model fitting was performed with FMRISTAT, and the residuals from this analysis were saved and used for the next stage.

Intersubject correlation maps were then constructed for every pair of subjects belonging to the same group (auditory or audiovisual). There were 12 subjects in each group, so there were 66 pairwise maps created for each group (Fig. 1*c*). These maps were created by a custom MATLAB program that computed the correlation coefficient r between residual timecourses obtained above at each voxel. The r statistic is not normally distributed, so it was converted to a normal distribution using the Fisher z transformation: $z = \log((1 + r)/(1 - r))/2$. In practice, this correction makes little difference for relatively small values of r such as were obtained in this study.

Group analyses were performed to discover at which voxels the intersubject correlations were significantly greater than zero. Note that under the null hypothesis, the expected value of r, and hence of z, is 0, because correlations would be positive or negative at random if the voxel in question is insensitive to the stimulus.

However, we were concerned that for each comparison we have 66 z scores, but only 12 subjects. To discover the distribution of the t statistic in this case, a null data set was created by shifting the data in time, such that time series were no longer aligned across subjects. The algorithm was run as above, except that at each voxel, the 2 time series being compared were both offset by a random number of volumes. For instance, supposing that a narrative was 50 volumes long, and the randomly chosen first volume was 10, then the volumes were rearranged in the order (10, 11, 12, . . ., 49, 50, 1, 2, 3, . . ., 8, 9). The 2 time series being compared were offset from one another by at least 5 volumes. Note that the discontinuity created by wrapping the data around does not significantly distinguish the null data from the real data, because temporal autocorrelation was very low in the residual data sets ($\Phi < 0.03$ in most voxels). This was confirmed based on simulations with randomly generated data based on autoregressive models with various parameters.

The null data set was analyzed with FMRISTAT to derive a t statistical parameter map, and we examined the distribution of the t statistic. We found that it was distributed approximately as t (65) (Fig. 1*d*). In particular, to threshold a t (65) map at voxelwise $P < 0.005$ requires a threshold of $t > 2.654$. The proportion of observations with $t > 2.654$ in the null data set was 0.0039. Finally, note that the observed distribution of the unshifted real data, also depicted in Figure 1*d*, is very different: many voxels were significantly correlated across subjects.

In sum, it appears that a t statistic generated based on the 66 pairwise images is distributed as approximately t (65) under the null hypothesis, and can be treated as such for the purpose of thresholding. Group analyses of the intersubject correlational maps were therefore performed as above, except t statistic images were thresholded at $t > 2.654$ (df = 65, $P < 0.005$ uncorrected) at the voxel level for each group, and at $t > 2.614$ (df = 130, $P < 0.005$ uncorrected) for the between-group comparison, then a minimum cluster size based on GRF theory was applied. Statistical parameter maps were displayed and tables created as described above.

Results

The group data for auditory-only speech comprehension are shown in Figure 2*a* and Table 1. The standard subtraction analysis (green outlines) revealed signal increases in bilateral superior temporal cortex, consistent with numerous previous studies of narrative comprehension (e.g., Mazoyer et al. 1993), as well as a speech motor region in the left precentral gyrus and central sulcus (Wilson et al. 2004). This analysis also revealed an extensive network of regions that were deactivated relative to rest (blue outlines). These included the anterior cingulate gyrus, posterior cingulate gyrus and precuneus, and bilateral angular gyri. These "default mode" areas have been observed in many previous studies contrasting a variety of tasks with resting or passive sensory baselines (Shulman et al. 1997; see Gusnard and Raichle 2001, for review).

The intersubject correlational analysis (red–yellow–white color scale) also demonstrated robust intersubject correlations in bilateral superior temporal cortex, paralleling the results of the standard analysis. However, numerous additional regions showed reliable intersubject correlations. These included several midline areas: the anterior cingulate gyrus, medial superior frontal gyrus, posterior cingulate, and precuneus, which were mostly deactivated relative to rest in the standard analysis. The intercorrelated regions in superior temporal cortex extended much more posteriorly and dorsally into the angular gyri in both hemispheres. There were extensive bilateral inferior frontal regions that were intercorrelated among subjects, extending into premotor cortex in the precentral gyrus.

For the subjects in the audiovisual speech comprehension group, the results were similar in many respects (Fig. 2*b*, Table 2). The most prominent differences were that activations, as well as reliable intersubject correlations, were observed in early visual areas and visual motion areas, reflecting the fact that the stimuli also involved the visual modality. Signal decreases, though only modest intersubject correlations, occurred in anterior occipital regions, where the peripheral visual field (which was not stimulated) is represented (Engel et al. 1994). Similar signal decreases have been shown to most likely reflect reduced neural activity in nonstimulated visual areas, perhaps a form of surround suppression (Shmuel et al. 2002).

As in the auditory-only condition, sizeable bilateral inferior frontal regions extending into premotor areas were intercorrelated across subjects. In this case, bilateral inferior frontal activity was also found relative to rest in the standard analysis, albeit considerably more circumscribed.

The audiovisual and auditory-only groups were then directly compared (Fig. 3, Tables 3 and 4). In the standard analysis, the only regions showing greater signal change in the audiovisual condition relative to the auditory condition were early visual and visual motion areas (Fig. 3*a*). The intersubject correlational analysis also showed significantly greater correlations across subjects in these areas, along with one additional region: the right posterior STS, previously implicated in perception of biological motion (Allison et al. 2000; Pelphrey et al. 2005).

Although in the standard analysis bilateral IFG activations were observed only for the audiovisual group, this difference between groups did not prove to be significant. No frontal regions were significantly more correlated among audiovisual subjects, but there were such areas that did not reach the minimum cluster size; peak coordinates were (−56, 16, 20; t = 3.0) in the left dorsal pars opercularis, and (42, 12, 24; t = 3.7) in the right inferior frontal junction.

The reverse comparison—auditory-only relative to audiovisual—is reported in Figure 3*b* and Table 4. The standard analysis showed greater activity relative to rest in the auditory group in bilateral primary auditory cortex in the transverse temporal gyri (Rademacher et al. 2001). The intersubject correlational analysis did not show reliable correlations across groups in the transverse temporal gyri, however, reliable differences in intersubject correlations were observed more ventrally, centered in the anterior STS, in both hemispheres. These STS regions extended as far anteriorly as the temporal pole; clusters extended from $y = -42$ to $y = 32$ on the left, and from $y = -36$ to $y = 24$ on the right. A number of premotor and prefrontal areas were also more closely correlated across auditory-only than audiovisual subjects: the left ventral precentral gyrus, left orbital gyrus, left inferior frontal sulcus/middle frontal gyrus and left anterior superior frontal gyrus, the right inferior frontal sulcus/middle frontal gyrus, and the right anterior superior frontal gyrus.

Because the standard analysis employed only a simple "boxcar" variable to model each narrative, we carried out further analyses including the RMS energy of the speech signal in both the auditory-only and the audiovisual conditions, and 2 additional variables quantifying the speed of motion of the actor's left and right hands for the audiovisual condition. The regions activated in these analyses are shown in Supplementary Tables 1 and 2. The sets of regions activated by the boxcar regressors in these fuller models were very similar to the analyses reported above where the boxcar regressors were the only explanatory variables in the models. The RMS energy of the speech signal was positively correlated with the transverse temporal gyri bilaterally in each group, reflecting activation of primary auditory cortex. The hand motion regressors were correlated with ipsilateral early visual areas (because, for instance, the actor's left hand appears in the subject's right visual field, which projects to left visual cortex), as well as bilateral visual motion areas, in some cases extending into the STS. Almost all voxels activated by these

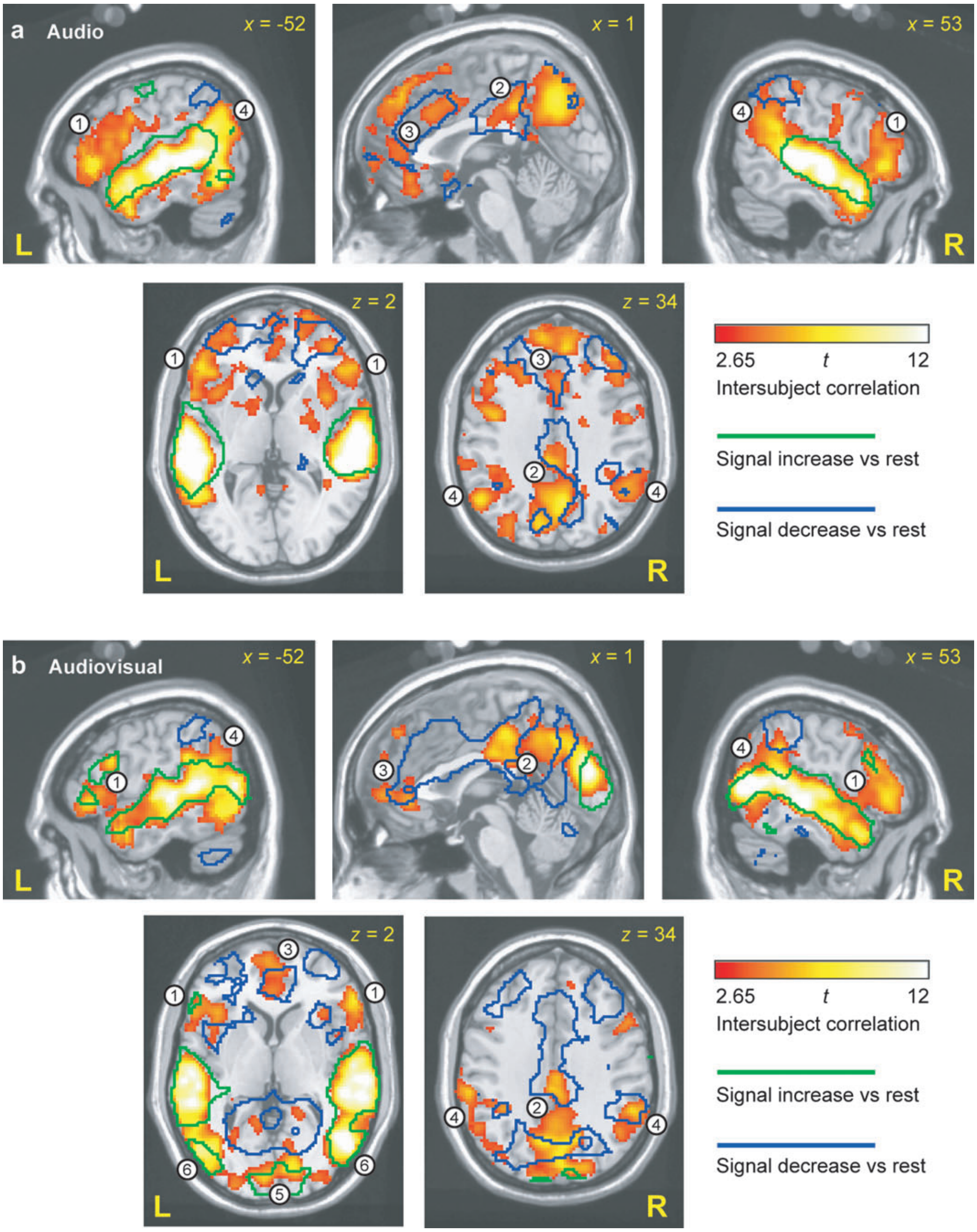

Figure 2. (*a*) Auditory speech comprehension. Five slices are shown with MNI coordinates provided in the top right of each slice. Images are displayed in neurological orientation with the left hemisphere on the left. Intersubject correlations are shown in the red–yellow–white color scale. The results of the standard subtraction analysis are shown as outlines. Activations relative to rest are shown in green, and deactivations relative to rest are shown in blue. Note that regions which are intercorrelated across subjects include activated regions, deactivated regions, and areas which were not significantly activated or deactivated in the standard analysis. Regions of interest: (1) IFG; (2) posterior cingulate and adjacent precuneus; (3) anterior cingulate and adjacent medial frontal cortex; (4) left and right angular gyri; (5) early visual areas; (6) visual motion areas. (*b*) Audiovisual speech comprehension.

additional regressors were also activated by the boxcar regressors, so although these variables confirmed the roles of various sensory regions, they did not reveal additional areas which may have been missed by the simple boxcar analyses.

Discussion

Both the standard analysis and the intersubject correlational analysis replicated the involvement of bilateral temporal areas in speech comprehension, which has been shown in numerous prior studies (for review see Hickok and Poeppel 2004). However, the intersubject correlational analysis also uncovered an extended network of areas involved in narrative speech comprehension including default mode areas (anterior cingulate and adjacent medial frontal cortex, posterior cingulate and adjacent precuneus), and the bilateral IFG and adjacent premotor areas. Many of these regions have rarely been reported in previous studies of auditory narrative comprehension (e.g., Mazoyer et al. 1993; Skipper et al. 2005), however, similar regions have been identified in studies of written narrative comprehension and in studies manipulating textual coherence (Fletcher et al. 1995; Maguire et al. 1999; St. George et al. 1999;

Table 1

Regions significantly correlated across subjects, or activated or deactivated relative to rest for auditory-only narratives

Area	Peak MNI coordinates (mm)			Extent (mm^3)	Max *t*	Cluster *P*
	x	*y*	*z*			
Intersubject correlational analysis						
Extensive bilateral fronto-tempero-parietal network				391 272	18.9	<0.0001
Left STG/STS/MTG	−62	−24	0		17.7	
Right STG/STS/MTG	48	−38	2		18.9	
Left anterior temporal lobe	−48	10	−30		12.3	
Right anterior temporal lobe	52	12	−28		12.5	
Right angular gyrus	38	−64	50		6.9	
Precuneus	4	−64	60		8.1	
Posterior cingulate	−2	−34	36		6.5	
Ventral anterior cingulate gyrus	0	40	4		3.8	
Ventral anterior cingulate gyrus	4	36	−12		4.7	
Left SFG (medial prefrontal)	−8	50	42		7.1	
Right SFG (medial prefrontal)	8	42	38		7.3	
Left IFG pars orbitalis	−50	28	−10		8.8	
Right IFG pars orbitalis	48	28	−4		9.2	
Left IFG pars triangularis/IFS	−46	32	16		7.6	
Right IFS	40	46	10		6.3	
Left ventral precentral gyrus	−40	−4	28		7.7	
Left precentral sulcus	−44	6	50		3.9	
Right precentral sulcus	46	6	48		5.4	
Left cerebellum	−22	−76	−36	13 104	11.5	<0.0001
Right cerebellum	26	−76	−34	10 536	10.2	<0.0001
Dorsal anterior cingulate gyrus	−10	14	42	8528	5.5	<0.0001
Left caudate/putamen	−26	−6	−14	3840	5.6	0.0093
Right fusiform and parahippocampal gyri	28	−34	−26	3368	5.3	0.018
Signal increases in standard analysis						
Left superior temporal				64 272	23.3	<0.0001
Left STG/STS	−52	−20	4		23.3	
Left anterior temporal lobe	−48	2	−14		9.5	
Left fusiform gyrus	−40	−42	−14		9.6	
Right superior temporal				48 880	14.6	<0.0001
Right STG/STS	50	−12	6		14.6	
Right anterior temporal lobe	50	12	−22		11.3	
Left precentral gyrus/central sulcus	−38	−6	58	3376	5.7	0.015
	−46	−6	50		5.2	
Signal decreases in standard analysis						
Midline structures, prefrontal cortex, and right parietal areas				174 800	13.6	<0.0001
Left precuneus	−8	−76	40		6.9	
Right precuneus	12	−70	40		5.6	
Posterior cingulate gyrus	−2	−32	38		8.6	
Dorsal anterior cingulate gyrus	2	32	26		9.1	
Right angular and supramarginal gyri	48	−46	50		13.6	
Left MFG (prefrontal)	−36	52	4		8.3	
Right MFG (prefrontal)	42	46	10		11.4	
Right MFG (prefrontal)	38	46	26		11.6	
Left cerebellum	−24	−40	−42	12 256	8.2	<0.0001
Left angular gyrus	−44	−54	50	6968	10.0	0.0005

Note: In this and other tables, where midline structures are listed without a hemisphere specified, activations were bilateral and separate peaks could not be distinguished. Abbreviations used in the tables: STG, superior temporal gyrus; MTG, middle temporal gyrus; SFG, superior frontal gyrus; MFG, middle frontal gyrus; IFS, inferior frontal sulcus.

Gallagher et al. 2000; Robertson et al. 2000; Ferstl and von Cramon 2002; Xu et al. 2005). Differences between intersubject correlations in the 2 groups were observed in the right posterior STS, which was more intercorrelated among audiovisual subjects, and the bilateral STS more anteriorly, along with premotor and prefrontal regions, which were more correlated across subjects in the auditory-only group.

Default Mode Network

A consistent set of brain regions are deactivated in multiple different active task conditions in comparison with passive or resting conditions. Regions commonly deactivated include the ventral anterior cingulate gyrus, dorsomedial frontal cortex, posterior cingulate cortex and the precuneus, and the angular gyrus (Shulman et al. 1997; Binder et al. 1999; Mazoyer et al. 2001; Gusnard and Raichle 2001; McKiernan et al. 2003).

In the standard analysis, deactivations relative to rest were observed in all of these regions in the present study (see Fig. 2, Tables 1 and 2). The most widely accepted explanation for these signal changes is that they represent the attenuation of a default mode involving processes such as monitoring of internal and external states, and “stream of consciousness” (Shulman et al. 1997; Binder et al. 1999; Gusnard and Raichle 2001; McKiernan et al. 2003).

A novel finding of the present study is that many of these regions were robustly correlated across subjects, as revealed in the intersubject correlational analysis. Data from the rest condition, as well as transitional volumes between rest and task, did not even enter into this analysis, so these correlations cannot reflect processes related to the resting state per se. Rather, the correlations must reflect modulation of these regions by the time-varying content of the narratives, and the linguistic, conceptual and affective processing which they

Table 2

Regions significantly correlated across subjects, or activated or deactivated relative to rest for audiovisual narratives

Area	Peak MNI coordinates (mm)			Extent (mm^3)	Max *t*	Cluster *P*
	x	*y*	*z*			
Intersubject correlational analysis						
Extensive network encompassing many regions				321 208	18.5	<0.0001
Left STG/STS/MTG	−52	−42	6		14.8	
Right STG/STS/MTG	50	−30	4		15.0	
Left anterior temporal lobe	−50	12	−24		9.2	
Right anterior temporal lobe	52	12	−28		9.3	
Left medial occipital cortex	−4	−90	14		13.2	
Right medial occipital cortex	8	−86	22		15.4	
Left middle temporal (MT)	−48	−72	8		14.4	
Right middle temporal (MT)	50	−68	6		18.5	
Left precuneus	−8	−66	34		7.1	
Right precuneus	8	−70	40		8.0	
Posterior cingulate gyrus	6	−34	40		7.5	
Left IFG pars orbitalis	−50	28	−6		6.7	
Right IFG pars orbitalis	56	32	0		7.1	
Left IFG pars opercularis	−54	14	24		6.8	
Right IFG pars opercularis/IFS	42	12	26		7.3	
Right precentral sulcus	50	4	46		5.6	
Left cerebellum	−22	−72	−36		6.3	
Right cerebellum	20	−76	−34		7.1	
Ventral anterior cingulate gyrus	0	36	−6	9744	5.3	<0.0001
Bilateral SFG				5488		0.0011
Left SFG (anterior prefrontal)	−20	34	44		5.5	
Right SFG (anterior prefrontal)	4	46	44		5.2	
Left precentral sulcus	−42	8	48	1128	4.7	0.02[a]
Signal increases in standard analysis						
Bilateral temporal cortex and occipital visual areas				176 912	21.9	<0.0001
Left STG/STS/MTG	−56	−20	4		17.9	
Right STG/STS/MTG	64	−18	−6		18.4	
Left anterior temporal lobe	−60	6	−12		7.6	
Right anterior temporal lobe	54	4	−16		9.1	
Right inferior temporal and fusiform gyri	48	−50	−22		9.9	
Left medial occipital cortex	−16	−96	20		21.9	
Right medial occipital cortex	14	−92	20		21.3	
Left visual motion area MT	−52	−70	8		12.3	
Right visual motion area MT	52	−68	6		13.2	
Right cerebellum	22	−76	−38		5.0	
Left inferior temporal and fusiform gyri	−46	−50	−18	4928	9.3	0.0027
Left IFG pars orbitalis, triangularis and opercularis	−54	32	0	5232	6.1	0.002
Right IFG pars opercularis	44	14	20	2632	7.4	0.041
Signal decreases in standard analysis						
Midline, bilateral prefrontal and bilateral parietal regions				335 832	13.5	<0.0001
Left lingual gyrus	−28	−58	−6		13.5	
Right lingual gyrus	12	−62	6		12.5	
Precuneus	−6	−76	50		11.8	
Left posterior cingulate gyrus	−6	−24	36		8.5	
Right posterior cingulate gyrus	8	−32	36		10.1	
Dorsal anterior cingulate gyrus	4	8	36		9.7	
Ventral anterior cingulate gyrus	−6	48	−2		6.6	
Left angular gyrus	−42	−50	46		7.1	
Right angular gyrus	44	−54	62		11.3	
Left MFG (anterior prefrontal)	−24	40	28		12.2	
Right MFG (anterior prefrontal)	30	34	26		13.4	
Right inferior temporal gyrus	58	−32	−24	3984	7.8	0.0073
Left cerebellum	−48	−64	−40	5000	7.2	0.0025
Right cerebellum	38	−46	−38	7112	8.1	0.0004

Note: STG, superior temporal gyrus; MTG, middle temporal gyrus; SFG, superior frontal gyrus; MFG, middle frontal gyrus; IFS, inferior frontal sulcus.

[a] This cluster was only significant when treated as an a priori hypothesized location.

entail. This demonstrates that default mode regions are not simply shut off in response to an active task. Instead, the data suggest 2 possible interpretations, which are not mutually exclusive. The first is that the narratives make differential demands as a function of time on the processes subserved by the default mode network. This appears likely given the evidence that semantic processing is one function of default mode areas (Binder et al. 1999; McKiernan et al. 2003). For instance, some parts of the narratives may be more semantically complex than other parts, so regions involved in semantic processing may be more active during the more complex stages of the narratives, consistently across subjects. The second interpretation is that the global level of engagement may vary in the narratives as a function of time, and this may contribute to the intersubject correlations observed in default mode areas. It has been shown that default mode regions are systematically downregulated as a function of task difficulty (Greicius and Menon 2004; McKiernan et al. 2006), so it is plausible that during parts of the narratives that are more engaging, default mode activity is more downregulated, which would result in correlations across subjects to the extent that subjects find the same parts of the narratives more or less engaging.

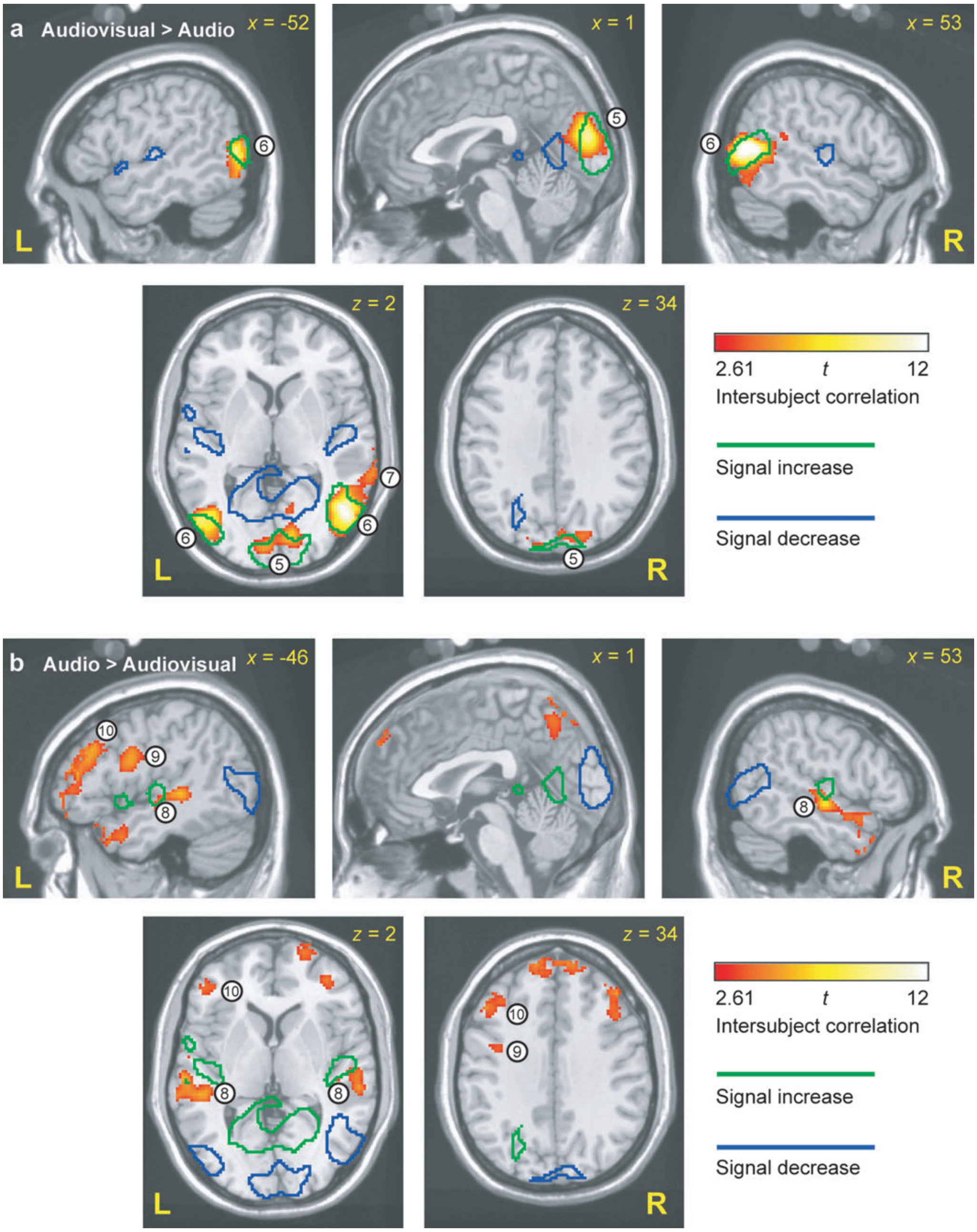

Figure 3. (*a*) Audiovisual speech comprehension relative to auditory speech comprehension. See caption to Figure 2 for explanation of conventions. The red–yellow–white color scale shows areas which were more correlated across subjects for audiovisual speech than for auditory-only speech. Similarly the green outlines show areas that were more activated relative to rest for audiovisual speech than auditory speech, and the blue outlines show areas that were less activated. Regions of interest: (5) early visual areas; (6) visual motion areas; (7) right STS. (*b*) Audio speech comprehension relative to auditory speech comprehension. The red–yellow–white color scale shows areas which were more correlated across subjects for auditory-only speech than for audiovisual speech. Similarly the green outlines show areas that were more activated relative to rest for auditory speech than audiovisual speech, and the blue outlines show areas that were less activated. Note that the blue and green outlines in this figure are simply the opposite of those in panel (*a*), where the reverse contrasts are depicted. Regions of interest: (8) superior temporal auditory areas; (9) left ventral precentral gyrus; (10) left prefrontal regions.

The functions of the various regions which make up the default mode network are not well understood, however, functional interpretations have been proposed for each area. The ventral, rostral section of the anterior cingulate gyrus appears to be involved with affective and emotional processes, whereas dorsal anterior cingulate cortex is more concerned with cognitive and motor functions (Bush et al. 2000). The adjacent dorsomedial prefrontal cortex is thought to be concerned with monitoring one's own internal state, as well as attributing mental states to others (Frith and Frith 1999), or with social processing more generally (Iacoboni et al. 2004). As for the posterior midline regions, Gusnard and Raichle (2001) have proposed that the role of these areas in the default mode network is to represent and monitor the external environment. Activations of posterior midline regions in narrative comprehension studies have been interpreted as reflecting the linking of incoming information with prior knowledge, and episodic memory retrieval (e.g., Xu et al. 2005).

Table 3

Regions which were significantly more correlated across audiovisual subjects than auditory-only subjects, or which were activated for audiovisual narratives relative to audio-only narratives

Area	Peak MNI coordinates (mm)			Extent (mm^3)	Max *t*	Cluster *P*
	x	*y*	*z*			
Intersubject correlational analysis						
Early visual areas and right higher-level visual areas				54 856	16.4	<0.0001
Left medial occipital cortex	−10	−94	20		7.8	
Right medial occipital cortex	8	−86	20		12.4	
Right visual motion area MT	50	−68	8		16.4	
Right posterior STS	70	−38	8		6.8	
Left visual motion area MT	−46	−72	8	13 872	12.9	<0.0001
Signal increases in standard analysis						
Early visual and visual motion areas				65 880	13.7	<0.0001
Left medial occipital cortex	−14	−96	16		11.4	
Right medial occipital cortex	12	−92	20		13.7	
Left visual motion area MT	−48	−82	8		8.1	
Right visual motion area MT	52	−68	6		10.7	
Signal decreases in standard analysis						
See Table 4 signal increases.						

Note: MT, middle temporal.

Table 4

Regions which were significantly more correlated across auditory-only subjects than audiovisual subjects, or which were activated for auditory-only narratives relative to audiovisual narratives

Area	Peak MNI coordinates (mm)			Extent (mm^3)	Max *t*	Cluster *P*
	x	*y*	*z*			
Intersubject correlational analysis						
Left anterior STS	−66	−36	−2	9976	6.6	<0.0001
Right anterior STS	48	14	−40	7960	8.7	<0.0001
Precuneus	−2	−64	50	5576	6.2	0.0009
Bilateral SFG				7168		0.0001
Left SFG (anterior prefrontal)	−6	54	40		6.0	
Right SFG (anterior prefrontal)	18	60	20		4.7	
Left IFS/MFG	−48	40	16	7344	5.3	0.0001
Right IFS/MFG	42	54	16	5360	5.0	0.0012
Left ventral precentral gyrus	−40	−2	26	3896	5.4	0.0089
Left orbital gyrus	−22	34	−12	3304	4.8	0.021
Left cerebellum	−22	−82	−56	4064	5.6	0.007
Signal increases in standard analysis						
Left transverse temporal gyrus	−50	−16	4	7248	5.5	0.0002
Right transverse temporal gyrus	48	−16	8	5640	5.8	0.001
Bilateral lingual gyri				38 776		<0.0001
Left lingual gyrus	−20	−54	2		7.3	
Right lingual gyrus	12	−62	6		9.5	
Signal decreases in standard analysis						
See Table 3 signal increases.						

Note: SFG, superior frontal gyrus; MFG, middle frontal gyrus; IFS, inferior frontal sulcus.

The dorsal part of the angular gyrus bilaterally was deactivated relative to rest in both the auditory and audiovisual groups, consistent with its part in the default mode network. However, unlike the other major default mode regions, significant intersubject correlations were not observed in this part of the angular gyrus. Importantly though, bilateral superior temporal regions showing correlations across subjects extended dorsally and posteriorly to include the posterior STS and the ventral part of the angular gyrus. This contrasted with the standard analysis, where these superior temporal activations did not extend so far back. Thus, there is a discrepancy between the 2 methods, in that the intersubject correlational analysis implies the involvement of posterior superior temporal and inferior parietal regions that are not more active than rest in the standard analysis. The results from the intersubject correlational analysis are more consistent with lesion studies, which have demonstrated that lesions to this region produce conduction aphasia (Green and Howes 1978). In general, this area has been argued to be important for auditory to articulatory mapping in language comprehension and production (Hickok and Poeppel 2000, 2004). We suggest that in the standard analysis the involvement of this region in speech comprehension is obscured, because it lies adjacent to the deactivated dorsal part of the angular gyrus. But the dorsal part of the angular gyrus that was deactivated relative to rest was not correlated across subjects and so appears to be concerned with internal processes that are not systematically modulated by linguistic input.

Previous studies of auditory narrative comprehension have rarely reported deactivations relative to baseline, and default mode regions have usually not been activated relative to baseline; exceptions in a handful of studies include the precuneus (Perani et al. 1998; Schmithorst et al. 2006) and regions in the vicinity of the angular gyrus (Perani et al. 1998; Crinion et al. 2003; Schmithorst et al. 2006).

Our results demonstrating intersubject correlations in default mode regions are at variance with those of Golland et al. (2007), who argued for a partition of cortical areas into an "extrinsic" system concerned with processing of sensory input, which was correlated across multiple presentations of the same time-varying audiovisual stimulus (a movie), and an "intrinsic" system important for internal processes, which was not correlated across multiple presentations of the same movie. The intrinsic system was argued to have much in common with the default mode network. Golland et al. (2007) defined the intrinsic system as voxels correlated with the timecourse of "seed" regions of interest in the inferior parietal cortex (IPC), which was chosen because it was the area which most consistently did not show correlations between repeated presentations of the same movie (similar to the angular gyri in our study). Significant intersubject correlations were not observed in the intrinsic system, which included most default mode areas with the exception of the anterior cingulate gyrus.

We propose 2 possible reasons for this discrepancy with our results. Firstly, Golland et al. (2007) assessed correlations between signal in response to 2 presentations of the same movie to each subject, rather than calculating correlations across subjects. If default mode regions are especially important for higher-level cognitive and affective processes, rather than more basic sensory processes, then it is logical that they respond differently to a movie which had already been seen recently. This might contribute to explaining the lack of correlations observed. In a previous study by the same group, correlated regions potentially in the default mode network were reported in the cingulate gyrus and retrosplenial cortex (Hasson et al. 2004).

A second major difference between our study and Golland et al. (2007) is that we used videos with constant linguistic content, whereas they presented subjects with a segment of a feature movie which contained language only some of the time. It is possible that the default mode regions we observed to be intercorrelated across subjects are especially involved in higher-level linguistic processes in particular, and are not engaged in such a consistent manner across individuals for different kinds of stimuli.

IFG and Premotor Cortex

Intersubject correlational analyses revealed extensive bilateral regions in the IFG and adjacent premotor cortex where there were significant intersubject correlations. This implies that these regions are sensitive to time-varying properties of the input and the computations entailed. The left IFG in particular (i.e., Broca's area) has been demonstrated to be involved in semantic, syntactic and phonological processes in both speech production and comprehension (Bookheimer 2002). Because the information content in each of these domains is constantly varying in the course of a narrative, the intersubject correlations in this region are not surprising. Left frontal activations have been observed in most previous studies of auditory narrative comprehension, though the precise regions reported have generally been much more circumscribed and have varied considerably from study to study. In the standard analysis in the present study, there were actually no significant activations in the IFG in either hemisphere in the auditory-only group. Although small clusters of voxels were observed in the pars triangularis of each hemisphere exceeding the threshold corresponding to $P < 0.005$, their cluster sizes were not close to significance: 800 mm^3 in the left hemisphere ($P = 0.78$), and 336 mm^3 in the right hemisphere ($P = 1.00$).

Activity in the right IFG was also shown to be highly significantly correlated across subjects, to a degree similar to the left IFG. Right IFG involvement has rarely been reported in previous studies of auditory narrative comprehension, with occasional exceptions (Dehaene et al. 1997; Tzourio-Mazoyer et al. 2004). However, right-hemisphere areas, including the IFG, are thought to play a role in a range of linguistic processes including prosody (Ross 1981; Wildgruber et al. 2005) and understanding of higher-level discourse (St. George et al. 1999; Robertson et al. 2000; Xu et al. 2005; see Bookheimer 2002; Jung-Beeman 2005, for review). We propose that the robust correlations across subjects that we observed in the right IFG reflect the sensitivity of the right IFG to modulation of such higher-level processes.

Why are inferior frontal activations in either hemisphere so much less extensive in previous studies of narrative speech comprehension, and in the standard analysis in the present study? High activity at rest or in passive conditions probably cannot account for the failure to observe bilateral IFG activity in narrative comprehension studies, because only parts of the left IFG (and not the right) have been suggested to belong to the default mode network (Shulman et al. 1997; Binder et al. 1999), and even the left IFG has not been identified in all studies (McKiernan et al. 2003; Greicius and Menon 2004). Rather, our results suggest that the left and right IFG do not exhibit a consistent signal increase or decrease during narrative comprehension, but rather they show a consistent signal fluctuation which tracks one or more aspects of the input. Precisely which aspects are tracked cannot be determined from our study, but recent reviews of the literature shed some light on the kinds of processes the left and right IFG might be concerned with (Bookheimer 2002; Jung-Beeman 2005).

The left ventral precentral gyrus (ventral premotor cortex), and bilateral regions spanning the inferior frontal sulcus and middle frontal gyrus, were more correlated across subjects in the auditory-only group than the audiovisual group. Comprehension of the narratives was considerably more difficult in the auditory-only condition, due to the lack of visual phonemic cues and the interference of the scanner noise with the auditory stimuli. This suggests that the differential recruitment of these frontal areas may reflect increased processing difficulty. In particular, we propose that frontal areas may play a role in generating top-down models of hypothesized linguistic structures, which would be assessed with respect to the acoustic input in superior temporal regions. A recent study has argued for a similar role for premotor cortex in low-level phonetic perception (Wilson and Iacoboni 2006). Under this view, increased intersubject correlations in the auditory-only group would reflect common modulations across subjects for parts of the narratives that were more difficult to understand and made increased demands on top-down processes.

Regions Differentially Implicated in Audiovisual Speech Perception

Besides early visual and visual motion areas, there was just one region that showed significantly greater correlations within the audiovisual group compared with the auditory group: the right STS. The STS, particularly in the right hemisphere, has been demonstrated in numerous studies to be important for perception of biological motion (Allison et al. 2000; Pelphrey et al. 2005). Our audiovisual stimuli contained movements of the arms, hands, head, mouth, and eyes. Another context in which the STS is often implicated is crossmodal binding in audiovisual speech perception (Calvert et al. 2000; Macaluso et al. 2004). In a previous study comparing audiovisual narrative comprehension with auditory-only narrative comprehension, Skipper et al. (2005) also reported greater activation of bilateral posterior superior and middle temporal regions for audiovisual speech.

Although there were no frontal regions which responded significantly more strongly to audiovisual narratives, nor that were more intercorrelated across subjects in the audiovisual condition, bilateral posterior inferior frontal areas were activated relative to rest in the standard analysis for the audiovisual group but not for the auditory-only group. Furthermore, the left dorsal pars opercularis and right inferior frontal junction (adjacent to the pars opercularis) showed greater intersubject correlations for the audiovisual subjects which did not reach the minimum cluster size criterion. These findings are consistent with a large body of research that has implicated regions in the IFG in the coding of actions (Rizzolatti and Craighero 2004), the actions in the present study being the speech-related gestures produced by the actor, as well as possibly the head, eye and mouth movements. Our identification of the dorsal pars opercularis in particular is consistent with recent data showing that this is the inferior frontal region most systematically implicated in action observation (Iacoboni et al. 2005; Molnar-Szakacs et al. 2005, 2006).

Superior Temporal Cortex

Both the standard analysis and the intersubject correlational analysis revealed greater involvement of superior temporal regions in the more difficult auditory condition relative to the audiovisual condition. However, the precise regions implicated were not identical across the 2 analyses. The standard analysis showed that there was greater activity in the transverse temporal gyri bilaterally, that is, primary auditory cortex. In contrast, the intersubject analysis did not reveal enhanced correlations between subjects in this area, but rather more ventrally and anterior in the STS, extending as far anteriorly as the temporal pole. It is likely that the more challenging auditory-only condition required increased auditory attention, which is known to increase signal in primary sensory areas (Pugh et al. 1996). However, because the temporal patterns of activity in these areas would simply reflect acoustic properties that are identical in the auditory-only and audiovisual conditions, there was no difference in the extent of intersubject correlations, even though there was more signal change in the auditory condition. On the other hand, activity in the anterior STS regions which showed increased intersubject correlations must reflect not only acoustic information but also linguistic processing, which we suggest would have had a qualitatively different temporal structure in the more heavily taxed auditory-only group. This constitutes evidence in support of a ventral, anterior route for speech perception in superior temporal cortex that has been proposed by several groups (Scott et al. 2000; Scott and Wise 2004; Liebenthal et al. 2005). It is noteworthy though that we observed increased intersubject correlations in the STS bilaterally, supporting the idea that the earliest stages of speech perception are bilateral (Hickok and Poeppel 2000, 2004).

Conclusion

Intersubject correlational analysis proved to be a useful complement to conventional subtraction analysis, as it revealed a wide network of regions involved in auditory or audiovisual narrative comprehension. Several "default mode" areas—ventral and dorsal anterior cingulate and adjacent medial frontal regions, and the posterior cingulate and adjacent precuneus—were modulated in a consistent manner across subjects by the narratives, despite being largely deactivated relative to rest. Extensive bilateral inferior frontal and premotor regions were also highly correlated across subjects. We propose that this network of regions beyond the superior temporal cortex is important for higher-level linguistic processes, and interfaces with extralinguistic cognitive, affective, and interpersonal systems.

Supplementary Material

Supplementary material can be found at: http://www.cercor.oxfordjournals.org/.

Notes

We thank Susan Arick, Lisa Aziz-Zadeh, Susan Duncan, Susan Goldin-Meadow, Amy Hubbard, Jonas Kaplan, and Roy Mukamel for their assistance in the design, implementation and analysis of this study, and 2 anonymous reviewers for their helpful comments. For generous support we thank the Brain Mapping Medical Research Organization, Brain Mapping Support Foundation, Pierson-Lovelace Foundation, The Ahmanson Foundation, William M. and Linda R. Dietel Philanthropic Fund at the Northern Piedmont Community Foundation, Tamkin Foundation, Jennifer Jones-Simon Foundation, Capital Group Companies Charitable Foundation, Robson Family, and Northstar Fund. The project described was supported by grants from the National Science Foundation (REC0107077), National Institute of Mental Health (MH63680), and grant numbers RR12169, RR13642, and RR00865 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH); its contents are solely the responsibility of the authors and do not necessarily represent the official views of NCRR or NIH. *Conflict of Interest*: None declared.

Address correspondence to Stephen M. Wilson, Laboratory for Cognitive Brain Research, Department of Cognitive Sciences, University of California, Irvine, CA 92697, USA. Email: stephenw@ucla.edu.

References

Ahmad Z, Balsamo LM, Sachs BC, Xu B, Gaillard WD. 2003. Auditory comprehension of language in young children: neural networks identified with fMRI. Neurology 60:1598-1605.

Alho K, Vorobyev VA, Medvedev SV, Pakhomov SV, Starchenko MG, Tervaniemi M, Naatanen R. 2006. Selective attention to human voice enhances brain activity bilaterally in the superior temporal sulcus. Brain Res 1075:142-150.

Allison T, Puce A, McCarthy G. 2000. Social perception from visual cues: role of the STS region. Trends Cogn Sci 4:267-278.

Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Rao SM, Cox RW. 1999. Conceptual processing during the conscious resting state. A functional MRI study. J Cogn Neurosci 11:80-95.

Bookheimer S. 2002. Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. Annu Rev Neurosci 25:151-188.

Bush G, Luu P, Posner MI. 2000. Cognitive and emotional influences in anterior cingulate cortex. Trends Cogn Sci 4:215-222.

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

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

Crinion JT, Lambon-Ralph MA, Warburton EA, Howard D, Wise RJ. 2003. Temporal lobe regions engaged during normal speech comprehension. Brain 126:1193–1201.

Crinion J, Price CJ. 2005. Right anterior superior temporal activation predicts auditory sentence comprehension following aphasic stroke. Brain 128:2858–2871.

Dehaene S, Dupoux E, Mehler J, Cohen L, Paulesu E, Perani D, van de Moortele PF, Lehericy S, Le Bihan D. 1997. Anatomical variability in the cortical representation of first and second language. Neuroreport 8:3809–3815.

Duvernoy HM. 1999. The human brain: surface, three-dimensional sectional anatomy with MRI, and blood supply. New York: Springer.

Engel SA, Rumelhart DE, Wandell BA, Lee AT, Glover GH, Chichilnisky EJ, Shadlen MN. 1994. fMRI of human visual cortex. Nature. 369:525.

Ferstl EC, von Cramon DY. 2002. What does the frontomedian cortex contribute to language processing: coherence or theory of mind? NeuroImage 17:1599–1612.

Fletcher PC, Happe F, Frith U, Baker SC, Dolan RJ, Frackowiak RS, Frith CD. 1995. Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. Cognition 57:109–128.

Frith CD, Frith U. 1999. Interacting minds—a biological basis. Science 286:1692–1695.

Gallagher HL, Happe F, Brunswick N, Fletcher PC, Frith U, Frith CD. 2000. Reading the mind in cartoons and stories: an fMRI study of "theory of mind" in verbal and nonverbal tasks. Neuropsychologia 38:11–21.

Giraud AL, Truy E, Frackowiak RS, Gregoire MC, Pujol JF, Collet L. 2000. Differential recruitment of the speech processing system in healthy subjects and rehabilitated cochlear implant patients. Brain 123:1391–1402.

Golland Y, Bentin S, Gelbard H, Benjamini Y, Heller R, Nir Y, Hasson U, Malach R. 2007. Extrinsic and intrinsic systems in the posterior cortex of the human brain revealed during natural sensory stimulation. Cereb Cortex 17:766–777.

Green E, Howes DH. 1978. The nature of conduction aphasia: a study of anatomic and clinical features and of underlying mechanisms. In: Whitaker A, Whitaker HA, editors. Studies in neurolinguistics. San Diego: Academic Press. p. 123–156.

Greicius MD, Menon V. 2004. Default-mode activity during a passive sensory task: uncoupled from deactivation but impacting activation. J Cogn Neurosci 16:1484–1492.

Gusnard DA, Raichle ME. 2001. Searching for a baseline: functional imaging and the resting human brain. Nat Rev Neurosci 2:685–694.

Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R. 2004. Intersubject synchronization of cortical activity during natural vision. Science 303:1634–1640.

Hickok G, Poeppel D. 2000. Towards a functional neuroanatomy of speech perception. Trends Cogn Sci 4:131–138.

Hickok G, Poeppel D. 2004. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition 92:67–99.

Holmes CJ, Hoge R, Collins L, Woods R, Toga AW, Evans AC. 1998. Enhancement of MR images using registration for signal averaging. J Comput Assist Tomogr 22:324–333.

Iacoboni M, Lieberman MD, Knowlton BJ, Molnar-Szakacs I, Moritz M, Throop CJ, Fiske AP. 2004. Watching social interactions produces dorsomedial prefrontal and medial parietal BOLD fMRI signal increases compared to a resting baseline. NeuroImage 21:1167–1173.

Iacoboni M, Molnar-Szakacs I, Gallese V, Buccino G, Mazziotta JC, Rizzolatti G. 2005. Grasping the intentions of others with one's own mirror neuron system. PLoS Biol. 3(3):e79.

Jung-Beeman M. 2005. Bilateral brain processes for comprehending natural language. Trends Cogn Sci 9:512–518.

Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA. 2005. Neural substrates of phonemic perception. Cereb Cortex 15:1621–1631.

Macaluso E, George N, Dolan R, Spence C, Driver J. 2004. Spatial and temporal factors during processing of audiovisual speech: a PET study. NeuroImage 21:725–732.

Maguire EA, Frith CD, Morris RG. 1999. The functional neuroanatomy of comprehension and memory: the importance of prior knowledge. Brain 122:1839–1850.

Mar RA. 2004. The neuropsychology of narrative: story comprehension, story production and their interrelation. Neuropsychologia 42:1414–1434.

Mazoyer BM, Tzourio N, Frak V, Syrota A, Murayama N, Levrier O, Salamon G, Dehaene S, Cohen L, Mehler J. 1993. The cortical representation of speech. J Cogn Neurosci 5:467–479.

Mazoyer B, Zago L, Mellet E, Bricogne S, Etard O, Houde O, Crivello F, Joliot M, Petit L, Tzourio-Mazoyer N. 2001. Cortical networks for working memory and executive functions sustain the conscious resting state in man. Brain Res Bull 54:287–298.

McKiernan KA, D'Angelo BR, Kaufman JN, Binder JR. 2006. Interrupting the "stream of consciousness": an fMRI investigation. NeuroImage 29:1185–1191.

McKiernan KA, Kaufman JN, Kucera-Thompson J, Binder JR. 2003. A parametric manipulation of factors affecting task-induced deactivation: an fMRI study. J Cogn Neurosci 15:394–408.

McNeill D. 1992. Hand and mind. Chicago: University of Chicago Press.

Molnar-Szakacs I, Iacoboni M, Koski L, Mazziotta JC. 2005. Functional segregation within pars opercularis of the inferior frontal gyrus: evidence from fMRI studies of imitation and action observation. Cereb Cortex 15:986–994.

Molnar-Szakacs I, Kaplan JT, Greenfield PM, Iacoboni M. 2006. Observing complex action sequences: the role of the fronto-parietal mirror neuron system. NeuroImage 33:923–935.

Papathanassiou D, Etard O, Mellet E, Zago L, Mazoyer B, Tzourio-Mazoyer N. 2000. A common language network for comprehension and production: a contribution to the definition of language epicenters with PET. NeuroImage 11:347–357.

Pelphrey KA, Morris JP, Michelich CR, Allison T, McCarthy G. 2005. Functional anatomy of biological motion perception in posterior temporal cortex: an FMRI study of eye, mouth and hand movements. Cereb Cortex 15:1866–1876.

Perani D, Paulesu E, Galles NS, Dupoux E, Dehaene S, Bettinardi V, Cappa SF, Fazio F, Mehler J. 1998. The bilingual brain. Proficiency and age of acquisition of the second language. Brain 121:1841–1852.

Pugh KR, Shaywitz BA, Shaywitz SE, Fulbright RK, Byrd D, Skudlarski P, Shankweiler DP, Katz L, Constable RT, Fletcher J, et al. 1996. Auditory selective attention: an fMRI investigation. NeuroImage 4:159–173.

Rademacher J, Morosan P, Schormann T, Schleicher A, Werner C, Freund HJ, Zilles K. 2001. Probabilistic mapping and volume measurement of human primary auditory cortex. NeuroImage 13:669–683.

Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL. 2001. A default mode of brain function. Proc Natl Acad Sci USA 98:676–682.

Rizzolatti G, Craighero L. 2004. The mirror-neuron system. Annu Rev Neurosci 27:169–192.

Robertson DA, Gernsbacher MA, Guidotti SJ, Robertson RR, Irwin W, Mock BJ, Campana ME. 2000. Functional neuroanatomy of the cognitive process of mapping during discourse comprehension. Psychol Sci 11:255–260.

Rodd JM, Davis MH, Johnsrude IS. 2005. The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. Cereb Cortex 15:1261–1269.

Ross ED. 1981. The aprosodias. Functional-anatomic organization of the affective components of language in the right hemisphere. Arch Neurol 38:561–569.

Schmithorst VJ, Holland SK, Plante E. 2006. Cognitive modules utilized for narrative comprehension in children: a functional magnetic resonance imaging study. NeuroImage 29:254–266.

Scott SK, Blank CC, Rosen S, Wise RJ. 2000. Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123:2400–2406.

Scott SK, Wise RJS. 2004. The functional neuroanatomy of prelexical processing in speech perception. Cognition 92:13–45.

Shmuel A, Yacoub E, Pfeuffer J, Van de Moortele PF, Adriany G, Hu X, Ugurbil K. 2002. Sustained negative BOLD, blood flow and oxygen consumption response and its coupling to the positive response in the human brain. Neuron 36:1195–1210.

Shulman GL, Fiez JA, Corbetta M, Buckner RL, Miezin FM, Raichle ME, Petersen SE. 1997. Common blood flow changes across visual tasks: II. Decreases in cerebral cortex. J Cogn Neurosci 9:648–663.

Skipper JI, Nusbaum HC, Small SL. 2005. Listening to talking faces: motor cortical activation during speech perception. NeuroImage 25:76–89.

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, et al. 2004. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 23:S208–S219.

St. George M, Kutas M, Martinez A, Sereno MI. 1999. Semantic integration in reading: engagement of the right hemisphere during discourse processing. Brain 122:1317–1325.

Tzourio-Mazoyer N, Josse G, Crivello F, Mazoyer B. 2004. Interindividual variability in the hemispheric organization for speech. NeuroImage 21:422–435.

Wernicke C. 1874. Der Aphasische Symptomencomplex. Breslau: Cohn and Weigert.

Wildgruber D, Riecker A, Hertrich I, Erb M, Grodd W, Ethofer T, Ackermann H. 2005. Identification of emotional intonation evaluated by fMRI. NeuroImage 24:1233–1241.

Wilson SM, Iacoboni M. 2006. Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. NeuroImage 33:316–325.

Wilson SM, Saygin AP, Sereno MI, Iacoboni M. 2004. Listening to speech activates motor areas involved in speech production. Nat Neurosci 7:701–702.

Wise RJ, Scott SK, Blank SC, Mummery CJ, Murphy K, Warburton EA. 2001. Separate neural subsystems within 'Wernicke's area'. Brain 124:83–95.

Worsley KJ, Liao C, Aston J, Petre V, Duncan GH, Morales F, Evans AC. 2002. A general statistical analysis for fMRI data. NeuroImage 15:1–15.

Xu J, Kemeny S, Park G, Frattali C, Braun A. 2005. Language in context: emergent features of word, sentence, and narrative comprehension. NeuroImage 25:1002–1015.