CHAPTER 4

Beyond superior temporal cortex: Intersubject correlations in speech comprehension

4.1 Abstract

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 an fMRI study, we presented 24 subjects with auditory or audiovisual narratives, and used model-free intersubject correlational analyses to reveal 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 involvement of superior temporal areas, however the intersubject correlational analyses also revealed an extended network of areas typically not reported in previous studies of narrative speech comprehension. Two novel 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 conceptual and affective representations.

4.2 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 & Poeppel, 2000, 2004; Wise et al., 2001; Narain et al., 2003; Scott & Wise, 2004), and convincing arguments have been presented based on imaging and lesion data that the earliest stages of speech perception are bilateral (Hickok & Poeppel, 2000, 2004; Poeppel, 2001).

However it is clear that language must interface with numerous other systems such as working memory, conceptual knowledge, emotion, and theory of mind. So we would expect that many brain regions beyond superior temporal cortex must be involved in speech comprehension. While activations restricted to the temporal lobe are not surprising in neuroimaging studies employing sophisticated control conditions to localize various aspects of prelexical processing (e.g. Liebenthal et al., 2005), it is striking that many studies which employ higher level linguistic structures such as sentences (e.g. Scott et al., 2000; Narain et al., 2003; Rodd et al., 2005) also produce activations predominantly restricted to superior temporal cortex. Even in studies of connected narratives, the only consistently activated region besides bilateral superior temporal cortex is the left inferior frontal gyrus (IFG) (e.g. Mazoyer et al., 1993; Dehaene et al., 1997; Skipper et al., 2005).

One possibility is that the necessity of comparing speech comprehension to 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 (e.g. Dehaene et al., 1997; Scott et al., 2000; Narain et al., 2003; Rodd et al., 2005), 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 (and potentially important) 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; Mazoyer et al., 2001; Gusnard & Raichle, 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). 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 to 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 visual processing. Our results revealed the involvement of numerous regions not typically reported in studies of 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.

4.3 Materials and methods

4.3.1 Participants

A total of 24 native English-speaking participants were scanned with fMRI. 12 subjects (3 males, mean age 24.2, range 19 to 33 years) listened to auditory cartoon narratives, and 12 subjects (6 males, mean age 24.7 years, range 20 to 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.

4.3.2 Experimental design

Our auditory and audiovisual stimuli consisted of cartoon narrations (McNeill, 1992). We showed our actor Looney Tunes cartoons from the video "Carrotblanca" (Figure 4.1a, Warner Brothers Family Entertainment) and she was videotaped while recounting the plot of various stories to a listener (Figure 4.1b). 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 seconds of rest (with blank screen)

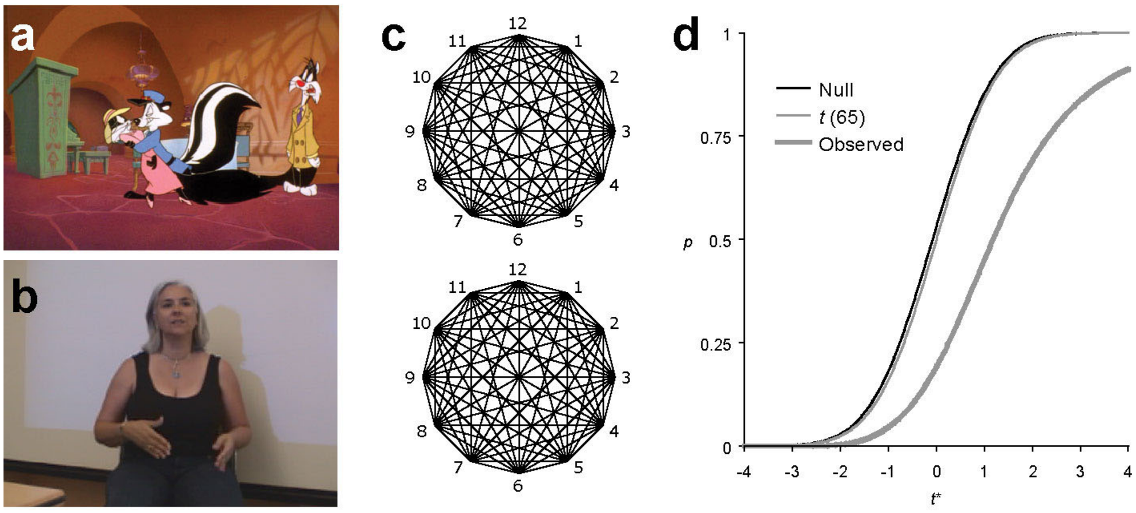

Figure 4.1 Materials and methods. (a) Frame from the movie "Carrotblanca". (b) Frame from stimulus video of the actor retelling the narrative. (c) Each group comprised 12 subjects, and 66 pairwise correlational maps were created for each group by correlating voxel timecourses for each pair of subjects. (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).

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 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).

4.3.3 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 (TR = 2000 ms; TE = 25 ms; flip angle = 90º; 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 two longer narratives, or 397 for the three shorter narratives. In addition, two 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).

4.3.4 Image processing

The fMRI data were preprocessed using tools from FSL (Smith et al., 2004). Skull stripping was performed with BET, motion correction was carried out with MCFLIRT, and the program IP was used to smooth the data with a Gaussian kernel (8mm FWHM) and to normalize mean signal intensity across subjects.

Functional images were aligned to high-resolution coplanar images using an affine transformation with 6 degrees of freedom. High-resolution coplanar images were then aligned to the standard MNI average of 152 brains using an affine transformation with 12 degrees of freedom.

4.3.5 Standard analysis

A standard subtraction analysis comparing auditory or audiovisual language comprehension to 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, in functional space. The boxcar design matrix was convolved with a hemodynamic response function modeled as a difference of two gamma functions. Temporal drift was removed by adding a cubic spline in the frame times to the design matrix (one covariate per 2 minutes of scan time), and spatial drift was removed by adding a covariate in the whole volume average. Six motion parameters (three 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 two runs within each subject were combined using a fixed effects model, then the resultant statistical images were registered to MNI space by concatenating the transformation matrices derived above.

Group analysis was performed for each of the two 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 theory were reported.

The two groups were also compared to 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 Gaussian random field theory as above.

Statistical parameter maps were displayed as overlays on a high-resolution single subject T1 image (Holmes et al., 1998) using AFNI (Cox, 1996) and custom software. In the tables of regions with 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 where two or more regions were contiguous, prominent local maxima were identified and tabulated separately.

4.3.6 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 minutes 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). Furthermore, the first 16 seconds 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 (Figure 4.1c). These maps were created by a custom MATLAB program that computed the correlation between residual timecourses obtained above at each voxel. The r statistic is not normally distributed, but it can be 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 dataset was created by shifting the data in time such that timeseries were no longer aligned across subjects. The algorithm was run as above, except that at each voxel, the two timeseries 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 two timeseries 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 datasets ($\Phi < 0.03$ in most voxels). This was confirmed based on simulations with randomly generated data based on autoregressive models with various parameters.

This dataset was analyzed with FMRISTAT to derive a *t* statistical parameter map, and we examined the distribution of the *t* statistic. Surprisingly, we found that it was distributed approximately as $t(65)$ (Figure 4.1d). 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 dataset was 0.0039, therefore it is slightly conservative to use the $t(65)$-based cutoff in thresholding the data. Finally, note that the observed distribution of the unshifted real data, also depicted in Figure 4.1d, 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 comparisons. Statistical parameter maps were displayed and tables created as described above.

4.4 Results

The group data for auditory-only speech comprehension are shown in Figure 4.2a and Table 4.1. 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). A homologous region was observed in the right hemisphere which did not reach the minimum cluster size criterion (peak: 60, –4, 50; t = 3.0). 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 & 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. Intercorrelations were observed in several midline areas: the anterior cingulate gyrus, medial superior frontal gyrus, posterior

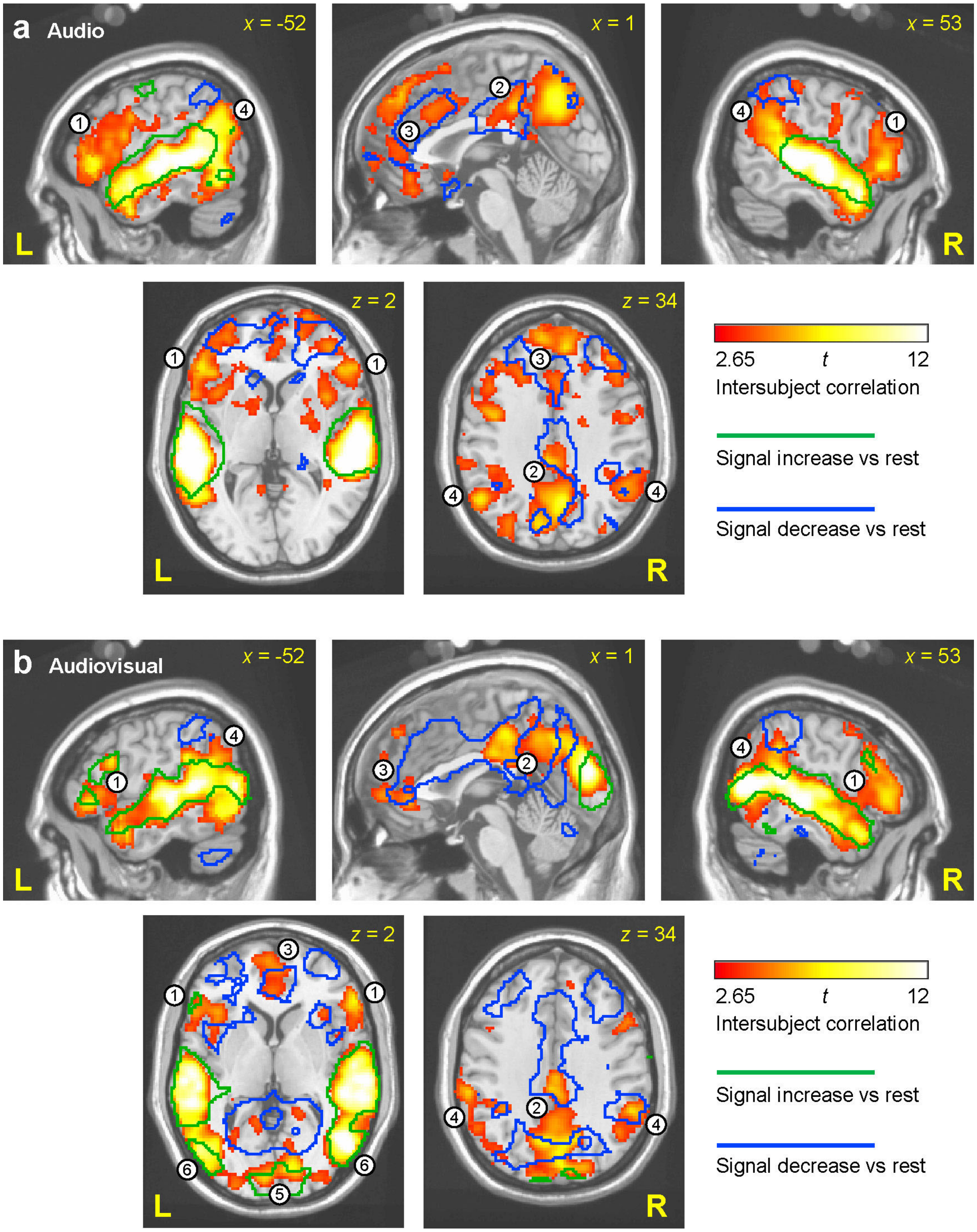

Figure 4.2 (a) Auditory speech comprehension. Five slices are shown with MNI coordinates provided in the top right of each slice. Images are displayed in neurological orientation with the left hemisphere on the left. Intersubject correlations are shown in the red-yellow-white color scale.

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) inferior frontal gyrus; (2) posterior cingulate and adjacent precuneus; (3) anterior cingulate and adjacent medial frontal cortex; (4) left and right angular gyri. (b) Audiovisual speech comprehension.

---

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 (Figure 4.2b, Table 4.2). 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 non-stimulated 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

Table 4.1 Regions significantly correlated across subjects, or activated or deactivated relative to rest for auditory-only narratives

Area	Peak MNI coordinates (mm)			Extent	Max *t*	Cluster *p*
	x	*y*	*z*	(mm^3)		
Intersubject correlational analysis						
Extensive bilateral fronto-tempero-parietal network				391272	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	13104	11.5	< 0.0001
Right cerebellum	26	–76	–34	10536	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				64272	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				48880	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				174800	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	12256	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: superior temporal gyrus (STG); superior temporal sulcus (STS); middle temporal gyrus (MTG); superior frontal gyrus (SFG); middle frontal gyrus (MFG); inferior frontal gyrus (IFG); inferior frontal sulcus (IFS).

frontal activity was also found relative to rest in the standard analysis, albeit considerably more circumscribed. Unlike in the auditory-only group, activity was not significant in the precentral gyrus/central sulcus in the standard analysis. However there were bilateral clusters in this vicinity which did not reach the minimum cluster size criterion; peaks were (–48, 2, 52; t = 6.8) on the left and (56, 4, 48; t = 4.2) on the right.

The audiovisual and auditory-only groups were then directly compared (Figure 4.3, Tables 4.3 and 4.4). In the standard analysis, the only regions which showed greater signal change in the audiovisual condition relative to the auditory condition were early visual and visual motion areas (Figure 4.3a). The intersubject correlational analysis also showed significantly greater correlations across subjects in these areas, along with one

Table 4.2 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				321208	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				176912	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 middle temporal (MT)	–52	–70	8		12.3	
Right middle temporal (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, 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				335832	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

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

additional region: the right posterior superior temporal sulcus (STS), previously implicated in perception of biological motion (Allison et al., 2000; Pelphrey et al., 2005).

Although in the standard analysis bilateral inferior frontal 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 4.3b and Table 4.4. The standard analysis showed greater activity relative to rest in the

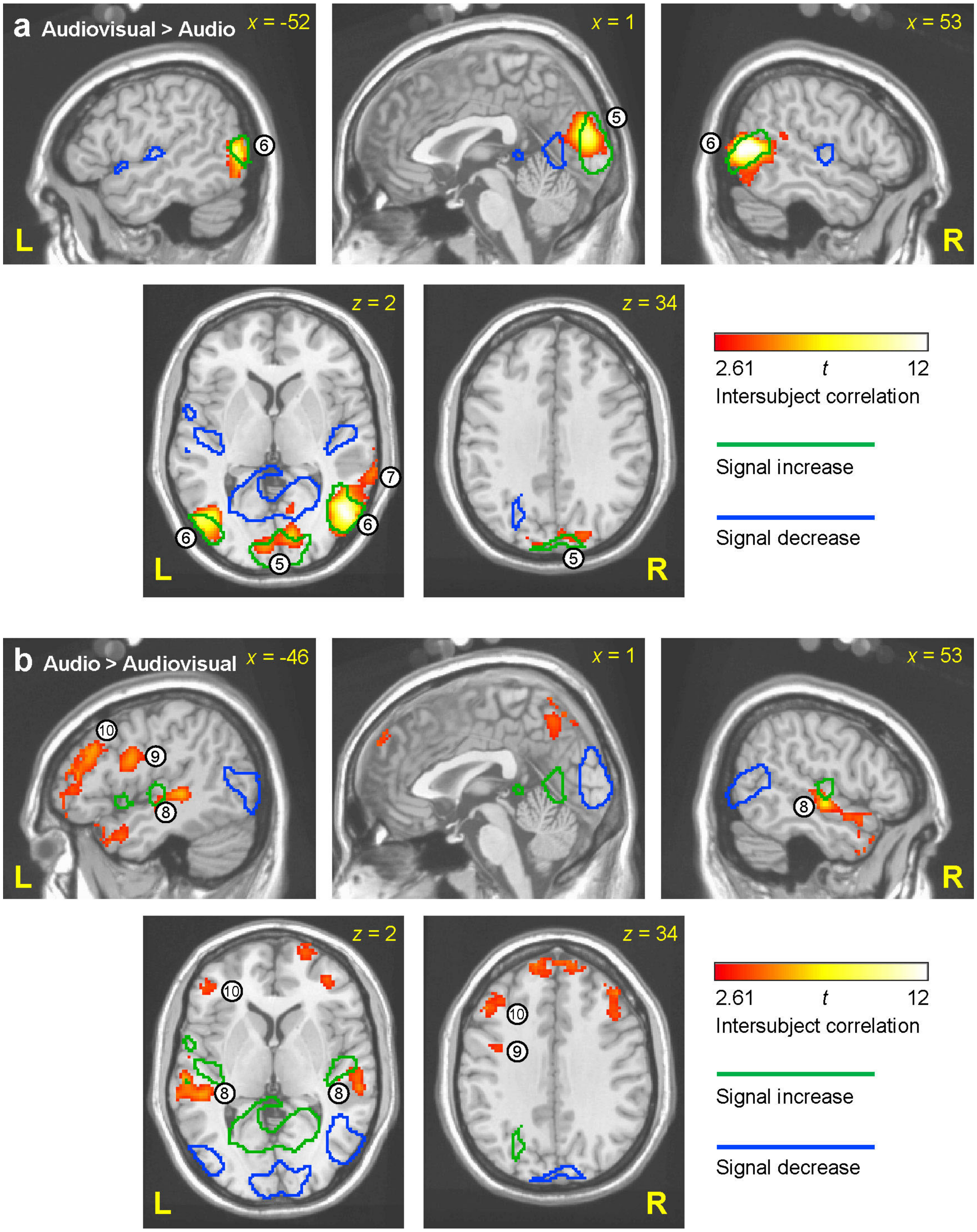

Figure 4.3 (a) Audiovisual speech comprehension relative to auditory speech comprehension. See caption to figure 4.2 for explanation of conventions. The red-yellow-white color scale shows areas which were more correlated across subjects for audiovisual speech than for auditory-only

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 superior temporal sulcus. (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.

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 role; 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.

Table 4.3 Regions which were significantly more correlated across audiovisual subjects than auditory-only subjects, or which were activated for audiovisual narratives relative to auditory-only narratives

Area	Peak MNI coordinates (mm)			Extent	Max *t*	Cluster *p*
	x	*y*	*z*	(mm^3)		
Intersubject correlational analysis				54856	16.4	< 0.0001
Early visual areas and right higher level visual areas						
Left medial occipital cortex	–10	–94	20		7.8	
Right medial occipital cortex	8	–86	20		12.4	
Right middle temporal (MT) visual motion area	50	–68	8		16.4	
Right posterior STS	70	–38	8		6.8	
Left middle temporal (MT) visual motion area	–46	–72	8	13872	12.9	< 0.0001
Signal increases in standard analysis						
Early visual and visual motion areas				65880	13.7	< 0.0001
Left medial occipital cortex	–14	–96	16		11.4	
Right medial occipital cortex	12	–92	20		13.7	
Left middle temporal (MT) visual motion area	–48	–82	8		8.1	
Right middle temporal (MT) visual motion area	52	–68	6		10.7	
Signal decreases in standard analysis						
See table 4 signal increases.						

4.5 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 & Poeppel, 2004). However the intersubject correlational analysis also uncovered an extended network of areas involved in speech comprehension including default mode areas (anterior cingulate and adjacent medial frontal cortex, posterior cingulate and adjacent precuneus), and the bilateral IFG and premotor areas. Many of these regions have not been implicated in previous studies

Table 4.4 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	Max *t*	Cluster *p*
	x	*y*	*z*	(mm^3)		
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				38776		< 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.						

of narrative speech comprehension. Differences between intersubject correlations in the two 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.

4.5.1 Default mode network

A consistent set of brain regions are deactivated in multiple different active task conditions in comparison to 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 & 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 Figure 4.2, Tables 4.1 and 4.2). 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 & 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 default mode 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 entail. This demonstrates that default mode regions are not simply shut off in response to an active task. Instead, the data suggest two possible interpretations, which may both be valid. 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 & 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. In the remainder of this section, we discuss the default mode regions in which intersubject correlations were observed, and their possible functional roles.

Intercorrelations were observed in the ventral anterior cingulate gyrus and adjacent medial prefrontal cortex in both the auditory and audiovisual groups. Very similar regions have been deactivated in previous default mode studies (see McKiernan et al., 2003 for review). This ventral, rostral section of the anterior cingulate gyrus is involved with affective and emotional processes (Bush et al., 2000). More specifically, it has been proposed that ventral medial areas are concerned with integration of emotional and cognitive processes (e.g. Bechara et al., 1997).

In the dorsal anterior cingulate, and adjacent medial prefrontal cortex, intersubject correlations were significant only in the auditory group. However, the between-groups comparison did not reveal any group differences in this region, as there were subthreshold correlations in the audiovisual group also. The coordinates of the correlated regions in the auditory group are similar to those of regions deactivated in prior studies, especially McKiernan et al. (2003). The dorsal anterior cingulate cortex is concerned with cognitive and motor functions as opposed to the affective functions of the ventral sector (Bush et al., 2000). This region is thought to play an executive attentional role and is especially concerned with monitoring and processing conflict (Botvinick et al., 1999). 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 & Frith, 1999).

Posterior medial regions including the posterior cingulate gyrus and the adjacent precuneus were highly intercorrelated across subjects in both auditory and audiovisual groups. The intercorrelated regions closely correspond with areas deactivated in prior studies (McKiernan et al., 2003). Gusnard & Raichle (2001) have proposed that the role of these areas in the default mode network is to represent and monitor the external environment, based in part on the fact that the visual periphery is represented along the dorsal midline.

The final prominent default mode region is the bilateral angular gyrus. In the present study, deactivations relative to rest were observed bilaterally in this region in both the auditory and audiovisual groups. However unlike the other three major default mode regions, significant intersubject correlations were not observed in the angular gyrus. Importantly though, bilateral superior temporal regions showing correlations across subjects extended dorsally and posteriorly approximately to the boundary of the angular gyrus regions that were deactivated in the standard analysis. This contrasted with the standard analysis, where these superior temporal regions did not extend so far back. Thus there is a discrepancy between the two 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 & Howes, 1978). In general, this area has been argued to be important for auditory to articulatory mapping in language comprehension and production (Hickok & 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 angular gyrus. But the parts of the angular gyri that are deactivated relative to rest and not correlated across subjects appear to be concerned with internal processes that are not systematically modulated by linguistic input.

Our results demonstrating intersubject correlations in default mode regions are at variance with those of Golland et al. (2006), 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. (2006) defined the intrinsic system as voxels correlated with the timecourse of "seed" ROIs 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 two possible reasons for this discrepancy with our results. Firstly, Golland et al. (2006) assessed correlations between signal in response to two 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 where intersubject correlations were first proposed, correlated regions were reported in the cingulate gyrus and retrosplenial cortex (Hasson et al., 2004). It is possible that these may relate to the default mode network, though it is difficult to determine because no group analysis was performed and flattened cortical maps were used, so MNI coordinates were not reported.

A second major difference between our study and Golland et al. (2006) 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.

4.5.2 Involvement of the bilateral inferior frontal gyrus in speech comprehension

Intersubject correlational analyses revealed extensive bilateral regions in the IFG 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 has been demonstrated to be involved in semantic, syntactic and phonological processes (Bookheimer, 2002). Since 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. However it is noteworthy that IFG activity in several studies which have compared narrative comprehension to rest or auditory baselines has been much more limited, and variable in location from study to study. A summary of IFG activity reported in studies of narrative comprehension was provided earlier in Table 1.7. Only two studies have failed to observe significant IFG activity, though in one, the subjects were children (Ahmad et al., 2003), and in the other, activity was reported more posteriorly in the left precentral gyrus (Perani et al., 1998).

The right IFG was also shown to be highly significantly correlated across subjects, to a degree similar to the left IFG. Only one previous study of auditory narrative comprehension has reported right inferior frontal involvement: Dehaene et al. (1997)

found activity in the right inferior frontal junction (where the inferior frontal sulcus meets the precentral sulcus). In that study, right inferior frontal junction activity was considerably weaker than homologous activity in the left hemisphere. However right hemisphere areas, including the IFG, are thought to play a role in a range of linguistic processes including prosody (Ross, 1981; Adolphs, 2002; Wildgruber et al., 2005) and understanding of higher level discourse (Xu et al., 2005; see Bookheimer, 2002, and 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 so much more circumscribed in previous studies of narrative speech comprehension, and in the standard analysis in the present study? Although inferior frontal areas are not prominent components of the default mode network, two studies have reported left IFG regions to be deactivated relative to rest: Shulman et al. (1997) reported a region on the border of Brodmann areas 47 and 10 with peak coordinates (–33, 45, –6) which tended to show reduced signal across a range of tasks, and Binder et al. (1999) observed signal decrease in Brodmann area 45 with peak (–51, 26, 14) when a tone task was compared to rest. Therefore one or more regions in the left IFG may be involved in cognitive processes during baseline conditions. Binder et al. (1999) proposed that the region they identified was involved in semantic processing, because it was not deactivated when a semantic task was compared to rest. However, the right IFG has not been claimed to belong to the default state network, and several studies have not even identified default mode areas in the left IFG (e.g. McKiernan et al., 2003;

Greicius & Menon, 2004). So with the possible exception of more anterior sectors of the left IFG, high activity at rest or in passive conditions probably cannot account for the failure to observe bilateral IFG activity in narrative comprehension studies.

Rather, our results suggest that the left and right IFG do not exhibit a consistent signal increase during narrative comprehension, but rather they exhibit 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 the literature cited above sheds light in the kinds of processes the left and right IFG might be concerned with (Bookheimer, 2002; Jung-Beeman, 2005). Functional imaging designs where a condition of interest is compared to a control condition clearly are unable to detect involvement of a region whose response consists of a time varying signal change which is sometimes positive relative to the control condition and sometimes negative.

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 lack of reinforcing visual 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. Under this view, increased intersubject correlations in the auditory-only group would reflect common modulations across subjects as parts of the narratives that were more difficult to understand than other parts required increased contributions from top-down processes.

4.5.3 Premotor cortex

The bilateral frontal regions which were correlated across subjects extended posteriorly and dorsally into premotor regions in both groups, but especially in the auditory-only group. The left ventral precentral gyrus (ventral premotor cortex) was significantly more correlated across subjects in the auditory-only group, and this region may also play a role in generation of top-down hypotheses in speech perception, perhaps at more of a prelexical, phonetic level than the prefrontal regions discussed above. A recent study has also argued for a similar role for premotor cortex in low-level phonetic perception (Wilson & Iacoboni, 2006).

More superior premotor regions were also identified in both groups at coordinates of approximately $z = 50$. The standard analysis revealed a significant left hemisphere activation in the precentral gyrus/central sulcus in the auditory group, a similar subthreshold activation in the right hemisphere, and bilateral subthreshold activations in the audiovisual group. The intersubject correlational analysis revealed similar regions which were generally somewhat anterior to those found in the standard analysis, centered around the precentral sulcus.

In previous studies we have classified similar regions as superior ventral premotor cortex (sPMv) (Wilson et al., 2004; Wilson & Iacoboni, 2006), because dorsal premotor cortex is not known to have any orofacial representations (Raos et al., 2003), and most such activations fall below the dividing line of $z = 51$ between dorsal and ventral premotor cortex proposed by Rizzolatti et al. (2002). It has been shown that this region responds not only to speech perception but also to speech production (Wilson et al., 2004; Wilson & Iacoboni, 2006). Similar activations have been observed in several studies of speech perception (Binder et al., 2000; Skipper et al., 2005; Uppenkamp et al., 2006) and in one study of auditory narrative comprehension (Crinion & Price, 2005).

4.5.4 Regions differentially implicated in audiovisual speech perception

Besides early visual and visual motion areas, there was just one region which showed significantly greater correlations within the audiovisual group compared to 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 and mouth and eyes. Previous studies have demonstrated a somatotopy in the STS, with mouth representations anterior to hand and eye representations (Allison et al., 2000; Pelphrey et al., 2005). In our study, the coordinates reported correspond to regions involved in perception of the mouth, but the intercorrelated region is contiguous with visual motion area MT posteriorly, so also encompasses STS regions involved in perception of hands and other body parts.

Another context in which the STS is often implicated is crossmodal binding in audiovisual speech perception (Calvert et al., 2000; Calvert, 2001; Raij et al., 2001; Macaluso et al., 2004). Anatomically, the STS is well situated for a role in crossmodal binding as it receives convergent projections from both visual and auditory areas (Jones & Powell, 1970; Bruce et al., 1981). In each of the audiovisual speech perception studies cited above, the left STS in particular was shown to be most important for crossmodal binding of linguistic stimuli. However the right STS appears to play a role also. For instance, the right STS exhibited a superadditive response to audiovisual stimuli greater than the sum of audio alone and visual alone, but it did not demonstrate a subadditive response to mismatched inputs as the left STS did (Calvert et al., 2000). In a previous study comparing audiovisual narrative comprehension to 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, it was the case that 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 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 & 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 (see Molnar-Szakacs et al., 2005 for meta-analysis; Iacoboni et al., 2005; Molnar-Szakacs et al., in press).

4.5.5 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 two analyses. The standard analysis showed that there was greater activity in the transverse temporal gyri bilaterally, i.e. 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 since 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 strong 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 & 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 & Poeppel, 2000, 2004).

4.5.6 Conclusion

Intersubject correlational analysis proved to be a useful complement to conventional subtraction analysis, as it revealed a network of regions involved in auditory or audiovisual speech comprehension which have not typically been reported in previous studies. 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 are important for higher level and top-down linguistic processes, and interfaces with conceptual and affective representations.