Behavioral/Systems/Cognitive

The Cortical Dynamics of Intelligible Speech

Alexander P. Leff,[1] **Thomas M. Schofield,**[1] **Klass E. Stephan,**[1,2] **Jennifer T. Crinion,**[1] **Karl J. Friston,**[1] and **Cathy J. Price**[1]

[1]Wellcome Trust Centre for Neuroimaging, University College London, London WC1N 3BG, United Kingdom, and [2]Branco-Weiss-Laboratory for Social and Neural Systems Research, Institute for Empirical Research in Economics, University of Zurich, CH-8006 Zurich, Switzerland

An important and unresolved question is how the human brain processes speech for meaning after initial analyses in early auditory cortical regions. A variety of left-hemispheric areas have been identified that clearly support semantic processing, although a systematic analysis of directed interactions among these areas is lacking. We applied dynamic causal modeling of functional magnetic resonance imaging responses and Bayesian model selection to investigate, for the first time, experimentally induced changes in coupling among three key multimodal regions that were activated by intelligible speech: the posterior and anterior superior temporal sulcus (pSTS and aSTS, respectively) and pars orbitalis (POrb) of the inferior frontal gyrus. We tested 216 different dynamic causal models and found that the best model was a "forward" system that was driven by auditory inputs into the pSTS, with forward connections from the pSTS to both the aSTS and the POrb that increased considerably in strength (by 76 and 150%, respectively) when subjects listened to intelligible speech. Task-related, directional effects can now be incorporated into models of speech comprehension.

Key words: intelligibility; speech; dynamic causal modeling; connectivity; language; fMRI

Introduction

Many studies have shown that intelligible or meaningful speech is processed within multimodal temporal lobe cortex and beyond, but little is known about how these areas interact and drive one another. To investigate directed influences, i.e., effective connectivity, among key speech comprehension areas within the dominant temporal and frontal lobes, we used dynamic causal modeling (DCM) and Bayesian model selection (BMS) to test a variety of alternative network architectures. DCM analyses have become popular recently to study various aspects of language processing (Bitan et al., 2005; Mechelli et al., 2005, 2007; Heim et al., 2007; Kumar et al., 2007; Sonty et al., 2007; Booth et al., 2008; Chow et al., 2008). However, there have been no investigations of directed influences among human brain areas in response to intelligible speech.

There is broad consensus that the "core" of Wernicke's area [including the posterior superior temporal sulcus (pSTS)] is important for auditory processing of speech, but its relationship to other important language areas, in both the temporal lobe and the inferior frontal gyrus (IFG), i.e., "Broca's area," is disputed. Some models emphasize the role of dominant pSTS in the processing of complex acoustic sounds, of which speech is one example, with anterior STS (aSTS) more involved in processing meaning (Scott et al., 2000). Other models state that both speech perception and post-perceptual (semantic) functions are performed by bilateral temporal regions working in concert, with a

reduced role for left aSTS in supporting meaning (Hickok and Poeppel, 2007). What these models have in common are bidirectional arrows connecting candidate regions. This is because, until relatively recently, researchers have tended to adopt correlational methods to investigate relationships between network elements identified by functional magnetic resonance imaging (fMRI) or positron emission tomography, providing evidence for task-related coactivation (Obleser et al., 2007) but not for "directed influences," or effective connectivity, between regions.

We used DCM and BMS to compare 216 alternative models of effective connectivity among key regions for auditory speech processing. Using this exhaustive model search procedure, we aimed to identify the most likely temporofrontal network architecture underlying speech comprehension. The models were fitted to fMRI measurements of volunteers listening to spoken stimuli with high semantic content but short enough to support an event-related fMRI analysis. Notably, BMS is not confounded by differences in model complexity or multiple comparisons (Penny et al., 2004); the validity of the winning model increases with the number of models tested.

Because no single acoustic feature predicts the intelligibility of a given auditory stimulus, a methodological challenge is to produce unintelligible stimuli that are matched acoustically. We used an established method, time-reversing the normal speech stimuli, because these sound like speech but are unintelligible; listeners often report that this sounds like a "foreign language". It has been argued that reversed speech contains phonetic information based on a relative intersubject consistency when subjects attempted to transcribe reversed words (Binder et al., 2000). Because speaker identity information is also retained in time-reversed speech, subjects were able to perform the same visually cued task (i.e., was the speaker male or female?) on all auditory stimuli, thereby keeping the task (and implicit cognitive set) or-

thogonal to the experimental manipulation (intelligible speech vs unintelligible speech).

Materials and Methods

Subjects, stimuli, and task

Twenty-six right-handed subjects with normal hearing, English as their first language, and no history of neurological disease participated (14 female; mean age, 27.3 years; range, 21–38 years). All subjects gave informed consent, and the study was approved by the local research ethics committee.

Auditory stimuli consisted of word pairs spoken by either a male or female English speaker. The task, which was orthogonal to the effect of interest, was a gender decision: was the speaker male or female? Half of the word pairs were idiomatic (e.g., "cloud nine," "mint condition"), whereas the other half were reordered idioms (e.g., "mint nine"). Both sets of stimuli were time reversed to create unintelligible stimuli but preserve acoustic complexity and voice identity. Reversal of speech does not alter the fundamental frequency of the acoustic signal, which is necessary for making gender decisions, but it destroys the intelligibility of the phrases while retaining the overall acoustic complexity. The idiom and rearranged idiom stimuli were recorded digitally in a soundproof room onto an Apple Macintosh laptop using Audacity 1.2.3 software. All stimuli were recorded twice, once by the male and once by the female speaker. The files were then edited for quality and length (any over 1080 ms were discarded; range, 677–1080 ms). Individual word-pair files were then loaded into Praat software version 4.2.21 (Boersma, 2001), time reversed, and amplified so that the male and female stimuli were of equivalent loudness.

In-scanner procedure

Three classes of auditory stimuli were played to subjects in an equal ratio: idioms, rearranged idioms and time-reversed stimuli. The stimuli were arranged in blocks of seven word pairs, with all stimuli within a block being of the same class with a variable ratio of male/female speaker (2:5, 3:4, 4:3, or 5:2). Subjects were asked to make a gender judgment on the auditory stimuli and to report their decision with a visually cued left-handed finger press. They were not told that some of the stimuli were idioms or that some were time reversed. At 1180 ms after an auditory stimulus was presented, a visual male/female response prompt was displayed for 2420 ms. To ensure a dense sampling of the hemodynamic response function, we used a stimulus onset asynchrony of 4050 ms for the auditory stimuli that was a non-integer multiple of the repetition time of the MRI data acquisition. Interleaved between every seven stimulus blocks, there was a block of no auditory stimulation for 12.6 s (subjects were asked to fixate a "∼" sign on the screen). There were nine blocks in each session, which lasted for 6.5 min. There were four sessions per subject, resulting in data being collected for 84 stimuli from each of the three classes. All blocks and male/female speaker ratios within blocks were pseudorandomized across subjects. No stimulus was repeated.

fMRI scanning and stimulus presentation

A Siemens Sonata 1.5 T scanner was used to acquire T2*-weighted echo-planar images with blood oxygenation level-dependent (BOLD) contrast. Each echo-planar image comprised 35 axial slices of 2.0 mm thickness with 1 mm interslice interval and 3 × 3 mm in-plane resolution. Volumes were acquired with an effective repetition time of 3150 ms per volume, and the first five volumes of each session were discarded to allow for T1 equilibration effects. A total of 127 volume images were acquired in four consecutive sessions, each lasting 6.5 min (accounting for the discarded volumes; 488 volumes in total were analyzed for each subject). After the functional runs, a T1-weighted anatomical volume image was also acquired.

Stimuli were presented binaurally using a custom-built electrostatic headphones based on Koss. The stimuli were presented initially at 85 dB/sound pressure level (SPL), and then subjects were allowed to adjust this to a comfortable level, while listening to the stimuli during a test period of echo-planar imaging scanner noise. The earphones provide ∼30 dB/SPL attenuation of the scanner noise.

Statistical parametric mapping

Statistical parametric mapping (SPM) was performed using the SPM5 software (Wellcome Trust Centre for Neuroimaging, available at http://fil.ion.ucl.ac.uk/spm). All volumes from each subject were realigned and unwarped, using the first as reference and resliced with sinc interpolation. For each subject, the mean functional image was coregistered to a high-resolution T1 structural image. This image was then spatially normalized to standard Montreal Neurological Institute (MNI) space using the "unified segmentation" algorithm available within SPM5 (Ashburner and Friston, 2005), with the resulting deformation field applied to the functional imaging data. These data were then spatially smoothed, with an 8 mm full-width at half-maximum isotropic Gaussian kernel to compensate for residual variability after spatial normalization and to permit application of Gaussian random field theory for multiple comparison correction. First, the statistical analysis was performed in a subject-specific manner. To remove low-frequency drifts, the data were high-pass filtered using a set of discrete cosine basis functions with a cutoff period of 128 s. Serial correlations were modeled as a first-order autoregressive process (Friston et al., 2002). Each stimulus was modeled as a separate delta function and convolved with a synthetic hemodynamic response function. In this study, our analyses focused on the main effect of intelligible relative to time-reversed speech, treating idiom and rearranged idioms as the same stimulus type, because initial analyses testing for the effect of idioms versus rearranged idioms found no significant effects.

Our first (within-subject) level statistical models included the realignment parameters (to regress out movement-related variance) and four session effects as covariates of no interest. The effects of interest (intelligible and time-reversed stimuli) were modeled in two separate columns of the design matrix, and corresponding parameter estimates were calculated for all brain voxels using the general linear model (Friston et al., 1995). Contrast images were computed for each subject and entered into a second (between-subject) level random effects analysis. We tested two orthogonal contrasts, the effect of all types of auditory stimulus and the effect of intelligibility; that is, the BOLD response specific to intelligible versus unintelligible stimuli. For the first contrast, we applied a correction for multiple statistical comparisons across the whole brain. For the second, we applied small volume corrections for spheres of 10 mm radius centered on coordinates based on previous language studies using implicit speech perception paradigms: posterior STS, [−52, −42, 6] (Binder et al., 2000); anterior STS, [−55, 8, −16] (Binder et al., 2000; Scott et al., 2000; Thierry and Price, 2006); and IFG, [−40, 24, −2] (Meyer et al., 2004; Thomsen et al., 2004; Yokoyama et al., 2006; Wu et al., 2007).

Dynamic causal modeling

DCM investigates how brain regions interact with one another during different experimental contexts (Friston et al., 2003). The strength and direction of regional interactions are computed by comparing observed regional BOLD responses with BOLD responses that are predicted by a neurobiologically plausible model. This model describes how activity in and interactions among regional neuronal populations are modulated by external inputs (i.e., experimentally controlled stimuli or task conditions) and how the ensuing neuronal dynamics translate into a measured BOLD signal. The parameters of this model are adjusted in an iterative manner (using Bayesian parameter estimation) so that the predicted responses match the observed responses as closely as possible; the parameter estimates are expressed in terms of the rate of change of neuronal activity in one area that is associated with activity in another (i.e., rate constants). The external inputs enter the model in two different ways: they can elicit responses through direct influences on specific regions ("driving inputs," e.g., sensory stimulation) or they can change the strength of coupling among regions ("modulatory inputs," e.g., stimulus properties, task effects, or learning). This distinction represents an analogy, at the level of neural populations, to the concept of driving and modulatory afferents in studies of single neurons (Sherman and Guillery, 1998). The modulatory terms are also referred to as "bilinear terms"; they capture how the coupling between the "source" and "target" regions varies as a function of the experimentally controlled manipulations.

Technically speaking, DCM for fMRI models neural dynamics in a system of interacting brain regions, representing neural population activity of each region by a single state variable (Friston et al., 2003; Stephan et al., 2007a). It models the change of this neural state vector x in time as a bilinear differential equation: for a single input u, this state equation can be written as follows:

$$\frac{dx}{dt} = (A + uB)x + Cu.$$

Here, the A matrix represents the fixed (context-independent or endogenous) strength of connections between the modeled regions, and the B matrix represents the modulation of these connections (e.g., attributable to intelligibility). Finally, the C matrix represents the influence of direct (exogenous) inputs to the system (e.g., sensory stimuli).

To explain regional BOLD responses, DCM for fMRI combines the model of neuronal population dynamics in the above equation with a biophysical forward model of hemodynamic responses. This hemodynamic model describes how neuronal population activity induces changes in regional vasodilation, blood flow, blood volume, deoxyhemoglobin content, and, eventually, a predicted BOLD signal (Friston et al., 2000; Stephan et al., 2007b). The incorporation of this biophysical model allows inferences to be made from coarsely sampled BOLD time series about neuronal events occurring at a much finer timescale. Together, the deterministic neural and hemodynamic state equations yield a likelihood model of BOLD data that includes confounds (e.g., signal drift) and measurement error. DCM uses a fully Bayesian approach to parameter estimation, with empirical priors for the hemodynamic parameters and conservative shrinkage priors (i.e., zero mean, small variance) for the coupling parameters (Friston et al., 2003).

Note that the neuronal state equation above does not account for conduction delays in either inputs or inter-regional influences; instead, each area exerts instantaneous effects on other areas. For fMRI data, modeling neuronal conduction delays is futile because, because of the large regional variability in hemodynamic response latencies, fMRI data do not posses enough temporal information to enable estimation of inter-regional axonal conduction delays, which are typically in the order of 10–20 ms (in contrast, conduction delays are important for DCMs of electrophysiological data). Critically, the differential latencies of regional hemodynamic responses are accommodated by region-specific biophysical parameters in the hemodynamic model described above (cf. Friston et al., 2003). In conclusion, the causal dependencies identified by DCM for fMRI do not describe a temporal sequence of activation. Instead, they rest on three things (Stephan et al., 2007a): (1) knowledge of external perturbations, i.e., when and where (sensory) inputs enter the system, (2) temporal dependencies within regional time series (embodied by the neuronal state equations), and (3) the combination of neuronal and hemodynamic models that enable inference about dependencies among neuronal time series despite large interregional variations in BOLD latencies. So far, implementations of DCM exist for fMRI (Friston et al., 2003), EEG/MEG (David et al., 2006), and invasively recorded local field potentials (Moran et al., 2007).

Selection of regional time series. The goal of DCM for fMRI is to explain regional BOLD responses in terms of inter-regional connectivity and its experimentally induced modulation. More specifically, DCM is a generative model of BOLD data that provides a mechanistic explanation for the experimentally induced variance in a number of interacting regions but does not account for variance that is unrelated to the experimental conditions (this is in contrast to methods such as vector autoregressive models or structural equation models, both of which are driven by white noise and usually do not incorporate experimentally controlled inputs). This means that a DCM should only contain regions whose dynamics are known to be driven by the experimental conditions; this is typically established by a conventional general linear model in which these regions are found to be "activated" by a particular statistical contrast (cf. Stephan et al., 2007a). On the contrary, it does not make sense to include regions that are not activated by any of the experimental manipulations but are chosen on the basis of anatomical considerations or previous experimental findings. In this study, the only regions showing a significant activa-

tion in the conventional SPM analysis were aSTS, pSTS, and IFG. We therefore constructed a series of alternative DCMs containing these areas, investigating how this differential activity could be optimally explained in terms of where auditory stimuli entered the network and which connections were modulated by intelligible versus unintelligible speech.

In each subject, 6 mm spherical volumes of interest (VOIs) were defined as the local maxima (contrast: intelligible vs unintelligible speech) in aSTS, pSTS, and IFG in the SPM{t} of each subject. The selection was guided by the results from the group analysis, such that the coordinates of the VOI did not differ from the group coordinates by >6 mm in any direction. This is a standard approach to ensure that subject-specific time series are both functionally and anatomically standardized (cf. Stephan et al., 2007c). Mean coordinates for the three VOIs were as follows: aSTS, [−54, 11, −17]; pSTS, [−53, −47, 7]; and IFG, [−48, 28, 4] (SD range, 3.2–4.3 mm).

Construction of DCMs. The main goals of this study were as follows: (1) to identify, in a first step, the basic architecture of the neural system underlying processing of intelligible speech, particularly with regard to the direction of influences among areas; and (2) to determine how connection strengths in this system were modulated by intelligible speech compared with unintelligible speech. This required us to formulate a set of DCMs representing all plausible alternatives of how the three regions identified by the SPM analysis could interact. We then tested which of these DCMs was an optimal explanation of the measured BOLD data.

The basic model, from which all alternative models were derived, is shown on the left of Figure 2. It assumes reciprocal connections among all pairs of areas (dotted arrows); this basic connectivity pattern is supported by anatomical tract tracing studies in the primate and human (Pandya and Yeterian, 1985). This basic model was then augmented by modulatory inputs that changed connection strengths as a function of the relevant experimental factors (i.e., intelligible vs unintelligible speech). Our finding that all three regions showed differential activity for this contrast provided two constraints for deriving a set of alternative models: each model was required to contain at least two modulatory inputs acting on connections between two different pairs of areas. Deriving all possible models under these constraints, the simplest models had only two modulatory inputs (12 models), whereas the most complex model had all six connections modulated by intelligibility (one model). In total, there were 54 different model structures. These were then crossed with four alternative ways of how auditory stimuli could enter the system, i.e., auditory inputs to IFG only, pSTS only, aSTS only, or to both temporal areas. In total, this resulted in 216 different DCMs per subject, which were fitted using a variational Bayes scheme using the so-called Laplace approximation (Friston et al., 2003, 2007). This means that the posterior distribution for each model parameter is assumed to be Gaussian and is therefore fully described by two values: a posterior mean [or maximum a posteriori (MAP) estimate] and a posterior variance. The fitted models were subsequently compared using Bayesian model selection as described in the next section.

Bayesian model selection. In this study, Bayesian model selection was used to decide which of the 216 alternative DCMs was optimal. Bayesian model selection not only takes into account the relative accuracy of competing models but also their relative complexity (i.e., the number of free parameters and the functional form of the generative model). It rests on the so-called "model evidence" [i.e., the probability $p(y|m)$ of the data y given a particular model m] (Raftery, 1995; Penny et al., 2004). Usually, the model evidence cannot be determined analytically; therefore, approximations are needed. For DCM, a number of different approximations have been suggested, including the Bayesian information criterion and Akaike information criterion. Following Penny et al. (2004), we used these two criteria for each model comparison and then chose the more conservative criteria to compute a Bayes factor, $p(y|m_1)/p(y|m_2)$, i.e., the evidence ratio of the two models being compared. Because model comparisons from different individuals are statistically independent, a group Bayes factor can be computed by multiplying the individual Bayes factors (Stephan et al., 2007b). For each subject of our group, we first performed pairwise comparisons between all models and then computed the group Bayes factor across subjects to select the best model. To make inferences on

the connections of the selected model, we used a conventional summary statistic approach, as described in the next paragraph.

Second-level analysis of model parameters. Once the best model was selected, we established intelligibility-related connection changes that were expressed consistently across subjects. To do this, we performed a classical second-level (between-subject) inference by applying a one-sample t test to the MAP estimates of the modulatory parameters from the subject-specific DCMs. We adopted a conservative procedure by using two-tailed t tests and a statistical threshold of $p < 0.05$ (Bonferroni's corrected). Note that, in DCM, the parameters of intrinsic connections and modulatory inputs are estimated using priors with zero mean (so-called shrinkage priors). This means that, in the absence of an effect, the MAP estimates of the coupling parameters shrink toward zero. At the group level, this warrants testing the null hypothesis that the parameter estimate of modulation by intelligibility is zero across subjects.

Results
Behavioral data
All subjects performed well on the gender-identity task, with an overall correct response rate of 99.77% (range, 97.61–100%). There was no accuracy difference in responses to intelligible (mean $\pm$ SE, 99.75 $\pm$ 0.10) and time-reversed stimuli (mean $\pm$ SE, 99.77 $\pm$ 0.12; $t_{(25)} = -0.15$; $p = 0.88$).

Imaging analyses (SPMs)
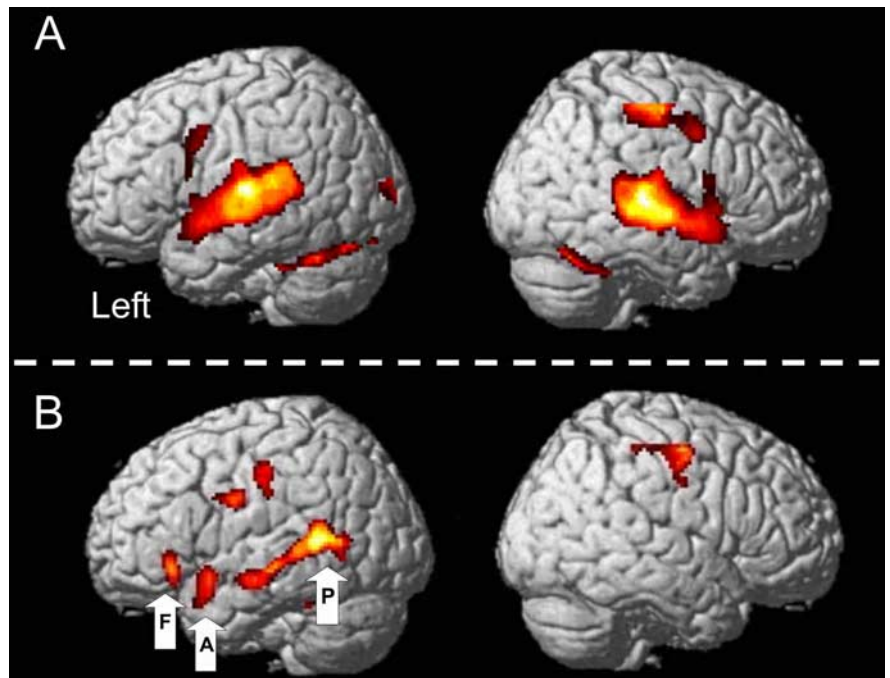*Main effects of all auditory stimuli*
BOLD responses to all auditory stimuli (contrasted with intervening periods of no auditory stimuli) were associated with significantly higher activation in the supratemporal plane bilaterally, with multiple peaks of activation in Heschel's gyrus, planum temporale, and superior temporal gyrus (STG) anterior and lateral to Heschel's gyrus (Fig. 1A, Table 1). The anatomical labels were assigned with reference to two studies on the boundaries of primary auditory cortex (Penhune et al., 1996) and planum temporale (Westbury et al., 1999). The auditory stimuli were always followed by the visual stimuli that prompted a left-handed button press; areas associated with this visual input and motor response were also activated: bilateral parafoveal visual cortex, right primary sensorimotor cortex, and bilateral rostral, paravermal cerebellum (lobule VI).

Main effect of intelligible relative to time-reversed speech stimuli
BOLD responses to intelligible speech were greater than for their time-reversed counterparts along the length of the STS on the left, and the anterior and inferior part of the left IFG, the pars orbitalis (POrb) (Fig. 1B, Table 1 for coordinates). There were several peaks of activation along the STS, including our two regions of interest in posterior and anterior STS.

DCM analysis of intelligible relative to time-reversed speech stimuli
We constructed an exhaustive set of 216 models using different combinations of directed and context-sensitive connections among the three areas. When comparing these models, across all subjects, with Bayesian model comparison, the optimal model

Figure 1. Results from the two SPM analyses. *A*, Main effects of all auditory stimuli; *B*, main effect of intelligible − unintelligible stimuli (intelligibility contrast). Results from analysis *A* show bilateral activation of Heschel's gyrus, planum temporal, and STG as well as cerebellar, visual, and right motor areas associated with making a decision on the gender of the speaker of the auditory stimuli. Results from analysis *B* show areas activated by intelligible auditory stimuli and include the length of the STS and part of the IFG on the left (POrb). Data from three of these areas (VOIs) were entered into the DCM analysis: A, aSTS; F, IFG (POrb); P, pSTS. *A* is thresholded at a voxel level of $p < 0.05$ (corrected for the search volume) and at a cluster level of 100 contiguous voxels; *B* is thresholded at a voxel level of $p = 0.01$ (uncorrected) and a cluster level of 50 contiguous voxels.

was found to be the model shown in the right panel of Figure 2. According to this model, intelligible speech modulates connections from pSTS to aSTS and, via a separate pathway, modulates connections from pSTS to the anterior part of the IFG (POrb). Furthermore, this model indicates that all auditory stimuli (whether intelligible or not) enter the system via the pSTS. Across the group, the evidence for the optimal model was considerably higher than that of the second-best model as reflected by a group Bayes factor of $\sim3.3 \times 10^5$ (this is the relative probability of the data under the two models, assuming all subjects are the same). The group Bayes factors for the top 10 models are shown in Figure 3.

In an additional analysis, we examined more closely the difference between the winning model and the second-best model; the latter was characterized by auditory input into aSTS only and modulation of the connections from aSTS to pSTS and IFG, respectively, by speech intelligibility. Because the group Bayes factor is based on a fixed effects analysis that can be distorted by outlier subjects (cf. Stephan et al., 2007c), we performed an additional analysis to exclude the possibility that the superiority of the first model was only attributable to such outliers. For this purpose, we used a novel random effects analysis method for group studies of DCMs (K. E. Stephan, W. D. Penny, J. Daunizeau, R. J. Moran, K. J. Friston, unpublished observations). This procedure, which is highly robust to outliers, uses a novel hierarchical model to optimize a probability density on the models themselves, given data from the entire group. It rests on a variational Bayesian approach for estimating the unknown parameters of a Dirichlet distribution that describes the probabilities of all models considered. Using this method, we found that there was an 81 versus 19% probability that our first model was

Table 1. Coordinates and associated *Z* and *p* values of the main peaks of task-related activity in both contrasts by region

| Region	MNI coordinates	*Z*	*p*
Analysis A: all auditory stimuli			
Temporal lobe			
L. HG	−44, −24, 8	7.34	<0.001
R. HG	50, −26, 6	7.52	<0.001
L. PT	−62, −32, 14	7.24	<0.001
R. PT	68, −32, 2	7.36	<0.001
L. STG	−58, 0, −8	7.01	<0.001
R. STG	56, 0, −8	7.09	<0.001
Parietal lobe			
R. SM	40, −18, 54	7.33	<0.001
Occipital lobe			
L. parafoveal V1	−4, −86, 2	5.82	<0.001
R. parafoveal V1	8, −86, 2	5.46	0.003
Cerebellum			
L. lobuleVI	−20, −48, −22	7.67	<0.001
R. lobuleVI	28, −62, −24	6.50	<0.001
Subcortical			
L. putamen	−26, −4, 8	6.36	<0.001
R. putamen	24, −6, 8	5.91	<0.001
Analysis B: forward − time-reversed stimuli			
Temporal lobe			
L. pSTS	−52, −48, 8	3.90	0.009
L. aSTS	−54, 10, −16	3.91	0.009
Frontal lobe			
L. POrb	−48, 28, −6	3.45	0.034

L, Left; R, right; HG, Heschel's gyrus; PT, planum temporale; SM, sensorimotor cortex; V1, primary visual cortex.
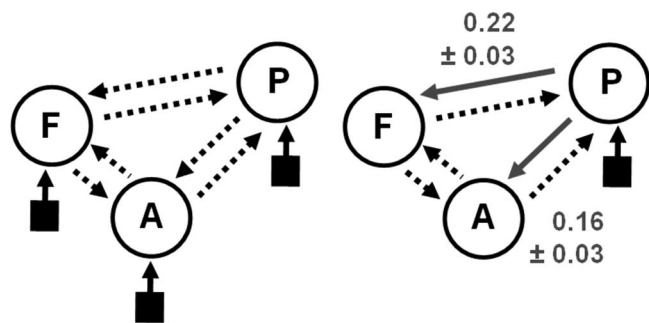
Figure 2. Results from the DCM analysis. The basic model structure for the DCM analysis is shown on the left, with possible bidirectional connections between each of the three regions shown as dotted arrows (P, pSTS; A, aSTS; F, IFG). Each model was crossed with four different inputs sites (black squares); into A only, P only, both A and P, or F only. Of the 216 models generated, the wining model is shown on the right. Connections modulated by intelligibility are shown as solid gray arrows with the average modulatory parameter estimates ±SEs over subjects shown alongside.

the more likely model compared with the second model. This corroborated the results from the initial model selection procedure based on the group Bayes factor and excluded that the first model was declared superior only because of a few outlier subjects.

Having selected the best model, we established the significance of its parameters using the DCM model parameters as subject-specific summary statistics in a conventional analysis. The results of one-sample *t* tests on the modulatory parameters from the optimal model are summarized in Figure 2. The intelligibility-related increases in coupling were $0.16 \pm 0.03\,\text{s}^{-1}$ (pSTS→aSTS; $t_{(25)} = 6.27$; $p < 10^{-6}$) and $0.22 \pm 0.03\,\text{s}^{-1}$ (pSTS → IFG; $t_{(25)} = 6.67$; $p < 10^{-7}$), respectively. These modulatory effects of speech properties on the connection strengths were strong, corresponding to an increase in connectivity of 76% from
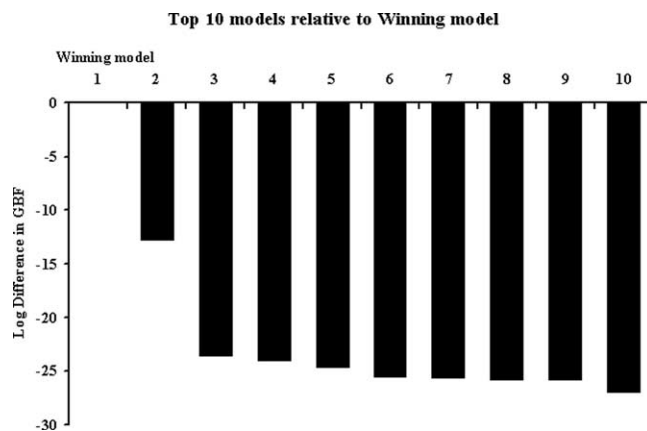
Figure 3. Group Bayes factors. Group Bayes factors for the top 10 models that best explain the BOLD data entered into the DCM compared with the winning model (of 216 in total). The winning model is on the left (rank 1, *x*-axis). The *y*-axis represents the difference in group Bayes factors between the winning model and the next nine best-fitting models. Note that, for display reasons, the differences in Bayes factors are logged.

pSTS to aSTS and 150% from pSTS to IFG during intelligible speech (relative to the intrinsic or "fixed" connection strengths). Furthermore, they were highly significant (both $p < 10^{-6}$), easily surviving Bonferroni corrections for multiple comparisons.

Note that each model of internal connections was crossed with four alternative ways that auditory inputs (both intelligible and time-reversed speech) could enter the STS. The optimal model, with the inputs into pSTS, was considerably better than the other versions of this model in which the inputs were to aSTS alone, both the pSTS and aSTS, or the IFG alone (all group Bayes factors $>10^8$).

Discussion

The primary aim of this study was to identify the direction of influences exerted among left temporal and frontal areas activated by listening to intelligible speech. We focused on three key regions that were more responsive to hearing intelligible speech: pSTS, aSTS, and the anterior portion of the IFG, which we refer to as POrb. Having identified these areas, we then used a DCM analysis to exhaustively test all mathematical possible models of connectivity among these regions. We found that the optimal model exhibited a "forward" architecture, in the sense that auditory inputs entered the system exclusively via the most posterior area in the model, pSTS, and our experimental manipulation (i.e., intelligible vs time-reversed speech) only modulated the forward connections from the pSTS to both the aSTS and the IFG.

The areas that were activated by the intelligible speech compared with the time-reversed stimuli were strongly left lateralized and were in "higher-order" or "multimodal" sensory cortex within the STS; that is, in brain regions activated by meaningful stimuli of both a linguistic and nonlinguistic nature [e.g., faces, biological motion (Puce and Perrett, 2003)]. This region is considered "heteromodal" in terms of language, because it is involved in processing both written and spoken inputs (Marinkovic et al., 2003; Spitsyna et al., 2006). Clearly this structure is important for processing meaning, and, although several current models stress its role in language processing linking posterior and anterior regions of the dominant temporal lobe (Scott et al., 2000; Spitsyna et al., 2006; Obleser et al., 2007), our results demonstrate for the first time that the processing stream for intelligible speech runs from posterior to anterior STS. This conclusion is drawn from the results of our model comparisons. Because each model

of interregional connections was crossed with three possible ways that the auditory stimuli (both meaningful and time-reversed stimuli) could enter the network, we were able to test where speech sounds are most likely to enter the system. As indicated by a group Bayes factor of 3.3×10^5, there was strong evidence (across the group) in favor of the pSTS as opposed to the aSTS alone, both pSTS and aSTS, or the IFG alone.

An important question is whether the architecture of the model identified as optimal enables one to infer the type of processing that occurs in the model system. For example, a dominant theme in many cognitive studies and in previous DCM analyses (Mechelli et al., 2005; Noppeney et al., 2006) is whether the identified architecture supports "bottom-up" (i.e., stimulus-driven) or "top-down" (i.e., driven by stimulus-unrelated variables such as task demands, cognitive set, etc.) processes. In the present study, the task demands were kept constant throughout the experiment (i.e., the subjects were asked to identify the speakers' gender), whereas the stimulus properties varied. Thus, the strong influence that speech intelligibility had on the pSTS → aSTS and pSTS → IFG connections (corresponding to a 76 and 150% increase in connection strength) represents a bottom-up, feedforward process.

Concerning the general structure of our models, it should be noted that, in models of effective connectivity such as DCM, the presence of a connection, or direct input, does not necessarily imply the existence of a direct (i.e., monosynaptic) anatomical connection; instead, a connection represents a causal influence, which can be mediated vicariously via intermediate regions not included in the model (e.g., local interconnected subregions). Both the pSTS and aSTS are connected to primary auditory cortex (Heschel's gyrus or A1) via polysynaptic pathways that spread out from A1 to secondary auditory cortex. In the primate literature, this progression is from the core auditory regions (A1) to surrounding lateral "belt" and thence to lateral "parabelt" regions in STG (Kaas and Hackett, 2000). This serial processing model has been recapitulated recently in man using a DCM analysis on fMRI data collected while subjects listened to complex (nonspeech) sounds, in which the spectral envelope was altered. The authors found that the optimum model for their data were serial or hierarchical, linking A1 to STS via the STG (Kumar et al., 2007).

The anatomical evidence for the structural pathways linking pSTS to aSTS comes primarily from studies in the 1970s and 1980 using anterograde and retrograde tracer techniques in rhesus monkeys and electrophysiological studies in macaques. These studies demonstrate that the upper bank of the STS receives inputs from multiple sensory cortical regions (visual, auditory, and somatosensory inputs: area TPO), whereas the lower bank is unimodal, receiving inputs from visual cortex only (area TEa) (Seltzer and Pandya, 1978; Baylis et al., 1987). The upper bank of rhesus STS has been subdivided into four regions of increasing cellular and laminar differentiation from anterior to posterior. These areas are reciprocally connected with both forward and backward connections, the former usually originate in supragranular layers of cortex and terminate in layer IV of the more anterior region; the longer, intermediate-range fibers connecting posterior and anterior regions of the upper bank of the STS run in the middle longitudinal fasciculus (Schmahmann and Pandya, 2006). There are at least two possible routes by which the human temporal lobe may be connected with the IFG, via either the uncinate fasciculus or a separate projection that passes through the extreme capsule (Seltzer and Pandya, 1989; Catani et al., 2005; Anwander et al., 2007).

Our finding that intelligible speech preferentially activates pSTS and this drives additional processing in the aSTS is in partial support of Hickok and Poeppel's model of the ventral processing steam; although their model does not have a direct connection between pSTS and aSTS, rather these areas influence each other through an intermediate region in inferolateral temporal cortex (Hickok and Poeppel, 2007), not identified in our study. In contrast to our result, Scott et al. (2000) found that the aSTS was more responsive than the pSTS to intelligible speech, with the pSTS being more involved in the processing of complex acoustic sounds. Although time-reversed speech has phonetic elements, there are clearly less native phonemes in these stimuli than in the non-time-reversed stimuli; thus, the pSTS region may be supporting both phonemic and semantic processing. Hickok and Poeppel have regions mediating phonological processing abutting those concerned with lexico-semantics at the junction of the pSTS/MTG, and there is evidence from other studies that parts of Wernicke's area, particularly in and around the pSTS, may have dissociable functions (Wise et al., 2001). Our results are consistent with cognitive models that predict complex auditory processing (in pSTS) precedes the recognition of intelligible speech (in aSTS).

The frontal region activated by intelligible speech in our study was in the most anterior and ventral part of the IFG, the POrb, equivalent to Brodmann area 47 (BA47). The POrb is distinct from Broca's area, which, immediately posterior, takes up the rest of the IFG (BA 45 and 44 or pars triangularis and pars opercularis, respectively) (Standring, 2004). The cytoarchitectonics of these regions shows that the POrb is more granular than the subdivisions in Broca's area (Bailey and Bonim, 1951) and has different projections from the mediodorsal nucleus of the thalamus (Fuster, 1997). The POrb is associated with semantic processing of both auditory and visually presented words (Poldrack et al., 1999; Vigneau et al., 2006), whereas more dorsal and posterior regions in Broca's area are involved in speech production. A recent study found that activation within left POrb, prompted by a semantic priming task, interacted with working memory demands such that there was more semantic task-related activity when the working memory component was low (Sabb et al., 2007). Our result is consistent with this because POrb activation was probably implicitly driven by the high semantic content of our intelligible speech stimuli compared with their time-reversed counterparts.

In summary, using a DCM analysis, we report for the first time how directed connection strengths among key regions within the dominant temporofrontal cortex are modulated by intelligible speech. A forward model emerges, in which auditory speech inputs drive activity in the pSTS; in turn, this activity influences activity in the more anterior areas, the aSTS and POrb, and the degree of this influence changes depending on which speech stimuli are being processed (intelligible vs unintelligible speech). This type of analysis, which combines dynamic system models and Bayesian model selection, is likely to prove fruitful in additional studies on normal subjects. For example, it would be of interest to perform similar analyses to manipulate task demands or expectations to investigate top-down influences on semantic processing. Also, studies of patients with auditory speech processing disorders acquired as a result of stroke could investigate how functional connectivity is reorganized after damage.

References

Anwander A, Tittgemeyer M, von Cramon DY, Friederici AD, Knösche TR (2007) Connectivity-based parcellation of Broca's area. Cereb Cortex 17:816–825.

Ashburner J, Friston KJ (2005) Unified segmentation. Neuroimage 26:839–851.

Bailey P, Bonim GV (1951) The isocortex of man. Urbana: University of Illinois.

Baylis GC, Rolls ET, Leonard CM (1987) Functional subdivisions of the temporal lobe neocortex. J Neurosci 7:330–342.

Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET (2000) Human temporal lobe activation by speech and non-speech sounds. Cereb Cortex 10:512–528.

Bitan T, Booth JR, Choy J, Burman DD, Gitelman DR, Mesulam MM (2005) Shifts of effective connectivity within a language network during rhyming and spelling. J Neurosci 25:5397–5403.

Boersma P (2001) Praat, a system for doing phonetics by computer. Glot Int 5:341–345.

Booth JR, Mehdiratta N, Burman DD, Bitan T (2008) Developmental increases in effective connectivity to brain regions involved in phonological processing during tasks with orthographic demands. Brain Res 1189:78–89.

Catani M, Jones DK, ffytche DH (2005) Perisylvian language networks of the human brain. Ann Neurol 57:8–16.

Chow HM, Kaup B, Raabe M, Greenlee MW (2008) Evidence of fronto-temporal interactions for strategic inference processes during language comprehension. Neuroimage 40:940–954.

David O, Kiebel SJ, Harrison LM, Mattout J, Kilner JM, Friston KJ (2006) Dynamic causal modeling of evoked responses in EEG and MEG. Neuroimage 30:1255–1272.

Friston KJ, Holmes AP, Worsley KJ, Poline JB, Frith CD, Frackowiak RSJ (1995) Statistical parametric maps in functional imaging: a general linear approach. Hum Brain Mapp 2:189–210.

Friston KJ, Mechelli A, Turner R, Price CJ (2000) Nonlinear responses in fMRI: the balloon model, Volterra kernels, and other hemodynamics. Neuroimage 12:466–477.

Friston KJ, Glaser DE, Henson RN, Kiebel S, Phillips C, Ashburner J (2002) Classical and Bayesian inference in neuroimaging: applications. Neuroimage 16:484–512.

Friston KJ, Harrison L, Penny W (2003) Dynamic causal modelling. Neuroimage 19:1273–1302.

Friston K, Mattout J, Trujillo-Barreto N, Ashburner J, Penny W (2007) Variational free energy and the Laplace approximation. Neuroimage 34:220–234.

Fuster JM (1997) The prefrontal cortex: anatomy, physiology, and neuropsychology of the frontal lobe, Ed 3. Philadelphia: Lippincott-Raven.

Heim S, Eickhoff SB, Ischebeck AK, Friederici AD, Stephan KE, Amunts K (2007) Effective connectivity of the left BA 44, BA 45, and inferior temporal gyrus during lexical and phonological decisions identified with DCM. Hum Brain Mapp. Advance online publication. Retrieved November 19, 2008. doi:10.1002/hbm.20512

Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev Neurosci 8:393–402.

Kaas JH, Hackett TA (2000) Subdivisions of auditory cortex and processing streams in primates. Proc Natl Acad Sci U S A 97:11793–11799.

Kumar S, Stephan KE, Warren JD, Friston KJ, Griffiths TD (2007) Hierarchical processing of auditory objects in humans. PLoS Comput Biol 3:e100.

Marinkovic K, Dhond RP, Dale AM, Glessner M, Carr V, Halgren E (2003) Spatiotemporal dynamics of modality-specific and supramodal word processing. Neuron 38:487–497.

Mechelli A, Crinion JT, Long S, Friston KJ, Lambon Ralph MA, Patterson K, McClelland JL, Price CJ (2005) Dissociating reading processes on the basis of neuronal interactions. J Cogn Neurosci 17:1753–1765.

Mechelli A, Allen P, Amaro E Jr, Fu CH, Williams SC, Brammer MJ, Johns LC, McGuire PK (2007) Misattribution of speech and impaired connectivity in patients with auditory verbal hallucinations. Hum Brain Mapp 28:1213–1222.

Meyer M, Steinhauer K, Alter K, Friederici AD, von Cramon DY (2004) Brain activity varies with modulation of dynamic pitch variance in sentence melody. Brain Lang 89:277–289.

Moran RJ, Kiebel SJ, Stephan KE, Reilly RB, Daunizeau J, Friston KJ (2007) A neural mass model of spectral responses in electrophysiology. Neuroimage 37:706–720.

Noppeney U, Price CJ, Penny WD, Friston KJ (2006) Two distinct neural mechanisms for category-selective responses. Cereb Cortex 16:437–445.

Obleser J, Wise RJ, Alex Dresner M, Scott SK (2007) Functional integration across brain regions improves speech perception under adverse listening conditions. J Neurosci 27:2283–2289.

Pandya DN, Yeterian EH (1985) Architecture and connections of cortical association areas. In: Association and auditory cortices (Peters A, Jones EG, eds), pp 3–61. New York: Plenum.

Penhune VB, Zatorre RJ, MacDonald JD, Evans AC (1996) Interhemispheric anatomical differences in human primary auditory cortex: probabilistic mapping and volume measurement from magnetic resonance scans. Cereb Cortex 6:661–672.

Penny WD, Stephan KE, Mechelli A, Friston KJ (2004) Comparing dynamic causal models. Neuroimage 22:1157–1172.

Poldrack RA, Wagner AD, Prull MW, Desmond JE, Glover GH, Gabrieli JD (1999) Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. Neuroimage 10:15–35.

Puce A, Perrett D (2003) Electrophysiology and brain imaging of biological motion. Philos Trans R Soc Lond B Biol Sci 358:435–445.

Raftery AE (1995) Bayesian model selection in social research. In: Sociological methodology (Marsden PV, ed), pp 111–196. Cambridge, MA: Blackwell.

Sabb FW, Bilder RM, Chou M, Bookheimer SY (2007) Working memory effects on semantic processing: priming differences in pars orbitalis. Neuroimage 37:311–322.

Schmahmann JD, Pandya DN (2006) Fibre pathways of the brain. New York: Oxford UP.

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.

Seltzer B, Pandya DN (1978) Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. Brain Res 149:1–24.

Seltzer B, Pandya DN (1989) Frontal lobe connections of the superior temporal sulcus in the rhesus monkey. J Comp Neurol 281:97–113.

Sherman SM, Guillery RW (1998) On the actions that one nerve cell can have on another: distinguishing "drivers" from "modulators." Proc Natl Acad Sci U S A 95:7121–7126.

Sonty SP, Mesulam MM, Weintraub S, Johnson NA, Parrish TB, Gitelman DR (2007) Altered effective connectivity within the language network in primary progressive aphasia. J Neurosci 27:1334–1345.

Spitsyna G, Warren JE, Scott SK, Turkheimer FE, Wise RJ (2006) Converging language streams in the human temporal lobe. J Neurosci 26:7328–7336.

Standring S, ed (2004) Gray's anatomy, Ed 39. London: Elsevier.

Stephan KE, Harrison LM, Kiebel SJ, David O, Penny WD, Friston KJ (2007a) Dynamic causal models of neural system dynamics:current state and future extensions. J Biosci 32:129–144.

Stephan KE, Weiskopf N, Drysdale PM, Robinson PA, Friston KJ (2007b) Comparing hemodynamic models with DCM. Neuroimage 38:387–401.

Stephan KE, Marshall JC, Penny WD, Friston KJ, Fink GR (2007c) Interhemispheric integration of visual processing during task-driven lateralization. J Neurosci 27:3512–3522.

Thierry G, Price CJ (2006) Dissociating verbal and nonverbal conceptual processing in the human brain. J Cogn Neurosci 18:1018–1028.

Thomsen T, Rimol LM, Ersland L, Hugdahl K (2004) Dichotic listening reveals functional specificity in prefrontal cortex: an fMRI study. Neuroimage 21:211–218.

Vigneau M, Beaucousin V, Hervé PY, Duffau H, Crivello F, Houdé O, Mazoyer B, Tzourio-Mazoyer N (2006) Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. Neuroimage 30:1414–1432.

Westbury CF, Zatorre RJ, Evans AC (1999) Quantifying variability in the planum temporale: a probability map. Cereb Cortex 9:392–405.

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

Wu J, Cai C, Kochiyama T, Osaka K (2007) Function segregation in the left inferior frontal gyrus: a listening functional magnetic resonance imaging study. Neuroreport 18:127–131.

Yokoyama S, Okamoto H, Miyamoto T, Yoshimoto K, Kim J, Iwata K, Jeong H, Uchida S, Ikuta N, Sassa Y, Nakamura W, Horie K, Sato S, Kawashima R (2006) Cortical activation in the processing of passive sentences in L1 and L2: an fMRI study. Neuroimage 30:570–579.