Parallel Systems in the Control of Speech

Anna J. Simmonds,^{1,*} Richard J.S. Wise,¹ Catherine Collins,¹ Ozlem Redjep,¹
David J. Sharp,¹ Paul Iverson,² and Robert Leech¹

¹Computational, Cognitive and Clinical Neuroimaging Laboratory (C3NL), Division of Brain Sciences, Imperial College London, United Kingdom

²Department of Speech, Hearing and Phonetic Sciences, Division of Psychology and Language Sciences, University College London, United Kingdom

◆=====◆
Abstract: Modern neuroimaging techniques have advanced our understanding of the distributed anatomy of speech production, beyond that inferred from clinico-pathological correlations. However, much remains unknown about functional interactions between anatomically distinct components of this speech production network. One reason for this is the need to separate spatially overlapping neural signals supporting diverse cortical functions. We took three separate human functional magnetic resonance imaging (fMRI) datasets (two speech production, one “rest”). In each we decomposed the neural activity within the left posterior perisylvian speech region into discrete components. This decomposition robustly identified two overlapping spatio-temporal components, one centered on the left posterior superior temporal gyrus (pSTG), the other on the adjacent ventral anterior parietal lobe (vAPL). The pSTG was functionally connected with bilateral superior temporal and inferior frontal regions, whereas the vAPL was connected with other parietal regions, lateral and medial. Surprisingly, the components displayed spatial anti-correlation, in which the negative functional connectivity of each component overlapped with the other component’s positive functional connectivity, suggesting that these two systems operate separately and possibly in competition. The speech tasks reliably modulated activity in both pSTG and vAPL suggesting they are involved in speech production, but their activity patterns dissociate in response to different speech demands. These components were also identified in subjects at “rest” and not engaged in overt speech production. These findings indicate that the neural architecture underlying speech production involves parallel distinct components that converge within posterior peri-sylvian cortex, explaining, in part, why this region is so important for speech production. *Hum Brain Mapp* 35:1930–1943, 2014. © 2013 Wiley Periodicals, Inc.

Key words: fMRI; ICA; connectivity; speech production; parietal operculum; planum temporale; perisylvian

◆=====◆

Additional Supporting Information may be found in the online version of this article.

Contract grant sponsor: Medical Research Council.

*Correspondence to: Anna Simmonds, Computational, Cognitive and Clinical Neuroimaging Laboratory (C3NL), Centre for Restorative Neurosciences, Division of Brain Sciences, Imperial College London, 3rd Floor, Burlington Danes Building, Hammersmith

Hospital, Du Cane Road, London, W12 0NN, UK. E-mail: anna.simmonds08@imperial.ac.uk

Received for publication 6 November 2012; Revised 24 January 2013; Accepted 19 March 2013.

DOI: 10.1002/hbm.22303

Published online 30 May 2013 in Wiley Online Library (wileyonlinelibrary.com).

INTRODUCTION

Spontaneous speech production depends on multiple processing stages that support intention through to articulation [Levelt, 1989]. The many intermediate stages encompass the retrieval, and ultimately the utterance, of word assemblies that convey appropriate meaning (semantics and syntax) and their correct pronunciation (phonology). It is unlikely that there is a close one-to-one correspondence between individual brain regions and the many processes established by psycholinguistic research. However, neurolinguistic meta-analyses of functional imaging data have attempted to relate individual cortical regions with specific core language processes, including those involved in speech production [Hickok and Poeppel, 2007; Indefrey and Levelt, 2004; Vigneau et al., 2006].

A closer anatomical-functional correspondence is becoming established for the final stage of speech, namely articulation accompanied by sensory (auditory and somatosensory) feedback. A left posterior perisylvian region (specifically planum temporale, extending up into the parietal operculum in some cases), area Spt, has been identified as a sensory-motor interface, critical for speech production, supporting the transformation of the encoded auditory form of a word into its motor form [Buchsbaum et al., 2005; Hickok et al., 2003; Hickok et al., 2009; Hickok et al., 2011; Pa and Hickok, 2008]. Area Spt has been defined using a series of covert speech production studies and is proposed to be active in the absence of actual motor or sensory activity, although claims about this area are not universally accepted.

In many imaging studies the default assumption is that the change in neural activation measured with functional MRI in a given region reflects a single underlying neural signal. By implication, discrete regions in the speech network can be given specific functional roles: e.g., perisylvian temporal and parietal regions may be involved in integrating sensory and motor information for speech production [Golfingopoulos et al., 2010; Guenther et al., 2006; Ventura et al., 2009]. However, the spatial resolution of fMRI means that the signal within any voxel reflects, at a minimum, the net activity of many tens of thousands of synapses on many thousands of neurons. An alternative view questions whether the measured neural activation evoked by speech within a given region reflects the summation of multiple distinct components that carry different functional information [Leech et al., 2012; Smith et al., 2012]. The most appropriate functional description may, therefore, not be at the level of brain regions but in terms of these components, which may span multiple regions or may overlap with other components. Importantly, at certain points in the brain, different components will converge and interact.

The aim of this study was to investigate the distinct functional components related to speech production that converge within left posterior perisylvian ‘language’ cortex, using native and non-native languages, as well as

comparing between overt and covert speech production. We investigated the hypothesis that activation within a region can be separated into distinct spatio-temporal components that make functionally different contributions to speech production, depending on the type of speech required, through the use of spatially constrained independent component analysis (ICA). Figure 1A demonstrates the inclusive boundary for the chosen region, one that encompasses posterior temporal and adjacent inferior parietal cortices and covering the left posterior perisylvian language area. This approach allows the theoretically motivated region to be partitioned into different spatially adjacent and overlapping components which can be shown to have different functional roles. This allows us to explore the functional organization of the perisylvian “language” cortex and understand how different types of information (e.g., somatosensory and auditory components) converge there.

We used fMRI data from an overt speech production study, requiring subjects to employ familiar and novel articulatory movements, to first define different posterior perisylvian components using spatially constrained independent components analysis. We then explored how these components were functionally connected with the rest of the brain and how they were modulated by the different speech conditions. We next investigated if these components could be replicated in independent datasets with different subjects: a second speech production dataset and a “rest” dataset. The rest fMRI dataset was used to investigate whether these components are only evoked by task, or are also present in the “resting” brain.

METHODS

Subjects

A different set of right-handed subjects was used in each of the three studies, all native speakers of English (Study 1 $n = 21$, ages 19–40 years, 10 females, all monolingual; Study 2 $n = 17$, ages 21–61 years, eight females; Study 3 $n = 16$, ages 26–58 years, eight females). Subjects had no history of neurological problems or hearing loss. The studies were approved by the local research ethics committee and all subjects gave informed written consent. The speech production datasets have not been published elsewhere.

Functional MRI Acquisition

Acquisition parameters were the same in each of the three datasets, except for the number of TRs and the use of sparse sampling. All MRI data were obtained using the same Philips Intera 3.0 Tesla scanner, using dual gradients, a phased array head coil, and sensitivity encoding with an undersampling factor of 2. Functional MR images were obtained using a T₂-weighted, gradient-echo, echoplanar

Figure 1.

A: The mask used for the original ICA. **B:** There were two components of theoretical interest in speech production. The superior temporal (pSTG) component is shown in purple and the ventral anterior parietal (vAPL) component in pink. The colored overlays are displayed on sagittal ($x = -45$ mm, left image), coronal ($y = -22$ mm, center image) and axial ($z = 12$ mm, right image) slices taken from a standard MNI brain template. **C:** The two components of theoretical interest in speech

production, as identified by ICAs with different dimensionalities. The temporal component is shown in purple and the parietal component in pink. A shows the two components from an ICA of 7 components (MNI: $x = -41$, $y = -22$, $z = 12$) and B shows them from an ICA of 15 components (MNI: $x = -39$, $y = -19$, $z = 8$). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

imaging (EPI) sequence with whole-brain coverage (TR = 2 s, echo time = 30 ms; flip angle = 90°). Thirty-two axial slices with a slice thickness of 3.25 mm and an interslice gap of 0.75 mm were acquired in ascending order (resolution, $2.19 \times 2.19 \times 4.00$ mm³; field of view, $280 \times 224 \times 128$ mm³). Quadratic shim gradients were used to correct for magnetic field inhomogeneities within the brain. High-resolution, T1-weighted images were also acquired for structural reference. Study 1 included three runs, each of 105 TRs. Study 2 used 120 TRs in a single run. Study 3 contained 300 TRs, again in a single run. For Studies 1 and

2 sparse acquisition was performed to minimize movement- and respiratory-related artifacts associated with speech studies [Hall et al., 1999], as well as to minimize auditory masking (this involved two seconds of acquisition followed by six (Study 1) or eight (Study 2) seconds of silence during which subjects spoke). Paradigms were programmed using E-Prime software (Psychology Software Tools) (Study 1) and Matlab Psychophysics toolbox (Psychtoolbox-3; www.psychtoolbox.org) (Study 2), and stimuli presented through an IFIS-SA system (In Vivo Corporation). Sounds were delivered through MR-compatible

headphones and speech was recorded using a fiber-optic noise-canceling microphone.

Functional MRI Experimental Procedures

Study 1

In the fMRI scanning session subjects listened to and repeated single native (English) nonwords and non-native words (from Mandarin, Spanish and German). Rest was included as the baseline condition. The non-native stimuli were specifically chosen to manipulate a different place and manner of articulation in each language (by manipulating either pitch, vowel sounds or consonants). The words were real bi-syllabic words with a consonant-vowel-consonant-vowel (CVCV) structure and were matched for number of phonemes, with the target phoneme and stress on the first syllable and the rest of the word easy to produce for native English speakers. The native stimuli consisted of bi-syllabic nonwords, also with a CVCV structure and matched for number of phonemes. Nonwords were used in order to avoid the confound of lexical processing, which would have been elicited by real native words. The native condition was included to investigate motoric production of familiar articulatory movements, in contrast to the unfamiliar articulatory movements involved in non-native speech. A list of stimuli is presented in Supporting Information Table S1. During speech trials visual cues indicated listen and repeat trials. In each of the three runs, there were 20 repeat trials for each of the non-native language groups, ten native repeat trials, and 15 rest trials. The order of these trials was pseudo-randomized and therefore different across runs. Audio recordings were taken to ascertain whether subjects were performing the task correctly. Speech trials in which the subjects failed to respond were excluded from the fMRI analyses. Also included in this paradigm were listening trials, which are not included in the analyses here (6 listen trials for each non-native language group and 2 native listen trials per run).

Study 2

There were four experimental conditions in a factorial design: overt and covert propositional and non-propositional speech, and a rest baseline. Propositional speech tasks required subjects to describe high frequency concrete nouns with high imaginability, selected from the MRC psycholinguistic database [Wilson, 1988]. A list of stimuli is presented in Supporting Information Table S2. There were no significant differences ($P > 0.2$) between the mean values of the two word lists (overt and covert) for concreteness, imaginability or frequency. Subjects were cued with a visually presented word to describe. Nonpropositional speech was tested with a counting task, during which subjects were required to count upward from one at a rate of approximately one number per second. This condition was

cued with the visually presented word, "Count". There were 25 trials for each of the experimental conditions and 20 rest trials. All tasks were preceded with an image that indicated whether the following task was to be performed overtly or covertly. Audio recordings were taken to ascertain whether subjects were performing the task correctly. Trials in which the subjects responded incorrectly were excluded from analyses.

Study 3

Subjects lay in the scanner with their eyes closed without exposure to stimuli and with no explicit task.

Data Analysis

Data preprocessing

For each of the three datasets, preprocessing included non-brain removal using BET (Brain Extraction Tool) [Smith, 2002]; spatial smoothing using a Gaussian kernel of FWHM 5mm; grand-mean intensity normalization of the entire 4D dataset by a single multiplicative factor; highpass temporal filtering (Gaussian-weighted least-squares straight line fitting, with $\sigma = 50.0s$) and motion correction using (MCFLIRT) [Jenkinson et al., 2002]. FMRI data were analyzed using voxel-wise time series analysis and the General Linear Model (GLM). Registration to high-resolution structural and standard space images was carried out using FLIRT (FMRIB's Linear Image Registration Tool) [Andersson et al., 2007; Jenkinson and Smith, 2001].

To remove non-neural noise, variance associated with motion (six variables), and the time courses of white matter and cerebro-spinal fluid (CSF) were removed from the whole brain functional data using ordinary least squares linear regression. To calculate the time course for the white matter and CSF, a 3-mm-radius sphere was created based on the coordinates MNI, -26, -22, 28, and MNI, 2, 10, 8 respectively, and the mean timecourse across the sphere calculated. Excluding variance from motion, white matter and CSF removes fluctuations in the timecourses that are not considered to be involved in specific regional correlations, thereby controlling for effects of physiological processes and is standard practice in many functional connectivity analyses [e.g., Fox et al., 2005].

Decomposing the left perisylvian language area into discrete components

The methods for the independent component analysis steps are summarized in Figure 2. In fMRI studies, particularly so for speech production tasks, noise is added to the data, both from head motion and from physiological fluctuations. ICA is able to separate components within functional brain networks from this noise. In addition, ICA reveals distinct networks originating from

Figure 2.

Methods used for finding components within the left perisylvian language area and finding whole-brain functional connectivity with these components.

anatomically adjacent or overlapping regions and is more sensitive than a standard subtraction analysis. We have previously used ICA on fMRI data collected with sparse-sampling [Geranmayeh et al., 2012] and the assumptions for ICA work for both continuous and sparse data. The temporo-parietal region was defined using the Harvard-Oxford probabilistic atlas within fslview, for the structures planum temporale, parietal operculum, Heschl's gyrus, middle temporal gyrus (temporo-occipital part) and superior temporal gyrus (posterior division) (see Fig. 1A for the mask). A temporal concatenation group ICA [Beckmann et al., 2005] was then run on the speech task data from Study 1 within this mask [Leech et al., 2012]. This approach produced spatial maps within the left temporo-parietal junction, identifying voxels that co-vary together. Each map corresponds to a different spatio-temporal pattern of neural activity or source of physiological or scanner noise, although these patterns can be partially spatially and temporally overlapping [Beckmann et al., 2005]. The timecourse for each spatial map was calculated by using a multiple regression model with the 4D fMRI data as the dependent variable and the ten spatial maps from the ICA as the independent variables. These timecourses represent the BOLD timecourse of that spatial component, similar to a timecourse extracted from a region of interest analysis,

except controlling for some of the variance explained by the other spatial maps. This technique identified ten independent **components** from within the perisylvian region for each subject. This was repeated with 7 and 15 components with qualitatively similar results, following [Leech et al., 2012]. The first speech production data was used for defining the different posterior perisylvian **components** as it was the largest dataset, both in number of subjects and in number of trials, allowing better modeling of within- and between-subject variability. However, Supporting Information Figure S1 shows the regions as identified from Studies 2 and 3, to demonstrate that similar components are revealed from separate datasets.

A second GLM was used, this time using whole-brain fMRI data. This enabled us to calculate correlations between each of these components with activity across the whole brain. The timecourses of the components calculated previously were simultaneously included in the design matrix to generate a set of whole-brain statistical maps. This approach is a variant on the dual regression approach [Zuo et al., 2010] and has now been used on a number of published datasets [Bonnelle et al., 2011; Leech et al., 2011; Sharp et al., 2011]. The statistical maps calculated by this analysis provided a whole-brain voxelwise measure of functional connectivity with each of the

components, controlling for variance from the other components. The timecourses were entered into a higher-level general linear model [Beckmann et al., 2003] to compute non-parametric statistics (using random permutation testing), showing whole-brain functional connectivity. Statistics were corrected for multiple comparisons using family-wise error (FWE) cluster corrections using a nominal t -value of 2.3. To understand the function of the separate components, the individual timecourses generated for each run for each subject were compared with the experimental timecourse (e.g., when specific speech sounds were produced). To do this, the individual timecourse for each component was entered as a dependent variable into a general linear model with the experimental timecourse as the design matrix. This was done separately for the two different speech datasets. This resulted in a beta value, quantifying how much each component's timecourse was modulated by the different task conditions. This measure of BOLD signal change with task was used in subsequent region of interest analyses.

Components were classed as artifacts and excluded from further analysis if the majority of voxels in the whole-brain functional connectivity patterns were in white matter, ventricles or outside the brain [Leech et al., 2012; Smith et al., 2009], which led to the exclusion of seven of the components. The need for excluding so many components is expected due to the nature of the dataset, which involved overt speech production tasks. Overt speaking results in motion-induced signal changes that confound the measured activity and lead to artifacts, as do individual differences in head size and physiological noise. One advantage of ICA over other approaches is that it explicitly models these confounds and allows for identification of components related to non-neural noise.

RESULTS

Study I

Twenty-one subjects were scanned with fMRI while performing a speech production task, producing bisyllabic native nonwords or non-native (Spanish, German and Mandarin) words that differed in place and manner of articulation (see Methods). We decomposed the fMRI data into ten components within the left temporo-parietal cortex using independent component analysis [Beckmann et al., 2005; Leech et al., 2012]. From those ten components, three were considered to be neural signal and not noise. Two of these components were defined as being of theoretical interest in speech production, replicating regions observed previously [Dhanjal et al., 2008; Simmonds et al., 2011]. One region was located primarily in superior temporal auditory cortex, including planum temporale (pSTG), and the other was predominantly within the ventral anterior parietal lobe (vAPL), including second-order somatosensory cortex in the parietal operculum (Fig. 1B). Although these two components were centered more in

either the temporal or parietal lobe, they had considerable overlap. The ICA was repeated with 7 and 15 components with qualitatively similar results (Fig. 1C). A third component displayed a pattern of functional connectivity consistent with the well-characterized default mode network and is discussed further in a later section.

To characterize these components, we investigated their functional connectivity with activity across the rest of the brain. We calculated subject-specific time courses for each component, and entered these simultaneously into a general linear model with the whole brain fMRI dataset [Beckmann et al., 2003; Leech et al., 2012]. This resulted in a functional connectivity statistical map for each component. One considerable benefit of this approach is the use of multiple regression, both in defining the subject specific time courses and the resultant statistical maps, thereby isolating functional connectivity specific to that component from competing components and non-neural noise [Zuo et al., 2010].

Although the regions associated with the superior temporal and ventral parietal components were spatially overlapping, the resulting whole-brain functional connectivity maps were spatially anti-correlated (Fig. 3). The pSTG component was positively correlated with activity in the superior temporal gyri, regions within the inferior parietal lobes (predominantly dorsal, extending up to the intraparietal sulci), and along most of the length of the inferior frontal gyri. The bilateral distribution appeared largely symmetrical across the cerebral hemispheres (Fig. 3A). Figure 3A shows some asymmetry in the frontal regions, with greater correlation in the right than the left, but this is dependent on the threshold used and activity in these frontal regions is more symmetrically distributed at a lower threshold. The parietal component positively correlated with the postcentral gyri, the posterior half of insular cortex, posterior inferior parietal cortex (predominantly the angular gyri), and midline posterior cortex, where posterior cingulate cortex, anterior precuneus and retrosplenial cortex lie in close spatial relationship to one another. Again, this bilateral component appeared largely symmetrical (Fig. 3B), although in the figure it appears stronger on the left dorsal central region than the right, for the thresholding reasons mentioned above. The negative functional connectivity results showed an anti-correlated spatial pattern: the temporal component with bilateral parietal opercular and postcentral cortex, anterior midline cortex, encompassing the supplementary motor area and anterior cingulate gyrus, and the cerebellum; and the parietal component with the superior temporal sulci, extending up into the inferior frontal gyri, inferior parietal cortex (predominantly dorsal, extending up to the intraparietal sulci) and the pre-supplementary motor area. Once again, the bilateral anti-correlated components appeared largely symmetrical.

The components can also be characterized in terms of how they were modulated by different task demands. There was increased activity for all speech production tasks relative to rest for both components, indicating their

Figure 3.

Whole-brain functional connectivity maps for the two components: the superior temporal (pSTG) component (**A**); and the ventral anterior parietal (vAPL) component (**B**). Positive functional connectivity maps are shown in warm colors and negative functional connectivity (anti-correlation) maps in cold colors. For each component, the overlays are displayed on slices from a standard MNI brain template, with two sagittal slices through the left ($x = -44$ mm) and right ($x = 44$ mm)

hemispheres, and two rows of axial slices (upper row, $z = 43$ – 15 mm, in 4 mm decrements, and lower row, $z = 11$ mm to -17 mm, again in 4 mm decrements). The statistical threshold for the overlays was set at $P < 0.01$, corrected for multiple comparisons using a correction for familywise error rate. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Figure 4.

Activity for the temporal (pSTG) and parietal (vAPL) components, as identified by the independent component analysis, in response to different speech production tasks. **A:** activity during the four conditions in Study 1 (one native and three non-native

speech production tasks). **B:** activity during the four conditions in Study 2 (two overt and two covert speech production tasks). Error bars represent standard error of the mean. Brackets indicate significant differences from post-hoc *t*-tests.

involvement in speech production. However, the components were differentially modulated by the specific types of speech produced; in particular, the type and place of articulation (Fig. 4A). This was evident from a significant interaction between the language produced and the region, with highest activity in pSTG for the Mandarin tones (producing novel pitch shifts), and highest activity in the vAPL for producing novel German vowels (involving subtle lip and tongue control). A 4 (one native and three non-native languages) $\times$ 2 (region) analysis-of-variance (ANOVA) demonstrated a main effect of language ($F_{(3,60)} = 4.94$, $P < 0.005$), a main effect of region ($F_{(1,20)} = 23.123$, $P < 0.001$) and a language $\times$ region interaction ($F_{(3,60)} = 8.423$, $P < 0.001$). For the temporal component a 4 $\times$ 1 (four languages) ANOVA revealed a significant main effect of language ($F_{(3,60)} = 12.930$, $P < 0.001$). To identify if the main effect of language was driven by differences between native and non-native languages, or whether there were differences between the three separate non-native languages, a further ANOVA was performed. A 3 $\times$ 1 (three non-native languages) ANOVA also revealed a significant main effect of language ($F_{(2,40)} = 13.854$, $P < 0.001$). There were no significant language effects in similar ANOVAs

for the parietal component. Two-tailed, post-hoc paired *t*-tests revealed that the temporal component showed significantly more activity when producing Mandarin compared to Native, German and Spanish (see Table 1). There were also significant differences, although less striking, between producing Spanish compared to Native and German.

There was also a third neural component, which displayed a pattern of functional connectivity consistent with the well-characterized default mode network (midline anterior and posterior cortices and lateral inferior parietal cortices, Fig. 5A). This demonstrated a relative task-evoked deactivation for all speech sounds (Fig. 5B).

Study 2

The pSTG and vAPL also showed highly distinct patterns of activation revealed by a second speech production dataset. In contrast to the first experiment, during which single words were produced in a constrained manner, 17 subjects produced self-generated overt and covert propositional speech cued by visually presented words (see Methods) and overt and covert nonpropositional speech

TABLE 1. Results from post hoc *t*-tests for Study 1

Component	Contrast (mean)	<i>t</i>	<i>df</i>	<i>p</i>
Temporal component	Mandarin (16.79) > Native (15.21)	6.247	20	0.000004
	Mandarin (16.79) > German (15.33)	5.148	20	0.00005
	Mandarin (16.79) > Spanish (15.97)	3.124	20	0.005
	Spanish (15.97) > Native (15.21)	2.393	20	0.027
	Spanish (15.97) > German (15.33)	2.228	20	0.038

Figure 5.

A: The default mode component identified by the ICA. **B:** activity during the four conditions in Study 1 (one native and three non-native speech production tasks). There were no significant differences between any of the languages. **C:** activity during the four conditions in Study 2 (two overt and two covert speech production tasks). Task (speaking, counting) by Type (overt, covert)

interaction: main effect of type $F_{(1,16)} = 21.037$, $P < 0.001$). Post-hoc t -test, 2-tailed paired t -tests: (mean CovertSpeaking = 0.375, OvertSpeaking = -2.476 , $t(16) = -4.488$, $P < 0.001$) (mean CovertCounting = 0.525, OvertCounting = -1.952 , $t(16) = -3.666$, $P = 0.002$). Error bars represent standard error of the mean. Brackets indicate significant differences from post-hoc t -tests.

(counting). Applying the same regions as defined with the single-word production task above resulted in a similar pattern of functional connectivity for the two studies (Fig. 6ii): functional connectivity in bilateral temporal regions and inferior frontal regions were shared in the two tasks. A more detailed and quantitative comparison of spatial maps across datasets showing similarity across tasks is presented in the section "Similarities between datasets and differences between components" below and in Figure 7.

A task by region ANOVA revealed a main effect of task ($F_{(3,48)} = 15.421$, $P < 0.001$), a main effect of region ($F_{(1,16)} = 19.961$, $P < 0.001$) and a task by region interaction ($F_{(3,48)} = 14.618$, $P < 0.001$). The activity within the pSTG component was higher relative to rest for both overt speech production tasks, propositional and nonpropositional (Fig. 4B). Two-tailed paired t -tests revealed that the temporal component was significantly more active for

overt speaking than covert speaking, and more active for overt counting than covert counting (see Table 2). In contrast, the parietal component showed a relative deactivation for speaking, both overt and covert, and activated only for overt, but not covert, counting (Fig. 4B). Two-tailed paired t -tests revealed that activation during overt counting was significantly higher than during covert counting and higher than overt speaking, and covert counting was significantly higher than covert speaking.

Study 3

Although the primary focus of the analysis was on regions involved in speech production, it is possible that the components are present even in the absence of active speaking. To investigate this, the same functional

Figure 6.

Whole-brain positive functional connectivity maps for the two components, temporal (**A**) and parietal (**B**) for the three datasets, i: Study 1, ii: Study 2, iii: Study 3. Images are corrected for multiple comparisons ($P < 0.05$, FWE) and presented on a standard rendered brain template (left hemisphere shown) and with the render clipped from the top to reveal activity in more medial regions. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

connectivity analysis, with the same regions, was conducted on a third dataset of 16 subjects scanned at “rest”. The subjects lay in the scanner with their eyes closed, without exposure to stimuli and with no explicit task. The resulting pattern of functional connectivity was, again, similar to that evoked by the speech production tasks (Fig. 6iii): for the pSTG component, bilateral temporal regions and inferior frontal regions; and for the vAPL component, bilateral basal parietal and medial and lateral parietal regions. This suggests that the superior temporal and the parietal components are present in some form, and have similar whole-brain functional connectivity, even in the absence of a speech task. The components in Study 3 are far more extensive than those in Study 1 and Study 2. One possibility is that this is due to the type of acquisition for each dataset, as, unlike the first two studies that used sparse acquisition, Study 3 used continuous acquisition and therefore consisted of many more trials with a consequent increase in power. In addition, the first two studies required participants to start and stop speaking as instructed, whereas in Study 3 participants may well have been covertly producing speech throughout the scan, either by thinking to themselves or mind-wandering.

Similarities Between Datasets and Differences Between Components

To provide a quantitative comparison of the similarity in the spatial pattern across the datasets, spatial correlation [Leech et al., 2012; Smith et al., 2009] showed that the

functional connectivity maps acquired from all three datasets were highly similar (pSTG component: all three r values > 0.5 , $P < 0.0001$; vAPL component: all three r values > 0.6 , $P < 0.0001$). In addition to assessing similarity across datasets, spatial correlation provides a quantification of the spatial similarity between the patterns of functional connectivity calculated for the parietal and temporal components. This analysis revealed that for all three datasets, the pattern of functional connectivity for the temporal component was significantly spatially anti-correlated with the functional connectivity for the parietal component (all three r values < -0.15 , $P < 0.01$). This confirms the qualitative pattern of anti-correlation seen in Figure 3. As a further demonstration of the similarity across datasets and then anti-correlation across components, peaks of the original functional connectivity maps were located and the Z-statistics across all three datasets plotted (Fig. 7). These figures again show both the broad similarity across the three datasets, and that in many regions the two components were anti-correlated.

DISCUSSION

The findings from these three studies have a number of implications for understanding the neural systems underlying speech production. First, in contrast to univariate analyses of perisylvian regions, there are two discrete but overlapping components centered around the posterior end of the Sylvian fissure: one centered on the planum

A: Peaks from vAPL Component

B: Peaks from pSTG Component

Figure 7.

Qualitative similarity of the whole-brain functional connectivity patterns. Peaks of the original functional connectivity maps were located and the z-statistics across all three datasets plotted. The plots demonstrate similarity across the three datasets and anti-correlation across the two components. Panel **A** shows plots from the peaks from the parietal component and Panel **B** shows

plots from the peaks from the temporal component. Signal change arbitrary coefficients are shown for the parietal component in light bars and for the temporal component in dark bars. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

temporale and the other on the parietal operculum. Although the former region is usually considered to be auditory association cortex and the latter to be somatosensory, these regions are probably not strictly unimodal [Smiley et al., 2007]. Second, these components appear to

be functionally different, evident from both the anti-correlated pattern of functional connectivity and the manner in which they were differentially modulated by speech tasks. Third, these components were present in 'resting scan' data, with similar patterns of functional connectivity in the

TABLE 2. Results from post hoc t-tests for Study 2

Component	Contrast (mean)	<i>t</i>	<i>df</i>	<i>p</i>
Temporal component	Overt Speaking (8.62) > Convert Speaking (1.76)	3.381	16	0.004
	Overt Counting (10.34) > Covert Counting (−1.80)	6.321	16	0.00001
Parietal component	Overt Counting (4.42) > Covert Counting (−0.28)	3.209	16	0.005
	Overt Counting (4.42) > Overt Speaking (−2.25)	3.186	16	0.006
	Covert Counting (−0.28) > Covert Speaking (−3.46)	2.585	16	0.02

absence of a speech task. This suggests that speech involves modulating signals that occur even in the absence of speech. This last finding has broader implications for how the brain accomplishes complex tasks in general. We consider each of these findings in turn.

We find that the pSTG and vAPL components overlap within left perisylvian cortex in a region encompassing area Spt, localized to the posterior half of the left planum temporale [Hickok et al., 2011]. This region has been invested with a central role in speech production and non-verbal vocal tract sounds [Hickok et al., 2011; Pa and Hickok, 2008], integrating the neural code of the auditory forms of words with the neural code for their motor (articulatory) forms. These functional imaging studies have investigated a one-to-one mapping between this specific cortical region and auditory-motor transformation during speech production. In contrast, in the present studies we have demonstrated that within area Spt there are at least two systems with quite different remote functional connections, both intra- and inter-hemispheric. This work is consistent with other studies indicating the role of multiple regions in the planum temporale, with dissociations of activity in response to auditory input and speech output [Tremblay et al., 2011]. Similarly, our findings accord with other studies that have demonstrated the context-dependent response of perisylvian second-order somatosensory association cortex to overt speech [Dhanjal et al., 2008; Simmonds et al., 2011].

The demonstration of the extensive perisylvian pSTG and vAPL networks is also consistent with the clinical literature on conduction aphasia [Buchsbaum et al., 2011]; infarction of the left temporo-parietal junction results in sound-based errors during speech production, repetition and reading aloud whilst retaining normal speech comprehension. The broad lesions resulting in conduction aphasia well match the extensive functional systems observed here.

As well as differences in distribution of functional connectivity, the two left posterior perisylvian components displayed an anti-correlated spatial pattern, with the negative functional connectivity of each component substantially overlapping the other component's positive functional connectivity in a region that included area Spt. The existence of anti-correlated networks is emerging as an important concept in understanding the brain's broad functional architecture; for example, the anti-correlation between the default mode network and task-positive attentional networks [Fox et al., 2005; Leech et al., 2012; Smith et al., 2012]. The datasets presented here demonstrate that the motor-sensory control of speech involves parallel signals converging on a large region of left temporo-parietal cortex that may act in competition (in that when one is active the other is deactivated).

One interpretation of this specific example of network anti-correlation is that the two sensory modalities alternate in importance during the generation of speech, depending on the type of utterance. This result is in line with recent

behavioral work which has found that there is individual preference for either auditory or somatosensory feedback during speech [Lametti et al., 2012]. For the single word production task (Study 1), all conditions, whether producing a non-native word or native nonword, led to increased activation for both components, implying that both are involved in speech production. However, the specific requirements of the different types of non-native sounds resulted in different patterns of activation for the two components, hinting at their different functional roles. Precisely which aspects of the speech production tasks result in the different signal modulations must be speculative without further studies, which will give greater understanding of distinct somatosensory, auditory, and motor contributions.

The functional dissociation evident from the anti-correlation of the components was accompanied by the further dissociation in how the components communicate with the rest of the brain. That the component centered on the left pSTG had functional connectivity with bilateral superior temporal (auditory) cortex, and the one centered on the vAPL had functional connectivity with bilateral somatosensory cortex in the postcentral gyri, relates to the previous discussion of the sensory modality that may dominate depending on the specific utterance. However, there were additional components of both functional connectivity patterns that were located in high-order inferior frontal and parietal cortex. Their contribution to speech control, by modulating activity in temporo-parietal cortex, cannot be determined from this study, although the contribution of inferior frontal and dorsal inferior parietal cortex may plausibly relate to attention. The **component** correlating with the vAPL region included posterior inferior parietal cortex and medial posterior cortex, including posterior cingulate cortex. These regions are typically thought to be part of the default mode network [Greicius and Menon, 2004]. However, the parietal opercular **component** reported here was significantly activated by speech production, unlike the deactivation expected if it formed part of the default mode network. There was a separate **component** in an expected spatial pattern for the default mode network with relative deactivation during the speech production tasks (Fig. 5). It would seem, therefore, that speech production involves communication with a local system within medial posterior cortex. This is consistent with the growing body of work suggesting the extensive functional heterogeneity of posterior cingulate cortex [Dastjerdi et al., 2011; Hagmann et al., 2008; Leech et al., 2012; Leech et al., 2011; Margulies et al., 2009; Seghier and Price, 2012].

The second study demonstrated that the vAPL component was deactivated during sentential volitional speech, both overt and covert, but not during counting. This result is consistent with previous work reporting deactivation in secondary somatosensory cortex during propositional speech [Dhanjal et al., 2008]. The finding of increased activity in vAPL for nonword and non-native single word speech production replicates previous research on non-

native propositional speech production [Simmonds et al., 2011]. The superior temporal component had a completely different pattern of modulation by task, whereby only overt speech resulted in increased activation. This suggests that activity in posterior perisylvian cortex is not independent of the speech task. Future studies incorporating speech tasks that variably manipulate dependency on feedback from sensory, linguistic and semantic systems will further refine our knowledge about the function of posterior perisylvian cortex.

By considering what happens at “rest”, we observed that the two neural components appear to occur even in the absence of overt speech production. This result contrasts with the normally implicit, and sometimes explicit, assumption that speech is governed by a well-described neural system that changes from being in an inactive state to becoming active when producing speech. Our result fits better with a view of speech production as a special case of the normal ongoing functioning of the brain. Under this view, neural regions, such as auditory association cortex, have specific functional roles in, for example, decoding the acoustic signal. However, these speech specific mechanisms are integrated with the ongoing processing of the other neural functions the brain performs. The temporal and parietal networks we identified here, although relatively stable and observable across tasks, are modulated differently by different task situations. We argue that it is this modulation that becomes apparent as increased activation evoked by a task. However, we cannot rule out the possibility that this pattern is the result of participants’ covert volitional speech production, i.e., thinking to themselves or mind-wandering.

The focus of these analyses has been on one region of the language network, albeit one that has received considerable emphasis in understanding speech. Future studies may concentrate on other regions involved in speech, such as anterior perisylvian cortex, including Broca’s area. In addition, we have only considered speech production and rest and not speech comprehension or nonspeech tasks. It will be of interest to discover how the components reported here are affected by other tasks.

REFERENCES

- Andersson JLR, Jenkinson M, Smith SM (2007): Non-linear optimisation. FMRIB technical report TR07JA1.
- Beckmann CF, DeLuca M, Devlin JT, Smith SM (2005): Investigations into resting-state connectivity using independent component analysis. *Philos Trans R Soc Lond B Biol Sci* 360:1001–1013.
- Beckmann CF, Jenkinson M, Smith SM (2003): General multilevel linear modeling for group analysis in fMRI. *Neuroimage* 20:1052–1063.
- Bonnelle V, Leech R, Kinnunen KM, Ham TE, Beckmann CF, De Boissezon X, Greenwood RJ, Sharp DJ (2011): Default mode network connectivity predicts sustained attention deficits after traumatic brain injury. *J Neurosci* 31:13442–13451.
- Buchsbaum BR, Baldo J, Okada K, Berman KF, Dronkers N, D’Esposito M, Hickok G (2011): Conduction aphasia, sensory-motor integration, and phonological short-term memory—An aggregate analysis of lesion and fMRI data. *Brain Lang* 119:119–128.
- Buchsbaum BR, Olsen RK, Kock P, Berman KF (2005): Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron* 48:687–697.
- Dastjerdi M, Foster BL, Nasrullah S, Rauschecker AM, Dougherty RF, Townsend JD, Chang C, Greicius MD, Menon V, Kennedy DP, et al. (2011): Differential electrophysiological response during rest, self-referential, and non-self-referential tasks in human posteromedial cortex. *Proc Natl Acad Sci USA* 108:3023–3028.
- Dhanjal NS, Handunnetthi L, Patel MC, Wise RJS (2008): Perceptual systems controlling speech production. *J Neurosci* 28:9969–9975.
- Fox MD, Snyder AZ, Vincent JL, Corbetta M, Essen DCV, Raichle ME (2005): The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc Natl Acad Sci USA* 102:9673–9678.
- Geranmayeh F, Brownsett SLE, Leech R, Beckmann CF, Woodhead Z, Wise RJS (2012): The contribution of the inferior parietal cortex to spoken language production. *Brain Lang* 121:47–57.
- Golfinopoulos E, Tourville JA, Guenther FH (2010): The integration of large-scale neural network modeling and functional brain imaging in speech motor control. *Neuroimage* 52:862–874.
- Gonzalez-Castillo J, Saad ZS, Handwerker DA, Inati SJ, Brenowitz N, Bandettini PA (2012): Whole-brain, time-locked activation with simple tasks revealed using massive averaging and model-free analysis. *Proc Natl Acad Sci USA* 109:5487–5492.
- Greicius MD, Menon V (2004): Default-mode activity during a passive sensory task: Uncoupled from deactivation but impacting activation. *J Cogn Neurosci* 16:1484–1492.
- Guenther FH, Ghosh SS, Tourville JA (2006): Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang* 96:280–301.
- Hagmann P, Cammoun L, Gigandet X, Meuli R, Honey CJ, Wedeen VJ, Sporns O (2008): Mapping the structural core of human cerebral cortex. *PLoS Biol* 6:e159.
- Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW (1999): “Sparse” temporal sampling in auditory fMRI. *Hum Brain Mapp* 7:213–223.
- Hickok G, Poeppel D (2007): The cortical organization of speech processing. *Nat Rev Neurosci* 8:393–402.
- Hickok G, Buchsbaum B, Humphries C, Muftuler T (2003): Auditory-motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. *J Cogn Neurosci* 15:673–682.
- Hickok G, Okada K, Serences JT (2009): Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *J Neurophysiol* 101:2725–2732.
- Hickok G, Houde J, Rong F (2011): Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron* 69:407–422.
- Indefrey P, Levelt WJM (2004): The spatial and temporal signatures of word production components. *Cognition* 92:101–144.
- Jenkinson M, Smith S (2001): A global optimisation method for robust affine registration of brain images. *Med Image Anal* 5:143–156.

- Jenkinson M, Bannister P, Brady M, Smith S (2002): Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17:825–841.
- Lametti DR, Nasir SM, Ostry DJ (2012): Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *J Neurosci* 32:9351–9358.
- Leech R, Kamourieh S, Beckmann CF, Sharp DJ (2011): Fractionating the default mode network: Distinct contributions of the ventral and dorsal posterior cingulate cortex to cognitive control. *J Neurosci* 31:3217–3224.
- Leech R, Braga R, Sharp DJ (2012): Echoes of the brain within the posterior cingulate cortex. *J Neurosci* 32:215–222.
- Levelt WJM (1989): *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.
- Margulies DS, Vincent JL, Kelly C, Lohmann G, Uddin LQ, Biswal BB, Villringer A, Castellanos FX, Milham MP, Petrides M (2009): Precuneus shares intrinsic functional architecture in humans and monkeys. *Proc Natl Acad Sci USA* 106:20069–20074.
- Pa J, Hickok G (2008): A parietal-temporal sensory-motor integration area for the human vocal tract: Evidence from an fMRI study of skilled musicians. *Neuropsychologia* 46:362–368.
- Seghier ML, Price CJ (2012): Functional heterogeneity within the default network during semantic processing and speech production. *Front Psychol* 3:10.3389/fpsyg.2012.00281.
- Sharp DJ, Beckmann CF, Greenwood R, Kinnunen KM, Bonnelle V, De Boissezon X, Powell JH, Counsell SJ, Patel MC, Leech R (2011): Default mode network functional and structural connectivity after traumatic brain injury. *Brain* 134:2233–2247.
- Simmonds AJ, Wise RJS, Dhanjal NS, Leech R (2011): A comparison of sensory-motor activity during speech in first and second languages. *J Neurophysiol* 106:470–478.
- Smiley JF, Hackett TA, Ulbert I, Karmas G, Lakatos P, Javitt DC, Schroeder CE (2007): Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. *J Comp Neurol* 502:894–923.
- Smith SM (2002): Fast robust automated brain extraction. *Hum Brain Mapp* 17:142–155.
- Smith SM, Fox PT, Miller KL, Glahn DC, Fox PM, Mackay CE, Filippini N, Watkins KE, Toro R, Laird AR, et al. (2009): Correspondence of the brain's functional architecture during activation and rest. *Proc Natl Acad Sci USA* 106:13040–13045.
- Smith SM, Miller KL, Moeller S, Xu J, Auerbach EJ, Woolrich MW, Beckmann CF, Jenkinson M, Andersson J, Glasser MF, et al. (2012): Temporally-independent functional modes of spontaneous brain activity. *Proc Natl Acad Sci USA* 109:3131–3136.
- Tremblay P, Deschamps I, Gracco VL (2011): Regional heterogeneity in the processing and the production of speech in the human planum temporale. *Cortex* doi:10.1016/j.cortex.2011.09.004.
- Ventura MI, Nagarajan SS, Houde JF (2009): Speech target modulates speaking induced suppression in auditory cortex. *BMC Neurosci* 10:58–69.
- Vigneau M, Beaucousin V, Herve PY, Duffau H, Crivello F, Houde O, Mazoyer B, Tzourio-Mazoyer N (2006): Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *Neuroimage* 30:1414–1432.
- Wilson MD (1988): The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2. *Behav Res Meth Instrum Comput* 20:6–11.
- Zuo XN, Kelly C, Adelstein JS, Klein DF, Castellanos FX, Milham MP (2010): Reliable intrinsic connectivity networks: Test-retest evaluation using ICA and dual regression approach. *Neuroimage* 29:2163–2177.