Behavioral/Cognitive

Auditory–Motor Interactions for the Production of Native and Non-Native Speech

ʻŌiwi Parker Jones,[1,2] **Mohamed L. Seghier,**[1] **Keith J. Kawabata Duncan,**[1] **Alex P. Leff,**[3] **David W. Green,**[4] **and Cathy J. Price**[1]

[1]Wellcome Trust Centre for Neuroimaging, University College London, London WC1N 3BG, United Kingdom, [2]Wolfson College, University of Oxford, Oxford OX2 6UD, United Kingdom, [3]Institute of Cognitive Neuroscience, University College London, London WC1N 3AR, United Kingdom, and [4]Cognitive, Perceptual, and Brain Sciences, University College London, London WC1E 6BT, United Kingdom

During speech production, auditory processing of self-generated speech is used to adjust subsequent articulations. The current study investigated how the proposed auditory–motor interactions are manifest at the neural level in native and non-native speakers of English who were overtly naming pictures of objects and reading their written names. Data were acquired with functional magnetic resonance imaging and analyzed with dynamic causal modeling. We found that (1) higher activity in articulatory regions caused activity in auditory regions to decrease (i.e., auditory suppression), and (2) higher activity in auditory regions caused activity in articulatory regions to increase (i.e., auditory feedback). In addition, we were able to demonstrate that (3) speaking in a non-native language involves more auditory feedback and less auditory suppression than speaking in a native language. The difference between native and non-native speakers was further supported by finding that, within non-native speakers, there was less auditory feedback for those with better verbal fluency. Consequently, the networks of more fluent non-native speakers looked more like those of native speakers. Together, these findings provide a foundation on which to explore auditory–motor interactions during speech production in other human populations, particularly those with speech difficulties.

Introduction

We do not only speak to communicate with others. The sound of our voices also helps us to adjust our verbal articulations in real time (Yates, 1963; Houde and Jordan, 1998; Guenther, 2006), in particular when learning a new language (Borden, 1979, 1980). However, attending to auditory feedback is potentially disadvantageous, such as when it distracts us from our surroundings or from formulating what we want to say next. Auditory processing of self-generated speech is therefore likely to involve a balance between using auditory feedback and suppressing it. In this context, the aim of our study was twofold. First, we tested how auditory feedback and suppression during overt picture naming and reading manifest in the causal connections between selected articulatory and auditory processing regions in the brain. We identified these regions with functional magnetic resonance imaging (fMRI) and analyzed their causal connections with dynamic causal modeling (DCM). Second, by testing both native and non-native speakers of English, we investigated how auditory–motor interactions reflect speaker fluency.

Suppression within the auditory cortex has been shown when responses were attenuated for self-produced speech relative to the speech of others (Paus et al., 1996; Numminen and Curio, 1999). Evidence for auditory feedback comes from behavioral studies that show alteration in speech production when auditory feedback is perturbed (Takaso et al., 2010). In addition, Tourville et al. (2008) used structural equation modeling (McIntosh and Gonzalez-Lima, 1994) and showed increased connectivity from posterior superior temporal regions to right inferior frontal and ventral premotor regions, when auditory feedback of the spoken response was perturbed by shifting formant frequencies.

In this study, we focused on auditory suppression and auditory feedback during normal (unperturbed) speech in native and non-native speakers of English who ranged in fluency. Auditory suppression was inferred when the connections from motor to auditory areas were negative. Auditory feedback was inferred when the connections from auditory to motor areas were positive. This is explained below (see Materials and Methods, Interpreting connection strengths).

To examine how speaker fluency influenced auditory suppression and auditory feedback, we included both native and non-native speakers of English. We also investigated how auditory–motor interactions were influenced by the type of task performed. This involved comparing connectivity for object naming

Received July 11, 2012; revised Nov. 29, 2012; accepted Dec. 2, 2012.

Author contributions: ʻŌ.P.J., M.L.S., K.J.K.D., A.P.L., D.W.G., and C.J.P. designed research; ʻŌ.P.J., M.L.S., K.J.K.D., A.P.L., D.W.G., and C.J.P. performed research; ʻŌ.P.J., M.L.S., K.J.K.D., A.P.L., D.W.G., and C.J.P. analyzed data; ʻŌ.P.J., M.L.S., K.J.K.D., A.P.L., D.W.G., and C.J.P. wrote the paper.

This work was funded by the Wellcome Trust. We thank Alice Grogan, Nilufa Ali, Christos Pliatsikas, and Kostas Filippopolitis for their contribution to data collection and the initial publication of the data in Parker Jones et al. (2012). The authors also acknowledge the use of the University College London Legion High Performance Computing Facility and associated support services in the completion of this work.

The authors declare no competing financial interests.

Correspondence should be addressed to ʻŌiwi Parker Jones, Wellcome Trust Centre for Neuroimaging, Institute of Neurology, 12 Queen Square, London WC1N 3BG, UK. E-mail: o.parkerjones@fil.ion.ucl.ac.uk.

DOI:10.1523/JNEUROSCI.3289-12.2013

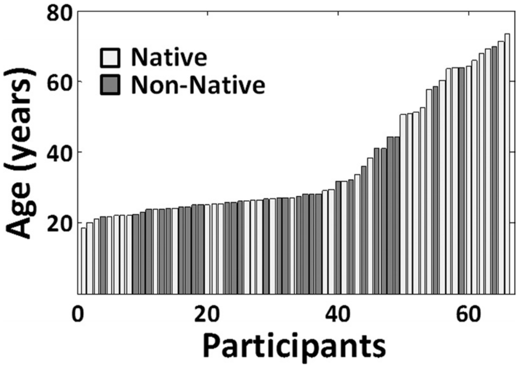

Figure 1. Participants by age. The participants ranged in age from 19 to 74 years. There was no significant difference in age between native (light gray) and non-native (dark gray) groups.

and reading to that for saying "1-2-3" to meaningless visual stimuli. Whereas naming and reading tasks are driven by the semantic content of the visual stimuli (pictures and words), saying 1-2-3 depends on the auditory representations of the intended sounds and not on the content of the visual stimuli.

We predicted that (1) auditory suppression would be greater for native speakers than non-native speakers, (2) auditory feedback would be greater for non-native speakers than native speakers, and (3) auditory–motor interactions would be greater during naming and reading relative to saying 1-2-3 repeatedly to meaningless stimuli. Moreover, we included two different auditory regions to investigate how the interaction of anterior and posterior auditory processing areas varied with speech task and speaker group.

Materials and Methods

The study was approved by the National Hospital for Neurology and Neurosurgery and by the Institute of Neurology Joint Ethics Committee. All participants gave informed consent to take part in this study.

Participants

A cohort of 36 monolingual native speakers of English (24 females) and 31 bilinguals who spoke English as a non-native language (13 females) were recruited from London (for details, see Parker Jones et al., 2012). All 67 participants were right-handed, MRI compatible, reported no neurological, hearing, or language impairments, and had normal or corrected-to-normal vision. Their mean age was 36.4 years, and there was a wide age range (19–74 years). Separated into groups, the mean age of the native English speakers was 39.7 years (range, 19–74 years), whereas the mean age of the non-natives speakers (excluding one missing case) was 32.3 years (range, 22–70 years). The practical reason for including participants with such a large age range was opportunistic: these were simply the participants to whom we had access. As shown in Figure 1, none of our participants (not even the oldest) was judged to be an outlier from the rest of the sample, because none of the participants' ages was more than 2 SDs from the aggregate mean (or, if separated into native and non-native groups, then none of the participants' ages was more than 2 SDs from the mean of the relevant group). There was no significant difference in age for the native and non-native speakers ($t_{(64)} = 1.84$, $p = 0.07$). Moreover, there was no significant difference in the connection strengths for older (than younger) participants, when participants were split around the median age or connection strengths were correlated across age.

All 67 participants were able to perform the in-scanner tasks with high accuracy. For native speakers, accuracy for picture naming was 96.3%, and accuracy for reading aloud was 99.7%. For non-native speakers, accuracy for picture naming was 90%, and accuracy for reading aloud was 98.3% (not including data lost from one of the 31 participants). Only picture-naming accuracy was significantly higher in the native than non-native speakers ($t_{(64)} = 6.5$, $p < 0.001$). All participants were able to respond with 1-2-3 to the meaningless visual stimuli without error. The high accuracy in the non-native as well as the native group was the consequence of our previous exclusion of any participant with poor in-scanner performance (for details, see Parker Jones et al., 2012). This approach allows us to avoid group differences that are confounded by accuracy.

All non-native English speakers were resident in the United Kingdom. Their native languages were Greek ($n = 21$), German ($n = 7$), Italian ($n = 1$), Dutch ($n = 1$), and Czech ($n = 1$). On average, they reported speaking 3.3 languages (range, 2–8), with a mean English age of acquisition of 9.2 years (range, 1–15 years). Their overall proficiency in English was tested using a battery of out-of-scanner tasks (for details, see Parker Jones et al., 2012). Here, we focus on the results of a standard letter fluency task that required participants to generate as many words as possible within 60 s that started with the same letter (e.g., "p") (Grogan et al., 2009). For 28 of 31 non-native speakers who completed the out-of-scanner fluency task, there was substantial variance in performance: the mean ± SE number of common nouns generated per minute was 13 ± 0.84 (range, 5–22), in which a larger score indicates better verbal fluency. We exploited this variation to investigate the influence of fluency on auditory–motor interactions within the non-native speakers. This within-group correlation effect could then be compared with the between-group comparison of native versus non-native speakers. We did not test the fluency of the native speakers, nor did we test their residual knowledge of other languages; neither measure was necessary for the current investigation. Our assumption that English fluency was better in the native than non-native group was supported in two ways. First, as detailed in our previous study (Parker Jones et al., 2012), picture-naming scores were significantly higher in the native than non-native participants ($t_{(64)} = 6.5$, $p < 0.001$). Second, the results of the DCM analyses show that group differences between native and non-native speakers (i.e., those assumed to be more vs less fluent), on connection strengths, were independently supported by corresponding effects in the correlation of good versus poor fluency scores within the non-native group.

Experimental design

The experimental design used to collect the raw data reported here and in a previous study (Parker Jones et al., 2012) has been described previously (for more on the stimuli in particular, see Hu et al., 2010). In brief, there were four experimental runs/sessions; two required overt speech production, and two required finger-press responses to indicate a semantic or perceptual judgment. We focus on data from the two scanning sessions requiring a spoken response. There were four different speech tasks during each of these sessions: (1) naming pictures of objects; (2) reading written object names; (3) saying 1-2-3 to unfamiliar (meaningless) strings of Greek letters; and (4) saying 1-2-3 to pictures of non-objects. For the 21 Greek speakers included in the non-native sample, the unfamiliar Greek letter strings (such as "δδδδ") were analogous to unfamiliar consonant letter strings in English (such as "tttt"). Thus, although each letter (δ) was familiar to the Greeks, the resulting string (δδδδ) did not constitute a familiar word. For the remaining native and non-native European participants, who did not speak Greek, we note that the unfamiliar Greek letter strings might be better described as unfamiliar false font strings. There was no evidence that Greek participants were activating the reading system when viewing unfamiliar Greek letter strings. For example, at a threshold of $p < 0.001$ (uncorrected), there were no voxels that were activated by both of the following: (1) Greek letters (non-objects in the Greek participants) and (2) reading words (Greek letters in the monolinguals). If the threshold was reduced to $p < 0.01$ (uncorrected), there were two voxels in the right cerebellum. If the threshold was reduced to $p < 0.05$ (uncorrected), there were 20 voxels distributed across six different regions.

The pictures and written words were derived from a set of 192 familiar objects with names that had three to six letters in English (e.g., bell, frog, camel, dagger). Each subject was presented with half of the objects as pictures (for object naming) and the other half as written words (for reading aloud).

Within each scanning session, there were four blocks of word reading, four blocks of object naming, and four blocks of saying 1-2-3 (two in response to meaningless letter strings that controlled for visual features in words and two in response to pictures of unfamiliar non-objects that controlled for visual features in pictures of objects).

The order of these conditions was counterbalanced within each session. After every two blocks of stimuli, there was 14.4 s of fixation. This provided a baseline measurement for activation. The duration of fixation (14.4 s), relative to the stimulus blocks and repetition time (TR), also allowed us to distribute the sampling of data across time by avoiding time locking of the stimulus onsets with the same acquisition slice (Veltman et al., 2002). Each block was preceded by 3.6 s of instructions. The instructions were "NAME," "READ," "1-2-3 SYMBOLS," or "1-2-3 PICTURES." Each block lasted 18 s, with 12 words per block presented every 4.32 s in groups of three called "triads," followed by 180 ms of fixation.

Grouping stimuli into triads enabled the participants to read or name the stimuli rapidly, thereby maximizing the efficiency of the experimental design. The same triads were presented as words and pictures across participants (to equate articulation responses). The triads were grouped with one stimulus above and two stimuli below, with one offset to the bottom left and the other to the bottom right, and with all three triad stimuli presented simultaneously. The semantic associations between items in the same triad were minimized in a pilot study, in which eight participants judged the semantic relatedness (or lack thereof) between items. In the speech-production tasks (which we used in this study), a triad might include the items "anchor," "carrot," and "broom," in which the pilot study indicated no similarity between these items. Although we excluded these stimuli in our current study, the full paradigm also included a semantic decision task in which the top item in a triad (e.g., anchor) was related to one of the items on the bottom ("ship" on the bottom left and "truck" on the bottom right). Hence, the pilot study also established which items were semantically related (i.e., anchor and ship, rather than anchor and truck). Our primary reason for excluding the semantic decision stimuli was that the relevant task required participants to press one of two buttons rather than produce speech.

Each participant undertook a short training session before the experiment, with printed sets of words and pictures that were different from those used in the scanner. They were instructed to read or name the three stimuli in the triad in a fixed order (i.e., top, bottom left, bottom right). During the 1-2-3 conditions, they were instructed to say 1-2-3 while looking consecutively at the top, bottom left, and then bottom right stimuli. During training, we also emphasized the need to keep the body, head, and mouth as still as possible.

In the scanner, we presented the stimuli via video projector and front-projection screen using a system of mirrors fastened to a head coil. We presented the words in lowercase Arial font to occupy ~4.9° (width) and 1.2° (height) of the visual field. Each picture was scaled to take ~7.3 × 8.5° of the visual field. Participants' verbal responses were recorded and filtered using in-house noise-cancellation equipment that allowed us to monitor in-scanner accuracy and distinguish correct and incorrect responses. However, the recordings were made independently of the presentation script and did not contain the stimuli onsets; therefore, we were unable to measure in-scanner naming or reading latencies.

Image acquisition

All images were acquired from a 1.5 T Siemens system (Siemens Medical Systems). Structural T_1-weighted images were acquired using a 3D modified driven equilibrium Fourier transform sequence and 176 sagittal partitions with an image matrix of 256 × 224 and a final resolution of 1 mm^3 [TR, 12.24 ms; echo time (TE), 3.56 ms; inversion time, 530 ms]. Functional T_2*-weighted echo-planar images comprised 40 axial slices of 2 mm thickness with 1 mm slice interval and 3 × 3 mm in-plane resolution (TR, 3600 ms; TE, 50 ms; flip angle, 90°; field of view, 192 mm; matrix, 64 × 64). With regard to this relatively low temporal resolution, we note that DCM adjusts for TR length by including temporal sampling information in its estimations (Kiebel et al., 2007). As Kiebel et al. (2007, p 1487) observe, "With a TR of several seconds, it is quite likely that some areas have slice-timing differences of >1 s." This was a problem for DCM in its original incarnation (Friston et al., 2003), which assumed that slices were acquired simultaneously. This turned out not to be a problem for slice timing differences up to 1 s. However, for the larger differences expected for our TR of 3.6 s, it is important to model these differences for valid results (as we have done). One hundred three volumes were acquired per session, leading to a total of 206 volume images across the two speech-production sessions. To avoid Nyquist ghost artifacts, a generalized reconstruction algorithm was used for data processing. After reconstruction, the first four volumes of each session were discarded to allow for the T_1 equilibration effect.

Data analyses

fMRI data preprocessing. Image processing and first-level statistical analyses were conducted using Statistical Parametric Mapping (SPM5; Wellcome Trust Centre for Neuroimaging, London, UK; http://www.fil.ion.ucl.ac.uk/spm/software/spm5/). Each participant's functional volumes were realigned and unwarped (Andersson et al., 2001), adjusting for residual motion-related signal changes that may cause geometric distortion in areas in which there are magnetic susceptibility artifacts in echo planar imaging data (e.g., in the oral cavity during speaking). Unwarping estimates and model field deformations with respect to subject position so that distortions from movement can be corrected during the realignment procedure (Andersson et al., 2001). This procedure therefore corrects for motion-induced artifacts during scanning. We use it in preference to including the realignment parameters as linear regressors in the first-level analysis because unwarping accounts for nonlinear movement effects by modeling the interaction between movement and any inhomogeneity in the blood oxygenation level-dependent (BOLD) signal. We further ensured that there was <3 mm movement for all participants during all scanning runs.

Realigned scans from the different participants were spatially normalized to Montreal Neurological Institute (MNI) space (voxel size, 2 × 2 × 2 mm^3) using unified segmentation and normalization of the structural image after it had been coregistered to the realigned functional images (Ashburner and Friston, 2005). The normalized functional images were then spatially smoothed with a 6 mm full-width half-maximum isotropic Gaussian kernel.

fMRI data analyses. We submitted each participant's preprocessed functional volumes to a participant-specific fixed-effect analysis, using the general linear model at each voxel. For each of the four speaking conditions, correct and incorrect responses were modeled as separate regressors (using event-related delta functions) with a duration of 4.32 s per trial and a stimulus onset interval of 4.5 s. In addition, we included a regressor that modeled all the instruction trials (regardless of condition). Each event was convolved with a canonical hemodynamic response function. To exclude low-frequency confounds, the data were high-pass filtered using a set of discrete cosine basis functions with a cutoff period of 128 s. At this point, we note that the methods were the same as in our previous study (Parker Jones et al., 2012), but from here on the studies diverged.

Regions of interest for the DCM analyses. Our DCM analysis of auditory–motor interactions during speech production focused on areas involved in the final stages of speech articulation (i.e., orofacial motor control and auditory processing). The influence of other speech production areas (e.g., involved in semantic analysis, lexical retrieval, phonological planning) were not explicitly modeled because they were not the focus of interest. Nevertheless, their contribution may indirectly influence the connectivity we measure between the occipital and precentral areas. This did not interfere with our ability to distinguish the effects of auditory suppression and auditory feedback, but it does mean that our DCM analysis was not designed to identify the brain regions that initiate or control auditory suppression. For example, any auditory suppression that we detect in the connectivity from motor to auditory regions could be controlled by premotor areas associated with auditory speech maps (Guenther et al., 2006).

Region identification (Richardson et al., 2010). In our speech-production tasks (i.e., naming pictures, reading words, and saying 1-2-3 in response to unfamiliar visual stimuli), activation co-occurred in areas involved in visual processing, all stages of speech articulation, and auditory processing. To functionally segregate auditory and orofacial motor areas, we therefore identified regions of interest (ROIs) using data from an independent study (Richardson et al., 2010).

Areas involved in the motor control of speech were identified as those that were activated during silent, non-verbal, mouth movements (lip pursing and tongue protrusion) relative to hand movements. Areas involved in auditory processing of speech were identified as those that were more activated when the participants listened to words and sentences relative to silently reading the same words and sentences. For both levels of processing (motor and auditory), activation was observed bilaterally. The current study focuses only on left hemisphere regions. Our reason for excluding the right hemisphere regions was essentially practical, because the computational cost of every region included in DCM analyses quickly becomes prohibitively high. A model with four regions required fitting and inverting >4000 models for each of 67 participants. To include right-hemisphere homologs of these four regions would require >16 million models for each of 67 participants. Moreover, as remarked on in the preceding subsection, this exclusion of regions did not interfere with our ability to distinguish the effects of auditory suppression and auditory feedback in the left hemisphere, but it does mean that our DCM analysis is not designed to identify undoubtedly important processes in the right hemisphere for the final stages of speech production.

The results (from a reexamination of the analysis reported by Richardson et al., 2010) identified the following.

(1) There was activation for mouth movements, relative to hand movements, in bilateral premotor/motor cortex. The left hemisphere peaks in MNI space were in the precentral gyrus (PrC) at coordinates (*x*, *y*, *z*) (−52, −12, 32) (*t* score > 18.0), extending dorsally to (−47 ± 11, −14, 42) (*t* score > 11.0).

(2) Listening to words and sentences, relative to the silent reading of the same words and sentences, activated bilateral auditory cortices associated with sound processing. The left hemisphere peaks in MNI space were at coordinates (−48, −24, 6) ($t =$ 11.9), which corresponds to Heschl's gyrus, extending posteriorly and dorsally into the planum temporale (PT) at (−58, −34, 12) (t score = 7.4) and anteriorly and ventrally into anterior regions in the superior temporal gyrus (aSTG) at (−52, −6, −2) (*t* score > 8.3).

None of the auditory processing or mouth movement areas referred to above were activated during silent reading relative to any other condition (e.g., fixation, mouth or hand movements), although activation for silent reading (relative to fixation, mouth and hand movements) was highly significant along the whole length of the superior temporal sulcus in which activation was also observed during auditory speech. Therefore, we are proposing that the absence of activation in Heschl's gyrus, PT, and aSTG during reading was a consequence of the lack of self-generated or stimulus-induced sound during silent reading.

ROI definitions from the data presented here and by Parker Jones et al. (2012). The identification of areas involved in mouth movements and auditory processing, using the data reported by Richardson et al. (2010), allowed us to delineate which parts of the activation profile reported for picture naming, reading aloud, and saying 1-2-3 in the current study were likely to follow from mouth movements and auditory processing. Specifically, we were able to confirm that the auditory and motor areas identified by Richardson et al. (2010) were also activated for naming, reading, or saying 1-2-3 relative to fixation in our native and non-native English speakers. Moreover, as reported by Parker Jones et al. (2012), non-native speakers of English had more activation than native speakers of English in three auditory–motor areas: (1) left PrC, (−48, −16, 42), associated with mouth movements; (2) left PT, (−56, −30, 14), associated with auditory processing during speech output (Dhanjal et al., 2008); and (3) left aSTG, (−60, −10, 2), associated with auditory processing during delayed auditory feedback (Takaso et al., 2010). Our regions also correspond to those associated with auditory–motor interactions by Tourville et al. (2008). Specifically, their posterior STG region at (−64, −30, 14) is remarkably close to our PT region at (−56, −30, 14), and their ventral primary motor cortex at (48, −10, 44) is almost the right-hemisphere homolog of our PrC region at (−48, −16, 42).

To conduct and report connectivity analyses in DCM, we defined volumes of interest from a group analysis of all participants and then extracted data for each participant from within each volume (PrC, PT, and aSTG). The dimensions of each volume were defined as the voxels that were activated in the group analysis ($p < 0.001$, uncorrected) by naming and reading more than fixation in (1) native speakers, (2) non-native speakers, and (3) non-native speakers more than native speakers. Criteria 1 and 2 ensured that we identified areas that were commonly activated by native and non-native speakers. The motivation for including criterion 3 was twofold. First, we needed a principled way to reduce the size of the large ROIs associated with motor and auditory processing by Richardson et al. (2010). Second, we wanted to ensure that any group differences in connectivity for one pair of regions, but not another, was not biased by group differences in activation level (e.g., greater for non-native than native in one region but not another). Because there were no areas in which brain activation was higher for native and non-native speakers, criterion 3 reduced bias in our region selection rather than increasing it. The resulting small volumes of interest provided precision and consistency across participants in the voxels that contributed to the DCM analysis. All three ROIs were larger than 60 voxels (with PrC comprising 77 voxels, aSTG 61 voxels, and PT 93 voxels).

Because all speech-production conditions were cued visually, we included a fourth ROI in the occipital lobe (Occ) to serve as an external input to the models. The Occ mask was extracted by finding the most significant group activation in the visual cortex. To limit the extensive activation in visual cortex to the most significant area, we set the statistical threshold for the contrast (naming and reading > fixation) to $t = 14$ and ensured that the same voxels were also activated by semantic word matching relative to fixation. We used an iterative procedure to pick this threshold, starting small and increasing *t* until the resulting activation was confined to the Occ (as visualized in SPM on the single-subject T_1 template, i.e., the "Colin brain"); we then selected the previous *t* value—effectively, the biggest cluster restricted to the Occ. This resulted in 347 voxels in the inferior occipital gyrus, which were converted into a binary image and subsequently used to extract the relevant data (see next section). Figure 2 illustrates the four ROIs. Table 1 records the peak coordinates and numbers of voxels.

Data extraction. For each participant, the principle eigenvariate (time series) was extracted from each of the four binary images (PrC, aSTG, PT, Occ) and adjusted to the participant's *F* contrast (effects of interest). The principal eigenvariates provide a method for summarizing the multivariate time series from all of the voxels in each ROI. The aim is to produce a single vector for each region. An alternative method would be to sum the voxels' time series, but, of course, positive and negative values would cancel. The principle eigenvector does not have this shortcoming. It is also standard practice to adjust the time series to each participants' *F* contrast (effect of interest). As Stephan et al. (2010, p 9) noted, ". . . it would be nonsensical to ask

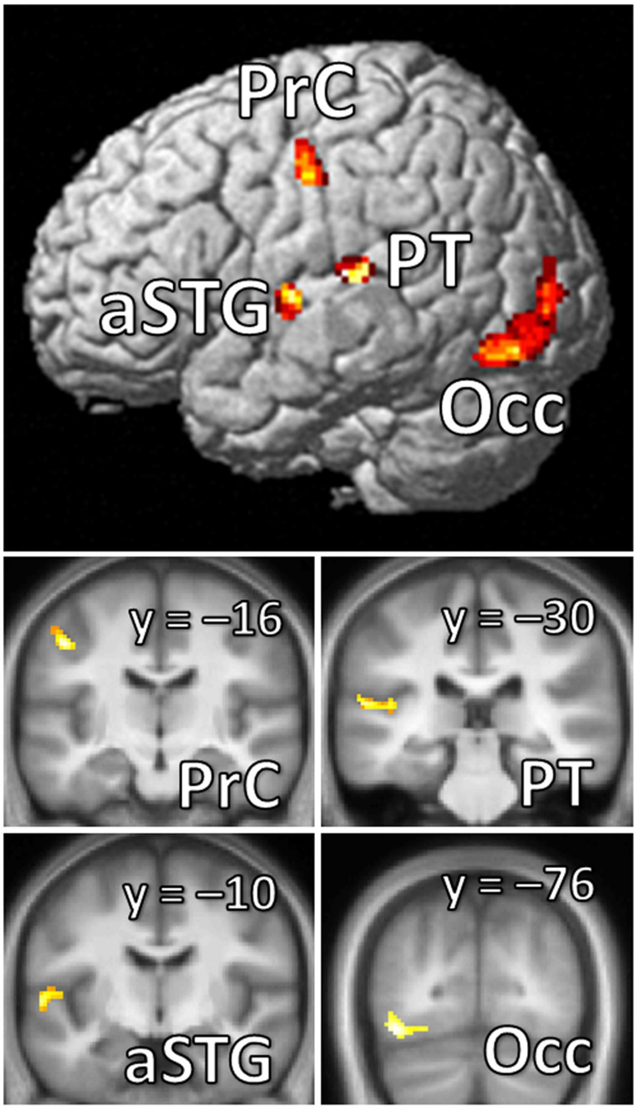

Figure 2. Anatomical illustration of our three left-hemisphere ROIs plus input area. The regions are the PrC, aSTG, PT, and Occ.

Table 1. ROIs

		Peak (MNI)		
ROI	Voxels	*x*	*y*	*z*
PrC	77	−48	−16	42
aSTG	61	−60	−10	2
PT	93	−56	−30	14
Occ	347	−42	−76	−12

Cluster sizes are given in voxels, and representative coordinates are given in MNI space.

this [connectivity] question of regional responses that did not show experimental effects. Generally one should use the most revealing *t*- or *F*-contrast for each region….” Thus, the contrast reveals the experimental effects, which we aimed to model.

These data were then concatenated across the two overt speech-production sessions and entered into the DCM analyses. No statistical threshold was imposed on activity within the ROI (i.e., the same set of voxels contributed to the extracted time series in all participants). The advantage of this approach is that it avoids the common practice of excluding participants from DCM analyses when they have weak activation in the ROIs. To put it another way, we were in this way able to include all of our participants in the DCM analyses. One potential disadvantage of this approach is that the principle eigenvariate from each ROI may reflect condition-independent noise, which would reduce the significance of the group-level inference and possibly yield false negatives. This was not a serious concern in the current study because the (relatively) large sample sizes produced highly significant results in the DCMs, which had a very low probability of being noise.

DCM. We provide a few more details here about DCM to elaborate on why we chose this technique and what it offers beyond other approaches.

DCM makes inferences about the causative (i.e., “effective”) connectivity of dynamic neural systems from neuroimaging data, including fMRI (Friston et al., 2003). It relies on an inference between observation (i.e., hemodynamic response) and biophysical modeling at the neuronal level and uses empirical Bayesian methods for model inversion and comparison (Friston, 2009).

A few general points when interpreting DCM results follow (Seghier and Price, 2010). First, to reiterate, connectivity parameters (endogenous and modulatory) are estimated at an inferred neuronal level, because neuronal activity is not directly accessible from the hemodynamic response in fMRI. Second, coupling between regions does not imply direct anatomical connections but can also result from polysynaptic connections between two ROIs. Third, the estimated model is context dependent. This means that interactions between regions are sensitive to model architecture (i.e., regions, connections, driving inputs, modulations). Changing any of these parameters may have a nonlinear effect on the results. Finally, it should be remembered that DCM can only find the best available model within the set of models (or “model space”) tested. This means that the inclusion of multiple competing models is an advantage and is not affected by the multiple-comparison problem. To put it another way, the data (extracted and adjusted times series) are identical across comparisons; it is only the models that vary.

DCM model space. We grouped all four speech tasks (picture naming, reading, saying 1-2-3 to meaningless letters and saying 1-2-3 to non-objects) as a single input that entered the system at Occ. Although the specific nature of the input differed by task type, we note that the system was driven in each case by external visual stimulation. Naming and reading were used as a modulatory input to estimate any changes in connectivity during these tasks versus saying 1-2-3 and to estimate any changes in connectivity during naming versus reading. We did not expect strong differences between task, but this approach allowed us to check this assumption while also exploring the overall effect of speech production on the system. To populate the model space, a fully connected DCM was specified for each participant (i.e., with connections within and between all four regions). We then generated models for every possible combination of between-region modulations, in effect fixing the endogenous (average) connectivity while varying the modulations for naming and reading (relative to everything else in the session, which included saying 1-2-3 and fixation) for every possible between-region connection (excluding self-connections). The result was a model space of 4095 DCMs for each participant (274,365 models overall), which we then analyzed. In general, a comprehensive model space such as this is advantageous during model comparison and averaging, because it allows multiple explanations of the data to be tested explicitly.

Parameter estimation. There were three sets of neural connectivity parameters to estimate in each model: (1) input parameters that quantify how brain regions respond to external stimuli; (2) endogenous parameters reflecting latent effective connectivity, characterizing average coupling strength between regions; and (3) modulatory parameters that measure task-induced changes in effective connectivity. In DCM, these parameters are expressed as rates of change in hertz between region-based time series at the neuronal level (Friston et al., 2000). Each model also included a fourth set of hemodynamic parameters, used to infer a predicted BOLD response from the estimated neural dynamics of each region (Buxton et al., 1998; Friston et al., 2000; Stephan et al., 2007a). For each model, all parameters were estimated using an expectation-maximization algorithm (Dempster et al., 1977; Friston et al., 2003). Because of the relatively large number

Table 2. Auditory–motor connections

		All participants			Native			Non-native			Group differences			Fluency task	
Connection		Statistics		Strength	Statistics		Strength	Statistics		Strength	Statistics			Statistics	
From	To	$t_{(66)}$	p	Hz	$t_{(35)}$	p	Hz	$t_{(30)}$	p	Hz	t	df	p	$r_{(28)}$	p
Endogenous															
Auditory suppression															
PrC	PT	−4.08	**<0.001**	−0.23	−5.23	**<0.001**	−0.35	−0.93	0.358	−0.08	−2.58	65	**0.012**	0.301	0.120
PrC	aSTG	−2.38	**0.020**	−0.15	−2.56	**0.015**	−0.23	−0.63	0.533	−0.05	−1.43	65	0.157	0.340	*0.077**
Auditory feedback															
PT	PrC	1.47	0.146	0.07	−0.37	0.717	−0.02	2.07	**0.047**	0.16	−2.04	65	**0.046**	−0.412	**0.030**
aSTG	PrC	0.43	0.666	0.02	0.92	0.362	0.06	−0.30	0.770	−0.02	0.83	65	0.412	0.301	0.119
Auditory interactions															
PT	aSTG	1.57	0.120	0.09	−0.34	0.740	−0.03	3.06	**0.005**	0.21	−2.28	65	**0.026**	−0.391	**0.039**
aSTG	PT	1.38	0.173	0.07	2.52	**0.016**	0.18	−0.67	0.506	−0.05	2.22	65	**0.030**	0.093	0.639
Modulatory															
Auditory suppression															
PrC	PT	−1.11	0.27	−0.03	−1.57	0.125	−0.01	−0.85	0.40	−0.04	0.61	31.3	0.55	0.170	0.387
PrC	aSTG	−2.82	**0.006**	−0.04	−2.52	**0.02**	−0.06	−1.34	0.19	−0.03	−1.02	65	0.31	−0.387	**0.042****
Auditory feedback															
PT	PrC	−0.74	0.465	−0.01	1.35	0.19	0.01	−1.54	0.13	−0.03	2.05	65	**0.04**	−0.052	0.794
aSTG	PrC	−1.48	0.145	−0.03	−1.20	0.24	−0.03	−0.86	0.4	−0.03	−0.23	65	0.82	−0.145	0.462
Auditory interactions															
PT	aSTG	−1.55	0.125	−0.02	−1.31	0.20	−0.03	−1.10	0.28	−0.01	−0.9	42.1	0.37	−0.015	0.960
aSTG	PT	−2.89	**0.005**	−0.07	−1.96	0.06	−0.07	−2.13	**0.04**	−0.07	0.04	65	0.97	0.056	0.776

Connectivity representing auditory suppression, auditory feedback, and auditory interactions. Endogenous or average connections are given in the top, whereas modulatory connections, which were additively stronger for naming and reading, are given in the bottom. Connection strengths are given in hertz. Significant p values ($p < 0.05$, two-tailed) are rendered in bold here and in subsequent tables. *To be clear, we assume that a trend equals $0.1 > p > 0.05$. So this correlation is a trend with a two-tailed test but would be significant if we had used a one-tailed test instead ($p_{\text{one-tailed}} = 0.0385$). Additional trends are italicized in subsequent tables. **When the effect of a single outlier was removed, this correlation disappeared ($r = -0.09$, $p = 0.66$).

of models in our analysis, the combined neural and hemodynamic parameters for all DCMs were estimated in parallel over a local cluster. To promote replicability and relevance, we used the most up-to-date, open-source code at the time of writing (i.e., DCM10 in SPM8; Wellcome Trust Centre for Neuroimaging; http://www.fil.ion.ucl.ac.uk/spm/software/spm8/).

Bayesian model averaging and statistics. Having estimated the free parameters, we analyzed the resulting DCMs in three ways. First, we considered all of the participants as a single group. In the second and third cases, we separated the total set of DCMs into disjoint subsets, one for the native English speakers and another for the non-native English speakers. In each case, we performed random-effects Bayesian model selection over the relevant model space to assess the posterior probability of each model generating the observed fMRI data series. As expected, no single model was found to be significantly better than the rest: the combined, native, and non-native model spaces were so big. Instead, we calculated a Bayesian model average (BMA; Penny et al., 2010) for each group (i.e., three BMAs in total), which revealed the weighted average connectivity of our system over each model space. BMA is the standard approach for large model spaces (Penny et al., 2010) and has been successfully applied in several recent DCM studies (Liang et al., 2011; Osnes et al., 2011; Richardson et al., 2011; Seghier et al., 2011).

Frequentist statistics were used to evaluate the significance of the resulting connection strengths from the BMA, for both the endogenous connections (i.e., average connection strengths across all correct trials) and modulatory connections (i.e., differences in the connection strengths for naming and reading vs saying 1-2-3, as well as for naming vs reading). Within each group, we used one-sample t tests to calculate the significance of each connection ($H_0 = 0$) and to evaluate self-connections ($H_0 = -1$). We note that DCM assumes different priors on connections between regions and self-connections. The first assumption is that there should be no causal relationship between regions, hence the prior and null hypothesis of 0. Because of the nature of Bayesian modeling, this prior thus exerts a powerful preference for 0, which means that the data really have to be strong to deviate from this; this is called a "shrinkage prior." The second assumption is that the "shrinkage prior" on self-connections is negative to force the model to quiet down over time. This is an implementation constraint: if we did not assume negative self-connections, then the model dynamics would be such that the patterns of activity could explode or become chaotic.

To compare connection strengths between native and non-native speakers, we used independent sample t tests. In this way, we were able to ascertain whether a specific connection was significantly stronger for one group relative to the other. For all t tests (all of which were two-tailed), we used a statistical threshold of $p < 0.05$. Because Student's t test assumes equal variances between samples, we also tested the equality of variances (Levene, 1960): when between-sample variances were heteroscedastic, we used an unequal variance t test rather than Student's t test (Welch, 1938, 1947). The effects of this more conservative test can be seen in the adjusted degrees of freedom (df), t values, and p values in our results (see the non-integer df in Tables 2 and 3). Incidentally, we note that the model and the choice of how to evaluate the model are independent. Whereas the model is Bayesian (for example, its parameters were estimated using the standard Bayesian expectation-maximization algorithm), the question of whether the estimated parameters differ from their priors can be evaluated with simple-to-interpret t tests.

We are aware that there has been some inconsistency in the DCM literature about whether or not to correct for multiple comparisons when evaluating connection strengths with frequentist statistics (Stephan et al., 2010; for a discussion on the use of Bonferroni's correction for multiple comparisons, see Seghier et al., 2010). Our position is that it would be strictly inaccurate to correct each connection for multiple comparisons as if they were independent, because the calculated strength of any connection depends on every other connection in the DCMs. The situation is analogous to using contrasts in fMRI analysis: one does not correct for the number of contrasts defined, selected, or reported. This is because no matter how many contrasts one might use, the data remain the same. Similarly, no matter how many connections we evaluate within the results of the BMA, the data likewise remain the same.

Table 3. Connections to and from Occ

Connection		All participants			Native			Non-native			Group differences		
		Statistics		Strength	Statistics		Strength	Statistics		Strength	Statistics		
From	To	$t_{(66)}$	*p*	Hz	$t_{(35)}$	*p*	Hz	$t_{(30)}$	*p*	Hz	*t*	df	*P*
Endogenous													
Occ	PrC	1.35	**<0.001**	0.36	7.17	**<0.001**	0.31	7.58	**<0.001**	0.41	1.43	65	0.158
Occ	aSTG	8.11	**<0.001**	0.32	6.39	**<0.001**	0.33	5.01	**<0.001**	0.31	0.30	65	0.767
Occ	PT	1.21	**<0.001**	0.36	7.69	**<0.001**	0.36	6.64	**<0.001**	0.36	−0.07	65	0.941
PrC	Occ	0.26	0.799	0.01	0.29	0.771	0.02	0.04	0.971	0.01	−0.19	65	0.850
PT	Occ	3.31	**0.002**	0.21	1.92	*0.064*	0.20	3.26	**0.005**	0.23	−0.21	59.32	0.832
aSTG	Occ	1.28	0.205	0.08	1.09	0.284	0.11	0.67	0.511	0.04	0.59	56.63	0.561
Modulatory													
Occ	PrC	0.80	0.424	0.01	0.38	0.707	0.01	2.24	**0.033**	0.01	−0.12	65	0.908
Occ	aSTG	−1.17	0.247	−0.01	−0.43	0.672	−0.01	−1.28	0.212	−0.02	0.61	65	0.545
Occ	PT	0.99	0.327	0.01	0.52	0.606	0.01	0.83	0.412	0.02	−0.45	65	0.654
PrC	Occ	2.66	**0.01**	0.06	2.49	**0.018**	0.03	1.99	*0.056*	0.09	−1.23	34.86	0.227
PT	Occ	1.88	0.064	0.03	2.61	**0.013**	0.06	−0.54	0.593	−0.01	2.44	60.92	**0.018**
aSTG	Occ	0.53	0.598	0.02	−0.04	0.972	−0.01	0.84	0.408	0.04	−0.62	65	0.536

Connections to and from the input region, Occ. Endogenous or average connections are given in the top, whereas modulatory connections, which were additively stronger for naming and reading, are given in the bottom. Connection strengths are given in hertz.

Interpreting connection strengths. The result of each of the three BMAs was an average model with representative connection strengths either across speakers, within native speakers, or within non-native speakers. As we stated in Introduction, auditory suppression was inferred when the connections PrC → PT or PrC → aSTG (i.e., from motor to auditory areas) were negative, and auditory feedback was inferred when the connections PT → PrC or aSTG → PrC (from auditory to motor areas) were positive; here we explain what exactly positive and negative connections mean in DCM.

In DCM, a negative connection means that higher activity in the source region causes a decreased rate of change in the activity of the target region. Technically, this relationship is modeled by a differential (evolution) equation (Friston et al., 2003):

$$\frac{dx}{dt} = \left(A + \sum_{j=1}^{m} u_j B^{(j)} \right) x + Cu,$$

where x, short for $x(t)$, is a vector of n regions, $x_1(t), \ldots, x_n(t)$, such that the equation relates (1) the strength of each region in x at time t to (2) the rate of change of each region in x at time t (which is written as the derivative dx/dt). "Source" and "target," then, are just different entries in the same activation state vector x. A negative connection causes the rate of change in the activity of the target region to decrease when the source is high (i.e., high x and negative dx/dt for the relevant entries in x): this is what we mean by "suppression." In contrast, a positive connection means that the rate of change in the activity of the target region increases when the source is high (i.e., high x but this time positive dx/dt, again for the relevant entries in x). A third possibility is that the connection strength does not differ from zero, in which case the rate of change dx/dt is flat and there is no significant causal relationship. Because a derivative is a slope (i.e., "rise over run"), dx/dt may thus take values from negative infinity to positive infinity (although in our models typical values range closer to zero, influenced by the shrinkage priors referred to above and representing a previous belief that there is no causal relationship between regions).

In this equation, the endogenous connection strengths are represented in the A matrix. These are average effects across all conditions in the experiment. The B matrices represent differences in connectivity for different conditions. For instance, if the experiment were to consist of naming pictures, reading words, and saying 1-2-3 to unfamiliar objects and letter strings, then one modulation would be to look at how the average connectivity differs for naming and reading compared with saying 1-2-3. In this way, one models context-sensitive connectivity by simply adding A and B matrices together, given that there may be multiple contexts expressed through multiple B matrices. [Our analysis included two modulations for (naming and reading > saying 1-2-3) and for (naming > reading).] The inputs (represented by u) enter the system through the connections specified in C. In our case, C ensures that the inputs enter through Occ. Therefore, all of the activity in the system is driven through this region (for more details, see Friston et al., 2003; Stephan et al., 2007b).

Finally, DCM models regional interactions at an inferred neuronal level. It is critical to model interactions at this neuronal level because hemodynamic responses vary across the brain such that surface interactions can lead to false inferences (Penny et al., 2004). In practice, these neural representations can be inferred as outlined above (see *Data extraction*). To reiterate, (1) an adjusted BOLD signal is extracted from a set of voxels within a region, (2) these time series are summarized by their principle eigenvariate (a single vector representing the region), and (3) an empirically validated hemodynamic model is used to infer an underlying and neuronal time series (Buxton et al., 1998; Friston et al., 2000; Stephan et al., 2007a). Each region is thus modeled by a neuronal time series and connections model interactions between these.

Correlating connection strengths with behavior. Twenty-eight of 31 non-native speakers completed the out-of-scanner verbal fluency task. We correlated these scores with specific connections from the third BMA (non-native speakers only) to help interpret the results and validate group differences between native and non-native speakers.

Results

A summary of the significant effects is provided for the endogenous connectivity in Figure 3 and the modulatory connectivity in Figure 4.

Auditory suppression (from motor to auditory regions)

Auditory suppression manifested as negative connectivity from motor to auditory regions (PrC → PT and PrC → aSTG). This was significant across groups and conditions (Table 2, top), which indicates that, when articulatory activity was higher, activity in auditory areas declined. Within groups, evidence for auditory suppression reached significance in the native speakers but not in the non-native speakers, and this group difference reached significance on the connection from PrC → PT. Within the non-native group, there was a trend for the negative influence of PrC → aSTG to be stronger in those with less good verbal fluency (for details, see Table 2, top).

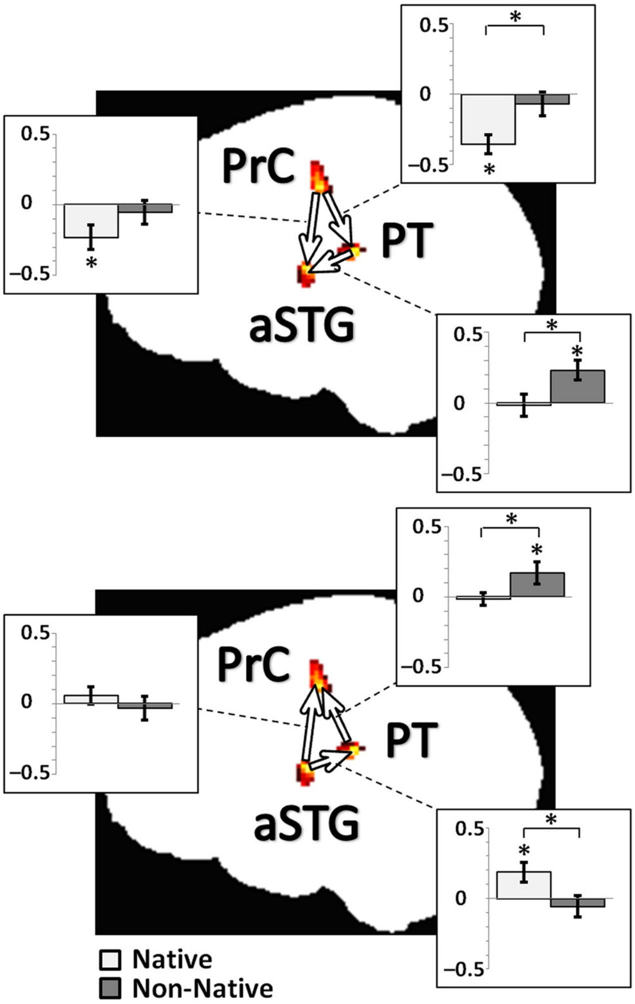

Figure 3. Differences in endogenous connectivity for native and non-native speakers. Top, The auditory suppression connections, from PrC to PT and aSTG, are rendered along with the "bottom-up" connection (in predictive coding), from PT to aSTG. Bottom, The two potential auditory feedback connections, from PT and aSTG to PrC, are rendered along with the "top-down" connection (for details, see Discussion), from aSTG to PT. Connection strengths are illustrated using bar plots, with native speakers (light gray) on the left and non-native speakers (dark gray) on the right. Error bars indicate SEs. Asterisks show significant p values at $p < 0.05$ (two-tailed).

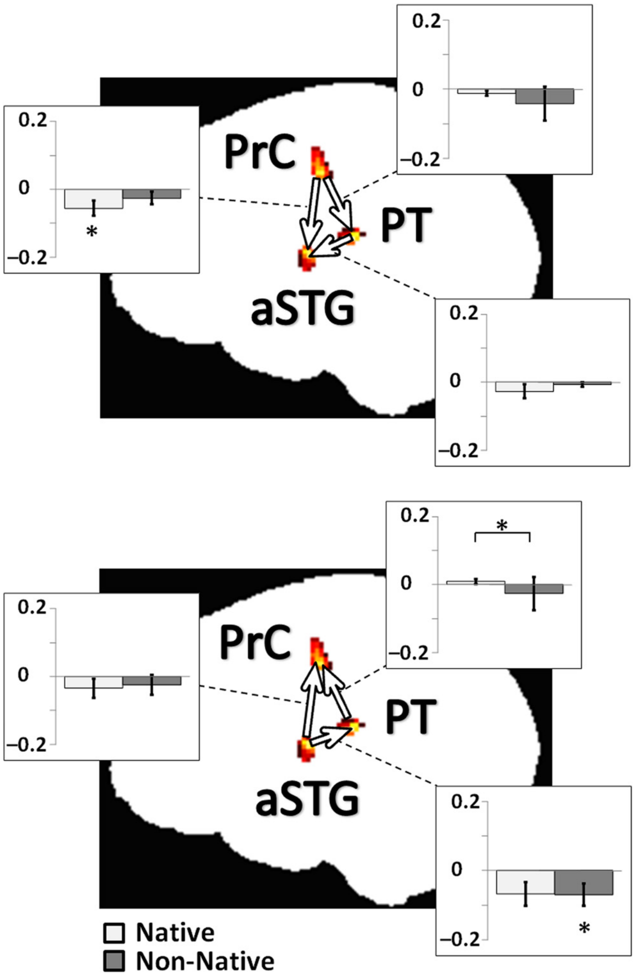

Figure 4. Differences in modulatory connectivity for native and non-native speakers. Top and Bottom, The parameter estimates for each group on the modulations (i.e., naming and reading > saying 1-2-3), between PrC, aSTG, and PT. For details, see Figure 3. To emphasize the modulation strengths, we have changed the scale of the y-axes from Figure 3. We note that aSTG → PT is a trend in the native group (Table 2, bottom). Error bars indicate SEs. Asterisks show significant p values at $p < 0.05$ (two-tailed).

The modulatory effect of task showed that the negative influence of PrC on auditory activity (i.e., auditory suppression) was greater for naming and reading than for repeating 1-2-3. This reached significance in the native group and across groups on the connection from PrC → aSTG. The strength of this modulation (PrC → aSTG) in the non-native group was not significantly related to verbal fluency when the effect of a single outlier was removed (Table 2, bottom).

Auditory feedback (from auditory to motor regions)

Auditory feedback was evidenced by significant positive connectivity from PT → PrC in non-native speakers (Table 2, top). The group difference on the strength of this connection was significant, indicating that auditory feedback was stronger in non-native than native speakers. This inference was supported by a significant negative effect of verbal fluency in the non-native speakers (Fig. 5). In other words, auditory feedback on PT → PrC was negatively correlated with fluency, so that feedback is stronger in non-native speakers with poor fluency.

There was no significant evidence for auditory feedback for aSTG → PrC in either native or non-native speakers. However, the difference in connection strengths for PT → PrC and aSTG → PrC was not significant either. The absence of a significant difference in the strength of these connections could be because either the difference in connection strengths was small or the intersubject variability was high. Because the difference in connection strengths was large (0.16 for PT → PrC and −0.02 for aSTG → PrC; see Table 2, top), we infer that the intersubject variability was high. Indeed, this is indicated by the significant effect of fluency on interparticipant variability in the strength of PT → PrC (Fig. 5). Together, the results indicate that auditory feedback from PT → PrC was driven by non-native speakers with poor verbal fluency with no evidence for a corresponding effect on aSTG → PrC.

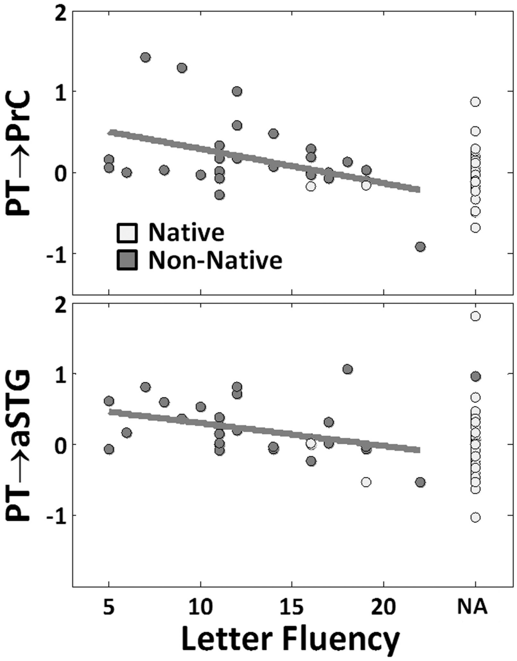

Figure 5. Connection strengths by fluency. Two (endogenous) connections were negatively correlated (gray line) with letter fluency in non-native speakers (dark gray data points): PT → PrC and PT → aSTG. The two available examples of native speakers (light gray data points) with letter fluency scores would appear to fit this pattern, and the remaining participants who did not have fluency scores (plotted over "NA" on the *x*-axis) were mostly native speakers and would appear to cluster (as expected) toward the more fluent end of the fluency distribution.

There was no significant influence of task (naming and reading vs saying 1-2-3) on either feedback connection in either group (Table 2, bottom); however, there was an interaction between group and the effect of modulation on the strength of PT → PrC. Because this was not qualified by significant effects in either group, we will not discuss it further.

Auditory interactions between PT and aSTG

The connectivity on PT → aSTG was significantly more positive for non-native than native participants and when verbal fluency was lower rather than higher (Table 2, top). In contrast, the connectivity on aSTG → PT was significantly more positive for native than non-native speakers. It was also more positive across speakers when saying 1-2-3 than naming or reading (Table 2, bottom). We suggest an interpretation of this in Discussion.

Self-connections and connectivity to and from Occ

All self-connections were significant, and, on PT, self-connections were stronger for non-native than native speakers (Table 4). All of the connections from Occ were significant, with no differences between groups or tasks (Table 3, top). The connection PrC → Occ was stronger for naming and reading than saying 1-2-3 for all participants (Table 3, bottom). Finally, the connection PT → Occ was significant for all participants and stronger for naming and reading than for saying 1-2-3 in the native speakers. We can only speculate that these connections back to Occ may indicate that visual processing is related to attention or time spent processing, because we do not have response time data for speech-production tasks.

Discussion

This study has provided novel evidence that activity in brain regions involved in articulation suppresses activity in brain regions involved in auditory processing (auditory suppression). Conversely, activity in brain regions involved in auditory processing influences activity in brain regions involved in articulation (auditory feedback). In addition, we have shown that, compared with native speakers of English, non-native speakers have (1) less auditory suppression, (2) more auditory feedback, and (3) altered coupling between the auditory regions PT and aSTG. The differences between native and non-native speakers were further supported by finding that, within non-native speakers, auditory feedback was higher when verbal fluency was poor. Put another way, better fluency in non-native speakers made the regional interactions more similar to those of native speakers.

Auditory suppression

The influence of articulatory activity on auditory activity is consistent with previous studies that reported less auditory processing during self-produced speech than the same speech spoken by another (Paus et al., 1996; Curio et al., 2000; Christoffels et al., 2007, 2011; Tourville et al., 2008; Ventura et al., 2009), with this attenuation of auditory processing being less in non-native than native speakers (Simmonds et al., 2011). The implication from these previous findings is that the act of producing speech sends messages to the auditory cortices, and these messages predict the expected sound and so reduce the response to it (Guenther et al., 2006; Tourville et al., 2008; Golfinopoulos et al., 2010). However, direct evidence for auditory–motor interactions can only be provided by demonstrating that activity in brain areas associated with speech production influences activity in brain regions associated with auditory processing and vice versa (Tourville et al., 2008). A novel contribution of our study is therefore the demonstration that differences in activation levels in the auditory cortex of native and non-native speakers is causally related to activity in speech-production regions.

In addition, by looking at task-dependent connectivity, we have been able to show that auditory suppression was significantly greater when participants were naming and reading than when they were repeating 1-2-3. This result suggests that auditory suppression is greater when speech production is driven by the semantic content of the visual stimuli than when it is driven by auditory representations of the intended sounds.

Auditory feedback

Our finding that auditory feedback (PT → PrC) was stronger in (1) the non-native group than the native group and (2) non-native speakers with poorer verbal fluency advances previous inferences from behavioral studies (Yates, 1963; Borden, 1979, 1980; Houde and Jordan, 1998; Guenther, 2006), computational simulations (Guenther et al., 2006; Guenther, 2006; Golfinopoulos et al., 2010; Houde and Nagarajan, 2011), and functional connectivity studies of perturbed speech (Tourville et al., 2008). Moreover, we were able to show that auditory feedback is greatest in the participants who had the least auditory suppression—namely, the non-native speakers.

It is also interesting to note that, despite evidence for auditory feedback on PT → PrC, there was no significant evidence for auditory feedback on aSTG → PrC in either group. The

Table 4. Self-connections (endogenous)

		All participants			Native			Non-Native			Group differences		
Connection		Statistics		Strength	Statistics		Strength	Statistics		Strength	Statistics		
From	To	$t_{(66)}$	p	Hz	$t_{(35)}$	p	Hz	$t_{(30)}$	p	Hz	t	df	p
PrC	PrC	−65.44	**<0.001**	−0.48	−53.98	**<0.001**	−0.48	−39.49	**<0.001**	−0.47	−0.77	65	0.443
PT	PT	−50.61	**<0.001**	−0.44	−32.63	**<0.001**	−0.43	−43.87	**<0.001**	−0.46	2.03	65	**0.046**
aSTG	aSTG	−70.57	**<0.001**	−0.46	−56.89	**<0.001**	−0.45	−44.82	**<0.001**	−0.47	1.49	65	0.140
Occ	Occ	−57.29	**<0.001**	−0.43	−40.97	**<0.001**	−0.42	−39.7	**<0.001**	−0.43	0.61	65	0.547

Self-connections from each region to itself. These were only defined for endogenous or average connections; there were no modulated self-connections. Connection strengths are given in hertz.

source of the auditory feedback, in PT, is consistent with Takaso et al. (2010) and the DIVA (Directions Into Velocities of Articulators) computational model (Guenther et al., 2006; Golfinopoulos et al., 2010). In their description of these models, the authors emphasize that auditory feedback plays an important role in speech acquisition but is not involved in fluent adult speech unless the online monitoring of speech output detects a mismatch between the anticipated sound and the sounds that are produced. Likewise, we did not find evidence for auditory feedback in native speakers or non-native speakers with good fluency, but we did find evidence for auditory feedback in non-native speakers with poorer fluency. Critically, this was observed during correct responses, generated in real time, which suggests that these participants may have been actively monitoring and correcting their articulation online. Our results, showing auditory feedback during typical object naming and reading tasks, therefore go beyond previous studies that have only demonstrated the effects of auditory feedback in expert adult speakers when speech is artificially perturbed (Christoffels et al., 2007; Tourville et al., 2008; Parkinson et al., 2012).

Auditory interactions between PT and aSTG

The influence of PT on aSTG activity was more positive in non-native than native speakers and in non-native speakers with poorer relative to better fluency. This effect mirrors that on the PT → PrC connection (more positive connectivity with lower fluency) and complements the effect on the PrC → PT connection (with lower fluency). Together, these results suggest that PT activity, which has not been suppressed during articulation, has an excitatory influence on both articulatory activity (PrC) and auditory activity in the anterior auditory processing stream (involving aSTG). In contrast, when PT activity is strongly suppressed in native speakers, the influence of PT on aSTG or PrC is not significant.

A more surprising finding was that aSTG → PT was significantly more positive in native than non-native speakers. This connection was also significantly more positive for naming or reading than saying 1-2-3. Plausibly, aSTG involves higher-level speech processing than PT and predicts the lower-level activity in PT. According to a predictive coding account (Friston and Kiebel, 2009), the representation of speech (in aSTG) and its lower-level auditory expectation (in PT) will be more reliable for native speakers who are more confident in their top-down predictions of how speech should sound. Stronger connectivity on aSTG → PT is therefore hypothetically proportional to the confidence in how well higher-level auditory predictions in aSTG can predict lower-level auditory activity in PT. In support of this speech-processing hierarchy, we note that aSTG is more sensitive to the acoustics of speech than non-speech, which is not true for PT (Leff et al., 2009; Rauschecker and Scott, 2009; Leaver and Rauschecker, 2010).

Such an interpretation of the determinants of aSTG → PT connectivity also provides a more refined interpretation of the determinants of PT → aSTG connectivity. As suggested above, PT → aSTG connectivity makes sense as being greater in non-native speakers who have lower confidence in their higher-level auditory representations. Therefore, it might be the case that, when top-down predictions on aSTG → PT are weak, the outputs from auditory processing in PT need more in-depth speech-recognition processing; in turn, this results in more potent connectivity on PT → aSTG. Conversely, increased confidence in top-down predictions on aSTG → PT (in native speakers) has, as predicted, less need for in-depth speech-recognition processing, producing less potent connectivity of PT → aSTG.

Limitations

First, as with any DCM study, significant connectivity between two regions does not imply direct anatomical connections between these regions. There may be many other intermediate areas that have not been included. Although a full understanding of regional interactions requires additional investigation into the participating regions, the absence of these regions in the current study does not undermine the importance of the conclusions we can draw from a discrete and carefully selected set of regions.

Second, each region in the DCM represents a different level of linguistic processing: PrC activity is representative of articulatory output, PT activity is representative (arguably) of lower-level auditory processing, and aSTG activity is likewise representative (arguably) of higher-order auditory processing. Our observation that there was a significant influence of PrC activity on auditory activity, and vice versa, does not mean that the interactions mediating these effects occurred within PrC or within PT or within aSTG. For example, it might be the case that an inferior frontal region (1) sends predictions of forthcoming events to PrC and the auditory regions, (2) receives feedback from both motor and auditory areas, and (3) updates subsequent predictions to both regions in accordance with the feedback (Price et al., 2011). To test how the auditory–motor interactions are more elaborately mediated will therefore require follow-up studies.

Third, as in any fMRI study, we can only interpret effects that are significant and therefore unlikely to occur by chance. However, we cannot conclude that there is no connection if we fail to find a significant coupling.

Conclusions

DCM has allowed us to demonstrate auditory suppression and auditory feedback at a neural level and then to show how these auditory–motor interactions change with the fluency of the speaker. Our novel conclusions are as follows: (1) auditory suppression is greater when speech production is driven by the semantic content of visual stimuli than when it is driven by auditory representations of the intended sounds; (2) feedback from auditory to motor regions is observed in adult speakers in the absence of artificially perturbing their speech; (3) the source of auditory feedback was located in PT rather than aSTG; and (4) non-native speakers, particularly those with poorer fluency, had less auditory predictions from higher-to-lower auditory areas, less auditory suppression, more auditory feedback, and stronger connectivity from lower-to-higher auditory areas. Additional studies are now needed to investigate (1) how additional regions, such as right-hemisphere homologs and frontal and parietal areas, are involved in mediating the auditory–motor interactions shown here, and (2) how the direction and strength of these auditory–motor interactions are affected in additional speaker populations, such as people with speech difficulties (e.g., auditory verbal hallucinations, dyslexia, or aphasia after stroke).

References

Andersson JL, Hutton C, Ashburner J, Turner R, Friston K (2001) Modelling geometric deformations in EPI time series. Neuroimage 13:903–919. CrossRef Medline

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

Borden GJ (1979) An interpretation of research on feedback interruption in speech. Brain Lang 7:307–319. CrossRef Medline

Borden GJ (1980) Use of feedback in established and developing speech. In: Speech and language: advances in basic research (Lass NJ, ed), pp 223–242. New York: Academic.

Buxton RB, Wong EC, Frank LR (1998) Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. Magn Res Med 39:855–864. CrossRef Medline

Christoffels IK, Formisano E, Schiller NO (2007) Neural correlates of verbal feedback processing: an fMRI study employing overt speech. Hum Brain Mapp 28:868–879. CrossRef Medline

Christoffels IK, van de Ven V, Waldorp LJ, Formisano E, Schiller NO (2011) The sensory consequences of speaking: parametric neural cancellation during speech in auditory cortex. PLoS One 6:e18307. CrossRef Medline

Curio G, Neuloh G, Numminen J, Jousmäki V, Hari R (2000) Speaking modifies voice-evoked activity in the human auditory cortex. Hum Brain Mapp 9:183–191. CrossRef Medline

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Ser B 39:1–38.

Dhanjal NS, Handunnethti L, Patel MC, Wise RJ (2008) Perceptual systems controlling speech production. J Neurosci 28:9969–9975. CrossRef Medline

Friston K (2009) Causal modelling and brain connectivity in functional magnetic resonance imaging. PLoS Biol 7:e33. CrossRef Medline

Friston K, Kiebel S (2009) Predictive coding under the free-energy principle. Philos Trans R Soc Lond B Biol Sci 364:1211–1221. CrossRef Medline

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. CrossRef Medline

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

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. CrossRef Medline

Grogan A, Green DW, Ali N, Crinion JT, Price CJ (2009) Structural correlates of semantic and phonemic fluency ability first and second languages. Cereb Cortex 19:2690–2698. CrossRef Medline

Guenther FH (2006) Cortical interactions underlying the production of speech sounds. J Commun Dis 39:350–365. CrossRef Medline

Guenther FH, Ghosh SS, Tourville JA (2006) Neural modeling and imaging of the cortical interactions underlying syllable production. Brain Lang 96:280–301. CrossRef Medline

Houde JF, Jordan MI (1998) Sensorimotor adaptation in speech production. Science 279:1213–1216. CrossRef Medline

Houde JF, Nagarajan SS (2011) Speech production as state feedback control. Front Hum Neurosci 5:82. CrossRef Medline

Hu W, Lee HL, Zhang Q, Liu T, Geng LB, Seghier ML, Shakeshaft C, Twomey T, Green DW, Yang YM, Price CJ (2010) Developmental dyslexia in Chinese and English populations: dissociating the effect of dyslexia from language differences. Brain 133:1694–1706. CrossRef Medline

Kiebel SJ, Klöppel S, Weiskopf N, Friston KJ (2007) Dynamic causal modeling: a generative model of slice timing in fMRI. Neuroimage 34:1487–1496. CrossRef Medline

Leaver AM, Rauschecker JP (2010) Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. J Neurosci 30:7604–7612. CrossRef Medline

Leff AP, Iverson P, Schofield TM, Kilner JM, Crinion JT, Friston KJ, Price CJ (2009) Vowel-specific mismatch responses in the anterior superior temporal gyrus: An fMRI study. Cortex 45:517–526. CrossRef Medline

Levene H (1960) Robust tests for equality of variances. In: Contributions to probability and statistics (Olkin I, ed), pp 278–292. Palo Alto, CA: Stanford UP.

Liang M, Mouraux A, Iannetti GD (2011) Parallel processing of nociceptive and non-nociceptive somatosensory information in the human primary and secondary somatosensory cortices: evidence from dynamic causal modeling of functional magnetic resonance imaging data. J Neurosci 31:8976–8985. CrossRef Medline

McIntosh AR, Gonzalez-Lima F (1994) Structural equation modeling and its application to network analysis in functional brain imaging. Hum Brain Mapp 2:2–22. CrossRef

Numminen J, Curio G (1999) Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. Neurosci Lett 272:29–32. CrossRef Medline

Osnes B, Hugdahl K, Specht K (2011) Effective connectivity analysis demonstrates involvement of premotor cortex during speech perception. Neuroimage 54:2437–2445. CrossRef Medline

Parker Jones ‘Ō, Green DW, Grogan A, Pliatsikas C, Filippopolitis K, Ali N, Lee HL, Ramsden S, Gazarian K, Prejawa S, Seghier ML, Price CJ (2012) Where, when and why brain activation differs for bilinguals and monolinguals during picture naming and reading aloud. Cereb Cortex 22:892–902. CrossRef Medline

Parkinson AL, Flagmeier SG, Manes JL, Larson CR, Rogers B, Robin DA (2012) Understanding the neural mechanisms involved in sensory control of voice production. Neuroimage 61:314–322. CrossRef Medline

Paus T, Perry DW, Zatorre RJ, Worsley KJ, Evans AC (1996) Modulation of cerebral blood flow in the human auditory cortex during speech: role of motor-to-sensory discharges. Eur J Neurosci 8:2236–2246. CrossRef Medline

Penny WD, Stephan KE, Mechelli A, Friston KJ (2004) Modelling functional integration: a comparison of structural equation and dynamic causal models. Neuroimage 23:S264–S274. CrossRef Medline

Penny WD, Stephan KE, Daunizeau J, Rosa MJ, Friston KJ, Schofield TM, Leff AP (2010) Comparing families of dynamic causal models. PLoS Comput Biol 6:e1000709. CrossRef Medline

Price CJ, Crinion JT, Macsweeney M (2011) A generative model of speech production in Broca's and Wernicke's areas. Front Psychol 2:237. CrossRef Medline

Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci 12:718–724. CrossRef Medline

Richardson FM, Thomas MS, Price CJ (2010) Neuronal activation for semantically reversible sentences. J Cogn Neurosci 22:1283–1298. CrossRef Medline

Richardson FM, Seghier ML, Leff AP, Thomas MS, Price CJ (2011) Multiple routes from occipital to temporal cortices during reading. J Neurosci 31:8239–8247. CrossRef Medline

Seghier ML, Price CJ (2010) Reading aloud boosts connectivity through the putamen. Cereb Cortex 20:570–582. CrossRef Medline

Seghier ML, Zeidman P, Neufeld NH, Leff AP, Price CJ (2010) Identifying abnormal connectivity in patients using dynamic causal modeling of fMRI responses. Front Syst Neurosci 4:142. CrossRef Medline

Seghier ML, Josse G, Leff AP, Price CJ (2011) Lateralization is predicted by reduced coupling from the left to right prefrontal cortex during semantic decisions on written words. Cereb Cortex 21:1519–1531. CrossRef Medline

Simmonds AJ, Wise RJ, Dhanjal NS, Leech R (2011) A comparison of sensory-motor activity during speech in first and second languages. J Neurophysiol 106:470–478. CrossRef Medline

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

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

Stephan KE, Penny WD, Moran RJ, den Ouden HE, Daunizeau J, Friston KJ (2010) Ten simple rules for dynamic causal modeling. Neuroimage 49: 3099–3109. CrossRef Medline

Takaso H, Eisner F, Wise RJ, Scott SK (2010) The effect of delayed auditory feedback on activity in the temporal lobe while speaking: a positron emission tomography study. J Speech Lang Hear Res 53:226–236. CrossRef Medline

Tourville JA, Reilly KJ, Guenther FH (2008) Neural mechanisms underlying auditory feedback control of speech. Neuroimage 39:1429–1443. CrossRef Medline

Veltman DJ, Mechelli A, Friston KJ, Price CJ (2002) The importance of distributed sampling in blocked functional magnetic resonance imaging designs. Neuroimage 17:1203–1206. CrossRef Medline

Ventura MI, Nagarajan SS, Houde JF (2009) Speech target modulates speaking induced suppression in auditory cortex. BMC Neurosci 10: 58. CrossRef Medline

Welch BL (1938) The significance of the difference between two means when the population variances are unequal. Biometrika 29:350–362. CrossRef

Welch BL (1947) The generalization of "Student's" problems when several different population varlances are involved. Biometrika 34:28–35. CrossRef Medline

Yates AJ (1963) Delayed auditory feedback. Psychol Bull 60:213–232. CrossRef Medline