frontiers in PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 16 September 2011
doi: 10.3389/fpsyg.2011.00237

A generative model of speech production in Broca's and Wernicke's areas

***Cathy J. Price*[1]*, Jenny T. Crinion*[2] *and Mairéad MacSweeney*[2]**

[1] *Wellcome Trust Centre for Neuroimaging, University College London, London, UK*
[2] *Institute of Cognitive Neuroscience, University College London, London, UK*

Edited by:
Albert Costa, University Pompeu Fabra, Spain

Reviewed by:
Xing Tian, New York University, USA
Pascale Tremblay, The University of Trento, Italy

****Correspondence:***
Cathy J. Price, Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, 12 Queen Square, London WC1N 3BG, UK.
e-mail: c.price@fil.ion.ucl.ac.uk

Speech production involves the generation of an auditory signal from the articulators and vocal tract. When the intended auditory signal does not match the produced sounds, subsequent articulatory commands can be adjusted to reduce the difference between the intended and produced sounds. This requires an internal model of the intended speech output that can be compared to the produced speech. The aim of this functional imaging study was to identify brain activation related to the internal model of speech production after activation related to vocalization, auditory feedback, and movement in the articulators had been controlled. There were four conditions: silent articulation of speech, non-speech mouth movements, finger tapping, and visual fixation. In the speech conditions, participants produced the mouth movements associated with the words "one" and "three." We eliminated auditory feedback from the spoken output by instructing participants to articulate these words without producing any sound. The non-speech mouth movement conditions involved lip pursing and tongue protrusions to control for movement in the articulators. The main difference between our speech and non-speech mouth movement conditions is that prior experience producing speech sounds leads to the automatic and covert generation of auditory and phonological associations that may play a role in predicting auditory feedback. We found that, relative to non-speech mouth movements, silent speech activated Broca's area in the left dorsal pars opercularis and Wernicke's area in the left posterior superior temporal sulcus. We discuss these results in the context of a generative model of speech production and propose that Broca's and Wernicke's areas may be involved in predicting the speech output that follows articulation. These predictions could provide a mechanism by which rapid movement of the articulators is precisely matched to the intended speech outputs during future articulations.

Keywords: speech production, auditory feedback, PET, fMRI, forward model

INTRODUCTION

Speech production is a complex multistage process that converts conceptual ideas into acoustic signals that can be understood by others. The stages include conceptualization of the intended message, word retrieval, selection of the appropriate morphological forms, sequencing of phonemes, syllables, and words, phonetic encoding of the articulatory plans, initiation, and coordination of sequences of movements in the tongue, lips, and laryngeal muscles that vibrate the vocal tract, and the control of respiration for vowel phonation and prosody. In addition to this feed forward sequence, auditory, and somatosensory processing of the spoken output is fed back to the motor system for online correction of laryngeal and articulatory movements (Levelt et al., 1999; Guenther et al., 2006). This self monitoring process is thought to be essential for learning to speak in a first (native) or second language but also plays a role in adult/fluent speech production, particularly when the auditory feedback is distorted. The sensorimotor interactions involved in monitoring the spoken response require an internal model of the intended speech to which the output can be matched (Borden, 1979; Paus et al., 1996; Heinks-Maldonado, 2005). The aim of the current study was to identify brain responses related to the internal model and to consider how these responses might predict auditory output prior to auditory or sensorimotor feedback.

The concept of internal models that predict the sensory consequences of an action is not specific to speech production. In the motor system, internal models that finesse motor control are referred to as "forward models" (Miall, 1993; Wolpert et al., 1995). More generally, forward models are examples of generative models that the brain may use for both perception (Helmholtz, 1866; MacKay, 1956; Gregory, 1980; Ballard et al., 1983; Friston, 2001, 2005) and active inference (Friston, 2010). The underlying principle of a generative model of brain function is that higher-level systems predict the inputs to lower-levels; and the resulting prediction error is then used to optimize future predictions – a scheme known as predictive coding.

Recent accounts of forward models in speech production have varied in how the auditory predictions and feedback are implemented (see **Figure 1**). For example, in the model proposed by Tian and Poeppel (2010), a motor efference copy is generated during motor planning and fed into a forward model of motor

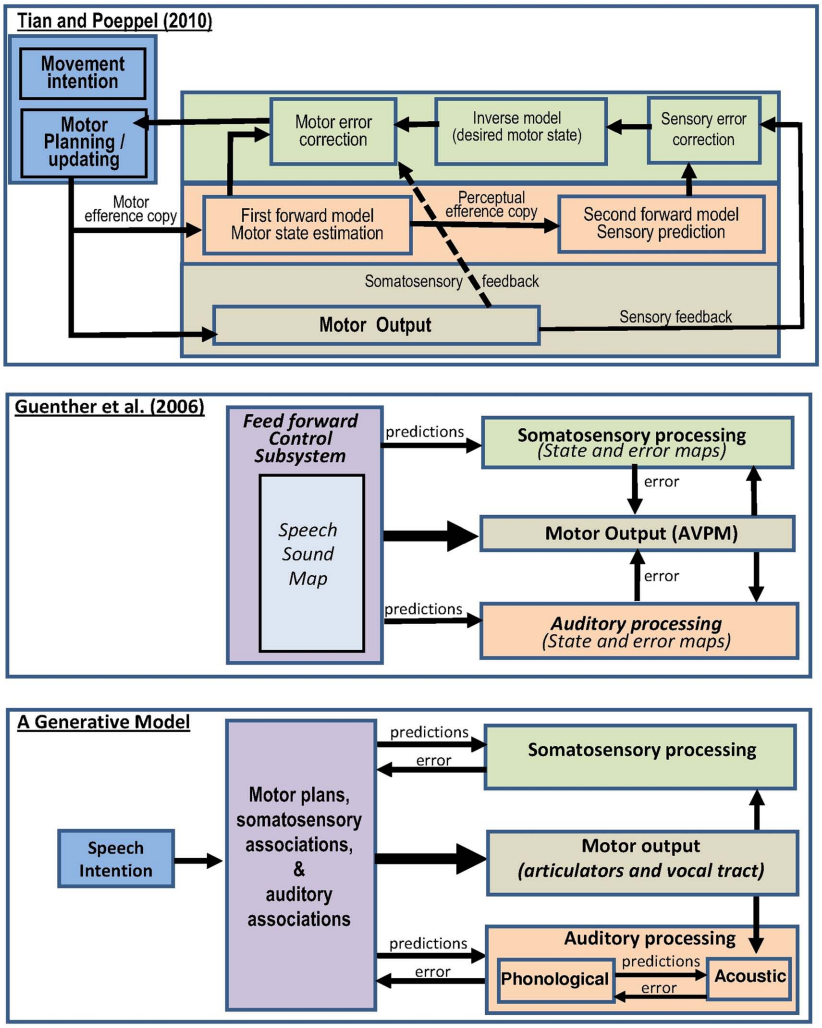

FIGURE 1 | Three different implementations of internal models of speech production. Top: Tian and Poeppel (2010) proposed model of motor control based on internal forward models and feedback. Middle: hypothesized processing stages involved in speech acquisition and production according to the DIVA model (directions into velocities of articulators), adapted from Guenther et al. (2006). AVPM, articulatory velocity and position maps. Below: a proposed generative model of speech production that is more consistent with the free energy and predictive coding framework (Friston, 2010).

processing which in turn feeds into a second forward model of sensory (auditory) processing (see first panel in **Figure 1**). This perspective differs from that proposed by Guenther et al. (2006) in which auditory and sensorimotor predictions are generated in parallel with motor commands (rather than in series as in the Tian and Poeppel, 2010 model), see second panel in **Figure 1**. The parallel processing of predictions and motor commands in Guenther et al. (2006) is more consistent with predictive coding accounts in generative models of active inference (Friston, 2010; Friston et al., 2010) where higher-level representations (i.e., prior knowledge of movements and their associations) drive the motor commands and predict the sensory responses in parallel (see third panel in **Figure 1**). However, in the Guenther et al. (2006)model, mismatches between the sensory response and the predicted sensory response (i.e., the prediction errors) are fed back to the motor system. This differs to the predictive coding in generative models (third panel of **Figure 1**) where the prediction errors are fed back to the source of the predictions (i.e., the high level representations) in order to optimize future predictions and minimize future prediction error. In addition, predictions in generative models are propagated in a hierarchical fashion through the system. For example, the third panel of **Figure 1** shows that higher-level representations predict phonological associations of words and phonological processing predicts acoustic associations of words, with potentially many intervening stages that are not illustrated.

Although, the importance of a forward model of speech output during articulation is well recognized (Heinks-Maldonado, 2005; Christoffels et al., 2007; Hawco, 2009), no previous functional imaging study has attempted to identify the anatomical location of brain activation related to the forward model of speech output during articulation. This requires an experimental paradigm that

activates speech production but controls for processing related to (a) auditory feedback and (b) movement of the articulators. Instead, previous functional imaging studies that have investigated the self monitoring of speech have primarily focused on activation related to auditory feedback rather than auditory predictions. This has involved altering rather than eliminating the auditory feedback (Paus et al., 1996; Hashimoto and Sakai, 2003; Ford et al., 2005; Fu et al., 2006; Christoffels et al., 2007; Toyomura et al., 2007; Tourville et al., 2008; Takaso et al., 2010). The results have highlighted activation changes in the superior temporal gyri but do not distinguish activation related to predicting speech from activation related to changes in auditory feedback. In contrast to this prior work, our study used a speech task that did not involve the generation of sound or auditory feedback because our aim was to identify brain activation that might be related to the internal model that predicts speech output during articulation.

To isolate brain activation associated with the internal model of speech output, we compared the production of speech to the production of non-speech mouth movements. The key difference between these conditions is that articulation of speech typically results in auditory speech processing whereas the production of non-speech mouth movements is not associated with auditory speech, although there may be some degree of acoustic association. In the speech condition, participants repeatedly articulated the words "one" and "three" without generating any sounds. This task places minimal demands on conceptualization of the intended message, word retrieval, the selection of the appropriate morphological forms, sequencing, respiration control, prosody, and auditory processing of the spoken output. However, silent articulation of words does not eliminate the experience of previously learnt auditory associations that have been tightly coupled with movement in the articulators during speech production (i.e., we have auditory imagery of the words "one" or "three" as they are silently articulated). These auditory images of speech may play a role in predicting the auditory consequences of speech production (Tian and Poeppel, 2010).

The words "one" and "three" were chosen because they have very distinct muscle movements that could be approximately matched in the non-speech mouth movement condition. Articulating "one" primarily involves lip pursing whereas articulating "three" primarily involves tongue protrusion and retraction. In the non-speech mouth movement condition, participants either pursed their lips (in a kissing action), protruded, and retracted their tongue or alternated between these movements. By including three different levels of non-speech mouth movements (lips repeatedly, tongue repeatedly, lips alternating with tongue), we were able to compare activation for different types of articulators (lips versus tongue) and also manipulate the complexity of the movements. For example, we were able to check whether increased activation for speech compared to non-speech was observed in areas where activation increased with the complexity of the movements (i.e., for alternating between different movements compared to repeatedly making the same movement).

Having controlled for auditory feedback and movement of the articulators, we predicted that activation related to the forward/generative model of auditory processing during speech production would be observed in the left ventral premotor cortex and/or the superior temporal gyrus/sulcus. These predictions are made on the basis of prior proposals by Guenther et al. (2006) who link the internal model of speech sound maps to the ventral premotor cortex; and Hickok et al. (2011) who link an internal model of motor processing to the premotor cortex; an internal model of auditory processing to the superior temporal gyrus/sulcus; and the translation between auditory and motor processing to the area they refer to as Spt (in the Sylvian fissure between the planum temporale (PT) and ventral supramarginal gyrus).

In addition to dissociating brain activation for speech and non-speech mouth movements, we also looked for activation that was common to both speech and non-speech mouth movements relative to finger tapping and visual fixation. Previous imaging studies have distinguished different systems involved in the motor control of speech: An articulatory "preparatory loop" that includes the inferior frontal, anterior insula, supplementary motor area, and superior cerebellum; an executive loop including the motor cortex, thalamus, putamen, caudate, and inferior cerebellum (Riecker et al., 2005) and a feedback loop including the postcentral gyri, the supratemporal plane, and the superior temporal gyri (Dhanjal et al., 2008; Peschke et al., 2009). The involvement of these regions in non-speech as well as speech mouth movements has already been demonstrated. For example, Chang et al. (2009) compared speech to non-speech orofacial movements and vocal tract gestures (whistle, cry, sigh, cough) and found common activations in the inferior frontal gyrus, the ventral premotor cortex, the supplementary motor area, the superior temporal gyrus, the insula, the supramarginal gyrus, the cerebellum, and the basal ganglia. This suggests a general role for these regions in orofacial movements and their auditory consequences.

By including a visual fixation baseline, we could also identify activation that was common to both finger and mouth movements; and control for inner speech that occurs independently of mouth movements during free thought.

MATERIALS AND METHODS

Functional imaging data were acquired using positron emission tomography (PET). For the current study of speech production there are two advantages of using PET rather than fMRI: the PET scanning environment is quieter for recording the presence or absence of speech output; and the regional cerebral blood flow (rCBF) signals are not distorted by air flow through the articulators. The study was approved by the local hospital ethics committee.

PARTICIPANTS

We scanned 12 right handed, native English speakers who had normal or corrected vision and hearing and no history of neurological disease or mental illness. All gave written informed consent. One participant was subsequently excluded for reasons given below. The remaining 11 subjects (10 male) had a mean age of 34 years (range 19–68). The predominance of male participants is a consequence of using PET scanning which is not appropriate for women of child bearing age. Our results did not change when the one female was removed ($n = 10$; mean age = 32 years, age range = 19–52) therefore we did not exclude the female participant. Inter-subject variability in our results was investigated and

reported (see **Figure 2**) to ensure consistency across participants, despite the wide range of ages and unequal distribution of males and females.

PARADIGM

There were four conditions: silent speech, non-speech mouth movements, finger tapping, and visual fixation. Each condition was repeated in three different blocks (with one block equivalent to one 90 s PET scan). In all 12 scans, a black circle, presented every 750 ms, was used as an external stimulus to pace movement production. During the three speech scans, participants were instructed to articulate the word "one" or "three" in time with the stimulus. They were specifically instructed to move their mouths as if they were speaking but without generating any sound (i.e., silent mouth movements). In one of the three speech scans, they articulated the word "one" on every trial; in a second, they articulated the word "three" on every trial and in the third, they alternated the articulation of "one" and "three," with one speech utterance per stimulus. In the three non-speech mouth movement scans, participants pursed their lips in time with the stimulus, protruded, and retracted their tongue, or alternated between pursing their lips and protruding and retracting their tongue. In the three-finger tapping scans, participants made a two-finger movement in one scan, a three-finger movement in another scan and alternated between the two-finger movement and three-finger movement in the third scan. The two-finger movement involved a tap of their index finger followed by a tap of their middle finger on a table placed under their arm in the scanner. The three-finger movement involved a tap of their index finger followed by a tap of their middle finger followed by a tap of their fourth finger. Participants practiced these movements before the scan and they were referred to as "double drum" and "triple drum" respectively. In the three baseline scans, participants were instructed "Please look at the flashing dot and try to empty your mind."

All responses, during all conditions were video recorded to ensure that the data collected were consistent with the experimental aims (e.g., mouth movements without sound during the speech condition). A scan/condition was repeated if the participants did not follow the instructions correctly. This only happened once for three different participants and in each case the repeated scan replaced the faulty scan. One subject (20-year-old male) did not follow the instructions in two different scans and was therefore excluded from the final analyses ($n = 11$). There was no further behavioral analysis because, in the final data sets, each condition was accurately performed (i.e., error free). Moreover, the functional imaging data showed no activation in the primary auditory cortex during any condition. This is consistent with the participants performing all conditions silently.

DATA ACQUISITION

Functional activation images were acquired using a SIEMENS/CPS ECAT EXACT HR+ (model 962) PET scanner (Siemens/CTI, Knoxville, TN, USA). Each participant had 12 or 13 PET scans (see previous section), to measure rCBF using bolus infusion of radioactively labeled water ($H_2{}^{15}O$). The dose received was 9 mCi per measurement, as approved by the UK Administration of Radioactive Substances Advisory Committee (ARSAC). Using statistical parametric mapping (SPM99), scans from each subject were realigned using the first as a reference, transformed into a standard MNI space (Ashburner and Friston, 1997) and smoothed

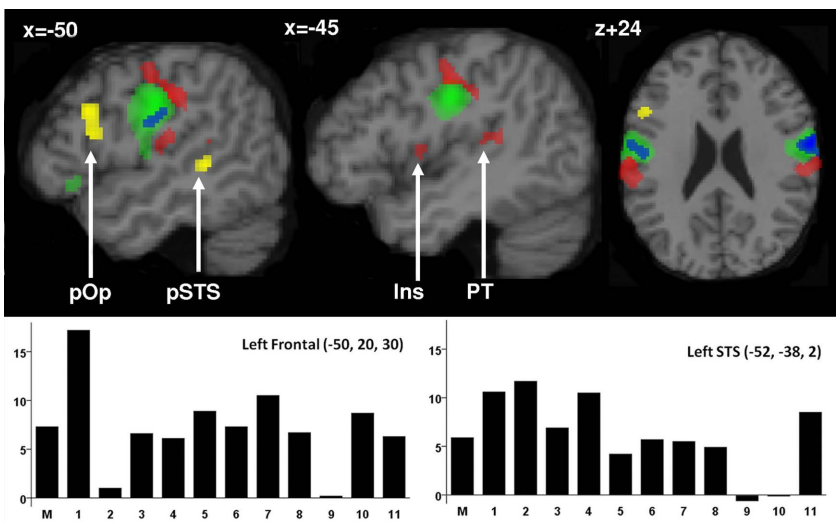

FIGURE 2 | Activation during silent articulation of speech. Top: Activation for speech more than non-speech mouth movements is illustrated in yellow in the pars opercularis (pOp) and the left posterior superior temporal sulcus (pSTS). Activation for speech and non-speech mouth movements relative to finger movements and fixation is illustrated in green. The blue area within this system corresponds to the location where activation was greater for tongue movements relative to lip movements. Activations for all movement tasks (mouth and finger) relative to fixation are illustrated in red. Within the red areas, we have marked activation that was located in the insula (INS) and the left planum temporale (PT). Statistical threshold was set at $p < 0.05$ after FWE correction for multiple comparisons across the whole brain in extent, see **Table 1** for details. Below: Activation for speech relative to non-speech mouth movements (percentage signal change on the *y* axis) in each participant (1–11 on the *x* axis) and the mean (M) at the peak co-ordinates for group activation in the frontal and temporal regions. This illustrates the consistency of the effect in the same voxels.

with a Gaussian kernel of 8 mm FWHM. Structural MRI images for each subject were obtained for coregistration with the PET data.

STATISTICAL ANALYSIS

Statistical analysis used standardized procedures (Friston et al., 1995). This involved ANCOVA with subject effects modeled and global activity included as a subject specific covariate. The condition and subject effects were estimated according to the general linear model at each voxel. The statistical model included 10 conditions: Fixation (summed over three scans), the three-finger tapping conditions, the three non-speech mouth movement conditions and the three speech conditions. The statistical contrasts of interest identified activation that was greater for (1) all speech than all non-speech mouth and finger conditions; (2) all speech than all non-speech mouth movements; (3) all speech and all non-speech mouth movements relative to all finger movements; and (4) all movement conditions relative to fixation; (4) non-speech tongue movements relative to non-speech lip movements or vice versa; and (5) alternating between movements or the same type (e.g., non-speech lip/tongue/lip) versus repetition of the same movement (e.g., non-speech tongue/tongue/tongue). The statistical threshold was set at $p < 0.05$ after family wise error (FWE) correction for multiple comparisons across the whole brain in height or extent. To ensure that activation in contrast (3) reflected common activation for all types of movement, we used the inclusive masking option on SPM to exclude voxels that were not significantly activated (at $p < 0.001$ uncorrected) by (6) speech > fixation, (7) non-speech mouth > fixation, and (8) finger movements > fixation. As the inclusive masking removes voxels from activation maps that are highly significant ($p < 0.05$ corrected), they make the results more conservative rather than less.

RESULTS

GREATER ACTIVATION FOR SILENT SPEECH THAN NON-SPEECH MOUTH MOVEMENTS

There were two areas where activation was significantly higher for silent speech than non-speech mouth movements: the left posterior superior temporal sulcus (pSTS) and the left dorsal pars opercularis within the inferior frontal gyrus extending into the left middle frontal gyrus. In each of these areas, activation was also higher for speech than finger movements and for speech relative to the visual fixation baseline. The loci and significance of these effects are shown in **Table 1** and **Figure 2**.

OTHER EFFECTS

Both speech and non-speech mouth movements resulted in extensive activation in bilateral pre-central gyri relative to finger tapping and visual fixation (see **Table 1** and green areas in **Figure 2** for details). In addition, activation that was common to speech, non-speech mouth movements, and finger tapping (relative to the

Table 1 | Location of activation for speech relative to non-speech mouth movements and finger movement; and for all movement tasks relative to fixation; at peaks that were significant at $p < 0.05$ after correction for multiple comparisons across the whole brain in height ($Z > 4.7$) or extent ($Z > 90$ voxels).

Location of speech activations	Speech > mouth and fingers					Speech > mouth	
	Co-ordinates (*x,y,z* in MNI)			*Z* score	*k*	*Z* score	*k*
Left posterior superior temporal sulcus	−52	−38	2	5.2	103	4.5	110
Left dorsal pars opercularis	−50	20	30	4.9	153	4.3	153
	Speech and mouth > fingers					**Tongue > lips**	
Left pre-central gyrus	−54	6	6	5.8	1658		
	−58	−2	14	7.4			
	−62	−6	26	7.5		4.5	139
	−48	−12	32	8.0			
Right pre-central gyrus	58	−4	10	6.0	1632		
	64	−6	26	8.1		5.8	234
	58	−8	30	8.0			
	Speech, mouth and fingers > fixation						
Left pre/post-central gyrus	−56	−12	16	6.7	1212		
	−58	−8	30	8.0			
	−48	−20	36	7.4			
	−46	−12	42	7.4			
Right post-central gyrus	+64	−14	30	5.6	576		
Left posterior cerebellum	−16	−60	−24	6.9	513		
Right posterior cerebellum	+28	−62	−24	7.7	903		
Left anterior insula	−38	2	+4	5.5	67		
Left planum temporale	−46	−38	+14	5.7	53		

k = number of voxels significant at p < 0.001 uncorrected.

visual fixation baseline) was observed bilaterally in the postcentral gyri, superior cerebellum, inferior cerebellum, putamen, with left lateralized activation in the thalamus, insula, supratemporal plane, and supplementary motor area (see **Table 1** and **Figure 2** which represents a subset of these regions in red). Common activation in these areas may relate to shared processing functions. For example, it has been proposed that activation in the anterior insula is related to the voluntary control of breathing during speech production (Ackermann and Riecker, 2010). It might therefore be the case that all three motor tasks (speech, non-speech mouth movements, and finger tapping) involve voluntary control of breathing in time with the motor activity. Alternatively, common activation might reflect different functions that could not be anatomically distinguished in the current study. As the current study is concerned with differential activation for speech relative to non-speech mouth movements, we do not discuss the common activations further.

The only other significant effect was observed when non-speech tongue movements were compared to non-speech lip movements. These effects are shown in blue in **Figure 2**. The MNI coordinates of this effect ($x = +64$, $y = -6$, $z = 26$; Z score = 5.9; and $x = -58$, $y = -6$, $z = 26$; Z score = 4.1) correspond to those previously reported for tongue movements (Corfield et al., 1999; Pulvermuller et al., 2006). The consistency of this effect with recent functional imaging (Takai et al., 2010) and early electrocortical mapping (Penfield and Rasmussen, 1950) provides reassuring support that our study had sufficient power to identify effects of interest with high precision. We did not see significantly increased activation for non-speech lip relative to mouth movements; nor did we see differential activation between any of the conditions that alternated between two movements (e.g., lips/mouth/lips) were compared to the corresponding conditions when the same movements was repeated continuously (e.g., lips/lips/lips or mouth/mouth/mouth).

DISCUSSION

Silently articulating the words "one" and "three" strongly activated left inferior frontal and superior temporal language regions compared to lip pursing, tongue movements, finger tapping, and visual fixation. The left inferior frontal activation was located in the left dorsal pars opercularis and therefore corresponds to classic Broca's area. The left superior temporal activation was located in the left pSTS and therefore corresponds to classic Wernicke's area. We suggest that, during speech production, activation in these classic language areas are related to covertly generated auditory associations that are evoked automatically, and in synchrony, with highly familiar mouth movements, previously intimately associated with sound production, and thus auditory feedback. In contrast, lip pursing, tongue, and finger movements are less practiced actions that are not intimately associated with speech sounds although they may have acoustic associations. The location and function of these activations is discussed below, in the context of generative models of perception and active inference (Friston, 2010; Friston et al., 2010). These data lead us to propose that Broca's and Wernicke's areas may play a role in predicting the auditory response during articulation, even in the absence of auditory feedback.

The activation in the dorsal pars opercularis extended anteriorly into the left inferior frontal sulcus (see **Figure 2**). It does not, therefore, correspond to the ventral premotor site of the speech sound maps proposed in the model by Guenther et al. (2006). It is also anterior to the more posterior premotor areas that respond during the observation of hand actions (Caspers et al., 2010), speech perception (Skipper et al., 2007; Callan et al., 2010), mirror neurons (Morin and Grezes, 2008; Kilner et al., 2009), and phonetic encoding during speech production (Papoutsi et al., 2009). Nevertheless, it does correspond to the area that is activated during both inner and overt speech tasks, for example, silent phonological decisions on written words (Poldrack et al., 1999; Devlin et al., 2003), lip reading (Fridriksson et al., 2009; Turner et al., 2009), overt speech production (Jeon et al., 2009; Whitney et al., 2009; Holland et al., 2011), and sentence comprehension (Bilenko et al., 2009; Mashal et al., 2009; Tyler et al., 2010). Moreover, it is not differentially activated by articulating words silently (as in the current study) or saying them aloud (see Price et al., 1996). Therefore the activation is more likely to reflect a fundamental property of speech production than atypical task-specific processing (e.g., the act of inhibiting the production of sounds following instructions to articulate silently). Given the minimal demands on conceptual, lexical, and auditory processing in the current study, we suggest that increased activation in the left dorsal pars opercularis for silently articulating words relative to non-speech mouth movements is related to higher-level representations of learnt words that predict the auditory consequences of well learnt speech articulations. Further we propose that these "predictions" are sent to auditory processing regions in the PT and the pSTS. Confirmation of this hypothesis requires a functional connectivity study with high temporal resolution to determine how activation in the left dorsal pars opercularis interacts with that in the superior temporal gyrus and sulcus.

The left pSTS activation that we observed during the silent articulation of speech is associated with phonological processing of speech sounds (Scott et al., 2000). The same STS area is also activated by written words in the absence of auditory inputs (Booth et al., 2003; Richardson et al., 2011). In addition, Leech et al. (2009) associated the left pSTS with learnt auditory associations. Specifically, they used a video game to train participants to associate novel acoustically complex, artificial non-linguistic sounds to visually presented aliens. After training, viewing aliens alone, with no accompanying sound, activated the left pSTS with activation in this area proportional to how well the auditory categories representing each alien had been learnt. As Leech et al. (2009) point out, part of what makes speech special is the extended experience that we have with it throughout development and this includes acoustic familiarity, enhanced audio–visual associations, and auditory memory in addition to the higher-level processing that is specific to speech (e.g., phonology and semantics). The activation that we observe in left pSTS may therefore reflect auditory associations of the articulated words. This might either be seen as a consequence of auditory predictions from the left dorsal pars opercularis and left pSTS may, in turn, play an active role in generating the predicted acoustic input during articulation (see the generative model in **Figure 1**). As acknowledged above, future functional connectivity studies using data

with high temporal resolution will be required to distinguish these alternatives.

We did not find speech-selective activation in the lower bank of the Sylvian fissure that has been referred to as the PT, left supratemporal plane (SPT), or Sylvian parietal temporal junction (Spt). The Sylvian fissure is the sulcus above the superior temporal gyrus but our speech-selective activation was in the pSTS which is the sulcus below the superior temporal gyrus. We did, nevertheless, confirm the involvement of PT/STP/Spt in speech production because we found common PT/STP/Spt activation for speech, mouth movements, and finger movements, relative to fixation. In other words, as shown previously (Binder et al., 2000), PT/STP/Spt was activated by speech but activation in this region was not specific to speech.

The observation of activation in PT/STP/Spt during finger tapping movements is surprising. Traditionally, PT has been considered to be an auditory association area that is important for speech but not more activated for speech than tone stimuli (Binder et al., 2000). More recent proposals suggest that the PT/STP/Spt region is an interface for speech perception and speech production (Wise et al., 2001; Hickok et al., 2009) and involved in anticipating the somatosensory consequences of movements in the articulators (Dhanjal et al., 2008). Our finding that PT/STP/Spt activation is observed for finger tapping and mouth movements might suggest an even more general role in sensorimotor processing. Alternatively, it might be the case that finger tapping and non-speech mouth movements have low level acoustic associations that are predicted during the movements that have previously been associated with such sounds. In other words, we are proposing that, during speech production, auditory predictions are generated at (a) the level of acoustic associations of any type of movement (in PT/STP/Spt) and (b) the phonology associated with learnt words (in pSTS), see lower part of **Figure 1**.

How do our results fit with the models illustrated in **Figure 1**? As emphasized above, the full answer to this question requires techniques with higher temporal resolution that can characterize how all the speech production areas interact and influence one another during articulation. Nevertheless, our data do allow us to test the anatomical hypotheses from the different models. Specifically, the Tian and Poeppel (2010) model suggests that the forward model of auditory processing is in the sensory cortex and the Guenther et al. (2006) model suggests their speech sound maps are in the ventral premotor cortex. In contrast, the effects that we observed for speech processing in the left dorsal pars opercularis and pSTS are in higher-level association areas, not in sensory areas or the ventral premotor cortex. The Spt activation that we observed for speech, non-speech, and finger tapping movements might plausibly correspond to the model proposed by Hickok et al. (2011) in which Spt translates an internal model of motor processing to an internal model of auditory processing. However, the Hickok et al. (2011) model does not provide an interpretation of our speech-selective activation in left dorsal pars opercularis or pSTS. Thus, the anatomical predictions of the previous models do not explain our data. We therefore propose a new anatomical model. Within the framework of the generative model, illustrated in **Figure 1**, we suggest that the activation we observed in the left dorsal pars opercularis corresponds to processing in higher-level areas that predicts the auditory and motor consequences of speech; and the pSTS activation corresponds to phonological processing that may be involved in predicting the auditory response in PT/STP/Spt. Future studies are now required to investigate the validity of this proposal and test how higher-level systems predict inputs to lower-levels; and how prediction error is used to optimize future predictions (Friston, 2010; Friston et al., 2010). We speculate that, during overt speech production, top-down predictions from higher-level areas optimize auditory processing of the heard response by minimizing the prediction error (i.e., the mismatch between the produced and predicted response). In parallel, the prediction error is fed back to the higher-level regions and used to optimize future motor commands and auditory predictions.

In conclusion, we found that regions corresponding to distinct parts of Broca's and Wernicke's areas were activated for mouth movements that have previously been learnt as words and therefore have well established auditory associations. We therefore suggest that the dorsal pars opercularis part of Broca's area and pSTS part of Wernicke's area are involved in predicting the auditory consequences of well rehearsed articulations. In addition, we propose that the left dorsal pars opercularis and pSTS areas may be involved in generating and maintaining a forward generative model of expected speech which can be used as a template for auditory prediction. Mismatches between the auditory predictions and auditory feedback can then be fed to the articulators to improve the precision of subsequent output. These audio–motor interactions are particularly important during speech acquisition in childhood, in those with hearing loss or when adults learn a new language. They are also needed to modify the intensity of speech output in noisy environments and when auditory feedback is altered (e.g., by delay on the telephone). We speculate that the devastating impact of damage to Broca's and Wernicke's areas on speech production may in part be related to the importance of dorsal pars opercularis and pSTS for auditory–motor integration of speech.

ACKNOWLEDGMENTS

This work was funded by the Wellcome Trust. We would like to thank Karl Friston for useful discussions and our radiographers (Amanda Brennan and Janice Glensman) for their help scanning the participants.

REFERENCES

Ackermann, H., and Riecker, A. (2010). The contribution(s) of the insula to speech production: a review of the clinical and functional imaging literature. *Brain Struct. Funct.* 214, 419–433.

Ashburner, J., and Friston, K. (1997). Multimodal image coregistration and partitioning – a unified framework. *Neuroimage* 6, 209–217.

Ballard, D. H., Hinton, G. E., and Sejnowski, T. J. (1983). Parallel visual computation. *Nature* 306, 21–26.

Bilenko, N. Y., Grindrod, C. M., Myers, E. B., and Blumstein, S. E. (2009). Neural correlates of semantic competition during processing of ambiguous words. *J. Cogn. Neurosci.* 21, 960–975.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., and Possing, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex* 10, 512–528.

Booth, J. R., Burman, D. D., Meyer, J. R., Gitelman, D. R., Parrish, T. B., and Mesulam, M. M. (2003). Relation between brain activation and lexical performance. *Hum. Brain Mapp.* 19, 155–169.

Borden, G. J. (1979). An interpretation of research of feedback interruption in speech. *Brain Lang.* 7, 307–319.

Callan, D., Callan, A., Gamez, M., Sato, M. A., and Kawato, M. (2010). Premotor cortex mediates perceptual performance. *Neuroimage* 51, 844–858.

Caspers, S., Zilles, K., Laird, A. R., and Eickhoff, S. B. (2010). ALE meta-analysis of action observation and imitation in the human brain. *Neuroimage* 50, 1148–1167.

Chang, S. E., Kenney, M. K., Loucks, T. M., Poletto, C. J., and Ludlow, C. L. (2009). Common neural substrates support speech and non-speech vocal tract gestures. *Neuroimage* 47, 314–325.

Christoffels, I. K., Formisano, E., and Schiller, N. O. (2007). Neural correlates of verbal feedback processing: an fMRI study employing overt speech. *Hum. Brain Mapp.* 28, 868–879.

Corfield, D. R., Murphy, K., Josephs, O., Fink, G. R., Frackowiak, R. S., Guz, A., Adams, L., and Turner, R. (1999). Cortical and subcortical control of tongue movement in humans: a functional neuroimaging study using fMRI. *J. Appl. Physiol.* 86, 1468–1477.

Devlin, J. T., Matthews, P. M., and Rushworth, M. F. (2003). Semantic processing in the left inferior prefrontal cortex: a combined functional magnetic resonance imaging and transcranial magnetic stimulation study. *J. Cogn. Neurosci.* 15, 71–84.

Dhanjal, N. S., Handunnetthi, L., Patel, M. C., and Wise, R. J. (2008). Perceptual systems controlling speech production. *J. Neurosci.* 28, 9969–9975.

Ford, J. M., Gray, M., Faustman, W. O., Heinks, T. H., and Mathalon, D. H. (2005). Reduced gamma-band coherence to distorted feedback during speech: when what you say is not what you hear. *Int. J. Psychophysiol.* 57, 143–150.

Fridriksson, J., Moser, D., Ryalls, J., Bonilha, L., Rorden, C., and Baylis, G. (2009). Modulation of frontal lobe speech areas associated with the production and perception of speech movements. *J. Speech Lang. Hear. Res.* 52, 812–819.

Friston, K. (2005). A theory of cortical responses. *Philos. Trans. Biol. Sci.* 360, 815–836.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138.

Friston, K. J. (2001). Dynamic representations and generative models of brain function. *Brain Res. Bull.* 54, 275–285.

Friston, K. J., Daunizeau, J., Kilner, J., and Kiebel, S. J. (2010). Action and behavior: a free-energy formulation. *Biol. Cybern.* 102, 227–260.

Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C., Frackowiak, R. S., and Turner, R. (1995). Analysis of fMRI time-series revisited. *Neuroimage* 2, 45–53.

Fu, C. H., Vythelingum, G. N., Brammer, M. J., Williams, S. C., Amaro, E. Jr., Andrew, C. M., Yágüez, L., van Haren, N. E., Matsumoto, K., and McGuire, P. K. (2006). An fMRI study of verbal self-monitoring: neural correlates of auditory verbal feedback. *Cereb. Cortex* 16, 969–977.

Gregory, R. L. (1980). Perceptions as hypotheses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 290, 181–197.

Guenther, F. H., Ghosh, S. S., and Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang.* 96, 280–301.

Hashimoto, Y., and Sakai, K. L. (2003). Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: an fMRI study. *Hum. Brain Mapp.* 20, 22–28.

Hawco, C. S. (2009). Control of vocalization at utterance onset and midutterance: different mechanisms for different goals. *Brain Res.* 1276, 131–139.

Heinks-Maldonado, T. H. (2005). Fine-tuning of auditory cortex during speech production. *Psychophysiology* 42, 180–190.

Helmholtz, H. (1866). "Concerning the perceptions in general," in *Treatise on Physiological Optics*, Vol. III, 3rd Edn. (translated by J. P. C. Southall 1925 Opt. Soc. Am. Section 26, reprinted New York: Dover, 1962).

Hickok, G., Houde, J., and Rong, F. (2011). Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron* 69, 407–422.

Hickok, G., Okada, K., and Serences, J. T. (2009). Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *J. Neurophysiol.* 101, 2725–2732.

Holland, R., Leff, A. P., Josephs, O., Galea, J. M., Desikan, M., Price, C. J., Rothwell, J. C., and Crinion, J. (2011). Speech facilitation by left inferior frontal stimulation. *Curr. Biol.* 21, 1–5.

Jeon, H. A., Lee, K. M., Kim, Y. B., and Cho, Z. H. (2009). Neural substrates of semantic relationships: common and distinct left-frontal activities for generation of synonyms vs. antonyms. *Neuroimage* 48, 449–457.

Kilner, J. M., Neal, A., Weiskopf, N., Friston, K. J., and Frith, C. D. (2009). Evidence of mirror neurons in human inferior frontal gyrus. *J. Neurosci.* 29, 10153–10159.

Leech, R., Holt, L. L., Devlin, J. T., and Dick, F. (2009). Expertise with artificial nonspeech sounds recruits speech-sensitive cortical regions. *J. Neurosci.* 29, 5234–5239.

Levelt, W. J., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–38; discussion 38–75.

MacKay, D. M. (1956). "The epistemological problem for automata," in *Automata Studies*, eds C. E. Shannon and J. McCarthy (Princeton, NJ: Princeton University Press), 235–251.

Mashal, N., Faust, M., Hendler, T., and Jung-Beeman, M. (2009). An fMRI study of processing novel metaphoric sentences. *Laterality* 14, 30–54.

Miall, R. C. (1993). Is the cerebellum a Smith predictor. *J. Mot. Behav.* 25, 203–216.

Morin, O., and Grezes, J. (2008). What is "mirror" in the premotor cortex? A review. *Neurophysiol. Clin.* 38, 189–195.

Papoutsi, M., and de Zwart, J. A., Jansma, J. M., Pickering, M. J., Bednar, J. A., and Horwitz, B. (2009). From rom phonemes to articulatory codes: an fMRI study of the role of Broca's area in speech production. *Cereb. Cortex* 19, 2156–2165.

Paus, T., Perry, D. W., Zatorre, R. J., Worsley, K. J., and Evans, A. C. (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.

Penfield, W., and Rasmussen, T. (1950). *The Cerebral Cortex of Man.* New York: The Macmillan Company.

Peschke, C., Ziegler, W., Kappes, J., and Baumgaertner, A. (2009). Auditory-motor integration during fast repetition: the neuronal correlates of shadowing. *Neuroimage* 47, 392–402.

Poldrack, R. A., Wagner, A. D., Prull, M. W., Desmond, J. E., Glover, G. H., and Gabrieli, J. D. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage* 10, 15–35.

Price, C. J., Moore, C. J., and Frackowiak, R. S. (1996). The effect of varying stimulus rate and duration on brain activity during reading. *Neuroimage* 3, 40–52.

Pulvermuller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., and Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proc. Natl. Acad. Sci. U.S.A.* 103, 7865–7870.

Richardson, F. M., Seghier, M. L., Leff, A. P., Thomas, M. S., and Price, C. J. (2011). Multiple routes from occipital to temporal cortices during reading. *J. Neurosci.* 31, 8239–8247.

Riecker, A., Mathiak, K., Wildgruber, D., Erb, M., Hertrich, I., Grodd, W., and Ackermann, H. (2005). fMRI reveals two distinct cerebral networks subserving speech motor control. *Neurology* 64, 700–706.

Scott, S. K., Blank, C. C., Rosen, S., and Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123, 2400–2406.

Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., and Small, S. L. (2007). Speech-associated gestures, Broca's area, and the human mirror system. *Brain Lang.* 101, 260–277.

Takai, O., Brown, S., and Liotti, M. (2010). Representation of the speech effectors in the human motor cortex: somatotopy or overlap? *Brain Lang.* 113, 39–44.

Takaso, H., Eisner, F., Wise, R. J., and Scott, S. K. (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.

Tian, X., and Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. *Front. Psychol.* 1:166. doi: 10.3389/fpsyg.2010.00166

Tourville, J. A., Reilly, K. J., and Guenther, F. H. (2008). Neural

mechanisms underlying auditory feedback control of speech. *Neuroimage* 39, 1429–1443.

Toyomura, A., Koyama, S., Miyamaoto, T., Terao, A., Omori, T., Murohashi, H., and Kuriki, S. (2007). Neural correlates of auditory feedback control in human. *Neuroscience* 146, 499–503.

Turner, T. H., Fridriksson, J., Baker, J., Eoute, D. Jr., Bonilha, L., and Rorden, C. (2009). Obligatory Broca's area modulation associated with passive speech perception. *Neuroreport* 20, 492–496.

Tyler, L. K., Shafto, M. A., Randall, B., Wright, P., Marslen-Wilson, W. D., and Stamatakis, E. A. (2010). Preserving syntactic processing across the adult life span: the modulation of the frontotemporal language system in the context of age-related atrophy. *Cereb. Cortex* 20, 352–364.

Whitney, C., Weis, S., Krings, T., Huber, W., Grossman, M., and Kircher, T. (2009). Task-dependent modulations of prefrontal and hippocampal activity during intrinsic word production. *J. Cogn. Neurosci.* 21, 697–712.

Wise, R. J., Scott, S. K., Blank, S. C., Mummery, C. J., Murphy, K., and Warburton, E. A. (2001). Separate neural subsystems within "Wernicke's area." *Brain* 124, 83–95.

Wolpert, D. M., Ghahramani, Z., and Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science* 269, 1880–1882.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 02 February 2011; accepted: 30 August 2011; published online: 16 September 2011.

Citation: Price CJ, Crinion JT and MacSweeney M (2011) A generative model of speech production in Broca's and Wernicke's areas. Front. Psychology ***2****:237. doi: 10.3389/fpsyg.2011.00237*

This article was submitted to Frontiers in Language Sciences, a specialty of Frontiers in Psychology.

Copyright © 2011 Price, Crinion and MacSweeney. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.