Neuropsychologia 50 (2012) 1380–1392

Contents lists available at SciVerse ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

# Individual differences in premotor and motor recruitment during speech perception

Gayaneh Szenkovits*, Jonathan E. Peelle[1], Dennis Norris, Matthew H. Davis

*Medical Research Council, Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, England, UK*

ARTICLE INFO

*Article history:*
Received 29 June 2011
Received in revised form
13 December 2011
Accepted 25 February 2012
Available online 12 April 2012

*Keywords:*
Speech perception
Dual stream model
Premotor cortex
Motor cortex
Individual differences
fMRI

ABSTRACT

Although activity in premotor and motor cortices is commonly observed in neuroimaging studies of spoken language processing, the degree to which this activity is an obligatory part of everyday speech comprehension remains unclear. We hypothesised that rather than being a unitary phenomenon, the neural response to speech perception in motor regions would differ across listeners as a function of individual cognitive ability. To examine this possibility, we used functional magnetic resonance imaging (fMRI) to investigate the neural processes supporting speech perception by comparing active listening to pseudowords with matched tasks that involved reading aloud or repetition, all compared to acoustically matched control stimuli and matched baseline tasks. At a whole-brain level there was no evidence for recruitment of regions in premotor or motor cortex during speech perception. A focused region of interest analysis similarly failed to identify significant effects, although a subset of regions approached significance, with notable variability across participants. We then used performance on a battery of behavioural tests that assessed meta-phonological and verbal short-term memory abilities to investigate the reasons for this variability, and found that individual differences in particular in low phonotactic probability pseudoword repetition predicted participants' neural activation within regions in premotor and motor cortices during speech perception. We conclude that normal listeners vary in the degree to which they recruit premotor and motor cortex as a function of short-term memory ability. This is consistent with a resource-allocation approach in which recruitment of the dorsal speech processing pathway depends on both individual abilities and specific task demands.

## 1. Introduction

A prevailing question in cognitive science is the degree to which the cognitive and neural systems engaged in a particular task are consistent across individuals. In most domains there is an assumption that these systems are relatively uniform, which enables the construction of generalisable neuroanatomically constrained models of cognitive processes in both health and disease. However, there is an increasing awareness of the role that individual differences in cognitive ability play in this process: because the availability of cognitive and neural resources varies across individuals, the particular instantiation of neural support for a given task will also vary (Seghier & Price, 2009). In the current study we examine how these individual differences manifest in speech processing, with a particular focus on the role of the motor system in speech perception. Although studies of speech perception generally reveal some involvement of premotor and motor regions (Devlin & Aydelott, 2009) there remains disagreement about whether this activity is an obligatory part of speech processing (D'Ausilio et al., 2009; Meister, Wilson, Deblieck, & Wu, 2007; Wilson, Saygin, & Sereno, 2004), or might instead reflect other associated processes (Hickok, 2010; Hickok et al., 2008; Scott & Wise, 2004; Scott, Mcgettigan, & Eisner, 2009). We hypothesised that inter-individual variability in cognitive ability may be one factor that contributes to the seemingly inconsistent pattern of results present in the literature. In the current study, we used a set of pseudoword processing tasks to investigate the role of motor areas in speech perception, and most importantly, whether the level of activity could be related to individual differences in behavioural measures.

Early models of spoken language processing suggested that sensory representations of speech interface with at least two systems: an articulatory motor system and a conceptual semantic system (e.g. Lichtheim, 1885/2006). This idea remains at the heart of contemporary neurocognitive models of speech processing (Hickok & Poeppel, 2007; Scott & Johnsrude, 2003). According to these dual-stream accounts, acoustic processing (in Heschl's gyrus and the superior temporal gyrus) is followed by at least two diverging processing streams. The *ventral stream* that projects towards

---

* Corresponding author. Tel.: +44 1223 273 631; fax: +44 1223 359 062.
*E-mail address:* gayaneh.szenkovits@mrc-cbu.cam.ac.uk (G. Szenkovits).

[1] Present address: Department of Neurology, University of Pennsylvania, Philadelphia, USA.

0028-3932/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2012.02.023

the left anterior and/or posterior inferior temporal regions, and serves as sound-to-meaning interface by mapping sound representations of speech onto conceptual representations. And the *dorsal stream* that projects towards the left posterior temporo-parietal junction, left supramarginal gyrus, premotor and motor cortices, and left inferior frontal gyrus, and serves as an auditory–motor interface by mapping speech sounds onto articulatory/motor representations. According to a traditional interpretation of this dual stream account, tasks selectively involving speech comprehension (e.g., listening to meaningful speech) are proposed to be primarily processed in the ventral stream, while tasks linking speech perception and production (e.g., repeating back heard speech) are thought to primarily engage the dorsal stream (Hickok & Poeppel, 2004, 2007; Saur et al., 2008; Scott and Johnsrude, 2003).

In recent years, however, a number of functional MRI studies have challenged this traditional view by reporting dorsal pathway involvement in speech perception tasks, even when no production component was required (Osnes, Hugdahl, & Specht, 2011; Pulvermüller et al., 2006; Wilson & Iacoboni, 2006; Wilson et al., 2004). These studies focused in particular on regions of left premotor and motor cortex that are the output of the dorsal speech pathway. For example, an influential study by Wilson et al. (2004) reported neural activation in the vicinity of the precentral gyrus (and premotor cortex) during passive listening to repeated consonant-vowel pseudowords compared to rest, whereas listening to a non-speech sound (a bell) did not trigger neural responses in the same regions. More direct evidence for premotor and motor cortex recruitment during speech perception comes from studies using Transcranial Magnetic Stimulation (TMS) to either facilitate (D'Ausilio et al., 2009; Fadiga et al., 2002; Watkins, Strafella, & Paus, 2003) or to temporarily disrupt (Meister et al., 2007; Möttönen & Watkins, 2009) processing in motor regions. Furthermore, TMS has been also shown to affect perceptual discrimination of speech sounds in an articulator-specific manner (D'Ausilio et al., 2009; Möttönen & Watkins, 2009). Taken together, the above findings have led some to conclude that regions of the dorsal pathway – namely, premotor and motor cortices – are critical for speech perception, and that when listening to speech both dorsal and ventral processing streams are necessarily recruited.

Despite these provocative findings, however, the degree to which premotor and motor processing is a necessary component of speech perception is still unclear, and recent reviews of the literature have called for caution when interpreting the above results (Hickok, 2008, 2009; Hickok and Poeppel, 2007; Lotto, Hickok, & Holt, 2009; Scott et al., 2009). One reason for caution is that the above findings are not always replicated; for example, in a repetitive TMS study of premotor regions, Sato, Tremblay, and Gracco (2009) reported slower phoneme discrimination only (requiring phonemic segmentation) but no effect on phoneme identification or syllable discrimination (Sato et al., 2009). In addition, several methodological points are consistently highlighted. First, the critical contrasts generally fail to show activation at a corrected level of significance. Second, functional imaging studies reporting motor activation have not made comparisons with a well-matched non-speech condition, and in studies using better controlled stimuli, there is little evidence for motor involvement (Obleser, Wise, Dresner, & Scott, 2007; Rodd, Davis, & Johnsrude, 2005; Rodd, Longe, Randall, & Tyler, 2010; Scott, Blank, Rosen, & Wise, 2000).

Perhaps most importantly, many studies that have found motor activation during speech perception have used phoneme identification or discrimination tasks (e.g. Möttönen & Watkins, 2009; Sato et al., 2009; Yuen, Davis, Brysbaert, & Rastle, 2009) which require attending to sublexical elements of speech (such as phonemes). These complex meta-phonological tasks (see Morais, Bertelson, Cary, & Alegria, 1986; Morais, Cary, Alegria, & Bertelson, 1979; Morais & Kolinsky, 1994) involve multiple processes (e.g. segmentation of speech into its constituent elements, decision making and/or categorisation), which may in turn depend on speech production or verbal short-term memory processes. For example, behavioural studies have found that articulatory suppression slows rhyme judgements but not homophony judgements of written words (Besner, 1987; Besner, Davies, & Daniels, 1981; Brown, 1987; Richardson, 1987), implying that subvocal articulatory processes may be involved in tasks involving manipulation of sublexical representations. More generally, it is well established that verbal short-term memory is supported by subvocal rehearsal processes (Baddeley & Hitch, 1974; Baddeley, Gathercole, Papagno, & Degli, 1998) and that these are mediated by the dorsal auditory pathway (Buchsbaum, Olsen, Koch, & Berman, 2005). Based on these findings, Hickok and colleagues suggest that the recruitment of subvocal rehearsal processes could explain much of the evidence for motor involvement in speech perception (Buchsbaum, Hickok, & Humphries, 2001; Hickok & Buchsbaum, 2003; Hickok, Buchsbaum, Humphries, & Muftuler, 2003).

Finally, consistent with this task-based explanation for premotor and motor activity, other studies have shown motor involvement in challenging listening conditions such as listening in noise, or using degraded or phonemically ambiguous stimuli (D'Ausilio et al., 2009; Davis & Johnsrude, 2003; Dufor, Serniclaes, Sprenger-Charolles, & Demonet, 2009; Osnes et al., 2011). Such challenging listening situations may similarly recruit short-term memory processes, and perhaps rely on a form of analysis-by-synthesis which recruits motor regions (Davis & Johnsrude, 2007; Skipper, van Wassenhove, Nusbaum, & Small, 2007; van Wassenhove, Grant, & Poeppel, 2005). In the same line, Callan, Callan, Gamez, Sato, and Kawato (2010) have reported that increased accuracy of phoneme identification in noise was associated with increased activation in the ventral part of the premotor cortex. Taken together, these findings might suggest that the function of the motor system in speech perception may be contingent upon perceptual ambiguity (Callan et al., 2010; Sato et al., 2009). While these additional task related cognitive processes are likely to recruit dorsal networks, they go beyond those brain regions associated with natural speech perception (Hickok & Poeppel, 2000; Osnes et al., 2011; Scott et al., 2009).

In this paper, we return to the question of whether premotor and motor cortices are necessarily activated during speech perception. To minimise ventral stream engagement and lexical/semantic effects, we use phonotactically legal pseudowords; thus, we operationally define speech perception as encompassing prelexical phonological processing. We avoid complex meta-phonological tasks and use a simple, one-back identity judgement with no requirement for either overt speech production or phonological segmentation, and minimal short-term memory load. We compare the neural activation required for this task to that in a one-back identity judgement involving complex non-speech stimuli, acoustically well matched to the speech stimuli on both spectral and amplitude characteristics. We also included two additional production tasks known to rely on the dorsal pathway (and for which we are therefore confident will show activation in premotor and motor cortices): reading aloud, and repeating heard pseudowords. These production tasks allow us to localise dorsal pathway regions in our cohort of participants and verify the efficacy of our general paradigm independently of our speech perception task.

Most important for the current study is our approach of explicitly examining individual differences in the degree of premotor and motor activation during speech perception. As discussed above, speech perception is often tested by phoneme identification or discrimination tasks. These tasks arguably involve segmentation, short-term memory, and subvocal rehearsal – likely supported by the dorsal stream. Moreover, all studies to date have conducted group analyses in which mean activation differences across subjects are compared to the null hypothesis of zero activation for speech compared to non-speech perception. We therefore aim to examine, for the first time, whether participants show linked variability between behaviour (as measured by tasks engaging the dorsal stream) and neural measure of speech perception. To do so, we will first characterise participants' behaviour separately on tasks that are suggested to rely on the dorsal stream: namely, phonological awareness and verbal short-term memory tasks. We will then relate the observed variability in performance to neural activation during the speech perception task that (as described above) is largely independent of segmental phonological awareness and short-term memory. Our prediction is that neural activity in the dorsal stream will reflect individual differences in cognitive ability.

## 2. Materials and methods

### 2.1. Participants

Twenty-one healthy, right-handed native speakers of British English participated in the study (9 men, average age 26.8 years, SD = 3.1, and 12 women, average age 22.8 years, SD = 7.9). They were recruited through the MRC Cognition and Brain Sciences Unit volunteer panel, and received £10 per hour for their participation. None of the participants reported any history of neurological, speech, or hearing disorder. All showed normal MRI structural scans. One participant was excluded from the analyses due to excessive head motion; the fMRI analyses reported here are on the remaining 20 participants. All participants were fully briefed and provided written informed consent. Ethical approval was granted by the Cambridge Psychology Research Ethics Committee.

### 2.2. Stimuli and experimental design

#### 2.2.1. Behavioural tasks to predict neural activity

We used a series of behavioural tests conducted outside the scanner to characterise participants' behaviour on complex speech processing and verbal short-term memory tasks frequently used in the literature, with the goal of linking individual abilities in these domains to neural activation in premotor and motor regions. Specifically, we assessed meta-phonological skills (the ability to consciously manipulate and evaluate speech segments) and verbal short-term memory capacity using the following tasks.

*Spoonerisms task.* We used the spoonerisms task to assess participants' meta-phonological ability. The spoonerisms task consisted of 40 pairs of spoken disyllabic English nouns with matching stress pattern. For half of the trials, participants were instructed to swap the initial sound of each word (e.g. 'chemist – leader' → [lɛmist – kiːdə ]) and for the other half to swap the final sound (e.g. 'fetish – scalpel' → [fetil – scælpəʃ ]). In both conditions participants were asked to maintain the original order of the words. Responses were recorded, and percent correct responses were averaged over the initial- and final-swap trials.

*Auditory and visual digit spans tasks*. The auditory digit span task is a computerised version of the WAIS-III subtest (Wechsler, 1998) which we used to help assess participants' verbal short term memory ability. Sequences of digits were presented aurally at a rate of 1 item per second. The task consisted of forward and backward repetitions. The sequence length is increased from two digits per sequence to nine digits per sequence in the forward condition and from two to eight digits in the backward condition. Participants were presented with two trials per length; the test finished when participants failed on both trials. The sum of all correct responses (sequences repeated correctly) provided participants' scores. The visual digit span task mirrored the auditory one, with the difference that the digits were presented in the middle of the screen for 500 ms, one by one, at a rate of 1 item every 500 ms. The same scoring was used as for the auditory digit task.

*Pseudoword repetition task.* As an additional measure of verbal short-term memory we used a pseudoword repetition task. In this task, participants had to repeat sequences of between two and six monosyllabic pseudowords. Items in the sequence were presented quickly, with an inter syllable interval of 100 ms. Since pseudowords lack long-term memory representations, pseudoword span should predominantly be driven by the capacity of phonological store and the efficiency of the articulatory rehearsal (Baddeley, 1998; Baddeley & Hitch, 1974) therefore providing a more accurate measure of verbal short-term memory capacity than digit or word span tasks (Cowan, 2001; Gathercole & Baddeley, 1990). Note that although pseudowords themselves have no long-term memory representations, memory for pseudowords is nevertheless sensitive to linguistic knowledge about the phonological structure of a given language (i.e., phonotactic probability). Previous studies have demonstrated that pseudowords composed of commonly co-occurring segments (high phonotactic probability) are repeated faster and more accurately than pseudoword composed of less common segments (low phonotactic probability) (Edwards, Beckman, & Munson, 2004; Vitevitch & Luce, 2005), and are also remembered better (Gathercole, Frankish, Pickering, & Peaker, 1999). In order to investigate the extent to which variability in short-term memory performance is influenced by language specific constraints, we included both high and low phonotactic probability pseudowords. We selected 52 consonant–vowel–consonant (CVC) pseudowords from (Gathercole et al., 1999). To assess the effects of information load on memory performance, and maximise variability on the task, sequence length was parametrically manipulated. Sequence length increased over successive presentation blocks beginning with sequences of two pseudowords and ending with sequences of six pseudowords. Each test consisted of 5 blocks of 12 trials (a total of 60 trials). At the end of each sequence participants heard an auditory cue to begin verbal recall. Participants were instructed to repeat each sequence in the correct order after the cue. In addition to online coding, spoken responses were also recorded. Participants were allowed to have a short break between the blocks. High and low phonotactic probability pseudoword repetitions were administered separately. Each test lasted approximately 15 min. The order of the tests as well as the order of the experimental session (before or after fMRI) was counterbalanced across participants.

All stimuli were recorded by a native female speaker of Southern British English, at 44,100 Hz sampling rate, and were edited offline using Adobe Audition (Adobe System Corporation, San Jose, CA). Experiments were programmed and run using E-Prime® (Psychology Software Tools, Inc., Pittsburgh, PA) and DMDX (Forster & Forster, 2003) software.

#### 2.2.2. Imaging tasks

In order to investigate the role of premotor and motor regions in speech perception, we used three tasks. Two of the tasks (reading and repetition) required speech production, which should involve obligatory premotor and motor cortex activation. The third task, a perception task using the same materials, was our critical test condition.

In order to minimise lexical and semantic effects and to tap networks underling phonological input and output systems, all stimuli in the imaging experiments were CVC pseudowords. A total of 360 pseudowords were created and were recorded by two speakers (one male and one female, both native speakers of British English) at 44,100 Hz sampling rate. For the non-speech baseline control conditions, the pseudoword recordings were passed through a single-channel pulse-train vocoder (Deeks & Carlyon, 2004) implemented using Praat software (www.praat.org). This procedure generated a buzzy sound (a pulse train) filtered to have the same long-term spectrum and amplitude envelope as the original pseudowords, and thus, well matched for relevant acoustic properties including the presence of pitch, harmonic spectral structure and a slowly fluctuating amplitude envelope. For each pseudoword, three control stimuli were constructed with an F0 of 100, 150 or 200 Hz, introducing pitch variability in addition to the intrinsic variability between different pseudowords in their amplitude envelope.

All tasks had the same timing characteristics and used the same blocked-design where test and control (or baseline) stimuli alternated in 12.6 s blocks as recommended for tasks involving overt speech production (Soltysik & Hyde, 2006, 2008). In addition, a silent inter-block interval of 2 s was included. Blocks consisted of 6 stimuli that were presented with a 2.1 s stimulus onset asynchrony. Scanning runs consisted of 40 blocks (20 test and 20 control blocks) and took approximately 10 min to complete. The specific tasks were as follows.

*Pseudoword reading vs. visual–motor baseline.* In the reading task participants had to read aloud short monosyllabic pseudowords. These were printed on the screen in black Times New Roman font, with 36-point size. In the control condition participants were presented with unpronounceable consonant strings (e.g. xtqs), and had to say 'yes' to acknowledge them. These consonant strings were matched in length to the average length of the pseudowords (4 letters). To make the consonant strings more salient, they were displayed on the screen in blue.

*Pseudoword repetition vs. auditory–motor baseline.* The repetition task was designed to engage both speech perception and production. Participants listened to a series of monosyllabic pseudowords and were instructed to repeat each back immediately. In the control task, participants heard matched non-speech buzzes (as used as a baseline in the speech perception task) and had to say 'yes' after each buzz.

*Speech perception vs. auditory baseline.* In the speech perception task, participants listened to short monosyllabic pseudowords and were instructed to press a button with their left hand when they detected two successive presentations of the same syllable (one-back task). Only 10% of the stimuli were repeated. In order to prevent participants relying on low-level acoustic information, the presentation of the auditory pseudowords alternated between a male and female voice. Hence, judgements of repetition depend on abstract phonological comparisons but do not require division of pseudowords into segments or other meta-phonological abilities. In the control condition, participants listened to the non-speech buzzes and again had to detect immediate repetitions (i.e. two successive stimuli with the same pitch and amplitude envelope) with a left hand button press.

Pseudowords, consonant strings and buzzes were pseudo randomly distributed and counterbalanced across the three tasks and participants such that no two stimuli were presented twice. All control tasks were well matched for stimulus and response characteristics. Participants' spoken responses were recorded with a FOMRI MRI safe noise-cancelling microphone (Opto-Acoustics Ltd., Or-Yehuda, Israel) for offline analysis. Prior to scanning, participants took part in a short practice session outside of the scanner, during which they were familiarised with the tasks. The practice session included four blocks of each tasks.

**Table 1**
Descriptive statistics of the behavioural experiments for 18 participants. CR = correct responses.

| | Descriptive statistics | | | | | |
|---|---|---|---|---|---|---|
| | *N* | Minimum | Maximum | Mean | Std. deviation | Variance |
| Participants | 18 | | | | | |
| Age | 18 | 19 | 35 | 23.5 | 4.13 | 17.08 |
| Men | 7 | 19 | 35 | 24.7 | 5.3 | 28.2 |
| Women | 11 | 19 | 29 | 22.7 | 3.2 | 10.4 |
| High phonotactic probability pseudowords | | | | | | |
| Length 2 | 17 | 87.5 | 100 | 98.28 | 3.914 | 15.32 |
| Length 3 | 17 | 83.33 | 100 | 94.28 | 5.51 | 30.35 |
| Length 4 | 17 | 50 | 95.83 | 75 | 12.84 | 164.93 |
| Length 5 | 17 | 25 | 73.33 | 51.08 | 12.15 | 147.55 |
| Length 6 | 17 | 18.06 | 59.72 | 38.97 | 11.12 | 123.57 |
| Low phonotactic probability pseudowords | | | | | | |
| Length 2 | 17 | 87.5 | 100 | 98.04 | 3.33 | 11.11 |
| Length 3 | 17 | 77.78 | 100 | 88.89 | 6.87 | 47.26 |
| Length 4 | 17 | 43.75 | 91.67 | 65.93 | 15.41 | 237.6 |
| Length 5 | 17 | 16.67 | 66.67 | 41.18 | 15.84 | 250.96 |
| Length 6 | 17 | 13.89 | 40.28 | 25.57 | 8.63 | 74.52 |
| Average of all high phonotactic probability pseudowords (% CR sequence) | 17 | 57.5 | 85.36 | 71.52 | 6.64 | 44.12 |
| Average of all low phonotactic probability pseudowords (% CR sequence) | 17 | 51.81 | 74.06 | 63.92 | 7.56 | 57.23 |
| Auditory digit | 18 | 14 | 29 | 19.67 | 3.97 | 15.76 |
| Visual digit | 18 | 12 | 27 | 18.055 | 3.9 | 15.23 |
| Spoonerisms first phoneme swap (% CR) | 17 | 35 | 100 | 70.29 | 18.32 | 3.4 |
| Spoonerisms last phoneme swap (% CR) | 17 | 20 | 95 | 61.18 | 20.58 | 4.2 |
| Spoonerisms first and last phoneme average (% CR) | 17 | 40 | 98 | 65.74 | 17.16 | 2.9 |

### 2.3. Image acquisition and preprocessing

The imaging data were acquired with a 3T Siemens Tim Trio MRI system with a 12 channel head coil. Stimuli were presented over high quality electrostatic headphones built into ear defenders (NordicNeuroLab, Bergen, Norway). Participants were instructed to stay as still as possible during the scan and to avoid excessive head movement while speaking. We acquired 312 echo planar (EPI) volumes in each of the three 10 minute sessions. Each volume consisted of 32 × 3 mm thick slices with 0.75 mm inter-slice gap, TR = 2000 ms, TA = 2000 ms, field of view 19.2 × 19.2 cm, acquisition matrix 64 × 64, echo time 30 ms, flip angle 78°, and in-plane resolution of 3 × 3 mm. The acquisition was transverse oblique, angled to avoid the eyes and to achieve whole-brain coverage including the cerebellum. In a few cases the very top of the parietal lobe was not covered; this did not affect coverage of motor cortex. High-resolution 1 × 1 × 1 mm MPRAGE anatomical images were collected for anatomic localisation and coregistration.

SPM5 was used for image preprocessing and data analysis (Wellcome Trust Centre for Neuroimaging, London, UK). After discarding 7 initial scans for each session to allow for T2 equilibrium, images for each participant were corrected for motion by spatial realignment to the first image in the series, using a least squares approach with 6 rigid body parameters (Friston et al., 1995). Following realignment, the images were corrected for differences in slice time acquisition and coregistered with the structural image (Ashburner & Friston, 1997) which was then segmented and normalised (using affine and smoothly nonlinear transformations) to a brain template in Montreal Neurological Institute (MNI) space (Ashburner & Friston, 2005). The resulting normalisation parameters were applied to all the coregistered EPIs. Finally, the EPI images were smoothed with a 10 mm full-width at half-maximum isotropic Gaussian kernel.

Data were first analysed separately for each participant, using a separate general linear model for each session (perception, reading and repetition). Low-frequency noise was removed with a 128 s high-pass filter. Individual stimuli and button presses were separately modelled using delta functions convolved with the canonical hemodynamic response function to create the regressors used in the model. The 6 motion parameters obtained during realignment were also included in the model as additional regressors of no interest. Trials with button press were modelled out in the analysis. In the perception run, we only analysed trials that resulted in a correct response; in the localiser runs (reading and repetition), all trials were included.

Contrasts of parameter estimates from the least mean-square fit of these single-subject analyses were then entered into the second level random-effects analyses (one sample t-tests). Contrasts of interest were: pseudowords vs. control buzzes in the listening and repeating tasks, and pseudowords vs. control consonant strings in the reading task. In the speech perception task, only trials without button press were included in the analysis. Unless otherwise specified, results are reported at a whole brain-corrected level of significance at pFWE <.05 (voxelwise). Family-wise correction was achieved by using Random Field Theory as implemented in SPM (Friston, Frith, Liddle, & Frackowiak, 1991). To ensure that critical results are not omitted, results for the perception task are also reported at whole brain uncorrected $p < .001$ and qFDR < .05 levels of significance (see Table 2 and Fig. 2). Data for the region of interest (ROI) analyses were extracted using MarsBar (http://marsbar.sourceforge.net).

## 3. Results

### 3.1. Behavioural tasks

Descriptive statistics for the behavioural tasks are presented in Table 1. Out of the 20 participants, 18 completed the behavioural tasks. Due to a technical problem, one participant did not complete the pseudoword repetition and the auditory digit span tasks, and another did not complete the spoonerisms task. In total, complete behavioural datasets were acquired for 16 participants. For the pseudoword repetition task, analyses were run on the percentage of words correctly repeated. For the digit span tasks, analyses were run on number of correctly recalled sequences.

Pseudoword repetition accuracy is shown in Fig. 1. To assess the effect of phonotactic probability and sequence length on repetition accuracy, we conducted a repeated-measures ANOVA with Phonotactic probability (high vs. low) and Length (5 levels) as within subject variables. This analysis revealed significant main effects of both Phonotactic probability $F(1,16) = 25.14$, $p < .001$ and Length $F(1,16) = 299.21$ $p < .001$. The interaction between Phonotactic Probability and Length was also significant $F(1,16) = 5.59$, $p < .01$. These results replicate previous findings (Gathercole et al., 1999; Vitevitch & Luce, 2005) indicating that memory for high phonotactic probability pseudowords is better then memory for low phonotactic probability pseudowords, and better for short sequences than for long ones. Post hoc comparisons with paired sample *t*-tests revealed that the interaction is driven by length 2, the only sequence length where performance did not differentiate between high and low phonotactic probability pseudowords ($t(16) = 0.187$, $p = .85$, two-tailed). For all the other lengths, participants performed better with high phonotactic probability pseudowords (length 3: $t(16) = 2.716$, $p = .015$; length 4: $t(16) = 3.03$, $p = .008$; length 5: $t(16) = 3.3$, $p = .004$; length 6: $t(16) = 5.212$, $p < .001$, all two-tailed).

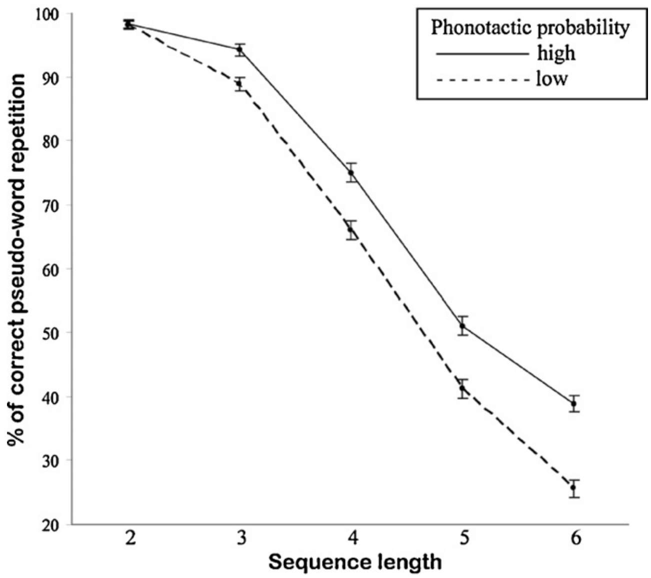

**Fig. 1.** Performance on high (solid line) and low (dashed line) phonotactic probability pseudoword repetitions as a function of sequence length. Error bars indicate standard error of the mean after between-subject variability has been removed, appropriate for repeated measures comparisons (Loftus & Masson, 1994).

In the spoonerisms tasks, the average accuracy (±SD) for initial sound swap was 70% (±17), and 61% (±19), for the final sound swap. Overall accuracy was 65.7% (±16). In the auditory digit span task the average score was 19.1 (±3.21), in the visual digit span task it was 18.05 (±3.79).

## 3.2. *Imaging results*

In order to make sure participants did not fall asleep and performed the tasks appropriately, behavioural performance was monitored online. This confirmed that participants followed task instruction in the localiser tasks (reading and repetition), and produced pseudowords and 'yes' answers as appropriate. Only behavioural data from the speech perception tasks (experiment of interest) was analysed further. The average detection accuracy for pseudowords was 91.2% (SD = 13.1%) and 90.4% (SD = 13.3%) for the buzzes, indicating that participants performed the speech perception and control task reliably.

### 3.2.1. *Whole-brain analysis*

We first examined brain areas that showed significant activation during reading, repetition, and perception tasks compared to their corresponding control conditions, shown in Fig. 2 and Table 2. Results were generally consistent with the findings of previous studies on speech perception and production (see Price, 2010; Price et al., 1996). For pseudoword reading, at pFWE <.05 statistical threshold, we observed extensive activation of bilateral motor and premotor cortices, left inferior frontal gyrus (LIFG), left inferior temporal gyrus (LIT), the right cerebellum and the supplementary motor area (SMA). In addition, a region in the superior temporal gyrus (STG) was activated, presumably reflecting a response to participants' own speech (Hashimoto & Sakai, 2003; Zheng, Munhall, & Johnsrude, 2010). The pseudoword repetition task revealed activation that overlapped with networks for both speech perception and production, namely bilateral premotor and motor cortices, left middle temporal gyrus (MTG), as well as LIFG and SMA. In addition the left putamen were also found to be active. *Speech perception* at pFWE <.05 statistical threshold showed a much more restricted pattern of activation encompassing portions of left inferior and middle temporal gyri. Because we wanted to ensure we were not missing sub-threshold effects for this critical contrast, we repeated this contrast at a voxelwise uncorrected threshold of $p < .001$ (which nonetheless exceeds qFDR <.05). At this threshold, we observed the bilateral MTG, LIFG (pars orbitalis), left precuneus, right angular gyrus, LIT and left fusiform gyus. Frontal responses during speech perception were confined to the LIFG (pars orbitalis) and did not extend into premotor or motor cortex.

### 3.2.2. *Region of interest analysis*

Following the failure to find whole-brain evidence for premotor involvement during speech perception, we conducted a follow-up region of interest (ROI) analysis. To avoid issues of nonindependence (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009), two sets of independent ROIs were determined to sample

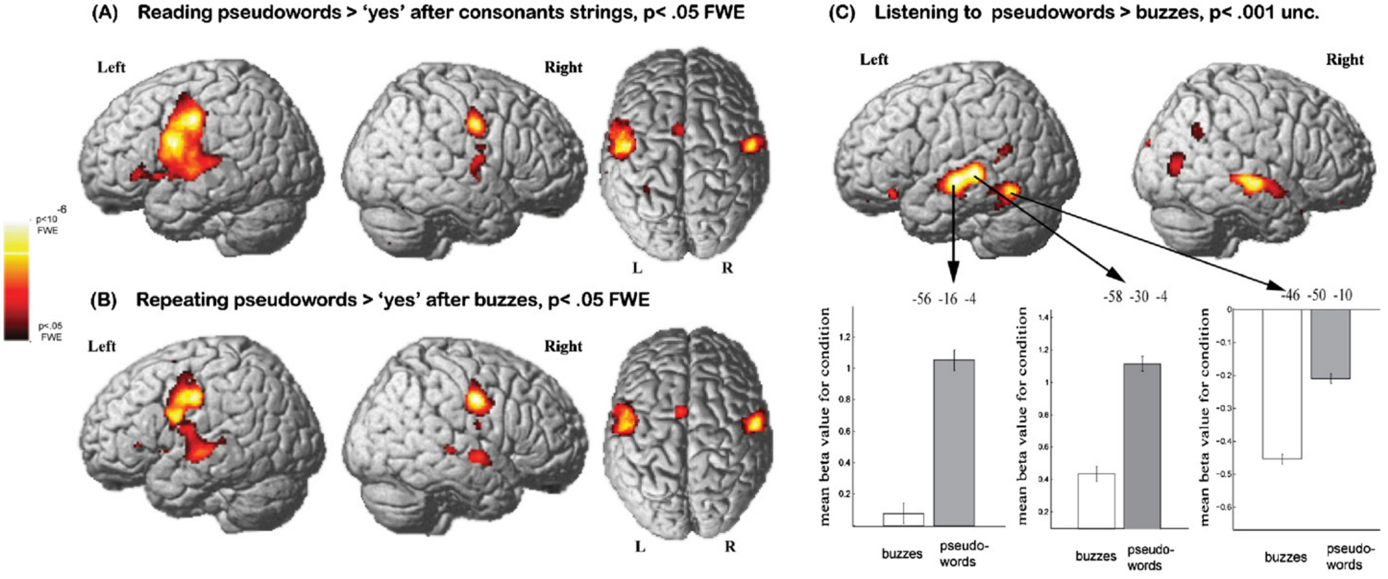

**Fig. 2.** Brain regions showing response differences between task (pseudowords) and basline (non-speech sounds/consonant strings). Activation maps are rendered onto the MNI canonical brain, thresholded at pFWE < .05 (A) and (B) and at $p < .001$ (uncorrected) (C). (C) Lower panel shows bar chart of the mean BOLD parameter estimate of the peak voxels in the left middle temporal gyrus and left infero-temporal gyrus that reached $p < .05$ FWE corrected level for pseudowords (grey) and buzzes (white). Error bars show standard error of the mean after between-subject variability has been removed, suitable for repeated measures comparisons (Loftus & Masson, 1994).

**Table 2**
MNI Coordinates for peak voxels of increased activity for all speech versus non-speech contrasts. We report a maximum of 20 local maxima that are more than 8 mm apart.

| Location | Cluster size (Voxels) | *Z*-scores | Coordinates | | |
|---|---|---|---|---|---|
| | | | *x* | *y* | *z* |
| (A) Reading pseudowords > 'yes' to consonant string | | | | | |
| Right Postcentral G. | 437 | 6.66 | 56 | −8 | 40 |
| Left superior temporal G. | 3656 | 6.56 | −58 | −8 | 0 |
| Left postcentral G. | | 6.54 | −50 | −10 | 38 |
| Left rolandic operculum | | 6.44 | −56 | −2 | 10 |
| Left precentral G. | | 6.23 | −50 | 2 | 26 |
| Left rolandic operculum | | 5.88 | −42 | −14 | 24 |
| LIFG – pars opercularis | | 5.79 | −54 | 10 | 12 |
| Left insula | | 5.59 | −36 | −8 | 18 |
| Left superior temporal G. | | 5.44 | −56 | −12 | 12 |
| Left superior temporal G. | | 5.39 | −58 | −24 | 8 |
| Left precentral G. | | 5.05 | −44 | −6 | 52 |
| Left insula | 852 | 6.04 | −26 | 18 | 6 |
| LIFG – pars triangularis | | 5.45 | −38 | 32 | −4 |
| Left pallidum | | 5.35 | −18 | 8 | 0 |
| Left pallidum | | 5.26 | −18 | 0 | −6 |
| Right cerebellum (8) | 108 | 5.9 | 24 | −70 | −50 |
| Left middle cingulum | 214 | 5.71 | −10 | 14 | 38 |
| SMA | | 5.28 | −6 | 4 | 58 |
| Left middle cingulum | | 4.85 | −6 | 6 | 44 |
| Right rolandic operculum | 116 | 5.62 | 66 | −6 | 12 |
| Right superior temporal G. | | 4.81 | 62 | −6 | −4 |
| Left inferior parietal lobule | 22 | 5.24 | −32 | −48 | 48 |
| Left thalamus | 46 | 5.14 | −14 | −14 | 6 |
| Right rolandic operculum | 36 | 5.09 | 46 | −6 | 18 |
| Right cerebellum (8) | 43 | 4.92 | 22 | −62 | −24 |
| Left inferior temporal G. | 15 | 4.85 | −44 | −52 | −10 |
| Inferior colliculus | 6 | 4.82 | −8 | −22 | −6 |
| (B) Repeating pseudowords > 'yes' to buzzes | | | | | |
| Right postcentral G. | 774 | 6.66 | 54 | −8 | 38 |
| Right rolandic operculum | | 5.28 | 46 | −10 | 20 |
| Left putamen | 1004 | 6.33 | −26 | 2 | −4 |
| Left putamen | | 6.06 | −24 | 10 | 6 |
| Left putamen | | 5.82 | −22 | −4 | 10 |
| Left caudate nucleus | | 5.57 | −14 | 6 | 12 |
| Left thalamus | | 5.13 | −10 | −12 | 4 |
| Right superior temporal G. | 156 | 6.3 | 62 | −6 | −4 |
| Left precentral G. | 2085 | 6.26 | −50 | −2 | 38 |
| Left postcentral G. | | 6.21 | −46 | −12 | 36 |
| Left middle temporal G. | | 6.06 | −64 | −12 | −4 |
| Left precentral G. | | 6.04 | −52 | 2 | 26 |
| Left precentral G. | | 5.65 | −42 | −2 | 28 |
| Left rolandic operculum | | 5.51 | −54 | −6 | 12 |
| Left rolandic operculum | | 5.41 | −42 | −2 | 18 |
| Left rolandic operculum | | 5.14 | −46 | −10 | 18 |
| Left precentral G. | | 4.96 | −46 | −4 | 52 |
| Left middle temporal G. | | 4.92 | −62 | −32 | 4 |
| SMA | 139 | 6.03 | −6 | 2 | 62 |
| Right middle temporal G. | 36 | 5.37 | 62 | −28 | 0 |
| SMA | | 5.3 | −8 | 16 | 44 |
| Right putamen | 63 | 4.94 | 24 | 4 | −4 |
| Right putamen | | 4.92 | 18 | 10 | 8 |
| LIFG – pars triangularis | 10 | 4.89 | −42 | 30 | 2 |
| Right pallidum | 1 | 4.75 | 26 | −2 | −4 |
| (C) Listening to pseudowords > buzzes | | | | | |
| **Left inferior temporal G.** | **643** | **5.34**$^{**}$ | −46 | −50 | −10 |
| Left inferior temporal G | | 4.37 | −42 | −46 | −16 |
| **Left middle temporal G.** | **934** | **5.08**$^{**}$ | −56 | −16 | −4 |
| **Left middle temporal G.** | | **5.06**$^{**}$ | −58 | −30 | 4 |
| Right superior temporal G. | 614 | 4.68 | 66 | −14 | −6 |
| Right middle temporal G. | | 3.78 | 60 | 0 | −12 |
| Right middle temporal G. | 107 | 4.33 | 52 | −76 | 6 |
| Left precuneus | 728 | 3.95 | −4 | −58 | 38 |
| Left precuneus | | 3.74 | −14 | −54 | 40 |
| LIFG – pars orbitalis | 47 | 3.49 | −40 | 30 | −14 |
| Left superior temporal G. | 47 | 3.42 | −64 | −54 | 20 |
| *Left superior temporal G.* | | *3.12 #* | −58 | −46 | 8 |
| Right angular G. | 33 | 3.35 | 60 | −60 | 34 |
| Right cuneus | 16 | 3.33 | 18 | −96 | 24 |
| Right middle temporal pole | 2 | 3.21 | 50 | 18 | −28 |
| Left fusiform G. | 8 | 3.2 | −38 | −24 | −18 |

Table 2 (*Continued*)

| Location | Cluster size (Voxels) | *Z*-scores | Coordinates | | |
|---|---|---|---|---|---|
| | | | *x* | *y* | *z* |
| Right fusiform G. | 4 | 3.17 | 38 | −40 | −20 |
| Right cerebellum (6) | 2 | 3.14 | 26 | −60 | −16 |
| *Right rectus* | 3 | *3.14 #* | 2 | 48 | −20 |
| *Left pallidum* | 1 | *3.12 #* | −24 | −12 | −4 |
| *Left cuneus* | 1 | *3.11 #* | −8 | -84 | 30 |

For the localiser tasks, reading (A) and repetition (B) results are shown at whole brain FWE corrected at $p < .05$ level. For the perception task (C), results are shown at an uncorrected $p < .001$. All peak voxels reach whole brain FDR correction at $p < .05$ except those in *italic* with a hash mark (#). Voxels marked in bold with two asterisks (**) reach whole brain FWE correction at $p < .05$ level. LIFG = left inferior frontal gyrus; SMA = supplementary motor area.

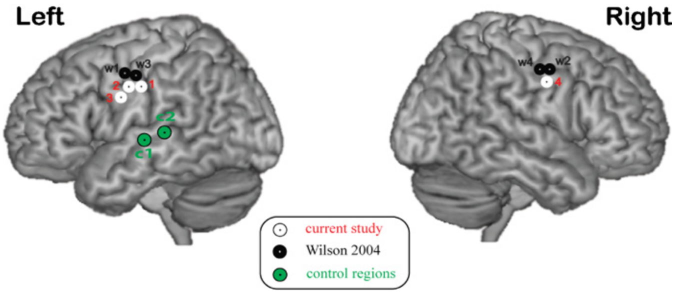

**Fig. 3.** Location of ROIs displayed on a template brain. ROIs were obtained by using independent contrasts from the present set of experiments (white and green targets) and peak voxels from Wilson et al. (2004) (black targets). All ROIs were spheres with a 7 mm radius around the peak coordinates. Peak coordinates are (1) [−48 −12 38], (2) [−50 −4 38], (3) [−50 0 26] and (4) [56 −8 38]; (c1) [−64 −12 −4] and (c2) [−60 −30 6]; (w1) [−50 −6 47], (w2) [55 −3 45], (w3) [−51 −11 46] and (w4) [56 −8 44] (see text for more details).

premotor and motor regions, illustrated in Fig. 3. These ROIs were then used to extract and assess the neural activation in the speech perception task. Like for the whole brain analysis, the ROI analyses were carried out on contrasts of each task compared to its respective baseline.

First, we used contrasts from the localiser runs (reading and repetition), by combining activations in these runs and choosing the first four peak activations. There were two maxima in the left postcentral gyrus and two in the bilateral precentral gyri. The second set of ROI coordinates were chosen from the existing literature on speech perception and have been previously reported to show activation during passive listening to simple syllables (Wilson et al., 2004). There were two pairs of bilateral peaks that were reported to show significant neural activation during passive listening to monosyllabic pseudowords, including two in the left hemisphere (one in the precentral and one in the postcentral gyrus), and two in the right precentral gyrus. Finally, we chose two additional regions from our independent scanning runs (reading and repetition combined), both in the left middle temporal gyrus, to serve as control regions.

For each peak coordinate, we extracted data from a 7 mm radius sphere around the peak, averaging over all voxels, and assessed the activation during speech perception using a one-sample, one-tailed *t*-test. We used the Bonferroni correction to control for the use of multiple ROIs (corrected separately for each set). Table 3 lists regions of interest and statistical significance of neural activation during speech perception.

Following correction for multiple comparisons, none of the ROIs showed evidence for significant premotor or motor activation during speech perception.[2] However, there were three regions in the left postcentral gyrus that showed a trend in this direction.

### 3.2.3. *Individual differences in neural activation within regions in the premotor and motor cortices*

The primary goal of the present study was to investigate the extent to which behavioural performance might be associated with differential neural recruitment within the premotor cortex during speech perception. To test this hypothesis, we selected the three ROIs whose activity during speech perception approached statistical significance (Bonferroni-corrected $p < .09$), and plotted the percent signal change (experimental task vs. control tasks) for each participant in each task for visual inspection (an example is shown in panel A of Fig. 4). These results suggested considerable variability across participants, particularly in the perception task.

To account for inter-individual variations in all three tasks, and to obtain a single measure representing the activation during listening compared to reading and repetition, for each participant we computed a BOLD ratio score corresponding to the observed neural activation during speech perception divided by the average neural activation in reading and repetition. We used these ratio scores instead of absolute activation level in the speech perception task to exclude differences in the overall magnitude of the BOLD response common to speech perception and production. Examination of these ratio scores revealed that approximately half of the participants showed increased neural activation in these regions during speech perception, and about half did not (see Fig. 4B). In order to test whether this observed variability in BOLD ratio signal was due to systematic variability based on individual cognitive ability (as opposed to noise), we carried out multiple regression analyses with BOLD ratios in each region as the dependent measure and behavioural scores on each test as predictors. These were scores on auditory and visual digit task, scores on high and low phonotactic probability pseudoword repetitions averaged over length, and scores on the spoonerisms task. Predictor variables were entered simultaneously.[3] The descriptive statistics (bivariate correlations) are shown in Table 3, and the multiple regressions in Table 4.

The results for the first ROI, located in the left precentral gyrus (centred at [−48 −12 38]), indicated that the model with all five predictors explained 72.7% of the variance ($R^2 = .727$, $F(5,10) = 5.34$, $p = .012$). Furthermore, repetition scores for the low phonotactic probability pseudowords significantly predicted neural activation in this ROI ($B = 5.44$, SE-B = 2.11, $\beta = .97$, $p = .028$). None of the other variables alone predicted neural responses in this region.

The results of the multiple regression for the second ROI within the left precentral gyrus (centred at [−50 −4 38]) showed that the five predictors all together explained 68.3% of the variance ($R^2 = .683$, $F(5,10) = 4.3$, $p = .024$). In this region, scores on high phonotactic probability pseudoword repetition predicted neural responses significantly ($B = -3.39$, SE-B = 1.4, $\beta = -.71$, $p = .04$), whereas scores on low phonotactic

---

[2] We also conducted the same ROI analysis on data extracted from the pseudowords contrast only (as opposed to pseudowords > baseline). Here, none of the ROIs showed significant activation, all $ps > .4$ (Bonferroni corrected for the number of ROIs).

[3] We ran several additional exploratory analyses with different methods for entering predictors (e.g., hierarchical). These exploratory analyses produced results that were qualitatively equivalent to those obtained with the method we report here (in which predictors were entered simultaneously).

**Table 3**

Statistical significance of the neural activation in the ROIs during speech perception, and correlation coefficients between the ratio scores (BOLD in perception/average BOLD in reading and repetition) and scores on the behavioural tests. Correlations were only run on ROIs whose neural activation in the speech perception task was close to reach significance.

| Cerebral regions | ROI centre coordinates | One-tailed $t$-test (mean, ±SD) | Correlations (one-tailed Pearson cor.) | | | | |
|---|---|---|---|---|---|---|---|
| | | | Auditory digit span | Visual digit span | Low phonotactic probability pseudoword repetition, all length collapsed | High phonotactic probability pseudoword repelition, all length collapsed | Spoonerisms |
| **Current study:** | | | | | | | |
| Left postcentral | −48 −12 38 | $t(19)=2.3$<br>$p$(bonf) =.064<br>(0.17, ±0.33) | $r=47$, $p=.032$ | $r=.73$, $p=.001$ | $r=.55$, $p=.014$ | $r=.09$, $p=.37$, ns | $r=.04$, $p=.44$, ns |
| Left postcentral | −50 −4 38 | $t(19)=2.16$<br>$p$(bonf) =.084<br>(0.17 ±0.37) | $r=36$, $p=.08$ | $r=.69$, $p=.002$ | $r=.43$, $p=.049$ | $r=-.077$, $p=39$, ns | $r=-.02$, $p=.47$, ns |
| Left precentral | −50 0 26 | $t(19)=0.63$<br>$p$(bonf) =1.4, ns<br>(0.04, ±0.26) | – | – | – | – | – |
| Right precentral | 56 −8 38 | $t(19)=0.42$<br>$p$(bonf) =1.32, ns<br>(0.03, ±0.37) | – | – | – | – | – |
| **Wilson study (2004):** | | | | | | | |
| Left precentral | −50 −6 47 | $t(19)=1.59$<br>$p$(bonf)=.25, ns<br>(0.11, ±0.32) | – | – | – | – | – |
| Left postcentral | −51 −11 46 | $t(19)=2.25$<br>$p$(bonf)=.072<br>(0.18, ±0.35) | $r=.64$, $p=.004$ | $r=.69$, $p=.001$ | $r=.54$, $p=.016$ | $r=.17$, $p=.26$, ns | $r=.18$, $p=.25$, ns |
| Right precentral | 55 −3 45 | $t(19)=-2.73$<br>$p$(bonf)=3.96, ns<br>(−0.24, ±0.4) | – | – | – | – | – |
| Right precentral | 56 −8 44 | $t(19)=-0.41$<br>$p$(bonf)=2.6, ns<br>(−0.03, ±0.037) | – | – | – | – | – |
| **Control regions:** | | | | | | | |
| Left middle temporal | −64 −12 −4 | $t(19)=6.64$<br>$p$(bonf) <.0001<br>(0.99, ±0.67) | $r=.12$, $p=.33$, ns | $r=.37$, $p=.08$ | $r=.0.8$, $p=8$, ns | $r=-.19$, $p=.24$, ns | $r=-.37$, $p=.08$ |
| Left middle temporal | −60 −30 6 | $t(19)=6.04$<br>$p$(bonf) <.0001<br>(−0.53, ±0.39) | $r=.09$, $p=.37$, ns | $r=.09$, $p=.36$, ns | $r=.25$, $p=.18$, ns | $x=-.05$, $p=.43$, ns | $r=.4$, $p=.06$ |

**Table 4**
Summary of the multiple regressions between the BOLD ratios and the behavioural scores on each test.

| Predicting neural activity from behaviour | | | | |
|---|---|---|---|---|
| ROI centre coordinates | *B* | SE *B* | *beta* | *p*-value |
| **ROI [−48 −12 38]** ($R^2$ = .727, adjusted $R^2$ = .6, $F$(5,10) = 5.34, $p$ = .012) | | | | |
| Constant | *−0.998* | *0.723* | | *0.19, ns* |
| High phonotactic probability pseudoword repetition | −3.3 | 1.83 | −0.5 | *0.1, ns* |
| Low phonotactic probability pseudoword repetition | **5.44** | **2.11** | **0.98** | ***0.028*** |
| Auditory digit | 0.039 | 0.035 | 0.29 | *0.29, ns* |
| Visual digit | 0.021 | 0.043 | 0.155 | *0.635, ns* |
| Spoonerisms | −1.036 | 0.73 | −0.35 | *0.19, ns* |
| **ROI [−50 −4 38]** ($R^2$ = .683, adjusted $R^2$ = .53, $F$(5,10) = 4.3, $p$ = .024) | | | | |
| Constant | *−0.091* | *0.55* | | *0.87, ns* |
| High phonotactic probability pseudoword repetition | **−3.31** | **1.4** | **−0.71** | ***0.04*** |
| Low phonotactic probability pseudoword repetition | **3.39** | **1.62** | **0.86** | ***0.06*** |
| Auditory digit | 0.002 | 0.027 | 0.016 | *0.95, ns* |
| Visual digit | 0.035 | 0.033 | 0.36 | *0.32, ns* |
| Spoonerisms | −0.37 | 0.56 | −0.18 | *0.53, ns* |
| **Wilson ROI [−51 −11 46]** ($R^2$ = .682, adjusted $R^2$ = .52, $F$(5,10) = 4.28, $p$ = .024) | | | | |
| Constant | *−1.42* | *0.668* | | *0.059* |
| High phonotactic probability pseudoword repetition | −1.55 | 1.7 | −0.275 | *0.38, ns* |
| Low phonotactic probability pseudoword repetition | **3.78** | **1.95** | **0.79** | ***0.08*** |
| Auditory digit | **0.066** | **0.032** | **0.574** | ***0.069*** |
| Visual digit | −0.001 | 0.04 | −0.009 | *0.98, ns* |
| Spoonerisms | −0.76 | 0.68 | −0.306 | *0.29, ns* |
| **Control ROI [64 −12 −4]** ($R^2$ = .389, adjusted $R^3$ = .08, $F$(5,10) = 1.27, $p$ = .35, *ns*) | | | | |
| Constant | *2.11* | *3.16* | | *0.52, ns* |
| High phonotactic probability pseudoword repetition | −7.21 | 8.01 | −0.37 | *0.39, ns* |
| Low phonotactic probability pseudoword repetition | 9.78 | 9.25 | 0.6 | *0.31, ns* |
| Auditory digit | 0.05 | 0.15 | 0.12 | *0.76, ns* |
| Visual digit | 0.04 | 0.19 | 0.1 | *0.84, ns* |
| Spoonerisms | −4.85 | 3.21 | −0.57 | *0.16, ns* |
| **Control ROI [−60 −30 6]** ($R^2$ = .380, adjusted $R^2$ = .07, $F$(5,10) = 1.23, $p$ = .36, *ns*) | | | | |
| Constant | *2.81* | *4.25* | | *0.52, ns* |
| High phonotactic probability pseudoword repetition | −19.11 | 10.77 | −0.74 | *0.1, ns* |
| Low phonotactic probability pseudoword repetition | 11.66 | 12.43 | 0.54 | 0.37, ns |
| Auditory digit | −0.11 | 0.2 | −0.22 | *0.6, ns* |
| Visual digit | 0.008 | 0.25 | 0.16 | 0.97, ns |
| Spoonerisms | 6.4 | 4.31 | 0.56 | *0.17, ns* |

probability pseudoword repetition showed a trend ($B$ = 3.39, SE-B = 1.61, $\beta$ = .86, $p$ = .06). None of the remaining variables alone predicted neural responses in this region.

For the third ROI (from Wilson et al., 2004), within the left postcentral gyrus (centred at [−51 −11 46]), the model explained 68.2% of the variance ($R^2$ = .682, $F$(5,10) = 4.28, $p$ = .024). Again, we found two variables showing a trend to predict neural activation significantly: (1) scores on low phonotactic probability pseudoword repetition ($B$ = 3.78, SE-B = 1.95, $\beta$ = .79, $p$ = .08), and (2) scores on auditory digit span task ($B$ = .066, SE-B = .032, $\beta$ = .574, $p$ = .069). None of the other variables alone predicted neural responses in this region.

Additionally, we also extracted data from the two control ROIs in left middle temporal gyrus, computed the ratios and ran the same model. In the first control region (centred at [−64 −12 −4]), the model explained 38.9% of the variance but did not reach significance ($R^2$ = .389, $F$(5,10) = 1.27, $p$ = .35), and none of the variables alone predicted neural responses in this region. We observed the same in the second control region (centred at [−60 −30 6]). The model explained 38% of the variance and just as before was not significant ($R^2$ = .38, $F$(5,10) = 1.23, $p$ = .36). Again, none of the variables alone predicted neural responses in this region. The multiple regressions thus suggest that the predictive value of the behavioural tests was not general, but specific to premotor and motor regions that support verbal short-term memory.

Finally, we ran another multiple regression analysis for each ROI including the average level of activation during reading and repetition as the dependent variable and the behavioural scores as predictors, shown in Table 5. This regression was used to test whether the previous correlations might result from general upregulation or whether they reflect a relationship that is specific to speech perception. The results for the first left precentral gyrus ROI (centred at [−48 −12 38]) indicated that the model with all five predictors explained 24.71% of the variance which was not significant ($R^2$ = .241, $F$(5,10) < 1, ns). None of the variables alone predicted neural responses in this region. For the second ROI in the left precentral gyrus (centred at [−50 −4 38]), results showed that all the predictors explained 38.6% of the variance but failed to reach significance ($R^2$ = .386, $F$(5,10) = 1.25, $p$ = .35, ns). Again, none of the variables alone predicted neural responses in this region. For the third ROIs within the left postcentral gyrus (Wilson et al., 2004) (centred at [−51 −11 46]), the model explained 12.7% of the variance, again without reaching significance ($R^2$ = .127, $F$(5,10) < 1, ns). As before, none of the variables alone predicted neural responses in this region. These results thus indicate that the behavioural score selectively predicted motor/premotor activation during speech perception only, independent of what would be predicted by the average activity alone.

## 4. Discussion

The neural mechanisms supporting speech perception and production, and the degree to which these may be linked, have been of longstanding interest to researchers, with a recently renewed focus on the role of the motor system in speech perception. In the current study we examined this issue from the standpoint of individual

**Table 5**
Summary of the multiple regressions between the average BOLD signal in the production runs (reading and repetition) and the behavioural scores on each test.

| Predicting neural activity from behaviour | | | | |
|---|---|---|---|---|
| ROI centre coordinates | *B* | SE B | beta | *p*-Value |
| **ROI [−48 −12 38]** ($R^2$ = .241, adjusted $R^2$ = −.14, *F*(5,10) = 0.63, *p* = .7, *ns*) | | | | |
| Constant | *−0.887* | *1* | | *0.4*, *ns* |
| High phonotactic probability pseudoword repetition | −1.9 | 2.5 | −0.35 | *0.46*, *ns* |
| Low phonotactic probability pseudoword repetition | 0.65 | 2.92 | 0.14 | *0.83*, *ns* |
| Auditory digit | −0.014 | 0.048 | −0.13 | *0.77*, *ns* |
| Visual digit | 0.015 | 0.06 | 0.13 | *0.8*, *ns* |
| Spoonerisms | 1.34 | 1.01 | 0.55 | *0.22*, *ns* |
| **ROI [−50 −4 38**] ($R^2$ = .386, adjusted $R^2$ = .08, *F*(5,10) = 1.25, *p* = .35, ns) | | | | |
| Constant | *2.01* | *0.99* | | *0.07* |
| High phonotactic probability pseudoword repetition | −2.72 | 2.5 | −0.45 | *0.3*, *ns* |
| Low phonotactic probability pseudoword repetition | 3 | 2.9 | 0.6 | *0.32*, *ns* |
| Auditory digit | −0.08 | 0.05 | −0.63 | *0.14*, *ns* |
| Visual digit | 0.15 | 0.06 | 0.12 | *0.8*, *ns* |
| Spoonerisms | 0.71 | 1 | 0.27 | *0.5*, *ns* |
| **Wilson ROI [−51 −11 46]** ($R^2$ = .127, adjusted $R^2$ = −.31, *F*(5,10) = 0,29, *p* = .9, *ns*) | | | | |
| Constant | *1.9* | *1.5* | | *0.24*, *ns* |
| High phonotactic probability pseudoword repetition | −4.12 | 3.9 | −0.52 | *0.3*, *ns* |
| Low phonotactic probability pseudoword repetition | 2.42 | 4.5 | 0.36 | *0.6*, *ns* |
| Auditory digit | −0.005 | 0.075 | −0.034 | *0.94*, *ns* |
| Visual digit | −0.011 | 0.09 | −0.07 | *0.9*, *ns* |
| Spoonerisms | 0.86 | 1.6 | 0.25 | *0.6*, *ns* |

differences in cognitive ability. Our purpose was twofold: First, to further investigate premotor and motor involvement during speech perception, using tasks that overcome methodological limitations of previous studies. Second, to determine whether individual differences in cognitive abilities are associated with differences in neural recruitment along the dorsal pathway during speech perception. Below we review our findings with respect to these two questions.

The results of the whole-brain analysis showed activation in the middle and superior temporal gyri during all three language tasks compared to matched non-language baseline tasks. However, it was only during tasks requiring overt speech production (reading and repeating) that we saw significant activation of premotor and motor cortices. The only frontal region exhibiting elevated activity during speech perception was the left inferior frontal gyrus (pars orbitalis). The failure to observe significant neural activation in premotor and motor cortices during speech perception compared to matched non-speech perception is in line with suggestions from Scott et al. (2009) and Lotto, Holt, and Hickok (2009) that the presence of premotor and motor activation in fMRI studies of speech perception might be a consequence of inappropriate baseline conditions, and the failure to adopt an appropriate level of control for false positives.

To ensure that we did not miss premotor or motor activation in our whole brain analysis, we performed a focused analysis, sampling the premotor and motor cortices with a set of independently defined regions of interest (ROIs). The results of this region of interest analysis confirmed that the majority of the sampled motor region did not show significant neural activation in the speech perception task. Nevertheless, three regions within the left postcentral gyrus exhibited responses that approached significance, after correcting for the number of ROIs. The failure to find significant effects at the group level was due to the presence of substantial individual variability. Importantly, between 68% and 72% of the observed variance could be explained by performance on the behavioural tasks. Moreover, we can be confident that the activation observed within the left premotor cortex is not due to motor processes involved in executing the button press because we modelled trials involving a button presses separately and only included trials without button presses. However, although we only analysed trials without button press, all stimuli had to be subjected to a decision process, and to action planning or action inhibition. In the current experiment, we cannot exclude the possibility that the decision process and/or action planning and inhibition could have influenced the pattern of neural activity in our sample of motor cortex (Hasson, Nusbaum, & Small, 2006). Even if these executive processes may contribute to the absolute magnitude of motor activity this does not explain the association that we see between individual differences and neural activitation during speech perception. The variability in neural activation in our sample of the motor region is specific to this region and is caused by task related upregulation within this region rather than by global changes in neural activity as it does not extend to the control regions within the middle temporal gyrus.

The factor that predicted inter-individual variability in motor activation alone during speech perception was repetition accuracy for low phonotactic probability pseudowords. Interestingly, performance accuracy for high phonotactic probability pseudowords did not correlate with neural activation in the ROIs. This might be because the restricted range of performance on high phonotactic probability pseudowords allows less room to detect correlations with individual differences in neural activation. It may also be that the greater difficulty of the low phonotactic probability pseudoword task draws more heavily on subvocal rehearsal. Furthermore, the speech production process itself is also influenced by phonotactic probability (Vitevitch, Armbrüster, & Chu, 2004) and therefore production of low phonotactic probability pseudowords may be more effortful.

For one of the ROIs within the left postcentral gyrus (from the current study, centred around [−50 −4 38]), scores on high phonotactic probability pseudowords predicted a significant decrease in neural activation (standardised $\beta = -0.71$). This type of inverse correlation can occur when there is a nonlinear relationship between the factors in a multiple regression. Note that the corresponding bivariate correlation was small and nonsignificant.

In contrast to the verbal short-term memory measures, performance on the spoonerisms task did not predict variability in neural activation during speech perception. Although the spoonerism task involves verbal short-term memory, it is a metalinguistic task which relies on overt awareness of speech sounds, and on the ability to separate and mentally manipulate these sounds. Functional imaging investigations of the neural systems supporting speech segmentation and manipulation processes showed that

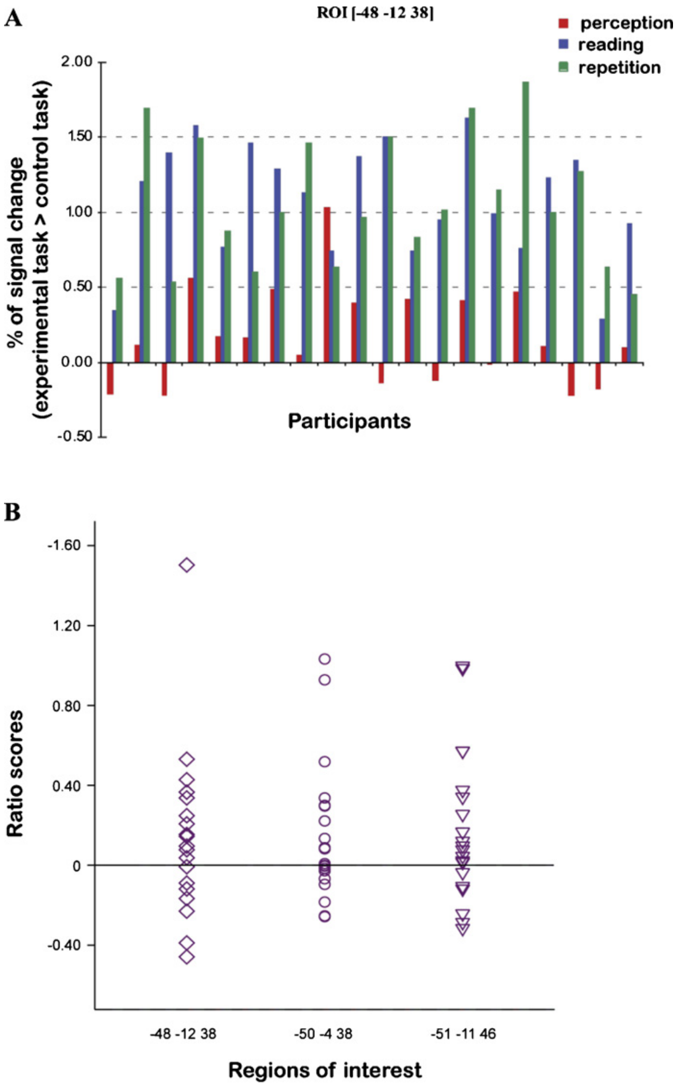

**Fig. 4.** Percent signal change (experimental task vs. control task) for each participant in perception (red), reading (blue), and repetition (green) tasks (A). (B) BOLD ratios for speech perception task in the three ROIs. The BOLD ratio was calculated as the parameter estimate for the perception task divided by the average parameter estimate for the reading and repetition tasks. Data are shown for ROIs centered at [−48 −12 38] (◊), [−50 −4 38] (◯), and [−51 −11 46] (▽). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

besides motor regions, more anterior regions of the prefrontal cortex were also recruited (e.g. LIFG; Burton, Small, & Blumstein, 2000). It may then be that individual differences in the ability to perform the spoonerism task are driven more by variability in processes involved in phonemic segmentation and manipulation, than by differences in the rehearsal component of verbal short-term memory. That is, while these tasks do have a motor component, the efficacy of rehearsal may have only a weak influence on overall performance. Whilst previous studies showing motor involvement in speech perception (Möttönen & Watkins, 2009; Yuen et al., 2009) have been criticised for their use of tasks that required speech segmentation, our results suggest that these tasks may be less closely linked to motor activation than tasks such as digit span, or pseudoword repetition tasks which directly tap into the efficiency of subvocal rehearsal.

Taken together, the results of the bivariate correlations and the multiple regressions indicate that neural activation in speech perception is linked to performance in verbal short-term memory but not phonological awareness. That is, participants with a better verbal short-term memory produced more neural activation in motor cortex when listening to speech. This is consistent with the long-established role of subvocal rehearsal in these verbal short-term memory tasks (Jarrold & Tam, 2010), and with the observation that subvocal rehearsal relies on intact pathways from visual or auditory speech processing to motor regions (Buchsbaum et al., 2005; Saur et al., 2008; Warrington & Shallice, 1969). In order to interpret this result, though, it is important to consider whether there is any possibility that the speech perception task used here (one-back identity monitoring) might itself involve either subvocal rehearsal or segmentation, both of which might potentially recruit motor regions. For example, speech segmentation has been shown to activate prefrontal regions (Burton et al., 2000) and has been argued to depend on auditory-motor pathways (Lotto et al., 2009; Scott et al., 2009). However, in our one-back task decisions are made on the identity of the entire syllable rather than on individual segments, so there is no requirement for segmentation. Although we are unable to completely rule out the possibility that participants in a one-back task might engage in subvocal rehearsal, this seems highly unlikely given that they need only remember a single item for a brief period with no intervening stimuli. We know of no evidence that would suggest that people might rehearse under these circumstances. Nonetheless, participants with better verbal short-term memory produce greater activation during speech perception in brain regions associated with subvocal rehearsal of verbal materials.

The most straightforward explanation of our data is therefore that participants with better short-term memory are more likely to engage the motor system in recoding speech input in order to perform the one-back task. This might be a purely strategic effect, in that participants with better short-term memory might simply find it easier to recode, even though the task does not strictly demand it. An alternative possibility is that there may be anatomical differences between participants (in grey matter volume, white matter connectivity, or both) which increase the likelihood of motor involvement in both speech and memory tasks. These explanations are not mutually exclusive. An initial anatomical difference could well be the driving force behind the development of different strategies, and persistent application of a rehearsal strategy could conceivably lead to longer-term changes in brain structure and connectivity. However, it might well be possible to separate these two accounts by further studies that explore the structure and connectivity of the auditory–motor pathway (Saur et al., 2008) and by exploring neural activity during verbal short-term memory tasks (Buchsbaum et al., 2005) in different populations. We note that our overall findings linking individual cognitive ability to motor and premotor activity during speech perception stand regardless of whether this relationship is mediated by differences in brain morphology, although this is an interesting possibility to consider in future studies.

In summary, we did not find evidence suggesting an obligatory role for premotor or motor regions in speech perception. Rather, our results support the idea that recruitment of these regions varies with individual differences in cognitive abilities that depend on auditory–motor pathways. These findings strengthen the evidence for a link between verbal short-term memory ability and speech perception, and suggest that individuals with impaired short-term memory may have particular difficulty recruiting auditory motor pathways in speech perception. More generally, our findings reinforce the view that speech perception relies on distinct dorsal and ventral pathways (cf. Hickok & Poeppel, 2007; Scott & Johnsrude, 2003). In our work, differential recruitment of these pathways depends on both the specific demands of the current listening situation and the underlying cognitive abilities of individual listeners.

## Acknowledgements

This work was supported by the Medical Research Council (MC_US_A060_0038). G. Szenkovits was supported by a postdoctoral fellowship from the Fyssen Foundation and by a Marie Curie Individual Fellowship (project number: 236662), funded by the European Commission as part of its Seventh Framework Program. We are grateful to our radiographers, Steve Eldridge, Diana Everson, and Helen Lloyd for assistance with data collection and to our volunteers for their participation. Finally we also thank the two anonymous reviewers for their comments.

## References

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage*, *26*, 839–851.

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

Baddeley, A. (1998). Recent developments in working memory. *Current Opinion in Neurobiology*, *8*, 234–238.

Baddeley, A., Gathercole, S., Papagno, C., & Degli, U. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*, 158–173.

Baddeley, A. D., & Hitch, G. I. (1974). *Working memory* New York: Academic Press.

Besner, D. (1987). Phonology, lexical access in reading, and articulatory suppression: A critical review. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *39*(3, Sect A), 467–478.

Besner, D., Davies, J., & Daniels, S. (1981). Reading for meaning: The effects of concurrent articulation. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *33a*(4), 415–437.

Brown, G. D. (1987). Phonological coding in rhyming and homophony judgement. *Acta Psychologica*, *65*(3), 247–262.

Buchsbaum, B. R., Hickok, G., & Humphries, C. (2001). Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science*, *25*, 663–678.

Buchsbaum, B. R., Olsen, R. K., Koch, P., & Berman, K. F. (2005). Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron*, *48*, 687–697.

Burton, M. W., Small, S. L., & Blumstein, S. E. (2000). The role of segmentation in phonological processing: An fMRI investigation. *Journal of Cognitive Neuroscience*, *12*(4), 679–690.

Callan, D., Callan, A., Gamez, M., Sato, M.-a., & Kawato, M. (2010). Premotor cortex mediates perceptual performance. *NeuroImage*, *51*, 844–858.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114.

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology*, 381–385.

Davis, M., & Johnsrude, I. (2003). Hierarchical processing in spoken language comprehension. *Stimulus*, *23*, 3423–3431.

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research*, *229*, 132–147.

Deeks, J. M., & Carlyon, R. P. (2004). Simulations of cochlear implant hearing using filtered harmonic complexes: Implications for concurrent sound segregation. *Journal of the Acoustical Society of America*, (115), 1736–1746.

Devlin, J. T., & Aydelott, J. (2009). Speech perception: Motoric contributions versus the motor theory. *Current Biology*, *19*(5), 198–200.

Dufor, O., Serniclaes, W., Sprenger-Charolles, L., & Demonet, J. F. (2009). Left premotor cortex and allophonic speech perception in dyslexia: A PET study. *NeuroImage*, *46*(1), 241–248.

Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic production accuracy and fluency in nonword repetition. *Hearing Research*,

Fadiga, L., Ferrara, Á., Biomediche, S., Umana, F., Parma, Á., Umana, F., et al. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *Neuroscience*, *15*, 399–402.

Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers*, (35), 116–124.

Friston, K., Ashburner, J., Frith, C., Poline, J.-B., Heather, J., & Frackowiak, R. (1995). Spatial registration and normalization of images. *Human Brain Mapping*, *2*, 165–189.

Friston, K. J., Frith, C. D., Liddle, P. F., & Frackowiak, R. S. (1991). Comparing functional, (PET) images: The assessment of significant change. *Journal of Cerebral Blood Flow & Metabolism*, *11*(4), 690–699.

Gathercole, S., & Baddeley, A. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology*, *81*, 439–454.

Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(1), 84–95.

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

Hasson, U., Nusbaum, H. C., & Small, S. L. (2006). Repetition suppression for spoken sentences and the effect of task demands. *Journal of Cognitive Neuroscience*, *18*(12), 2013–2029.

Hickok, G. (2008). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of Cognitive Neuroscience*, 1229–1243.

Hickok, G. (2009). Speech perception does not rely on motor cortex: Response to D'Ausilio et al. *Current Biology*,. Available at: http://www.cell.com/current-biology/comments_Dausilio

Hickok, G. (2010). The role of mirror neurons in speech perception and action word semantics. *Language and Cognitive Processes*, *25*, 749–776.

Hickok, G., & Buchsbaum, B. (2003). Temporal lobe speech perception systems are part of the verbal working memory circuit: Evidence from two recent fMRI studies. *Behavioral and Brain Sciences*, *26*, 740–741.

Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory–motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience*, *15*(5), 673–682.

Hickok, G., Okada, K., Barr, W., Pa, J., Rogalsky, C., Donnelly, K., et al. (2008). Bilateral capacity for speech sound processing in auditory comprehension: Evidence from Wada procedures. *Brain and Language*, *107*, 179–184.

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, *4*, 131–138.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*, 67–99.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature*, *8*, 393–402.

Jarrold, C., & Tam, H. (2010). *Rehearsal and the development of working memory. In cognitive development and working memory: A dialogue between neo-piagetian and cognitive approaches*. Hove: Psychology Press.

Kriegeskorte, N., Simmons, W., Bellgowan, P., & Baker, C. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience*, (12), 535–540.

Lichtheim, L. (1885/2006). On aphasia. *Brain*, 1347–1350.

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*(4), 476–490.

Lotto, A. J., Hickok, G. S., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences*, 110–114.

Meister, I. G., Wilson, S. M., Deblieck, C., & Wu, A. D. (2007). The essential role of premotor cortex in speech perception. *Current Biology*, *17*, 1692–1696.

Morais, J., Bertelson, P., Cary, L., & Alegria, J. (1986). Literacy training and speech segmentation. *Cognition*, *24*, 45–64.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, *7*, 323–331.

Morais, J., & Kolinsky, R. (1994). Perception and awareness in phonological processing: The case of the phoneme. *Cognition*, *50*, 287–297.

Möttönen, R., & Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *The Journal of neuroscience: The official journal of the Society for Neuroscience*, *29*, 9819–9825.

Obleser, J., Wise, R. J. S., Dresner, M. A., & Scott, S. K. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *The Journal of Neuroscience*, *27*(9), 2283–2289.

Osnes, B., Hugdahl, K., & Specht, K. (2011). Effective connectivity analysis demonstrates involvement of premotor cortex during speech perception. *NeuroImage*, *54*(3), 2437–2445.

Price, C. J. (2010). The anatomy of language: A review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences*, *1191*, 62–88.

Price, C. J., Wise, R. J. S., Warburton, E. A., Moore, C. J., Howard, D., Patterson, K., et al. (1996). Hearing and saying. The functional neuro-anatomy of auditory word processing. *Brain*, *119*, 919–931.

Pulvermüller, F., Huss, M., Kherif, F., Moscoso, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences*, *103*, 7865–7870.

Richardson, J. T. (1987). Phonology and reading: The effects of articulatory suppression upon homophony and rhyme judgments. *Language and Cognitive Processes*, *2*(3–4), 229–244.

Rodd, J., Davis, M., & Johnsrude, I. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex*, *15*, 1261–1269.

Rodd, J., Longe, O., Randall, B., & Tyler, L. (2010). The functional organisation of the fronto-temporal language system: Evidence from syntactic and semantic ambiguity. *Neuropsychologia*, *48*, 1324–1335.

Sato, M., Tremblay, P., & Gracco, V. (2009). A mediating role of the premotor cortex in phoneme segmentation. *Brain and Language*, *111*(1), 1–7.

Saur, D., Kreher, W. B., Schnell, S., Ku, D., Vry, M.-S., Umarova, R., et al. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences*, *105*.

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

Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, *26*, 100–107.

Scott, S. K., Mcgettigan, C., & Eisner, F. (2009). A little more conversation, a little less action – Candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience*, *10*.

Scott, S. K., & Wise, R. J. S. (2004). The functional neuroanatomy of prelexical processing in speech perception. *Cognition*, *92*, 13–45.

Seghier, M. L., & Price, C. J. (2009). Dissociating functional brain networks by decoding the between-subject variability. *NeuroImage*, *45*, 349–359.

Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex*, *17*, 2387–2399.

Soltysik, D. A., & Hyde, J. S. (2006). Strategies for block-design fMRI experiments during task-related motion of structures of the oral cavity. *NeuroImage*, *29*(4), 1260–1271.

Soltysik, D. A., & Hyde, J. S. (2008). High spatial resolution increases the specificity of block-design BOLD fMRI studies of overt vowel production. *NeuroImage*, (41), 389–397.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(4), 1181–1186.

Vitevitch, M. S., Armbrüster, J., & Chu, S. (2004). Sublexical and lexical representations in speech production: Effects of phonotactic probability and onset density. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 514–529.

Vitevitch, M. S., & Luce, P. A. (2005). Increases in phonotactic probability facilitate spoken nonword repetition. *Journal of memory and language*, *52*, 193–204.

Warrington, E. K., & Shallice, T. (1969). The selective impairment of auditory short term memory. *Brain*, *92*, 885–896.

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Electromyography*, *41*, 989–994.

Wechsler, D. (1998). *The Wechsler Adult Intelligence Scale-third edition (WAIS-III)*. The Psychological Corporation.

Wilson, S. M., & Iacoboni, M. (2006). Neural responses to non-native phonemes varying in producibility: Evidence for the sensorimotor nature of speech perception. *Scanning*.

Wilson, S. M., Saygin, P., & Sereno, M. I. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, 3–4.

Yuen, I., Davis, M. H., Brysbaert, M., & Rastle, K. (2009). Activation of articulatory information in speech perception. *Proceedings of the National Academy of Sciences of the United States of America*, 2–11.

Zheng, Z., Munhall, K., & Johnsrude, I. (2010). Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. *Journal of Cognitive Neuroscience*, *22*(8), 1770–1781.