Neural Correlates of Sublexical Processing in Phonological Working Memory

Carolyn McGettigan[1], Jane E. Warren[2], Frank Eisner[1],
Chloe R. Marshall[3], Pradheep Shanmugalingam[1],
and Sophie K. Scott[1]

Abstract

■ This study investigated links between working memory and speech processing systems. We used delayed pseudoword repetition in fMRI to investigate the neural correlates of sublexical structure in phonological working memory (pWM). We orthogonally varied the number of syllables and consonant clusters in auditory pseudowords and measured the neural responses to these manipulations under conditions of covert rehearsal (Experiment 1). A left-dominant network of temporal and motor cortex showed increased activity for longer items, with motor cortex only showing greater activity concomitant with adding consonant clusters. An individual-differences analysis revealed a significant positive relationship between activity in the angular gyrus and the hippocampus, and accuracy on pseudoword repetition. As models of pWM stipulate that its neural correlates should be activated during both perception and production/rehearsal [Buchsbaum, B. R., & D'Esposito, M. The search for the phonological store: From loop to convolution. *Journal of Cognitive Neuroscience, 20,* 762–778, 2008; Jacquemot, C., & Scott, S. K. What is the relationship between phonological short-term memory and speech processing? *Trends in Cognitive Sciences, 10,* 480–486, 2006; Baddeley, A. D., & Hitch, G. Working memory. In G. H. Bower (Ed.), *The psychology of learning and motiva-* *tion: Advances in research and theory* (Vol. 8, pp. 47–89). New York: Academic Press, 1974], we further assessed the effects of the two factors in a separate passive listening experiment (Experiment 2). In this experiment, the effect of the number of syllables was concentrated in posterior–medial regions of the supratemporal plane bilaterally, although there was no evidence of a significant response to added clusters. Taken together, the results identify the planum temporale as a key region in pWM; within this region, representations are likely to take the form of auditory or audiomotor "templates" or "chunks" at the level of the syllable [Papoutsi, M., de Zwart, J. A., Jansma, J. M., Pickering, M. J., Bednar, J. A., & Horwitz, B. From phonemes to articulatory codes: an fMRI study of the role of Broca's area in speech production. *Cerebral Cortex, 19,* 2156–2165, 2009; Warren, J. E., Wise, R. J. S., & Warren, J. D. Sounds do-able: auditory–motor transformations and the posterior temporal plane. *Trends in Neurosciences, 28,* 636–643, 2005; Griffiths, T. D., & Warren, J. D. The planum temporale as a computational hub. *Trends in Neurosciences, 25,* 348–353, 2002], whereas more lateral structures on the STG may deal with phonetic analysis of the auditory input [Hickok, G. The functional neuroanatomy of language. *Physics of Life Reviews, 6,* 121–143, 2009]. ■

INTRODUCTION

In phonological working memory (pWM), the "word length effect" has been taken to show an important role for articulation: The more syllables there are in a word, the fewer of such words can be accurately rehearsed in a list (via subvocalization; Baddeley, Thompson, & Buchanan, 1974). Later developments of this work showed that subsyllabic properties of speech are also important in serial recall (Caplan, Rochon, & Waters, 1992). For example, among bilingual Welsh/English speakers, it is common to have a shorter span for Welsh digits than for English digits (Murray & Jones, 2002). Although the Welsh digits are shorter in acoustic duration than the English digits, they are more complicated to pronounce, which detrimentally affects their covert rehearsal. Both of these findings lend support

to the predictions made in the Baddeley model of working memory (Baddeley & Hitch, 1974), in which the phonological or articulatory loop was primarily involved in subvocalization as a means of refreshing representations of verbal material in a phonological store.

Since the 1990s, neuroimaging has been used in an attempt to find the neural loci of the phonological loop and store, the key components of Baddeley's model. In PET, Paulesu, Frith, and Frackowiak (1993) identified the left supramarginal gyrus as the phonological store, and later studies also suggested loci for this component in other sites in parietal cortex (e.g., Smith, Jonides, Marshuetz, & Koeppe, 1998; Awh et al., 1996; Smith, Jonides, & Koeppe, 1996). The articulatory loop, in contrast, was thought to involve frontal structures such as the left inferior frontal gyrus (LIFG), precentral gyrus, and SMA (Wager & Smith, 2003). The finding that a network of brain areas might come together to support articulatory processes in pWM offers

[1]University College London, UK, [2]Imperial College London, UK,
[3]City University, London, UK

little challenge to the notion of a "loop" that refreshes phonological representations during rehearsal. However, when considering the "unitary" phonological store, the matching of model to brain was not so straightforward (Buchsbaum & D'Esposito, 2008). The Baddeley model specifies that auditory verbal input must gain obligatory access to the phonological store, yet inferior parietal sites are not commonly activated in studies of speech perception (e.g., Scott, Rosen, Lang, & Wise, 2006; Binder et al., 2000; Scott, Blank, Rosen, & Wise, 2000). On the other hand, the model also stipulates that the contents of the phonological store should be abstract and subsequent to acoustic–phonetic processes, which poses problems for a locus in early auditory cortex. Taken together with data from short-term memory patients showing lesions in temporo-parietal sites with some evidence of speech perception deficits (Buchsbaum & D'Esposito, 2008) and the considerable neuropsychological and behavioral evidence for cross-talk between phonological input and output systems in pWM (Jacquemot & Scott, 2006), the problem of housing the phonological store in a single site became intractable. Instead, several authors now view pWM as an emergent property of speech input and output streams (Hickok, 2009; Buchsbaum & D'Esposito, 2008; Jacquemot & Scott, 2006; Postle, 2006; Wilson, 2001; Hickok & Poeppel, 2000). Employing this approach, the current study aimed to investigate pWM within the context of speech production and perception tasks.

Several recent studies have repeatedly shown increased activation during both passive listening and covert rehearsal for speech in two posterior sites on the temporal lobe: one in the posterior lateral STS/STG and one in a posterior–medial region of the left planum temporale (PT: Buchsbaum, Olsen, Koch, & Berman, 2005; Hickok, Buchsbaum, Humphries, & Muftuler, 2003; Buchsbaum, Hickok, & Humphries, 2001). Adding to this basic finding, Hickok et al. (2003) asked participants to perform covert rehearsal of two stimulus types: "Jabberwocky" nonsense sentences in which nouns and verbs had been replaced by pseudowords, and simple tonal melodies. They found auditory and rehearsal responses of very similar magnitude to both stimulus types, indicating that the PT is not generally specialized for rehearsal of speech over other sounds. Furthermore, by using separate "listen only" runs, the authors were able to show that the response in the PT was much greater under conditions of rehearsal than during passive listening, thus supporting the interpretation of the PT as a site for audiomotor transformations rather than sensory imagery alone (see also Hickok, Okada, & Serences, 2009). Buchsbaum et al. (2005) extended this finding by showing that the lateral STG/STS site shows a preference for auditory information during rehearsal that showed decay after 4–6 sec, whereas the posterior–medial site showed no modality preference and more sustained rehearsal-related activity.

Hence, a system emerges in which a lateral STS performs phonetic/phonological analysis on speech, whereas the posterior–medial PT performs crucial audiomotor transformations supporting conversion to speech output. Such a "sensory–motor model" was explicitly proposed in a recent review by Hickok (2009). Other authors would agree with a role for the posterior–medial PT in constraining motor output in response to sound, including mental imagery or for repetition and rehearsal (covert or overt) of speech (Warren, Wise, & Warren, 2005). This view has implicated the PT as part of a dorsal "doing" pathway linking perception to action, and as an important linking structure in a dual-stream model of speech processing, as detailed by several authors (Rauschecker & Scott, 2009; Hickok & Poeppel, 2000, 2004, 2007; Scott & Johnsrude, 2003).

It is worth emphasizing that there is considerable evidence from passive listening studies that the PT does not show an enhanced response to speech relative to other sounds (Griffiths & Warren, 2002; Binder, Frost, Hammeke, Rao, & Cox, 1996), and rehearsal studies using speech and music have also not shown any selective activation by speech rehearsal (Hickok et al., 2003). This might even be interpreted as indication that "pWM" is not specific to phonological input at all (Hickok, 2009; Jones & Macken, 1996). However, despite a lack of selectivity for speech, it is still possible that the PT would be sensitive to structure in speech at a syllabic or segmental level, albeit in an abstract, general form. For example, Griffiths and Warren (2002) see the PT as a "computational hub," where incoming sounds are separated, segmented, and matched onto stored templates for known sounds. Pointing out that responses in the PT can show less sensitivity to phonetic features (e.g., voice onset time) than exhibited by primary auditory cortex, Griffiths and Warren suggest that this may reflect a role for the PT in "the processing of stored representations over hundreds of milliseconds rather than the faithful temporal representation of the incoming stimulus" (p. 350). Indeed, in the literature, sublexical responses to speech have mainly been associated with activity in the lateral STG (Obleser & Eisner, 2009), whereas responses to intelligible speech, when compared with complex acoustic baselines, have been associated with activity lateral and anterior to primary auditory cortex in the STS (Scott et al., 2000, 2006). However, more posterior supratemporal plane sites have also been implicated in some aspects of phonological processing. Jancke, Wustenberg, Scheich, and Heinze (2002) observed that the left lateral PT was more active during perception of consonant–vowel (CV) syllables beginning with voiceless stop consonants than with voiced consonants, whereas Jacquemot, Pallier, LeBihan, Dehaene, and Dupoux (2003) carried out a study showing increased lateral PT activity in the left hemisphere for phonologically salient acoustic changes in speech compared with acoustic changes bearing no phonological relevance (although this activation did extend along the STG). Raizada and Poldrack (2007) found weak categorical responses to a /ba/–/da/ continuum of speech sounds in the PT. Obleser, Zimmermann, Van Meter, and Rauschecker (2007) showed that the magnitude of PT response was no greater for

consonants than for complex acoustic controls; however, PT activity did correlate with particular acoustic properties of the speech sounds. This contradictory evidence in the literature, to date, warrants further investigation—the current study aimed to bring the question forward by explicitly testing sublexical sensitivities in the PT within a task that would emphasize its proposed audiomotor function.

The main objective of the current study was to explore the neural correlates of two phenomena in pWM—the word length effect and effects of phonetic complexity—through orthogonal manipulation of sublexical properties of spoken items. We also aimed to explore the roles of lateral and medial posterior sites in the supratemporal plane and assess these regions' sensitivity to sublexical phonetic information during active maintenance of heard speech (via covert rehearsal during a short poststimulus delay). We employed pseudoword (or nonword) repetition, a pWM task that has been identified as a purer measure of this system, as it limits the use of overt semantic or linguistic strategies or semantic representations to assist task performance (Jacquemot & Scott, 2006; Gathercole, Willis, Baddeley, & Emslie, 1994). Traditionally, the task involves immediate repetition of individual nonsense items (Gathercole et al., 1994), and so does not clearly engage the phonological loop in the same way as other tasks such as digit span. In the current study, we modified the pseudoword repetition task to incorporate a delay phase and allow for the measurement of neural activity related to the two factors of interest, specifically during subvocalization (Experiment 1). We chose a delay phase of 6–7 sec, in order that a trial would be comparable in duration to a trial from the more traditional digit-span task; this allowed us to capture the BOLD response to perception of the auditory items and the early part of active maintenance of the stimulus.

Several previous studies have investigated structural manipulations in the context of pseudoword rehearsal and repetition (Papoutsi et al., 2009; Riecker, Brendel, Ziegler, Erb, & Ackermann, 2008; Strand, Forssberg, Klingberg, & Norrelgen, 2008; Bohland & Guenther, 2006; Klein, Watkins, Zatorre, & Milner, 2006). Of these, a subset used auditory presentation of items (Papoutsi et al., 2009; Strand et al., 2008; Klein et al., 2006). Strand et al. (2008) presented participants with pseudowords comprising five, seven, or nine syllables (e.g., si–li–bo–na–la), which the participant was asked to covertly rehearse before performing a delayed match-to-sample button press on a visually presented item. A control condition involving passive listening to temporally reversed versions of the stimuli was used, on the basis that these had no phonetic or linguistic content. The authors found that covert rehearsal of pseudowords, when contrasted with the baseline, gave left-lateralized activation in superior frontal gyrus (SFG), LIFG, posterior STS, and putamen. However, they found no interaction with the number of syllables. This may have been due to a lack of power in the analysis, which involved 19 different conditions, and the use of a reversed speech baseline, which may not be intelligible but certainly retains aspects of the acoustic and

phonetic information in speech. Papoutsi et al. (2009) and Klein et al. (2006) used shorter pseudowords, of two and four syllables, which the participant was asked to covertly rehearse and then produce after a delay. At both two and four syllables, Klein et al., using PET, also manipulated "articulatory difficulty" through addition of consonant clusters, whereas Papoutsi et al., who used fMRI, employed two levels of phonotactic frequency. In both cases, the pseudoword items were rather more word-like in their structure than those used by Strand et al., although Papoutsi et al. did have to draw upon relatively unusual (although still phonotactically legal) phoneme combinations in English to construct their low-probability items. No baselines or listen-only trials were used in the analyses reported by Papoutsi et al. or Klein et al., and the results were collapsed activity across all phases of the task (perception, rehearsal, repetition). Both studies saw increased activity in a network of superior temporal and precentral sites, with the motor activity showing strong left-lateralization in Klein et al., and a left dominance in Papoutsi et al. A reduction in phonotactic frequency in the Papoutsi et al. study gave increased activity in the bilateral IFG, left precentral gyrus, and left SMA, whereas increased activity was detected in the study of Klein et al. in the bilateral cerebellum and left thalamus for pseudowords with added consonant clusters.

To satisfy the requirements for membership of the "store" or input systems in pWM, a brain region should respond during both perception and rehearsal of verbal material (Buchsbaum & D'Esposito, 2008; Jacquemot & Scott, 2006; Becker, MacAndrew, & Fiez, 1999). In line with this, and with the extra intention of completely separating conditions of "perception + active maintenance" from those of perception alone, we ran a second experiment (Experiment 2) in which the pseudowords were presented to a completely new group of participants and neural activity was sampled during passive listening, without rehearsal or repetition. We predicted that increasing the number of syllables and the number of consonant clusters would give increased activity in a generalized speech production network of frontal premotor and superior temporal sites for perception + active maintenance, but that these effects would be restricted to the temporal lobes for basic perception (in the absence of any readiness to repeat). We expected the greatest commonality between "perception + active maintenance" and "perception only" to occur in the posterior supratemporal plane—within this, in line with previous findings, we predicted a distinction between lateral regions sensitive to segmental structure and medial regions performing auditory–motor template matching at the syllable level.

METHODS

Participants

Experiment 1

Seventeen adult speakers of British English (8 men; mean age = 24 years 11 months, SD = 60.4 months, range =

19 years 1 month to 36 years 2 months) participated in the study. All had healthy hearing and reported no neurological history, nor any problems with speech or language. Participants were recruited from the UCL Psychology Subject Pool and were paid £15 for their participation. The study was approved by the UCL Department of Psychology Ethics Committee.

Experiment 2

Participants were 15 adult speakers of British English (9 men; mean age = 23 years 5 months, SD = 49.4 months, range = 19 years 7 month to 33 years 11 months). All were selected and recruited as described for Experiment 1. None of the participants had taken part in the previous experiment. The study was approved by the UCL Department of Psychology Ethics Committee.

Materials

Pseudowords were constructed in a full 2 × 2 factorial design with the factors number of syllables (2 vs. 4) and number of consonant clusters (0 vs. 2). There were 40 items in each cell. The aim was to manipulate difficulty without necessarily inducing errors, therefore, a strong emphasis was placed on creating items that held no meaning, yet sounded natural.

Forty different basic pseudoword forms were constructed, with two syllables and no clusters, of the form $C_1V_1C_2V_2C_3$ (C = consonant, V = vowel), where the first syllable is stressed (in accordance with the default pattern for English). Each basic pseudoword form was then manipulated in three ways:

1. A consonant added after C_1 to create an onset cluster, and a different consonant added before C_2 to create a code cluster: $C_1\underline{C}V_1\underline{C}C_2V_2C_3$ (2 syllable, 2 cluster condition)
2. Two extra syllables added to create $C_1\underline{VCVC}V_1C_2V_2C_3$ (4 syllable, 0 cluster condition). Primary stress was on the third syllable, secondary stress on the first syllable, in accordance with the default pattern for English.
3. Application of both Steps 1 and 2 above to create $C_1\underline{CVCVC}V_1\underline{C}C_2V_2C_3$ (4 syllable, 2 cluster condition). Primary stress was on the third syllable, secondary stress on the first syllable.

As an example: fIp@l, frIsp@l, fOt@mIp@l, frOt@mIsp@l (where "I" is the short vowel in "hit," "O" the short vowel in "hot," and "@" the centralized schwa vowel (e.g., the last vowel in the word "information"). The added consonants and vowels were varied across the item set to create 160 novel pseudowords. Occasionally, the vowels had to be altered to prevent the creation of a real word within the pseudoword.

The experimental materials were recorded by a female speaker in a sound-proof, anechoic chamber. Recordings were made on a digital audio tape recorder (Sony 60ES; Sony UK Limited, Weybridge, UK) and fed to the S/PDIF digital input of a PC soundcard (M-Audio Delta 66; M-Audio, Iver Heath, UK). The files were downsampled at a rate of 44,100 Hz to mono .wav files with 16-bit resolution, then were further edited into separate .wav files for each item using Cool Edit 96 (Syntrillium Software Corporation, USA), and normalized for peak amplitude in PRAAT (Boersma & Weenink, 2007). A further set of four simple tones was constructed, in PRAAT, for use in a baseline condition. We chose simple tones on the basis that these can be actively maintained and repeated with ease, unlike higher-order controls such as rotated speech. Moreover, the PT responds well to a range of complex sound categories (Griffiths & Warren, 2002), and as it was our primary interest to explore the effects of varying the number of syllables and clusters across pseudoword conditions, we wished to avoid subtracting away too much of the signal in the PT by using an unnecessarily complex baseline sound. However, we did wish to account for the variability in acoustic durations used in the pseudoword conditions, particularly given the dramatic effects of adding extra syllables on this parameter. Therefore, four 350-Hz tones, of durations 0.660, 0.808, 1.003, and 1.119 sec (to match the mean durations of items in the four pseudoword conditions), were used in the final baseline condition. Each tone included a cosine ramp at the onset and offset over a 0.05-sec window, and was normalized in peak amplitude to that of the normalized pseudoword stimuli.

Functional Imaging

Design and Procedure

Experiment 1.   Functional imaging data were acquired on a Siemens Avanto 1.5-Tesla scanner (Siemens AG, Erlangen, Germany) in a single run of 203 echo-planar whole-brain volumes (TR = 11 sec, TA = 3 sec, TE = 50 msec, flip angle = 90°, 35 axial slices, 3 mm × 3 mm × 3 mm in-plane resolution). A sparse-sampling routine (Hall et al., 1999) was employed, in which each stimulus was presented 4.5 sec (with jittering of ±500 msec) before acquisition of the next scan commenced (Figure 1).

Before entering the scanner, participants were told that they would hear "funny, made-up words" that they would be asked to repeat accurately after a delay. They were encouraged not to make any overt speech movement during the delay but that they should "think about" how they would produce the pseudoword. The participants were also told that they would occasionally hear a tone or beep instead of a pseudoword, and for these trials, they should sing or hum the tone after the delay. In order to ensure that the participants would engage fully with the tones, we emphasized that these would vary in acoustic duration across the experiment. We avoided a direct instruction to subvocalize or rehearse the heard items as we wanted the task to approximate the demands of other commonly used
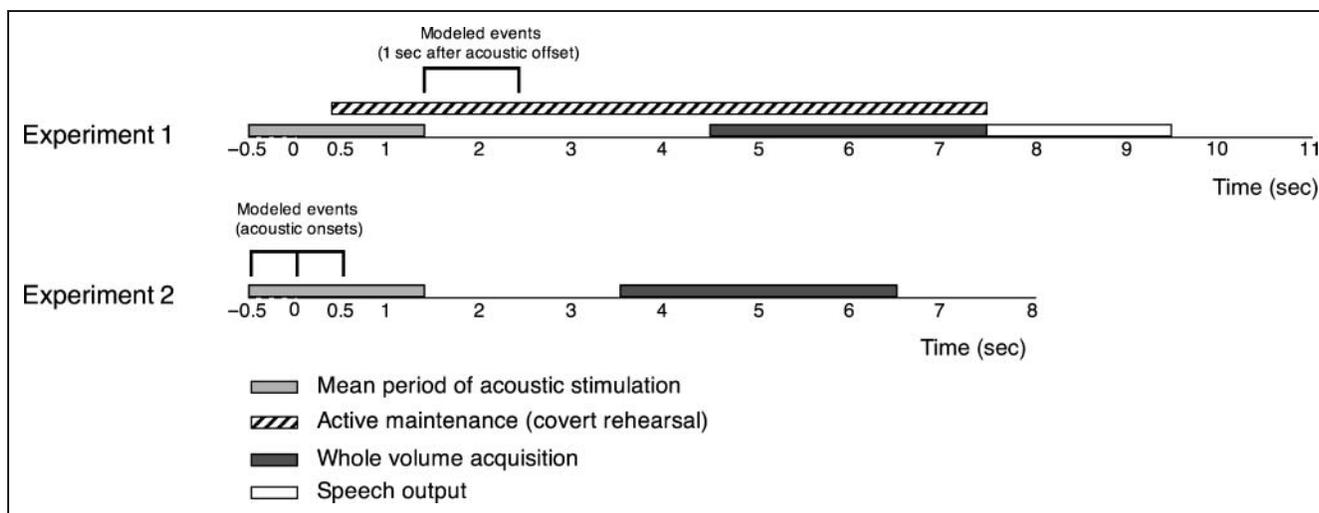
Figure 1. A comparison of the average trial structures in Experiments 1 and 2. "Modeled events" indicate the time range of event onsets as entered in the SPM design.

working memory measures such as digit span, in which it is assumed that the phonological loop will be engaged but in which subvocalization is not explicitly instructed. This was also done to ensure that listeners would not overtly mouth or whisper the pseudowords during the "active maintenance" part of the task. The in-scanner trial structure took advantage of the sparse sampling by using the offset of the scanner acquisition noise as the cue for the participant to give their spoken response. A short simulation of the task, using monosyllabic real words and tones, and including a recording of the scanner noise, was run outside the scanner before the experiment.

In the scanner, the order of presentation of the conditions was pseudorandomized, with each condition being represented once in every five trials. There were 40 trials for each of the pseudoword conditions and 40 baseline tone trials. Participants wore electrodynamic headphones fitted with an optical microphone (MR Confon GmbH, Magdeburg, Germany). Auditory stimuli were delivered using MATLAB (Mathworks Inc., Natick, MA) with the Psychophysics Toolbox extension (Brainard, 1997), via a Denon amplifier (Denon UK, Belfast, UK). The participants' spoken responses were recorded for later scoring using Audacity http://audacity.sourceforge.net).

After the functional run was complete, a high-resolution T1-weighted anatomical image was acquired (HIRes MP-RAGE, 160 sagittal slices, voxel size = 1 mm^3). The total time in the scanner was around 50 min.

Experiment 2. The experiment comprised four functional runs of 86 EPI volumes as described for Experiment 1, however, now with a repetition time of 8 sec. Stimuli were presented in MATLAB using the Cogent 2000 toolbox (Cogent 2000 Team, London, UK), with each sound being played 3.5 sec (with jittering of ±500 msec) before the onset of the next volume acquisition (Figure 1). In this experiment, the participant was told in advance that they would hear

"funny, made-up words" and some tones, and that they should simply listen carefully to the sounds. There were 64 presentations from each of the four pseudoword conditions and the tones baseline (with each token occurring twice throughout the experiment), plus 16 silent rest trials. One participant completed only three of the four functional runs. Visual prompts at the beginning and end of the functional runs were projected from a specially configured video projector (Eiki International, Inc., Rancho Santa Margarita, CA) onto a custom-built front screen, which the participant viewed via a mirror placed on the head coil. Auditory stimuli were delivered via headphones and amplifier as in Experiment 1.

After the functional runs were completed, a high-resolution T1-weighted anatomical image was acquired as described above. The total time in the scanner was around 60 min.

Analysis of fMRI data

Data were preprocessed and analyzed in SPM5 (Wellcome Trust Centre for Neuroimaging, London, UK). Functional images were corrected for slice-timing errors, realigned, co-registered with the anatomical image, normalized using parameters obtained from segmentation of the anatomical image, and smoothed using a Gaussian kernel of 8 mm FWHM. Event-related responses for each event type were modeled as a canonical hemodynamic response function. For Experiments 1 and 2, each condition was modeled as a separate regressor in a generalized linear model, with event onsets modeled at 1 sec after the offset of the acoustic stimulus in Experiment 1, and at the onsets of the acoustic stimuli in Experiment 2. In this way, the data in Experiment 1 reflect responses to the perception and early part of the maintenance phase of the repetition task, whereas the data in Experiment 2 correspond to basic auditory perception (see Figure 1 for a comparison of the trial structure and event modeling in the two experiments). Six

movement parameters (3 translations, 3 rotations) were included as regressors of no interest. At the first level (single subject) in Experiment 1, a contrast image of all pseudowords > tones baseline was generated for later use in an individual-differences analysis. Four further contrast images were created in both experiments for the comparison of each individual pseudoword condition with the tones baseline. These four images from each participant were entered in a random effects, 2 × 2 repeated measures ANOVA group model with factors syllables and clusters. Additional *T*-contrasts of interest were set up within this group model to assess the main positive effect of condition, and the positive and negative effects of the two main factors. The MarsBaR toolbox in SPM (Brett, Anton, Valabregue, & Poline, 2002) was used to construct ROI plots of percentage signal change.

RESULTS

For contrasts measuring responses to all pseudowords > baseline and effects of number of syllables, images were thresholded at a corrected (family-wise error) probability of $p < .05$, with a cluster extent threshold (k) of 20 voxels. For contrasts of number of clusters and the individual-differences analyses, where we anticipated weaker effects, the threshold was dropped to an uncorrected level of $p < .005$, with a cluster extent (k) of 10 voxels. All stereotactic coordinates are reported in MNI space (Montreal Neurological Institute, Canada).

Experiment 1

Behavioral Results

Participants' spoken responses in the scanner were scored with 1 for correct and 0 for incorrect. The group's scores were entered into a 2 × 2 repeated measures ANOVA in SPSS (v.16.0; SPSS Inc., Chicago, IL), with within-subjects factors of number of syllables and number of clusters. Repetition accuracy was significantly reduced for items of four syllables compared with two-syllable items [$F(1,16) = 57.49, p < .001$]. There was a marginally significant cost to accuracy for items with two consonant clusters [$F(1,16) = 4.34, p = .054$], and a nonsignificant interaction of the two factors [$F(1,16) = 8.88, p = .088$]. The results are plotted in Figure 2.

For each participant, a mean percentage accuracy score for all pseudoword conditions was calculated for use in an individual-differences correlation analysis with neural activity in the task.

Functional Imaging

Figure 3A shows a *T*-contrast image for the positive effect of all pseudoword conditions contrasted with the tones baseline. Perception and active maintenance of pseudowords activated the bilateral PT, with peaks in the lateral
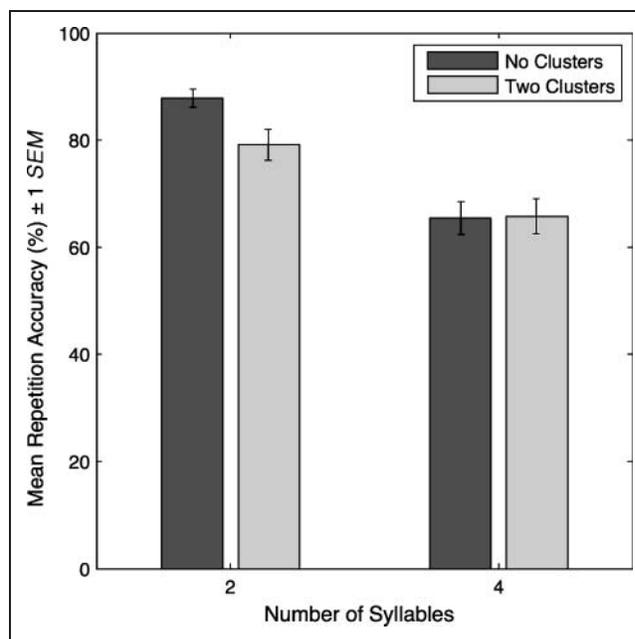
Figure 2. Plot showing pseudoword repetition accuracy by condition, as recorded from participant responses in the scanner during Experiment 1.

PT on the left and right, and an additional region of activation in the left posterior–medial PT. This activation extended slightly anterior to Heschl's gyrus along the STG, although not far beyond the anterior commissure line in either hemisphere. There was also increased activity in the left precentral gyrus (see Table 1 for coordinates and statistics). Figure 3B shows the results of a *T*-contrast for the positive effect of syllables (4 > 2) during perception and active maintenance. The contrast shows several peaks in the supratemporal plane in both hemispheres, extending posterior and medial on the PT, and with temporal activation now including a peak in the anterior STG on the left. Sites on the left precentral gyrus and the right cerebellum also show increased activity (see Table 1). There were no suprathreshold activations for the negative effect of syllables (2 > 4).

Given the marginal effect of clusters in the behavioral task, an uncorrected level of $p < .005$ (with cluster extent $k = 10$) was adopted for the *T*-contrast measuring the positive effect of consonant clusters (2 > 0). This contrast revealed a single area of activation in the left presupplementary motor area (pre-SMA; see Figure 3C).

In order to further investigate the positive effects of clusters, an ROI analysis was carried out on the peak activations from the positive effect of syllables *T*-contrast. To avoid problems of nonindependence, we adopted a hold-one-out approach in which the ROIs for each individual participant were generated from a group contrast of four syllables > two syllables for the other 16 participants. Five spherical ROIs with 4-mm radius (giving a diameter equal to the smoothing FWHM used in preprocessing) were constructed around the two left premotor peaks and the
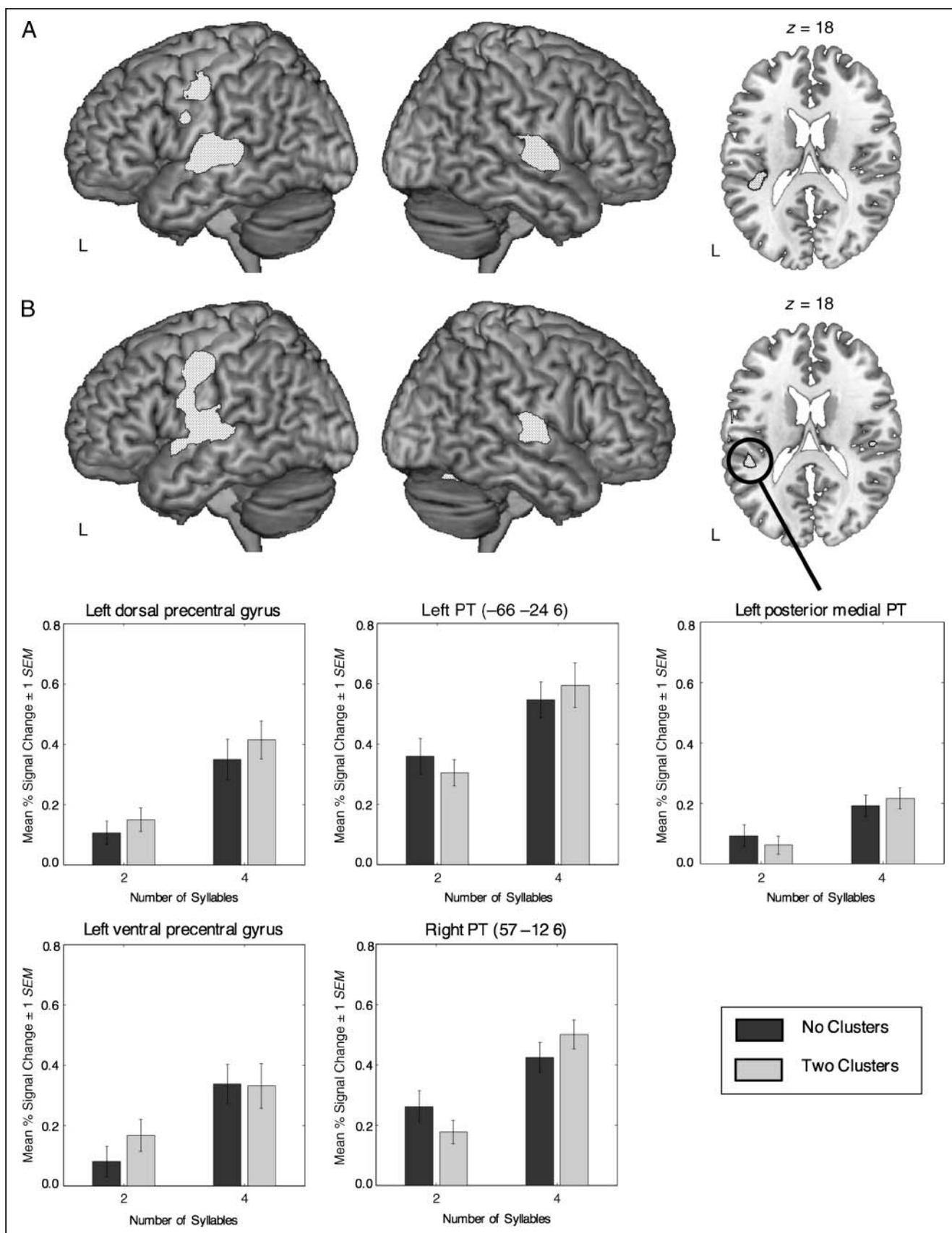
Figure 3. Neural responses to pseudoword rehearsal in Experiment 1: (A) response to all pseudowords > tones baseline, (B) positive correlates of increasing number of syllables, (C) positive correlates of increasing number of consonant clusters, (D) negative correlates of increased number of consonant clusters. All coordinates are reported in MNI space. PT = planum temporale; pre-SMA = presupplementary motor area; STG = superior temporal gyrus.
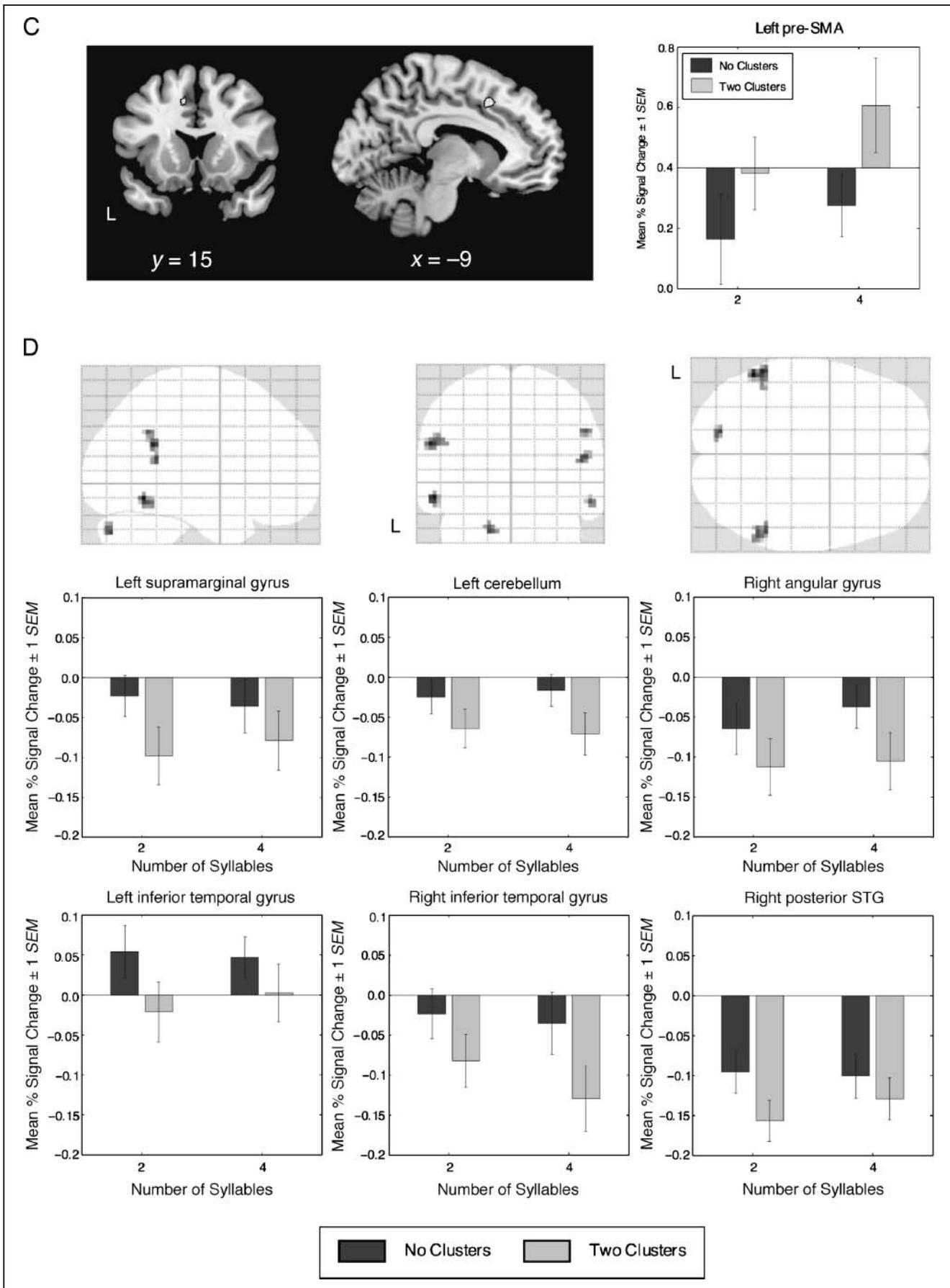
Figure 3. (*continued*)

Table 1. MNI Coordinates and Statistics for Peak and Subpeak Voxels from Contrast Images Obtained in the Analysis of Experiment 1

Contrast	No. of Voxels	Region	Coordinates			T	Z
			x	y	z		
All Pseudowords > Tones	182	Left PT	−66	−27	6	11.01	>8
		Left PT	−60	−9	−3	8.83	7.11
	196	Right PT	63	−9	0	9.58	7.52
	69	Left precentral gyrus	−54	−9	48	7.66	6.43
	49	Left PT	−42	−36	18	6.41	5.61
		Left PT	−36	−27	15	6.24	5.49
4 syllables > 2 syllables	308	Left precentral gyrus (BA 6)	−51	−6	48	8.78	7.08
		Left precentral gyrus (BA 6)	−60	0	24	7.79	6.51
		Left PT	−66	−24	6	7.38	6.25
		Left superior temporal gyrus	−54	6	−6	7.05	6.04
		Left PT	−63	−12	12	6.74	5.84
		Left PT	−63	−15	3	6.68	5.80
		Left PT	−60	−6	3	6.58	5.73
		Left PT	−51	−18	3	5.31	4.82
	154	Right PT	57	−12	6	7.81	6.52
		Right PT	48	−18	12	7.65	6.42
	57	Right cerebellum	18	−63	−21	6.70	5.81
		Right cerebellum	33	−60	−24	6.21	5.47
	22	Left PT	−48	−36	18	6.25	5.50
2 clusters > No clusters*	13	Left presupplementary motor area (pre-SMA)	−12	15	45	3.50	3.33
No clusters > 2 clusters*	17	Left inf. temporal gyrus	−54	−54	−9	3.51	3.34
	28	Left supramarginal gyrus	−57	−48	27	3.43	3.27
	13	Left cerebellum	−15	−78	−33	3.27	3.13
	13	Right posterior sup. temp. gyrus	54	−45	18	3.26	3.12
	12	Right angular gyrus	54	−51	36	3.17	3.04
	12	Right inf. temporal gyrus	57	−51	−12	3.17	3.04
Positive correlation with behavior*	23	Right angular gyrus	60	−54	24	5.91	4.19
	19	Left hippocampus	−24	−21	−9	3.83	3.15

PT = planum temporale; inf. = inferior.

*Clusters obtained at an uncorrected threshold of $p < .005$ (cluster extent threshold $k = 10$ voxels).

peak PT activation from the left fronto-temporal activation, and the peak voxels from each of the other two sites of significant activation (right PT and left posterior–medial PT), using the MarsBaR toolbox in SPM5 (Brett et al., 2002). The ANOVA analyses of percent signal change values obtained from these ROIs revealed a significant main effect of clusters in the left dorsal premotor [$F(1,16) = 7.91$, $p = .013$] and ventral premotor peaks [$F(1,16) = 5.15$, $p = $

.037]. A significant interaction of syllables and clusters was observed in the right PT [$F(1,16) = 5.91$, $p = .027$], which reflected a marginally significant positive effect of clusters at four syllables [$t(16) = −2.08$, $p = .054$] and no effect at two syllables [$t(16) = 1.59$, $p > .10$].

Although it was not predicted, a negative response to increased clusters (no clusters > two consonant clusters) was observed at the lower threshold, in a collection of sites

in the temporal and parietal lobes (see Figure 3D and Table 1), including the bilateral inferior temporal gyrus (ITG), the left supramarginal gyrus, the right angular gyrus (AG), and the right temporo-parietal junction. ROI plots of percent signal change indicate that, in several of these sites, the pseudowords produced *deactivation* relative to the tone baseline, while there appeared to be no modulation of activity by length in syllables.

In order to assess the functional correlates of individual variation in pseudoword repetition, we ran a random effects regression analysis on a contrast of all pseudowords > baseline generated at the single-subject level, with individual mean accuracy scores from the in-scanner task as a covariate. At a threshold of $p < .005$ (uncorrected; $k = 10$ voxels), this analysis revealed two loci of significant activation in the right AG and the left hippocampus (see Figure 4 and Table 1). Extraction of percent signal change data from 4-mm spherical ROIs around the peak voxels allowed these sites of positive correlation to be plotted and explored (see Figure 4).

Experiment 2

Functional Imaging

Figure 5A shows the positive effect (*T*-contrast) of pseudowords over the tones baseline in Experiment 2, with acti-

vation confined to the bilateral PT, extending posterior and medial in both hemispheres (see Table 2 for coordinates and statistics). Figure 5B shows the results of a *T*-contrast for the positive effect of increased number of syllables when participants were asked to listen to the pseudoword stimuli without rehearsal or repetition. We observed strong activation bilaterally along the supratemporal plane, extending posterior and medial to primary auditory cortex (see also Table 2). In Experiment 2, there were no significant voxels at the reduced threshold in a *T*-contrast for positive effects of added clusters (two clusters > no clusters), nor were there any indications from ROI data within the positive syllables contrast of any statistically significant sensitivity to added clusters, either as a main effect or in an interaction, in the activated regions (using 4-mm-radius ROIs built around each of 4 peak voxels generated by a hold-one-out approach—left lateral PT, left medial PT, right lateral PT, right medial PT).

As in Experiment 1, a number of activations showed an unpredicted negative effect of clusters (no clusters > clusters; see Figure 5C and Table 2), this time including a large cluster extending from the right middle temporal gyrus (MTG) to the middle occipital gyrus. Additional regions demonstrating a significant negative effect of complexity were located in the left MTG, bilateral fusiform gyrus, left primary visual cortex, and right precentral gyrus. Again, as
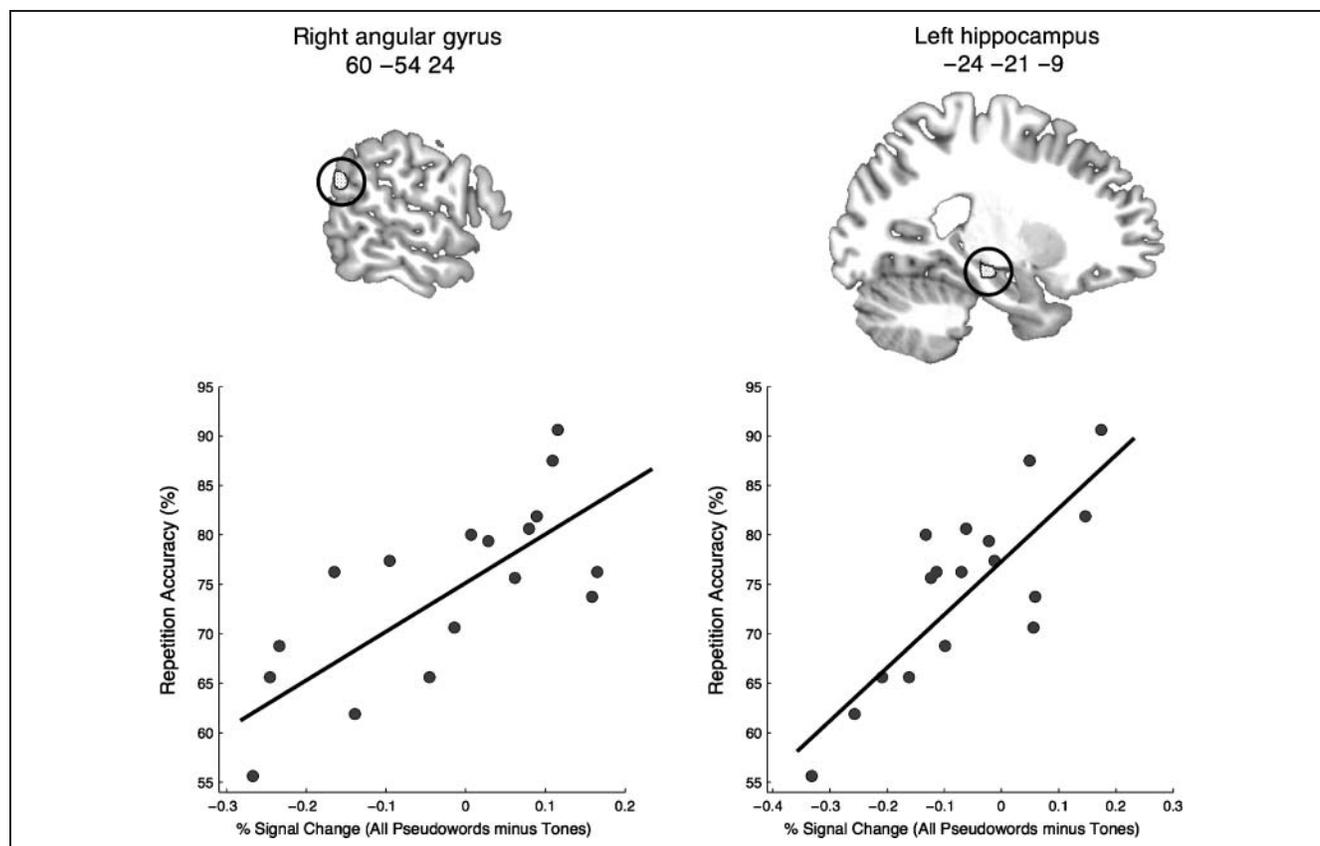
Figure 4. Brain areas showing positive correlation of activity in the all pseudowords > tones baseline contrast and mean behavioral performance on the repetition task. Coordinates are reported in MNI space.
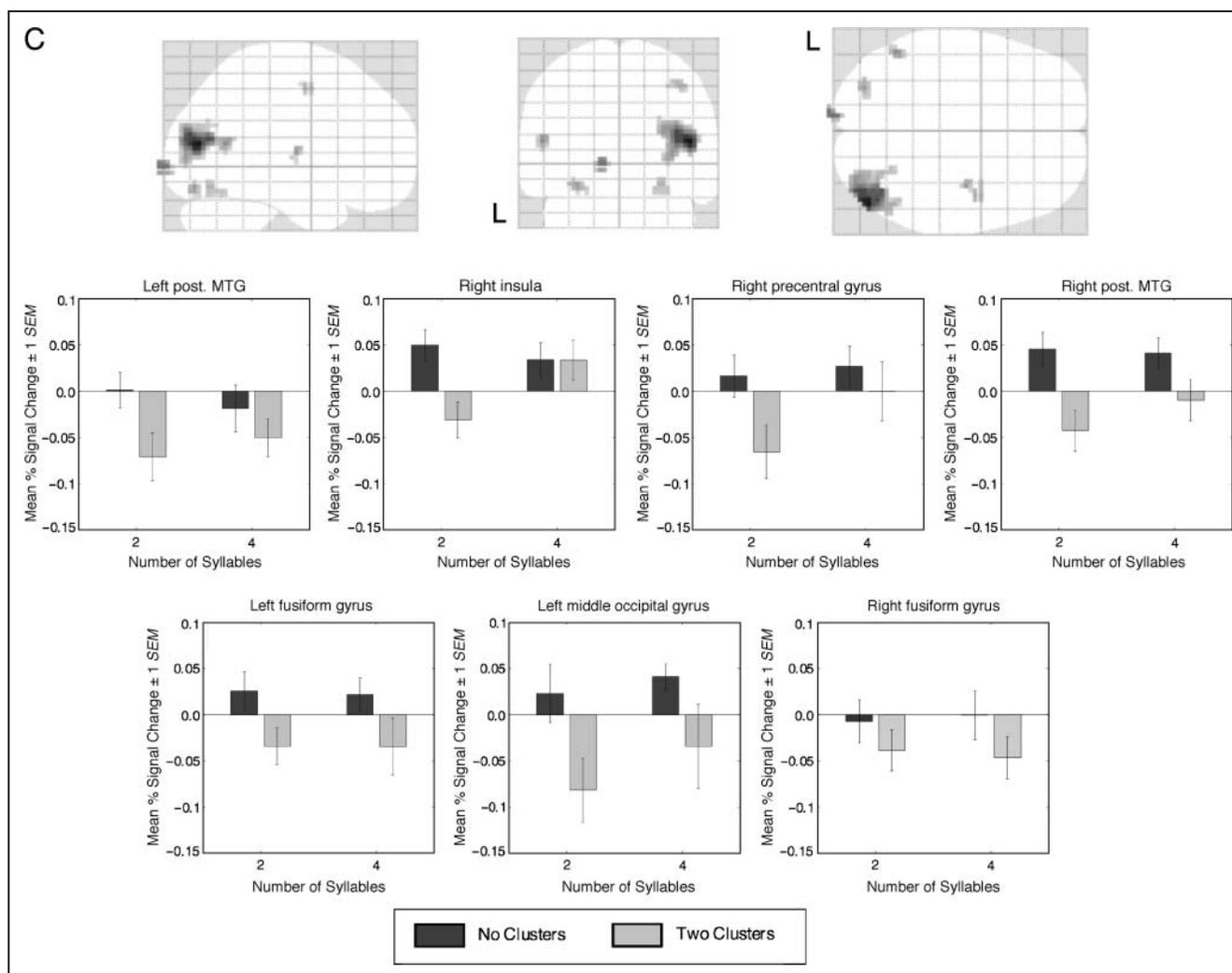
Figure 5. Neural responses during passive listening to pseudowords in Experiment 1: (A) response to all pseudowords > tones baseline, (B) positive correlates of increasing number of syllables, (C) negative correlates of increased number of consonant clusters. Coordinates are reported in MNI space. PT = planum temporale; post. = posterior; MTG = middle temporal gyrus.

Figure 5. (*continued*)

in Experiment 1, many of the ROI plots for the peak voxels in these regions show deactivation relative to the tones baseline.

DISCUSSION

During a pWM task involving delayed pseudoword repetition, there was a positive effect of increasing the number of syllables in left motor cortex, and in the left and right supratemporal plane, extending bilaterally into the medial PT. A distinct site of activation in this region lay in the functionally defined site in the left posterior–medial PT that has been described as a crucial locus for audio-motor transformations, and a key structure in pWM (Hickok, 2009; Buchsbaum & D'Esposito, 2008; Buchsbaum et al., 2001). Based on the previous literature, this site is a likely candidate for the "phonological store," although further work exploring longer rehearsal phases may be needed to functionally separate its role from that of more lateral PT sites (Buchsbaum et al., 2005). A contrast for the positive effect of consonant clusters demonstrated a single activation site in the left pre-SMA, whereas ROI analyses of signal

change for all conditions in the positive syllables contrast showed that there is a main effect of adding consonant clusters in the left precentral gyrus, plus some evidence of a positive effects of clusters (for longer items only) in the right lateral PT. In contrast, an experiment involving passive listening to the same pseudowords showed positive activation associated with increasing the number of syllables that was limited to bilateral superior temporal regions, extending into posterior–medial sites on the PT. There was no evidence, through main effects or interactions, for any positive effect of consonant clusters during passive listening without active maintenance or repetition.

In being completely separate from the first experiment, our second experiment allowed for the independent assessment of the effects of syllables and clusters during passive listening, uncontaminated by the effects of any readiness to repeat. For the general contrast of pseudowords over tones, and for positive effects of increasing the number of syllables, we found that the greatest overlap between active maintenance and listening contexts occurred in posterior portions of the supratemporal plane, extending bilaterally into posterior–medial portions of the

Table 2. MNI Coordinates and Statistics for Peak and Subpeak Voxels from Contrast Images Obtained in the Analysis of Experiment 2

Contrast	No. of Voxels	Region	Coordinates			T	Z
			x	y	z		
All Pseudowords > Tones	279	Left PT	−60	−18	6	13.03	>8
		Left PT	−42	−27	12	7.13	5.99
		Left PT	−45	−39	12	6.01	5.26
	245	Right PT	57	−18	6	8.44	6.75
		Right PT	63	−12	3	8.19	6.61
		Right PT	42	−27	12	8.17	6.60
4 syllables > 2 syllables	214	Left PT	−54	−18	6	9.64	7.37
		Left PT	−36	−33	12	8.20	6.62
		Left PT	−42	−27	9	7.72	6.34
	149	Right PT	51	−15	3	8.09	6.56
		Right PT	42	−27	12	7.33	6.11
No clusters > 2 clusters*	269	Right MTG (BA 39)	48	−75	15	5.46	4.87
		Right middle occipital gyrus	39	−81	21	4.58	4.20
		Right MTG (BA 39)	39	−66	18	4.35	4.02
		Right middle occipital gyrus	36	−84	6	3.37	3.20
		Right MTG	48	−60	18	3.19	3.04
		Right superior occipital gyrus	27	−63	27	2.70	2.61
	18	Left mid. occipital gyrus (BA 17)	−9	−99	0	3.95	3.70
	20	Left MTG	−51	−57	15	3.52	3.33
	11	Right insula	36	−9	9	3.40	3.23
	19	Left fusiform gyrus	−27	−78	−12	3.37	3.20
	17	Right precentral gyrus (BA 6)	42	0	51	3.19	3.05
	22	Right fusiform gyrus	27	−66	−12	3.07	2.94

MTG = middle temporal gyrus.

*Cluster obtained at an uncorrected threshold of $p < .005$ (cluster extent threshold $k = 10$ voxels).

PT. Hence, outside of any requirement to reproduce the pseudowords, posterior temporal regions show a strong response to pseudoword items, thus supporting the earlier indications from listening runs within a rehearsal study (Hickok et al., 2003). Where other authors found the posterior STG/STS, however, the main focus of our activations in the lateral PT was in the STG. In our study, the only evidence for a sensitivity to added consonant clusters in the PT occurred in the right lateral PT in Experiment 1 only, and only for items of four syllables in length. Overall, the data suggest that, in passive listening, medial and lateral PT sites encode phonetic structure, without the context of semantic processing, in a suprasegmental fashion. When the task is more demanding, it is only the lateral PT that shows a significant magnitude-based sensitivity to segmental manipulations; this supports previous findings by Jacquemot et al. (2003) and Jancke et al. (2002). This find-

ing also fits in with the model described by Hickok (2009), in which lateral superior temporal regions perform phonetic analysis on auditory input, whereas medial sites are more concerned with audiomotor conversion. We suggest that the medial PT may store templates, in this case at the level of the syllable, to which the incoming signal is matched and transformed into motor representations or plans (Warren et al., 2005).

Despite the above, inspection of ROI plots from both experiments shows a small increase in signal with the addition of clusters across many of the peak voxels responsive to increasing number of syllables. This could indicate that, although the overall pattern is one in which posterior temporal areas code the incoming speech according to combinations of familiar segments or articulations (at the level of the syllable), some subpopulations may perform analysis on the speech input that is more faithful to finer-grained

phonetic information (Hickok & Poeppel, 2007). In a recent review, Obleser and Eisner (2009) outline the research, to date, that has found evidence for prelexical abstraction of speech in auditory cortex. With regard to the PT, they acknowledge the difficulty in obtaining magnitude-based indicators of phonological or categorical speech processing, although this is readily observed in the STS. They put forward the argument that the lack of a speech advantage in the PT may not necessarily indicate a lack of sensitivity to prelexical speech structures, but that categorization information may be transmitted from the PT by means of a distributed pattern of activation to a later site in the processing stream, for example, the STS. A similar argument is also presented by Raizada and Poldrack (2007): It is possible that focal populations of cells in the PT are sensitive to phonetic/phonological changes in speech, but that this effect is swamped by the general lack of sensitivity in the region. Both sets of authors therefore propose that future work may better benefit from the advent of techniques based in multivariate pattern analysis. Indeed, a recent paper (Hickok et al., 2009) showed evidence from pattern analysis for a separation of the neural populations in the PT that respond during listening from those involved in later stages of covert rehearsal, but they did not investigate responses to the phonetic structure of the stimuli. We have presented numerical indications of sensitivity to consonant clusters in the lateral and medial PT—future experiments may well benefit from pattern-based analysis strategies.

Previous studies have shown a wider network of activity in response to articulatory/phonetic complexity than observed in the current study (Papoutsi et al., 2009; Riecker et al., 2008; Bohland & Guenther, 2006; Klein et al., 2006). All were able to identify positive correlates of increased complexity (via addition of consonant clusters or reduction of phonotactic probability) at the whole-brain level in motor areas such as the insula, SMA, and cerebellum. We also found a positive effect of added complexity in the left SMA, when the threshold was lowered to an uncorrected level of $p < .005$. There may be several reasons why we did not see whole-brain responses to added consonant clusters at a higher threshold. The complexity manipulations in previous studies were quite dramatic, such that some of the combinations used would be very unlikely to occur in real English utterances (Papoutsi et al., 2009; Bohland & Guenther, 2006). It is thus likely that these manipulations will have placed much greater loading on articulatory mechanisms. As our emphasis was on naturalness and relative ease of production, we were expecting the syllable number contrast to provide the larger effects (as in the behavioral literature on working memory), with the clusters/complexity contrast likely to require an ROI approach. Another important difference from previous studies is that their analyses included BOLD data corresponding to overt speech responses, either because the task involved frequent spoken responses with short intertrial intervals (Riecker et al., 2008), because several responses were collected in individual PET scans (Klein et al., 2006), or because fMRI data

from several stages of the task had been used together in the analysis (Papoutsi et al., 2009). As our interest was in investigating active maintenance processes as they might occur during a classic test of short-term memory (e.g., digit span), we intentionally avoided sampling BOLD responses to overt movement.

Unexpectedly, both datasets in the current study showed, at an uncorrected threshold, a network of cortical areas showing decreased activation in response to increased phonetic complexity. As none of the activities involved lay in regions of cortex associated with the early processing of speech or phonetic structure, it seems unlikely that the effects seen are acoustic or phonetic. However, the involvement of regions, such as the ITG and the AG (Experiment 1), and the fusiform gyrus (Experiment 2), suggests that these effects may reflect some attempt at semantic processing of the pseudoword stimuli, or a strategy in which participants engaged in visual imagery of possible written equivalents of the pseudowords. The ITG has previously been implicated in the visuo-semantic processing of words (Heim et al., 2009; Fiebach, Friederici, Muller, & von Cramon, 2002), and it may be that for the simpler phonotactic structures, the participants in the current study were making some attempt to map the heard items onto real-word neighbors to aid maintenance in pWM (Experiment 1). Raettig and Kotz (2008) presented data supporting this interpretation—they found activation in sites including the ITG, MTG, and AG related to the extent of engagement in lexical processing in an experiment involving words and pseudowords.

It is also apparent from Figures 2D and 5C that responses to the pseudowords in these contrasts often showed overall deactivation compared with the tones baseline. It may be that this inverse effect of added complexity reflects diversion of blood flow from noncritical sites for the task (e.g., primary visual cortex in Experiment 2) to centrally involved regions for those more taxing or complex pseudoword items (McKiernan, Kaufman, Kucera-Thompson, & Binder, 2003). Recent data indicate that the results in the fusiform gyrus may reflect the contribution of selective attention to heard speech in this region (Yoncheva, Zevin, Maurer, & McCandliss, 2010). Importantly for our contrasts of interest, these negative effects of added complexity indicate that the introduction of two extra consonant clusters to the simple two- and four-syllable pseudowords was enough to cause significant changes in neural activity in several neural sites, and thus, the relative lack of positive complexity-related activity was unlikely to be due to an insufficiently strong manipulation. Further explorations may need to address the possibility that, despite the lack of meaning in pseudowords, their inherent "wordlikeness" in phonotactic structure will lead to the brain attempting to process them as real words, which may, in turn, interact with basic segmental manipulations as indicated by the negative effects of increased complexity seen here. A further hint at this may come from the more anterior distribution of auditory cortex activation on the left than on the right for the posi-

tive effect of increased length in Experiment 1. Participants may employ a strategy of semantic processing of real words along the ventral stream for intelligible speech (Scott et al., 2000, 2006) in order to support the active maintenance of similar-sounding pseudowords. It would be interesting to explore the time course of these negative effects of added consonant clusters on the BOLD signal, as we would hypothesize that any strategic effects would happen at a lag after initial perception of the stimuli. Unfortunately, the limitations of sparse sampling routines meant that such an analysis was not possible from the current dataset.

An important point made by Buchsbaum and D'Esposito (2008), with reference to the earlier studies of pWM, is that, just because these studies identified the "phonological store" in a location incongruent with the psychological models of pWM (i.e., inferior parietal cortex), does not mean that such regions are not involved in some way in working memory processes. As we could only obtain a simple accuracy score from the participants' spoken output in Experiment 1, the summary behavioral score cannot be sensitive to the exact source of an error in the pseudoword repetition process (e.g., perception, encoding, maintenance, preparation for motor output, speech production)—this is a classic problem for nonword repetition and similar pWM tasks (Gathercole et al., 1994). However, the identification of neural correlates of overall task accuracy speaks to the attentional set and basic task strategies adopted by participants in performance of the task, which is important for relating the operation of the "core" working memory system to behavior in similar real-world scenarios, for example, holding a person's phone number in working memory while you find a pen to write it down. In the current study, we identified two regions that showed a significant positive correlation in activity (for all pseudowords minus tones) with accuracy on the in-scanner repetition task—right AG and left hippocampus. The AG activity may relate to semantic processing—the left AG is a structure that is activated when participants are asked to make explicit semantic decisions on spoken material (Sharp, Scott, & Wise, 2004), and has been implicated in the semantic processing of degraded speech (Obleser, Wise, Dresner, & Scott, 2007). However, Strand et al. (2008) cite numerous findings in the literature of AG involvement in the interpretation of orthographic forms during working memory tasks—it may be that the better listeners in our study are those who made better use of an orthographic strategy to visualize the written forms of the auditory items they were asked to rehearse. Alternatively, this right parietal activity may reflect attentional mechanisms, where those participants most keenly engaged with the task were those who made fewer errors. The hippocampus has long been associated with long-term memory formation. The current result could be interpreted within a word-learning framework—Davis, Di Betta, Macdonald, and Gaskell (2009) found greater responses in the hippocampus to completely novel words compared with previously trained novel words and real words, and within this, a positive correlation between hippocampal activity

and subsequent performance on a recognition test for the untrained novel words. In the Davis et al. (2009) experiment, as in ours, it appears that the hippocampus is important in the acquisition of new words—the more faithfully this is done, the better participants are at remembering and repeating them across a range of time scales.

We have described a pair of fMRI experiments assessing neural responses to pseudoword structure during the maintenance phase in a pWM task and in passive listening. Perception and active maintenance of auditory pseudowords recruits auditory and motor areas, with both regions showing increased responses to longer words (i.e., with more syllables), whereas increased activity for items of greater phonetic complexity (i.e., with more consonant clusters) is largely limited to motor regions. The greatest overlap between the two tasks occurred in the posterior PT for basic comparisons of pseudowords over tones, and for increasing the number of syllables in the pseudoword, and thus, this area emerges as a likely candidate region for the "phonological store." In contrast, analysis of individual differences showed that maintenance-related activity in regions outside auditory cortex that are classically associated with semantic and memory tasks is positively correlated with accuracy on pseudoword repetition.

REFERENCES

Awh, E., Jonides, J., Smith, E. E., Schumacher, E. H., Koeppe, R. A., & Katz, S. (1996). Dissociation of storage and rehearsal in working memory: PET evidence. *Psychological Science, 7,* 317–331.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York: Academic Press.

Baddeley, A. D., Thompson, N., & Buchanan, M. (1974). Word length effect and structure of short-term-memory. *Bulletin of the Psychonomic Society, 4,* 240.

Becker, J. T., MacAndrew, D. K., & Fiez, J. A. (1999). A comment on the functional localization of the phonological storage subsystem of working memory. *Brain and Cognition, 41,* 27–38.

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

Binder, J. R., Frost, J. A., Hammeke, T. A., Rao, S. M., & Cox, R. W. (1996). Function of the left planum temporale in auditory and linguistic processing. *Brain, 119,* 1239–1247.

Boersma, P., & Weenink, D. (2007). Praat: Doing phonetics by computer (Version 5) [Software]. Retrieved from: www.praat.org on 10 December 2007.

Bohland, J. W., & Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. *Neuroimage, 32,* 821–841.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436.

Brett, M., Anton, J. L., Valabregue, R., & Poline, J. B. (2002). Region of interest analysis using an SPM toolbox. *Neuroimage, 16.*

Buchsbaum, B. R., & D'Esposito, M. (2008). The search for the phonological store: From loop to convolution. *Journal of Cognitive Neuroscience, 20,* 762–778.

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 phonological working memory. *Neuron, 48,* 687–697.

Caplan, D., Rochon, E., & Waters, G. S. (1992). Articulatory and phonological determinants of word-length effects in span tasks. *Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology, 45,* 177–192.

Davis, M. H., Di Betta, A. M., Macdonald, M. J. E., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience, 21,* 803–820.

Fiebach, C. J., Friederici, A. D., Muller, K., & von Cramon, D. Y. (2002). fMRI evidence for dual routes to the mental lexicon in visual word recognition. *Journal of Cognitive Neuroscience, 14,* 11–23.

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The Children's Test of Nonword Repetition: A test of phonological working memory. *Memory, 2,* 103–127.

Griffiths, T. D., & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends in Neurosciences, 25,* 348–353.

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., et al. (1999). "Sparse" temporal sampling in auditory fMRI. *Human Brain Mapping, 7,* 213–223.

Heim, S., Eickhoff, S. B., Ischebeck, A. K., Friederici, A. D., Stephan, K. E., & Amunts, K. (2009). Effective connectivity of the left BA 44, BA 45, and inferior temporal gyrus during lexical and phonological decisions identified with DCM. *Human Brain Mapping, 30,* 392–402.

Hickok, G. (2009). The functional neuroanatomy of language. *Physics of Life Reviews, 6,* 121–143.

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,* 673–682.

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

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 Reviews Neuroscience, 8,* 393–402.

Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S., & Dupoux, E. (2003). Phonological grammar shapes the auditory cortex: A functional magnetic resonance imaging study. *Journal of Neuroscience, 23,* 9541–9546.

Jacquemot, C., & Scott, S. K. (2006). What is the relationship between phonological short-term memory and speech processing? *Trends in Cognitive Sciences, 10,* 480–486.

Jancke, L., Wustenberg, T., Scheich, H., & Heinze, H. J. (2002). Phonetic perception and the temporal cortex. *Neuroimage, 15,* 733–746.

Jones, D. M., & Macken, W. J. (1996). Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 369–381.

Klein, D., Watkins, K. E., Zatorre, R. J., & Milner, B. (2006). Word and nonword repetition in bilingual subjects: A PET study. *Human Brain Mapping, 27,* 153–161.

McKiernan, K. A., Kaufman, J. N., Kucera-Thompson, J., & Binder, J. R. (2003). A parametric manipulation of factors affecting task-induced deactivation in functional neuroimaging. *Journal of Cognitive Neuroscience, 15,* 394–408.

Murray, A., & Jones, D. M. (2002). Articulatory complexity at item boundaries in serial recall: The case of Welsh and English digit span. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 594–598.

Obleser, J., & Eisner, F. (2009). Pre-lexical abstraction of speech in the auditory cortex. *Trends in Cognitive Sciences, 13,* 14–19.

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. *Journal of Neuroscience, 27,* 2283–2289.

Obleser, J., Zimmermann, J., Van Meter, J., & Rauschecker, J. P. (2007). Multiple stages of auditory speech perception reflected in event-related fMRI. *Cerebral Cortex, 17,* 2251–2257.

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

Paulesu, E., Frith, C. D., & Frackowiak, R. S. J. (1993). The neural correlates of the verbal component of working memory. *Nature, 362,* 342–345.

Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience, 139,* 23–38.

Raettig, T., & Kotz, S. A. (2008). Auditory processing of different types of pseudo-words: An event-related fMRI study. *Neuroimage, 39,* 1420–1428.

Raizada, R. D. S., & Poldrack, P. A. (2007). Selective amplification of stimulus differences during categorical processing of speech. *Neuron, 56,* 726–740.

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience, 12,* 718–724.

Riecker, A., Brendel, B., Ziegler, W., Erb, M., & Ackermann, H. (2008). The influence of syllable onset complexity and syllable frequency on speech motor control. *Brain and Language, 107,* 102–113.

Scott, S. K., Blank, C. 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., Rosen, S., Lang, H., & Wise, R. J. S. (2006). Neural correlates of intelligibility in speech investigated with noise vocoded speech—A positron emission tomography study. *Journal of the Acoustical Society of America, 120,* 1075–1083.

Sharp, D. J., Scott, S. K., & Wise, R. J. (2004). Monitoring and the controlled processing of meaning: Distinct prefrontal systems. *Cerebral Cortex, 14,* 1–10.

Smith, E. E., Jonides, J., & Koeppe, R. A. (1996). Dissociating verbal and spatial working memory using PET. *Cerebral Cortex, 6,* 11–20.

Smith, E. E., Jonides, J., Marshuetz, C., & Koeppe, R. A. (1998). Components of verbal working memory: Evidence from neuroimaging. *Proceedings of the National Academy of Sciences, U.S.A., 95,* 876–882.

Strand, F., Forssberg, H., Klingberg, T., & Norrelgen, F. (2008). Phonological working memory with auditory presentation of pseudo-words: An event related fMRI Study. *Brain Research, 1212,* 48–54.

Wager, T. D., & Smith, E. E. (2003). Neuroimaging studies of working memory: A meta-analysis. *Cognitive, Affective & Behavioral Neuroscience, 3,* 255–274.

Warren, J. E., Wise, R. J. S., & Warren, J. D. (2005). Sounds do-able: Auditory–motor transformations and the posterior temporal plane. *Trends in Neurosciences, 28,* 636–643.

Wilson, M. (2001). The case for sensorimotor coding in working memory. *Psychonomic Bulletin & Review, 8,* 44–57.

Yoncheva, Y., Zevin, J. D., Maurer, U., & McCandliss, B. D. (2010). Auditory selective attention to speech modulates activity in the visual word form area. *Cerebral Cortex, 20,* 622–632.