

ORIGINAL ARTICLE

Non-perceptual Regions in the Left Inferior Parietal Lobe Support Phonological Short-term Memory: Evidence for a Buffer Account?

Qiuhai Yue¹, Randi C. Martin¹, A. Cris Hamilton¹ and Nathan S. Rose²¹Department of Psychology, Rice University, MS-25, P.O. Box 1892, Houston, TX 77251-1892, USA and²Department of Psychology, University of Notre Dame, Notre Dame, IN 46556, USAAddress correspondence to Randi C. Martin, Department of Psychology, Rice University, MS-25, P.O. Box 1892, Houston, TX 77251-1892, USA.
Email: rmartin@rice.edu

Abstract

Buffer versus embedded processes accounts of short-term memory (STM) for phonological information were addressed by testing subjects' perception and memory for speech and non-speech auditory stimuli. Univariate and multivariate (MVPA) approaches were used to assess whether brain regions recruited in recognizing speech were involved in maintaining speech representations over a delay. As expected, a left superior temporal region was found to support speech perception. However, contrary to the embedded processes approach, this region failed to show a load effect, or any sustained activation, during a maintenance delay. Moreover, MVPA decoding during the maintenance stage was unsuccessful in this region by a perception classifier or an encoding classifier. In contrast, the left supramarginal gyrus showed both sustained activation and a load effect. Using MVPA, stimulus decoding was successful during the delay period. In addition, a functional connectivity analysis showed that, as memory load increased, the left temporal lobe involved in perception became more strongly connected with the parietal region involved in maintenance. Taken together, the findings provide greater support for a buffer than embedded processes account of phonological STM.

Key words: fMRI, inferior parietal lobe, multivariate pattern analysis, phonological short-term memory, speech perception

Introduction

Phonological short-term memory (STM) refers to the capacity for retaining speech sounds over a short time period—a capacity that is critical for language acquisition in children (Gathercole et al. 1997) and new word learning in adults (Baddeley et al. 1998). Traditionally, cognitive models of STM have assumed the existence of a specialized buffer for maintaining phonological information (Baddeley et al. 1984; Martin et al. 1999). More recently, however, researchers have postulated “embedded processes” models for the short-term retention of all types of information. In the embedded processes approach, STM is assumed to consist of the results of perceptual processing and the corresponding activation of long-term memory representations (Cowan 2001; Oberauer and Lange 2009). Thus, in the speech domain, long-term

representations for speech units (e.g., phonemes, syllables, and words) would be activated during speech recognition and the persisting activation of these units would constitute STM (Martin and Saffran 1997).

To distinguish the dedicated buffer and embedded processes models, 1 approach from a cognitive neuroscience perspective is to determine whether the same or different neural substrates support speech recognition and STM. The earliest evidence in favor of the buffer approach came from studies by Warrington and colleagues (Warrington and Shallice 1969; Warrington et al. 1971) who reported patients with impaired auditory verbal STM but preserved word perception, leading to the claim of a separate storage buffer for STM (Shallice and Warrington 1977). Lesion overlap for patients localized the

critical region for phonological STM in the inferior parietal lobe, specifically in the left supramarginal gyrus (SMG) (Shallice and Vallar 1990; Vallar and Papagno 1995; see also Paulesu et al. 2017), which was distinct from the left posterior superior temporal lobe thought to support spoken word recognition (Price 2010, 2012).

Other researchers have claimed that patients with phonological STM deficits do have subtle deficits of speech perception that might be the source of their STM deficit (Allport 1984; Belleville et al. 2003), though there is evidence that these subtle deficits are not sufficient to cause the profound STM deficits (Martin and Breedin 1992). Two recent large-sample studies using voxel-based lesion-symptom mapping methods have provided evidence that seems to support the conclusion that phonological STM depends on speech perception regions, demonstrating a relation between phonological STM and superior temporal lobe damage (Leff et al. 2009; Baldo et al. 2012). However, in the Leff et al. study, nonword repetition ability was partialled out of the lesion-behavior correlations, which may have removed an important component of phonological STM given prior evidence that nonword repetition reflects pSTM capacity (Gathercole and Baddeley 1989; Gupta 2003; Gupta et al. 2005; Majerus 2013). In the Baldo et al. study, correlations with damage to both inferior parietal and superior temporal regions were observed for pseudoword and word span; however, speech perception abilities were not controlled for, which may account for the relation to temporal damage.

Evidence from neuroimaging studies of healthy subjects has also been mixed regarding the storage buffer versus embedded processes debate about phonological STM. In functional magnetic resonance imaging (fMRI) studies, a delayed recognition task has usually been used to tap short-term retention (Sternberg 1966; Curtis and D'Esposito 2003), with researchers testing for regions showing sustained activity during the delay (Goldman-Rakic 1995; Curtis and D'Esposito 2003; Leavitt et al. 2017) or a load effect—i.e., greater activity for a higher memory load (Cohen et al. 1997; Jonides et al. 1997). Early neuroimaging work found these effects for phonological STM tasks in the left inferior parietal lobe (Paulesu et al. 1993; Salmon et al. 1996; Martin et al. 2003), providing some evidence for the buffer account, whereas others did not (Buchsbaum et al. 2001; Ravizza et al. 2011). For example, Ravizza et al. (2011) failed to find any sustained activation relative to baseline during maintenance in a left parietal region (i.e., temporo-parietal junction, TPJ), but found such activation in an a priori region-of-interest (ROI) in the left superior temporal gyrus. However, this ROI was quite posterior and seemingly more posterior than regions typically thought to be involved in speech perception (Price 2010, 2012). Also, the Ravizza et al. (2011) study, and many other neuroimaging studies (Braver et al. 1997; Rypma and D'Esposito 1999; Rypma et al. 1999; Ravizza et al. 2004; Narayanan et al. 2005; Romero et al. 2006; Lewis-Peacock et al. 2012; Langel et al. 2014) have used visual rather than auditory presentation of their verbal stimuli. Although there is substantial evidence that subjects phonologically recode visually presented verbal stimuli, one might expect more direct and consistent activation of these codes with auditory presentation. Using auditorily presented letters, Cowan et al. (2011) found a region in the left intra-parietal sulcus (IPS) that showed a load effect during the maintenance period of a verbal STM task. However, the same parietal region showed a load effect in a visual-spatial STM task (e.g., in maintaining 4 colored squares vs. 2 colored squares). Thus, Cowan et al. proposed that the IPS stores modality-independent abstract information or perhaps plays a role in allocating attention to modality-specific information.

These fMRI studies of phonological STM employed a univariate analysis approach (i.e., examining the overall hemodynamic response averaged/smoothed across a region). Recent studies on visual STM using a multivariate approach (e.g., Multivariate Pattern Analysis, MVPA) have complicated the interpretation of the presence or absence of delay period activity (LaRocque et al. 2014; Sreenivasan et al. 2014; D'Esposito and Postle 2015; Rose et al. 2016). MVPA takes advantage of the covariance of activation across multiple voxels and relates the activation pattern to different classes of stimuli (Norman et al. 2006; Haxby et al. 2014) and this method may be more sensitive for detecting neural activation. One study of visual STM, using a delayed recognition task for moving dot patterns (Riggall and Postle 2012), found that motion information (e.g., direction) could be decoded from a sensory region involved in encoding the motion even though no sustained delay period activity was apparent with a univariate analysis. In contrast, the region that showed sustained delay period activity (i.e., superior parietal cortex) failed to allow for the decoding of such information. Thus, this evidence suggested that the sensory cortex that appeared silent during the delay did play a role in STM whereas the parietal region that showed sustained delay period activity, but no successful decoding, was performing some other process, such as directing attention to the appropriate sensory information. Similar results have been reported in other studies using the MVPA approach (Harrison and Tong 2009; Serences et al. 2009; Emrich et al. 2013). These findings favor the embedded processes model for visual STM.

In contrast, however, other MVPA fMRI studies have instead found that it was possible to decode different types or features of visual stimuli during the delay period from non-sensory cortical regions argued to be involved in short-term maintenance, such as the parietal lobe (Christophe et al. 2012, 2015), the prefrontal lobe (Lee et al. 2013) or both (Ester et al. 2015). Additionally, a recent study found that while visual information could be decoded in both visual-sensory cortex and parietal cortex (e.g., superior IPS) during a delay period when no distracting stimuli were presented during the delay, such decoding was only possible in the parietal lobe and not sensory cortex when distraction was present (Bettencourt and Xu 2016). These authors suggested that since sensory regions need to be available to process other incoming stimuli, storage in the parietal lobe is needed to maintain a representation during distraction. This finding suggests a central role for the parietal cortex in retaining visual STM representations, consistent with a previous univariate fMRI study (Xu and Chun 2006). Thus, some MVPA findings are more consistent with a claim that regions downstream from sensory areas are used to store information during distraction.

Although most MVPA studies of short-term maintenance have focused on the retention of visual non-verbal representations, a few have examined the retention of auditory or phonological information. Lewis-Peacock et al. (2012) examined the brain regions supporting the retention of phonological, semantic and non-verbal visual information, using visual stimuli as input (nonwords, words, and line patterns, respectively) in a short-term recognition task. However, the region displaying the most important role in decoding the nonwords against the other stimuli during the delay period was in the occipital cortex, making it unclear whether subjects were maintaining phonological information or, rather, visual or orthographic information. Linke et al. investigated non-verbal auditory STM (e.g., tones or environmental sounds) (Linke et al. 2011; Linke and Cusack 2015). Linke et al. (2011) found that in the encoding stage, auditory

sensory regions (e.g., Heschl's gyrus) could decode tone frequency information. However, during the maintenance stage, the same sensory region was suppressed for representing that information, which is inconsistent with the claim that sensory cortex actively maintains neural representations in STM, as had been found in the visual domain. Instead, this observation suggests that regions beyond sensory regions maintain the information, most likely in a representational format different from that in sensory regions (see consideration of dynamic recoding in the Discussion). It is unclear, however, if the same mechanism would be observed for auditory verbal materials. The different findings from Linke et al. versus those from studies using visual stimuli (see Lee and Baker 2016 for a review) highlight the importance of using auditory input for testing phonological STM.

While Lewis-Peacock et al. (2012) demonstrated successful decoding of different types of information during the delay period, they also provided evidence suggesting that active maintenance was not necessary for STM performance. That is, they found that when attention was directed away from the stimuli to-be-remembered during the delay, MVPA decoding fell to chance levels, only to rise back to significance when attention was redirected to the stimulus. These results, as well as others (Watanabe and Funahashi 2014; Rose et al. 2016), suggest that some activity-silent mechanism, such as the modification of synaptic weights (Mongillo et al. 2008), supports STM. The interpretation of these and related findings is complex and under active debate, however (Barbosa 2017; Xu 2017). The issue of retention without neural activity will be returned to in the discussion.

Thus, despite considerable neuropsychological and neuroimaging work directed at the issue, the neural substrate for phonological STM remains unclear. In this study, we used fMRI to test the embedded processes versus storage buffer account by investigating the brain regions involved in the processing of speech and non-speech stimuli during either perception, STM or both. In contrast to most prior studies, auditory rather than visual verbal stimuli were used to ensure that phonological coding was involved, and both speech perception and STM were examined in the same subjects. We wished to determine if the regions identified in perception showed persistent activation or MVPA decoding evidence during maintenance, or whether additional regions are recruited during phonological STM. Finally, we explored the functional connectivity between the identified perceptual regions and other activated brain regions, and assessed its sensitivity to memory load to try to distinguish the role of regions found beyond perceptual regions. Specifically, we predicted that according to a buffer approach, connectivity between perceptual and potential buffer regions would be enhanced during encoding as items were encoded, but would not persist throughout the delay once the information had been consolidated (Xu 2017). On the other hand, if a non-sensory region carries out a sustained attentional process in order to keep attention focused on the to-be-maintained perceptual representations, then connectivity should be enhanced throughout the delay (see details in Methods).

Materials and Methods

Participants

Fifteen undergraduate students (19–26 year olds, mean: 20.4 year olds, 6 females) from Rice University participated in the imaging experiment for credit toward course requirements. All subjects were native English speakers, right handed, and reported no hearing, neurological, or psychiatric disorder.

Informed consent was obtained according to procedures approved by the Rice University Institutional Review Board.

Stimuli

Both speech and non-speech auditory stimuli were used. For the speech materials, nonwords (e.g., “bift”, “dast”) were used to minimize the influence of semantics on activation patterns. The nonwords were produced by a female native English speaker and recorded at a sampling rate of 44.1 kHz by using Audacity software (<http://www.audacityteam.org>). The average duration of the nonwords was about 450 ms. The non-speech materials were 2-note chords created by using MuseScore software (<http://musescore.org>) to play 8 different instruments (piano, violin, harp, banjo, timpani, trumpet, saxophone, and flute) at 4 different root keys (F below middle C, middle C, E above middle C, and G above middle C) with 4 types of musical intervals (third, fourth, fifth, and sixth) above each of the root keys. Thus, there were 16 different chords played by each of the instruments. Pilot testing indicated that it was necessary to have the non-speech materials differ on 3 dimensions in order to match performance on the STM task for the speech and non-speech stimuli (see Supplementary Material). The average duration of the chords was 450 ms.

Tasks

In the perception task (Fig. 1A), each trial began with a yellow fixation cross for 500 ms. Two sounds (either speech or non-speech) separated by an SOA of 667 ms were played to the subjects within 2 s, followed by a question mark. Subjects judged immediately if these 2 sounds were identical by pressing buttons with their left or right thumb. In the STM task (Fig. 1B, C), a blue fixation cross was shown for 500 ms, and subjects heard a list of either 1 or 3 speech or non-speech stimuli within 2 s, followed by a 10-s silent interval. For the 1-item lists, 2 white noise stimuli were also included to match the sound input for the 1 versus 3-item lists. Subjects were instructed to remember the speech or non-speech stimuli during the delay period. Then a probe was played and a 3-s question mark was shown. Subjects pressed the buttons to indicate if the probe was in the preceding list or not. In the STM task, half of the trials were in the high-load memory condition and half in the low-load.

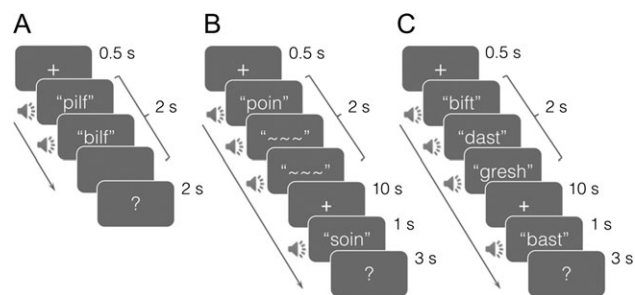


Figure 1. Examples of trials: (A) A speech non-matching trial in the perception task; (B) A non-matching trial for speech low memory load condition in the STM task (the wave line represents the white noise); (C) A non-matching trial for speech high memory load condition in the STM task. For the speech stimuli, in the high memory load STM trial, the non-matching probe (e.g., “bast” in Fig. 1C) overlapped with the onset (i.e., the initial consonant or consonants, “b” in Fig. 1C) of 1 list item and the rhyme portion (i.e., vowel plus final consonant(s), “ast” in Fig. 1C) of another list item. In the perception trial and the low memory load STM trial, the non-matching probe overlapped with either the onset or the rhyme portion (e.g., “ift” in Fig. 1A and “oin” in Fig. 1B) of the list item.

For both the perception task and the STM task, half of the trials were matching and half non-matching. The stimuli for both the speech and non-speech tasks were constructed to ensure that subjects had to maintain precise perceptual information for each to perform the task. For the speech stimuli, the non-matching speech probes contained some phonemes overlapping with those in the list items (see Fig. 1 caption for details). For the non-matching trials for the non-speech stimuli, the probe differed in either key or interval from all of the list items but might overlap on the other dimension and on the instrument. The probe-related list items in the non-matching trial and the probe-matching item in the matching trial were equally distributed across the 3 serial positions in the list. Thus, with this design, participants needed to maintain the identities of individual items in the list to correctly accept the matching probe or reject the non-matching probe. The difficulty of speech and non-speech conditions was matched in terms of accuracy in both the perception task and the STM task (see Supplementary Material). Speech and non-speech stimuli were randomly assigned to the perception, low-load and high-load conditions in creating the lists, and stimuli were counter-balanced across conditions.

fMRI Procedure and Data Acquisition

In the fMRI experiment, trials were grouped into speech or non-speech blocks with a 1-s cue word (“nonwords” or “chords”) at the beginning of each block indicating stimulus type. Within a block, 2 perception trials, 2 low-load STM trials, and 2 high-load STM trials were randomly presented with an average 5.5-s inter-trial interval of fixation. There were 2 speech and 2 non-speech blocks in a run, resulting in 4 trials of each of the 3 conditions for both the speech and non-speech conditions in a run. The order of blocks in a run was randomized, with a 2.5-s inter-block interval of fixation. There was a 10-s period of fixation at the beginning and the end of each run respectively. Thus, each functional run lasted 7 min 48 s. Every subject underwent 6 functional runs and thus completed 24 trials for each condition for the speech and non-speech stimuli.

The auditory stimuli were presented on E-prime 2.0 software (Psychology Software Tools) and the sounds were played to the subjects binaurally via MRI-compatible earphones. The foam canal tips were used with earphones for maximum noise reduction. Before the functional runs, a short auditory test run (about 40 s) including 2 sample trials was administered to make sure that subjects could clearly hear the stimuli against the noise of EPI scanning and the sound volume was adjusted to a comfortable level for each subject. The visual fixation crosses and cues were projected onto a screen and presented to the subjects via a mirror.

Functional MRI scans were conducted at the Core for Advanced Magnetic Resonance Imaging (CAMRI) at Baylor College of Medicine. Images were obtained on a 3 T Siemens Magnetom Tim Trio scanner equipped with a 12-channel head coil. Foam pads were used to keep subjects’ heads stabilized during the scanning. The functional run was obtained with Echo Planar Imaging sequences as follows: TR = 2.0 s, TE = 30 ms, FA = 90°, matrix size = 64 × 64, FoV = 220 mm, voxel size = 3.4 × 3.4 mm². Each run had 234 volumes and for each volume 34 axial 4-mm thickness slices were acquired to cover the whole brain. A high-resolution 3-dimension anatomical image was also acquired for each subject with MPRAGE sequence in axial plane (TR = 1200 ms, TE = 2.66 ms, FA = 12°, matrix size = 256 × 256, voxel size = 1 × 1 × 1 mm³).

Data Analysis

Preprocessing

Image preprocessing and univariate analysis were conducted using the AFNI software (Cox 1996). For each run, the first 3 EPI volumes were discarded to minimize effects that distort magnetic equilibrium. The image data were corrected for slice timing and head motion. Then, each participant’s EPI data were aligned to that individual’s anatomical image. A 6-mm full-width half-maximum Gaussian kernel was applied to spatially smooth the EPI data (except for multivariate analyses). Linear and quadratic trends were removed for each voxel time series and voxel-based signal scaling was also calculated for each run. The functional image data were resampled at 3 × 3 × 3 mm³.

Univariate Analyses

Preprocessed EPI data were then analyzed using a general linear model. Experimental regressors (perception, low-load, and high load for either speech or non-speech), as well as 6 head motion correction parameter regressors and 2 finger response regressors as covariates, were included in the model. For perception trials, the regressors were modeled by convolving the onset and duration (2 s) of nonwords and chords with a single parameter model of hemodynamic response functions. For memory trials, the regressors were modeled by using a single parameter hemodynamic response function to estimate the activation coefficients for encoding (2 s), maintenance (10 s) and retrieval (1 s) stages separately. We used a multiple parameters shape-free hemodynamic response model (i.e., “TENT” in AFNI) to estimate the amplitude of the signal change at each time point covering the whole memory trial from the onset to 24 s later. For every subject, the correct trials and incorrect trials were modeled separately, and only correct trials were included in further analysis. Each subject’s coefficients maps and signal change time courses for all conditions were warped and registered to a template (TT_N27) in Talairach standard space (Talairach and Tournoux 1988) for group analysis.

For voxel-wise group analysis, we compared speech versus non-speech in the perception task by using a paired t-test to find the region or regions specific to speech perception, and used this region as the ROI to test the memory data for effects of speech versus non-speech, memory load, and their interaction, using the signal change time courses from the STM task. For ROI analysis, the amplitudes of responses were averaged across time points 10 s, 12 s, and 14 s after stimulus onset as the signal change for the maintenance stage and the amplitude of response at the time point 6 s after stimuli onset was taken as the signal change for the encoding stage, and the signal changes for each memory condition were averaged within each ROI. If neither main effect nor interaction was significant, the activation for each condition (e.g., speech high load) relative to baseline was tested. Beyond analyses in the perception regions, we also carried out a whole brain voxel-wise repeated measures ANOVA during the maintenance stage in the STM task to test the memory load effect, to determine if there were any regions involved in the retention of phonological information beyond those involved in perception. The time course of signal changes were also examined in each of these clusters. For the whole brain analysis threshold, the activated areas were determined based on a voxel-wise $P < 0.001$ and then corrected at $\alpha < 0.05$ with family-wise approach based on Monte Carlo simulation in AFNI (3dClustSim) to determine the cluster size threshold. (Due to a recent issue raised by Eklund et al.’s (2016) paper, we chose a conservative voxel-level $P < 0.001$, and used updated

3dClustSim program (AFNI version:17.0.09) to determine the cluster size threshold. In addition, we adopted a new simulation approach of 3dClustSim, which has been shown to effectively control the false positive rate under 5% (Cox et al. 2017). Specifically, the program randomized the signs of the datasets by using the residuals of group testing (e.g., t-test) to simulate 10000 null hypothesis statistics. Then 3dClustSim is run with those null 3D results to generate the cluster threshold.)

Multivariate Analyses

To analyze the data with MVPA, images were preprocessed with the same procedures as in the univariate approach, except that no spatial smoothing was done to preserve spatial variances. Also, to reduce spatial interpolation, masks for regions of interest which were defined in the univariate group analysis (e.g., the left STG located by the perception task) were obtained for each subject by using an inverse transformation to warp back from the Talairach standard space to individual native space. Thus, all MVPA analyses were conducted in each subject's native space.

Classification analyses were conducted with Princeton MVPA toolbox (<https://github.com/princetonuniversity/princeton-mvpa-toolbox>) and custom scripts in MATLAB (MathWorks). Before carrying out the MVPA analyses, the preprocessed time series were z-scored separately within each run for each voxel. In the MVPA analyses, we wished to determine if patterns trained on the speech-nonspeech distinction would successfully discriminate the stimulus types during the delay period. As shown in Riggall and Postle's (2012) paper, it may be possible to decode the stimuli over the delay period in perceptual regions, even if no sustained delay period activity above baseline is observed. Thus, we tested if speech versus non-speech could be decoded in the speech perception region (i.e., left STG), during the delay period. A second region of primary interest was the left SMG which was defined based on the delay period activation (e.g., load effects) in the current study as this region has been argued to support the phonological STM buffer (Martin 2005). In line with the buffer approach, it should be possible to decode speech versus non-speech in this region, whether or not the regions show sustained neural activity during the delay period from the univariate approach. Other regions showing delay period sustained activity were tested with the MVPA approach as well.

We used the perception task data to train a classifier (perception classifier) to discriminate speech versus non-speech by using L2-regularized logistic regression algorithm as Riggall et al. used in their paper and setting the penalty value at 25 which has been shown to maximize the overall decoding performance (Riggall and Postle 2012). The perception trials were measured based on the signal at the third TR (i.e., 6 s) from the onset of the presentation of stimuli, to take into account the hemodynamic response lag. Then the trained classifier was applied to the STM task data to decode speech versus non-speech for the low and high memory load conditions separately. The decoding procedure was repeatedly conducted on each TR across a STM trial (e.g., from the onset to the following 22 s). A TR trial was considered correctly classified if the decoding evidence for the correct condition was higher than for the other condition (i.e., winner-take-all classification). Classification accuracy was scored by counting the proportion of correctly classified trials. Significance was assessed by comparing classification accuracy against a chance level (50%) by using a 1 tailed t-test. Using this classification procedure, the training and decoding data were independent.

One potential problem with the procedure described above was that the perception classifier might be not sensitive enough to decode the STM data beyond the left superior temporal region, if there was no clear evidence of involvement of other regions in perception. To address this issue, we also trained 2 other classifiers: one based on the encoding stage (TR = 3) (encoding classifier), and another based on the maintenance stage (TR = 6) (maintenance classifier) of the STM trials, and applied the trained classifiers to decode the STM trials across all TRs by using a leave-one-trial-out cross-validation procedure. This temporal-generalization analysis procedure was similar to one used in Riggall and Postle's (2012) study and is well-suited for decoding dynamic mental representations from neural activity (King and Dehaene 2014; Grootswagers et al. 2017). In this analysis, the classifier was trained on the data from all but 1 trial and then tested on the left-out trial, iteratively until all trials had been left out and tested once. The classification accuracy and significance testing were obtained as above.

Functional Connectivity Analyses

Finally, we conducted a context-dependent correlation analysis (i.e., generalized psychophysiological interaction, generalized-PPI) (McLaren et al. 2012; Cisler et al. 2014) to explore the effect of memory load on the connectivity between the temporal perceptual region and other regions of the brain (e.g., parieto-frontal maintenance regions). A positive (or negative) PPI effect could be interpreted as the psychological contrast of interest (e.g., high load vs. low load, or the increase in memory load) tends to increase (or decrease) the effect of the seed region on the target region (i.e., functional connectivity between the seed region and other regions). We chose a seed region in the left STG, which was defined by the contrast of speech versus non-speech in the perception task and was proposed to be involved in speech perception (see results section below). Both the buffer and embedded processes accounts predict that additional regions are involved in STM, with the buffer account implying that a region will play a role in storage, but with the embedded processes account claiming a role for attentional processes; thus, a positive PPI would be expected by both accounts. However, according to the buffer account, as more items need to be maintained in the buffer, a stronger connection between the perceptual region and the buffer area would be expected, but potentially only during the encoding period. Once the STM representation has been transferred and represented in the buffer, there is evidence that the connection to the perceptual region is no longer necessary during the delay (Xu 2017). On the other hand, the embedded processes account might predict a positive PPI effect that persists during the delay because more attention needs to be allocated to the perceptual region to keep the STM representations active over the delay, especially for the high-load condition. Thus, we tested the speech and non-speech load effects (i.e., high load vs. low load) during the encoding and the delay periods, respectively.

To implement this seed-based analysis, 2 types of additional regressors were added into the general linear model for individual subjects: the seed region time series (i.e., physiological signal) and its interaction with each psychological condition of interest (i.e., psychophysiological interaction). Specifically, these were twelve regressors of interaction terms for speech high load, speech low load, non-speech high load and non-speech low load during the encoding, maintenance, and retrieval periods respectively. Thus, all variances were well accounted for in the model. The regression coefficients for interaction terms were used as

dependent variables in the group level analysis. To correct for multiple comparisons, the voxel-wise p threshold was set at 0.005 and the cluster-wise α threshold was set at 0.05.

For display and reference, the group level results were mapped onto the FreeSurfer (Dale et al. 1999; Fischl et al. 1999) brain surfaces of the template (TT_N27) in Talairach space by using SUMA in AFNI (Saad and Reynolds 2012).

Results

Behavioral Results in the Scanner

During the fMRI scanning, participants' accuracy and RT were recorded. The RT analyses were conducted only on the data from correct trials.

Accuracy

In the perception task, accuracy was slightly higher on the non-speech condition (99.7%) than on the speech condition (97.5%), a difference which reached significance ($t_{14} = 2.48$, $P = 0.03$) (Fig. 2A). In the STM task, load effects were observed in both the speech condition and non-speech conditions, with better performance on low-load trials than on high-load trials. Performance on the speech and non-speech conditions did not differ from each other. These observations were confirmed by a 2×2 repeated measures ANOVA on accuracy with type of materials (speech, non-speech) and memory load (high, low) as within-subject factors. The main effect of memory load was significant [$F_{(1,14)} = 87.71$, $P < 0.001$]. The main effect of material type was not significant [$F_{(1,14)} = 0.05$, $P = 0.83$] nor was the interaction of material type and load [$F_{(1,14)} = 0.9$, $P = 0.36$].

Reaction Time

Response times were recorded from the onset of the second item in the perception task or the probe item in the STM task. As shown in Figure 2B, in the perception task, RT was longer for the speech condition (1029 ms) than for the non-speech condition (907 ms), a difference which reached significance ($t_{14} = 4.69$, $P < 0.001$) (There might be a concern that it was the longer RT that resulted in greater activation for speech than for non-speech in the perception task (see below). However, additional analyses using RT as a covariate showed that this was not the case. Controlling RT did not quantitatively change volumes or locations of the speech perception clusters. Even with the cluster derived from the analyses controlling for RT, the same pattern of results as those reported in the text were obtained (see Supplementary Material)). In the STM task, the average RT was longer for high-load conditions (speech: 1274 ms; non-speech: 1210 ms) than for low-load conditions (speech: 1029 ms; non-speech: 976 ms), and in each memory load level, RT was longer

for speech than for non-speech. These observations were confirmed by a 2×2 repeated measures ANOVA on RT. There were significant main effects of material type [$F_{(1,14)} = 4.86$, $P = 0.04$] and memory load [$F_{(1,14)} = 76.76$, $P < 0.001$]. The type by load interaction was not significant [$F_{(1,14)} = 0.06$, $P = 0.8$].

In summary, in contrast to the results of a pilot test outside the scanner (see Supplementary Material), there was evidence from both RTs and accuracy that the speech discrimination task was slightly harder than the non-speech discrimination task. It seems likely that the somewhat greater difficulty for speech than non-speech was due to the scanner noise interfering more in the speech conditions due to the smaller acoustic differences among the speech than the non-speech stimuli (e.g., with greater changes in pitch among chords presented on a given trial than for the nonsense syllables, which were all spoken by the same speaker).

fMRI Results

Defining Speech Perception Regions and Testing STM in these Regions

As shown in Figure 3A, 2 clusters showed more activation for the speech than for the non-speech condition, including the left superior temporal gyrus (STG) and the right superior temporal sulcus (STS) (voxel $P < 0.001$, corrected $\alpha < 0.05$, cluster size > 15 voxels). The pattern of activation was left lateralized, as the number of activated voxels in the left hemisphere was larger than that in the right hemisphere (Table 1). No region was activated more for the non-speech than the speech condition.

We tested phonological STM within the speech perception regions, as defined above. Based on claims about the neural substrates of STM derived from the embedded processes model (Jonides et al. 2005; Postle 2006), we would expect to observe sustained activation with the univariate analysis or MVPA decoding evidence with the multivariate analysis during the delay period in a typical speech perception region (e.g., the left STG). However, for the univariate analyses as shown in Figure 3B, during the maintenance stage of the STM task, there was no main effect of stimulus type or load ($P_s > 0.5$) nor any interaction ($P > 0.5$) in this ROI. We also failed to observe sustained activation relative to baseline for any condition ($P_s > 0.2$) during the maintenance stage in this ROI. During the encoding stage, this ROI showed a significant main effect of material type, with greater activation for speech than non-speech [$F_{(1,14)} = 50.49$, $P < 0.001$] and a significant main effect of memory load, with greater activation for the high than the low-load conditions [$F_{(1,14)} = 28.40$, $P < 0.001$]. The interaction was marginally significant [$F_{(1,14)} = 3.46$, $P = 0.08$]. The greater activation in the high-load condition can be attributed to the fact that a larger number of relevant stimuli were perceived in the high-load conditions. Similar patterns were observed in the right STS ROI (see Fig. 3C and Supplementary Material). Thus, using a univariate approach within the speech perception regions, which were defined by our perception localizer task, neither sustained activity relative to baseline nor a load effect was observed, particularly in the left superior temporal cortex that was presumed to be a typical speech perception region.

With the MVPA approach, we determined whether patterns of activation in the left superior temporal cortex could be distinguished for speech versus non-speech throughout the encoding, delay, and retrieval periods. Successful decoding would be expected under any theoretical account for the encoding and retrieval periods since speech perception occurs during those periods for the memory stimuli and the probe, respectively. In line with expectations, during the encoding stage (i.e., $TR = 3$),

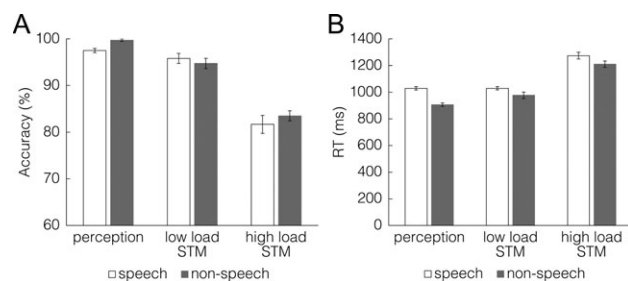


Figure 2. In-scanner behavioral results: (A) Accuracy for the perception task and the STM task; (B) Response times for the perception task and the STM task. Error bars represent the standard error of the mean.

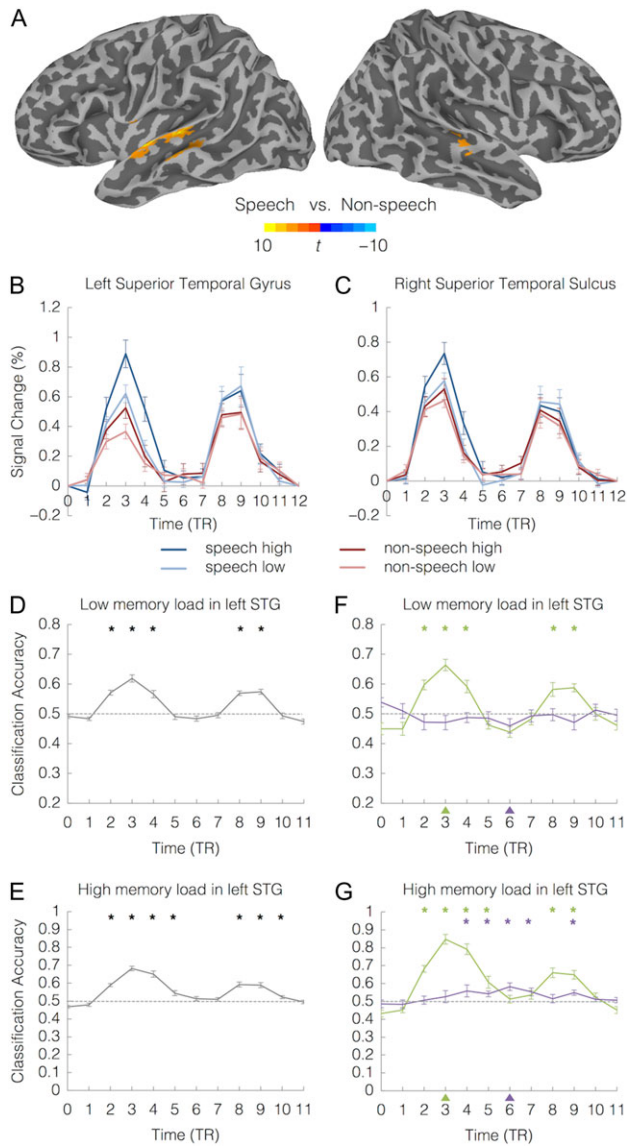


Figure 3. (A) The activation map for speech vs. non-speech in the perception task. (B, C) Estimated signal changes from univariate analyses for the STM task in (B) the left superior temporal cluster and (C) the right superior temporal cluster. (D–G) Classification accuracy from MVPA analyses for the STM task in the left superior temporal gyrus. In (D) and (E), the classifiers were trained on the perception data (perception classifier) and used to decode the STM data. In (F) and (G), the green and purple triangles indicate the TR of the STM task on which the classifier was trained, and the green and purple lines show the classification accuracy time series for each classifier when tested on each TR of the STM trial. Error bars represent the standard error of the mean. Dashed lines indicate the chance level. Asterisks indicate the significance of $P < 0.05$. STG: superior temporal gyrus.

classification accuracy was significantly higher than chance, both with the perception classifier [low load: accuracy = 0.62, $t_{(14)} = 9.7$, $P < 0.001$; high load: accuracy = 0.68, $t_{(14)} = 14.4$, $P < 0.001$] and the STM-encoding classifier [low load: accuracy = 0.66, $t_{(14)} = 8.6$, $P < 0.001$; high load: accuracy = 0.85, $t_{(14)} = 12.89$, $P < 0.001$] in this STG region. Those 2 classifiers could also significantly decode stimulus type relative to chance ($P_s < 0.003$) during the retrieval stage (i.e., TR = 9). However, during the delay period, as shown in Figure 3D–G, classification accuracy for speech versus non-speech was not significantly greater than chance ($P_s > 0.1$), for either type of classifier (Fig. 3D, E: perception classifier; Fig. 3F, G: STM-encoding classifier, green triangles and lines) for either memory load condition (Fig. 3D, F: low-load condition; Fig. 3E, G: high-load condition). When we trained a classifier from the STM maintenance period, there was no significant above chance level of decoding accuracy during either the delay period or encoding period in the low-load condition (accuracy < 0.5). However, in the high-load condition, this maintenance classifier significantly decoded stimulus type during the delay period [accuracy = 0.58, $t_{(14)} = 3.87$, $P = 0.001$] but not during the encoding period [accuracy = 0.53, $t_{(14)} = 0.79$, $P = 0.22$]. Thus, within this speech perception region, no significant multivariate evidence was detected during the delay period, except for the high memory load condition when using a classifier trained during the delay period (Fig. 3G, purple triangle and line).

Exploring STM Beyond the Speech Perception Regions

In the STM task, a voxel-wise repeated measures ANOVA revealed a set of regions showing a significant main effect of memory load (voxel $P < 0.001$, corrected $\alpha < 0.05$, cluster size > 15 voxels) beyond the speech perception regions discussed above (Fig. 4). These regions included a cluster in the left inferior parietal gyrus (i.e., supramarginal gyrus), a large cluster covering the posterior part of the left inferior frontal gyrus and left precentral gyrus (To reveal the potentially different functions of these 2 anatomically distinct regions in working memory (Price 2012; Nee et al. 2013), anatomical masks (from the anatomy atlas in AFNI) were used to intersect with this cluster. Thus, this large frontal cluster was divided into a left precentral cluster and a left posterior inferior frontal cluster for following analyses.), and a cluster in the right precentral gyrus. Moreover, some clusters in the left putamen, left supplementary motor area and bilateral cerebellum also showed this main effect of memory load (see Table 2 for details). All of these regions showed greater activity in the high than the low memory load conditions. None of these regions showed a significant main effect of material type nor any significant interaction between memory load and material type during the delay period.

We tested the time course of signal change with the univariate approach in the left supramarginal gyrus cluster, the left inferior frontal and precentral gyri cluster, and the right

Table 1 Activated regions for the contrast of speech vs. non-speech in the perception task

Regions	Coordinates (Talairach)			Peak t value	Cluster size (voxel)
	x	y	z		
Speech vs. non-speech					
Left superior temporal gyrus	-58.5	-25.5	5.5	8.27	127
Right superior temporal sulcus	52.5	-22.5	-0.5	5.44	16

precentral gyrus cluster. In the left SMG cluster (Fig. 5A), during the maintenance stage, there was significantly greater activation for the high load than the low-load condition [$F_{(1,14)} = 49.56$, $P < 0.001$]. While there was somewhat greater activation for non-speech than speech stimuli, the difference failed to reach significance [$F_{(1,14)} = 2.99$, $P = 0.11$]. The interaction was not significant [$F_{(1,14)} = 1.19$, $P = 0.29$]. A similar pattern was observed during the encoding stage, with significantly greater activation for the high than the low-load conditions [$F_{(1,14)} = 11.7$, $P = 0.004$] and a trend toward greater activation for non-speech than speech conditions [$F_{(1,14)} = 3.05$, $P = 0.10$]. The interaction was not significant [$F_{(1,14)} = 0.35$, $P = 0.56$]. Similar results as those in the SMG were observed in the 3 other clusters during the delay period (see Fig. 5B–D and Supplementary Material).

If phonological representations are maintained by a given region, successful decoding of speech versus non-speech by MVPA should be possible during the delay period. For the left SMG, as shown in Figure 5F, in the high memory load condition, decoding accuracy during the delay period (TR = 6) was above chance for both the encoding classifier [accuracy = 0.57, $t_{(14)} = 4.24$, $P < 0.001$] and the maintenance classifier [accuracy = 0.56, $t_{(14)} = 2.96$, $P = 0.005$]. Decoding accuracy during the encoding period (TR = 3) was also higher than the chance level for these 2 classifiers [encoding classifier: accuracy = 0.59, $t_{(14)} = 3.08$, $P = 0.004$; maintenance classifier: accuracy = 0.56, $t_{(14)} = 3.26$, $P = 0.003$]. In the low memory load condition, decoding accuracy was not above chance for either the encoding classifier or the maintenance classifier during either the encoding or delay period. In other regions showing delay period memory load effects with univariate analyses, the MVPA results basically resembled the ones in left SMG (see Fig. 5G–L and Supplementary Material).

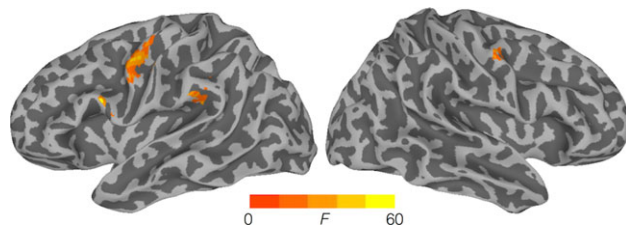


Figure 4. The activation map for the main effect of memory load during the delay period in the STM task. The supplementary motor area and cerebellum clusters (see Table 2 for details) are not shown in this lateral view figure.

Generalized Psychophysiological Interaction Analysis

Whole Brain Analyses with the Speech Perception Region as the Seed.

Encoding period

When the left superior temporal cortex that was defined by the contrast of speech versus non-speech in the perception task was chosen as a seed region, for the speech load effect (i.e., the contrast of speech high load vs. speech low load), we observed positive PPI effects in the left inferior parietal lobe (which partially overlapped with the SMG cluster found in the STM task in the univariate analyses), the right insular gyrus, and the right middle frontal gyrus, but no negative effects (voxel $P < 0.005$, corrected $\alpha < 0.05$, cluster size > 45) (Fig. 6A and Table 3). For the non-speech load effect (i.e., the contrast of non-speech high load vs. low load), we observed positive PPI effects in the right middle cingulate gyrus and the right insular gyrus during the encoding period (Fig. 6B and Table 3).

Maintenance period. For the speech load effect, we only observed a negative PPI effect in the right superior parietal lobule and precuneus. No significant PPI effect was observed during the maintenance period for the non-speech load effect (Fig. 6C and Table 3).

ROI-based Analyses

In addition to the seed-based whole brain approach, we also assessed PPI effects in specific regions, such as the left SMG that showed memory load effects during the delay period in the univariate activation analysis. This SMG cluster was generally anterior and inferior to the left parietal area that showed the positive PPI effect for the speech load effect during the encoding period (Fig. 6A), though partially overlaps with it. To validate the involvement of functional connectivity with the perceptual region, we chose this SMG cluster as a target region. We also chose the left PreCG and the left IFG clusters as other target regions to provide control sites. These 2 clusters showed delay period memory load effects as well. The regression coefficients of the psychophysiological interaction term for each condition (e.g., speech high load) were averaged across voxels within each target region, and were tested for the memory load effect (i.e., high vs. low) at a group level. The SMG target region showed a significant positive PPI effect (high $>$ low) for speech [$t_{(14)} = 4.96$, $P < 0.001$] and a marginally significant positive PPI effect for non-speech [$t_{(14)} = 1.83$, $P = 0.09$] during the encoding period. A direct comparison showed that this positive PPI effect for speech was greater during the encoding period than the delay period [delay period for speech: $t_{(14)} = 0.24$, $P = 0.8$; encoding vs. delay: $t_{(14)} = 3.78$, $P = 0.002$] (Fig. 6D). However, the left PreCG and the left IFG showed

Table 2 Activated regions for the repeated measures ANOVA during the delay period in the STM task

Regions	Coordinates (Talairach)			Peak F value	Cluster size (voxel)
	x	y	z		
Memory load main effect					
Left inferior frontal gyrus (p. opercularis) and precentral gyrus	-58.5	7.5	23.5	51.2	146
Left supplementary motor area	-4.5	-4.5	59.5	56.78	51
Left putamen	-22.5	4.5	14.5	35.45	27
Left cerebellum	-22.5	-55.5	-18.5	29.48	28
Left supramarginal gyrus	-52.5	-34.5	23.5	25.63	22
Right cerebellum	28.5	-49.5	-18.5	58.86	101
Right precentral gyrus	52.5	-10.5	44.5	32.49	23
Material type main effect					
n.s. region					
Interaction					
n.s. region					

no significant PPI effect for either speech ($P_s > 0.1$) or non-speech ($P_s > 0.5$). Taken together, during the encoding stage, the left inferior parietal lobe, including the SMG cluster which showed a delay period memory load effect and stimulus decoding evidence, revealed a positive memory load effect on the functional connectivity with the left superior temporal cortex.

Discussion

In the current study, we tested the buffer versus embedded processes accounts of phonological STM. We used a perception task and a STM task together with the same set of auditory verbal materials and the same subjects to locate speech perception regions and to identify additional regions involved in short-term storage. In the following, we begin by summarizing the observations within and beyond the region for phonological

processing, as well as highlighting the differences between them. Then, we argue that converging results from different methods provide more support for a buffer than an embedded processes account of phonological STM. Finally, we discuss the implications of the current study for theories of the neural basis of working memory, and acknowledge some open questions for future study.

Perception Regions and their Response to STM Tasks

Consistent with previous literature on speech perception, our perception task revealed more univariate activation for speech than non-speech in the left superior temporal gyrus (Binder et al. 2000; Scott and Johnsrude 2003; Hickok and Poeppel 2004; Vigneau et al. 2006; Price 2010; Turkeltaub and Coslett 2010). Prior findings suggest that the greater activation for speech

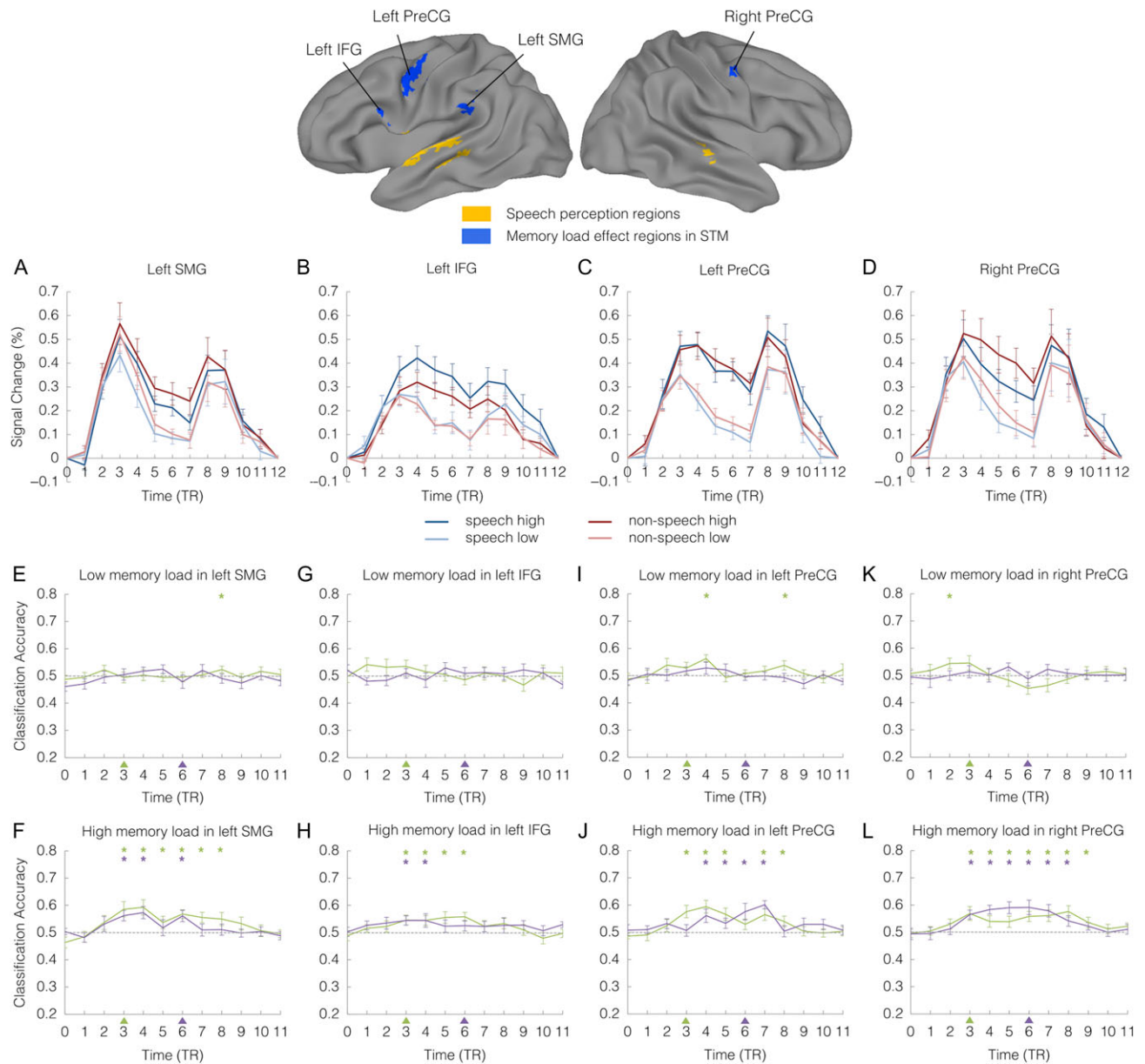


Figure 5. (A–D) Estimated signal changes from univariate analyses and (E–L) classification accuracy from MVPA analyses for the STM trials in clusters shown in color on the lateral view of brain map: (A, E, F) left SMG, (B, G, H) the posterior of left IFG, (C, I, J) left PreCG, (D, K, L) right PreCG. In (E–L), green and purple triangles, lines, and asterisks have the same indications as in Fig. 3. Error bars represent the standard error of the mean. SMG, supramarginal gyrus; IFG, inferior frontal gyrus; PreCG, precentral gyrus.

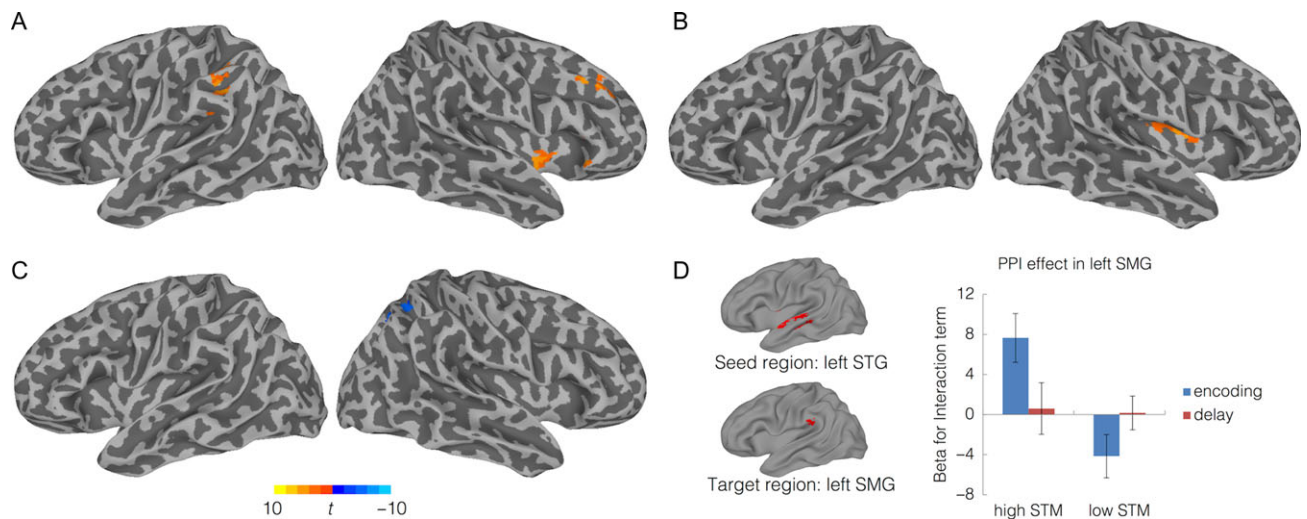


Figure 6. Activated regions for generalized-PPI analyses under (A) the speech load effect during the encoding period, (B) the non-speech load effect during the encoding period and (C) the speech load effect during the maintenance period of the STM task with a seed region in the left superior temporal gyrus. (D) The generalized-PPI effects for speech in the left SMG during the encoding and delay periods. Error bars represent the standard error of the mean.

Table 3 Activated regions showing PPI effects (seed region: left STG) under the load effects within speech and non-speech condition in the STM task

Regions	Coordinates (Talairach)			Peak t value	Cluster size (voxel)
	x	y	z		
Speech load contrast					
During the encoding period					
Left inferior parietal lobe	-46.5	-37.5	32.5	4.64	65
Right insular gyrus	40.5	7.5	-3.5	5.57	52
Right middle frontal gyrus	37.5	31.5	38.5	5.05	49
During the delay period					
Right superior parietal lobule/precuneus	19.5	-61.5	41.5	-6.28	81
Non-speech load contrast					
During the encoding period					
Right middle cingulate cortex	1.5	-7.5	26.5	6.47	114
Right insular gyrus	37.5	-4.5	17.5	5.13	80
During the delay period					
n.s. region					

than non-speech (i.e., chords) in the left superior temporal gyrus reflects aspects of phonological processing, such as the identification of phonetic features, phonemes or syllables (Formisano et al. 2008; Obleser et al. 2010; Mesgarani et al. 2014; Arsenault and Buchsbaum 2015). Our MVPA analyses showed that speech versus non-speech information could be decoded from the left superior temporal region during the encoding and retrieval stages of the STM task. Thus, both univariate and multivariate evidence showed that the speech perception region defined in the present study was involved in mapping acoustic representation onto speech categories.

In contrast to the predictions from the embedded processes approach, no elevated or sustained activity was observed during the delay stage, let alone a load effect. No multivariate evidence was detected during the delay stage in this region for classifiers trained on the perception or STM-encoding stage data. Significant decoding during the delay period was observed in the high-load condition when using the maintenance classifier.

Differing from our observations, some previous studies have reported sustained activity in the left superior temporal gyrus/sulcus during the delay phase of a verbal STM task (Postle et al. 1999; Buchsbaum et al. 2001, 2005; Ravizza et al. 2011). For

example, Ravizza et al. (2011) observed elevated activity during the maintenance stage for visually presented English letters in the left posterior superior temporal ROI chosen to encompass peak activations from previous findings in the literature. However, this ROI was quite posterior and close to the inferior parietal region in our study that showed load-sensitive activity, but was outside of our speech perception region (see Supplementary Material for details) and the region identified in the meta-analysis of speech perception by Turkeltaub and Coslett (2010). The differential location from Ravizza et al. (2011) may be due to the inclusion of studies using visual rather than auditory presentation, which obfuscates the localization of phonological processing. Clearer conclusions can be drawn by using auditory verbal input and the use of a speech localizer to isolate speech perception regions, as was done in the current study.

The results from the MVPA analyses in the perceptual region show a different pattern from those presented by Riggall and Postle (2012) who found that the direction information from moving dots could be decoded during the delay from the sensory areas (specifically, area MT) even if no sustained activity was observed during the same period. The differing patterns might be due to the different nature of visual versus auditory representations. Prior

evidence suggests that individuals use mental imagery to retain visual information over a delay, and this mental imagery serves to activate sensory regions (see Lee and Baker 2016 for an overview). Given the concrete and distinct representation of visual features in different early visual regions (King and Nelken 2009), this use of mental imagery may have resulted in decodable representations of motion in this sensory area. Primary auditory areas do not have the same organization by sensory feature (King and Nelken 2009) and, moreover, it seems unlikely that individuals use auditory imagery for the complex acoustic patterns making up speech sounds in order to maintain information over a delay, instead relying on articulatory rehearsal (Salamé and Baddeley 1982; Vallar and Baddeley 1984).

Significant decoding was observed in the left STG region during the delay period with a maintenance classifier, but not by an encoding classifier. This discrepancy goes against claims from some versions of the embedded process approach claiming that maintenance depends upon the persisting activation of representations established during perception or encoding (Jonides et al. 2005; Postle 2006). Instead, these findings suggest that the activation pattern in this region changed from the encoding to the delay period. Linke et al.'s study on maintenance of tones also found a change from encoding to delay, where encoding was suppressed during the delay (Linke et al. 2011). The authors claimed that this suppression was due to the need to protect STM representations from interference from newly encoded task-irrelevant sounds. It is possible that a similar suppression mechanism is applied to auditory verbal materials, resulting in decoding success in differentiating suppressed versus non-suppressed representations. Alternatively, it is possible that feedback from downstream processing areas maintaining perhaps more abstract representations of the speech sounds (e.g., categorical speech representations; Turkeltaub and Coslett 2010) changed the nature of any representations in the superior temporal gyrus (see Mendoza-Halliday et al. 2014 for related findings in the visual domain).

Phonological STM Beyond Speech Perception Regions

In contrast to the results in temporal cortex, we found a set of left hemisphere regions along the dorsal stream, including a left inferior parietal region, showing elevated and sustained activity, as well as memory load effects (i.e., high load vs. low load) during the delay period of our STM task. The location in the inferior parietal region was very close to the coordinates reported in earlier studies showing elevated and sustained activity and load effects in the parietal lobe, such as Paulesu et al. (1993) (Talairach: $x = -44$, $y = -32$, $z = 24$) (see also Salmon et al. 1996; Jonides et al. 1998; Postle et al. 1999; Ravizza et al. 2004; Buchsbaum et al. 2005). This region is usually described as the supramarginal gyrus (SMG; Brodmann Area 40), which lies just superior to the posterior end of the Sylvian fissure. Some left frontal regions, including the precentral gyrus, the posterior part of the inferior frontal gyrus and the supplementary motor area were activated and showed a load effect during the delay period, which is consistent with previous verbal working memory studies reporting sustained activity in prefrontal regions (Rypma et al. 1999; Chein and Fiez 2001; Ravizza et al. 2011) and showing load-sensitive responses (Braver et al. 1997; Narayanan et al. 2005; Zarahn et al. 2005). In contrast to the inferior parietal lobe which has been assumed to have a storage function (Smith and Jonides 1998; Martin 2005), the inferior frontal gyrus, the premotor area and the supplementary motor area have been proposed to make up a subvocal rehearsal system (Smith et al. 1998; Chein and Fiez 2001), or to support executive processes (Smith and Jonides 1998), important for maintaining the contents

of verbal working memory. For example, according to Chein and Fiez (2001), the posterior inferior frontal gyrus (or the dorsal BA 44 in their paper) is involved in planning and organizing the initial rehearsal for novel stimuli, and the supplementary motor area is involved in implementing motor processes throughout the maintenance stage. Thus, the load effects observed in these frontal regions arguably reflect an increasing demand for these processes in the high than low-load conditions.

Turning to the multivariate analyses, we also found above chance multivariate decoding accuracy during the delay period in the left SMG, as well as in the bilateral precentral gyrus. Whereas the decoding evidence in the left STG was limited to a classifier trained during maintenance, the decoding evidence in the left SMG was observed continuously from encoding to maintenance periods for classifiers trained during either encoding or maintenance (Kalm and Norris 2014). These patterns suggested that the neural representations for phonological STM were represented in the SMG during the encoding period and were maintained over the delay period. The observed decoding evidence in bilateral precentral gyrus may reflect the different motor representations evoked by the articulatory strategies used for speech (e.g., rehearsing) and non-speech (e.g., humming). The concurrence of decoding evidence observed in the left SMG and bilateral precentral gyri are also consistent with findings from a recent study which found that visual STM representations were stored across a widely distributed fronto-parietal cortical network (Ester et al. 2015).

Finally, the psychophysiological interaction analyses showed that memory load had a significant positive effect on the functional connectivity between the left superior temporal cortex and the left inferior parietal cortex in the speech condition during the encoding period. As the number of the to-be-remembered non-words increased, the speech perception region connected more strongly with the left SMG, which was presumed to be a phonological STM buffer region. This observation is also consistent with a proposal that from the beginning of the encoding period, the phonological STM representations were being transferred to a non-sensory area. Previous studies have also found this memory load-sensitive increase in functional connectivity in STM tasks for different domains, such as the human face (Rissman et al. 2008) and visually presented verbal materials (Fiebach et al. 2006). However, both of these results were observed during the maintenance stage of their STM tasks, which contrasts with our result that was observed during the encoding stage. The fact that the PPI effect was observed only during the encoding but not during the delay period suggests that once the STM representations have been transferred and represented into the buffer area, the perceptual region was less essential for STM (Xu 2017). PPI effects have also been observed between the left superior temporal cortex and the right frontal and insular gyri observed in the current study. These effects might reflect executive control functions (Aron et al. 2014) because in the high memory load condition subjects need to allocate their attention among the multiple items as opposed to only 1 in the low-load condition.

Overall, converging results from the univariate activation analyses, a multivariate decoding approach, and functional connectivity analyses support the conclusion that the left SMG plays an important role in the short-term storage of phonological representations.

The Buffer Account versus the Embedded Processes Account of STM

The embedded processes account assumes that STM reflects the transient activation of the results of perceptual processing in a

given domain. Thus, regions involved in perceptual processing should show evidence for sustained activation or successful decoding of stimulus information during a delay period (see Wager et al. 2004; Lee and Baker 2016 for reviews; also see Xu 2017 for a discussion). Regions outside these perceptual regions may also show sustained activity, but this activity is attributed to attentional or executive processes (Cowan et al. 2011). Our failure to find sustained activation or evidence of decoding in speech perception regions argues against the embedded processes account of phonological STM. The fact that the SMG showed such sustained activation and evidence for speech decoding is consistent with other evidence suggesting that this region serves as a phonological buffer (Paulesu et al. 1993; Salmon et al. 1996; Kalm and Norris 2014). The PPI evidence is also consistent with this claim. Thus, we are arguing that a region downstream from perceptual regions serves the function of maintaining phonological information in STM.

One argument about the left SMG is that it may play an attentional role in working memory. It has been proposed that a set of fronto-parietal regions, including the intra-parietal sulcus (IPS) and the temporo-parietal junction (TPJ), serve as a crucial hub in an attention control network, with the IPS supporting goal-directed attention and the TPJ supporting stimulus-driven attention. Given the proximity of the TPJ to the SMG, one might question whether the SMG activation uncovered in the present study was related to processing in this attention control network. The proposals regarding the role of the IPS and TPJ in attention have been derived mainly from studies in visual non-verbal domains, with the parietal regions being right lateralized (Corbetta and Shulman 2002). For instance, Todd et al. (2005) showed that as memory load increased in a visual non-verbal STM task, activation was increasingly suppressed in the right TPJ whereas it increased in the IPS. Majerus et al. (2012) found a similar pattern in the left hemisphere for short-term recognition memory of visually presented letters; that is, they found that the left TPJ showed greater deactivation in response to increasing memory load during the delay period and argued that, as the memory load increased, the stimulus-driven attention region was depressed while the goal-directed attention region was enhanced, preventing interference from task-irrelevant stimuli. Given that we found greater activation with increasing memory load in the SMG whereas these studies found greater deactivation in the TPJ, it seems unlikely that we are tapping into the same attentional function in the SMG region. In addition, the left TPJ region reported in the Majerus et al.'s study was close to the angular gyrus (Brodmann Area 39, see Tables 1 and 2 in Majerus et al. 2012) and more posterior to the left SMG reported in the current study. Thus, there was little correspondence between their findings and ours.

Nonetheless, one might argue that the SMG is playing an attentional role different from the TPJ, more in line with sustaining goal-directed activity. Such a role would be consistent with sustained activity during the delay as was observed in the intra-parietal sulcus in Riggall and Postle's (2012) study. However, if the SMG was playing an attentional role during the delay, one would have to explain how this activation of attention failed to activate sensory regions in the STG such that decoding would be possible using the perception or encoding stage classifiers. Moreover, the PPI effect in the left SMG only for the speech condition during the encoding stage would seem difficult to accommodate with this attentional explanation. One would have to assume that different subregions exist for directing attention to speech and non-speech stimuli, which give rise to our ability to discriminate speech from non-speech in the SMG—that is, this region is active for directing attention

to speech, but does not represent speech codes per se. The representational format for different stimulus types in this region need to be tested in future studies designed to be analyzed with multivariate approaches such as representational similarity analysis (Kriegeskorte et al. 2008) or item-level decoding methods (LaRocque et al. 2017), in order to more strongly support the contention that speech-specific representations were stored in the SMG.

The reasoning regarding the roles of speech perception and non-perceptual regions as supporting the buffer model depends on the assumption that sustained neural activity during a delay, which may be revealed through univariate or more sensitive multivariate analyses, serves as a signature for STM (Goldman-Rakic 1995; LaRocque et al. 2014; Sreenivasan et al. 2014). This hypothesis has been challenged recently by findings suggesting that STM representations can be maintained in the absence of neural activity—for example, through short-term synaptic change (Mongillo et al. 2008; Stokes 2015; Myers et al. 2017). Both human and animal data support the suggestion of activity-silent maintenance (Lewis-Peacock et al. 2012; Watanabe and Funahashi 2014; Rose et al. 2016) showing that spiking activity or fMRI signals of activation may be absent during a delay when attention is directed away from the stimulus to be remembered, but such activity may be reinstated when attention is switched back. However, more recent neurophysiological studies (Bolkan et al. 2017; Kamiński et al. 2017) have shown that lower sustained neuronal activity is associated with poor STM performance, supporting the assumption that persistent activity contributes to keeping mnemonic information active during the delay (Bray 2017). Some argue that this evidence has complicated the interpretation of results regarding activity-silent retention (Schneegans and Bays 2017), and the sustained spiking versus delay-silent hypothesis of retention is still actively debated (Barbosa 2017; Xu 2017). It should be noted that the current study cannot rule out the possibility that phonological information may be maintained in the STG via a delay-silent mechanism, with such latent representation not being measured by the hemodynamic recording of fMRI. Only a design with a re-direction of attention as in the Lewis-Peacock et al. and Rose et al. studies could address this issue. Recent work has shown that activity-silent representations in working memory can be reactivated with transcranial magnetic stimulation (Rose et al. 2016) or a non-specific perceptual impulse (Wolff et al. 2017). If the phonological representations are maintained in the STG via a delay-silent mechanism, future work should be directed at establishing how this mechanism gives rise to dynamic changes of neural representations in the perceptual area, and how these changes contribute to behavioral performance (See Supplementary Material for preliminary evidence regarding the behavioral importance of MVPA decoding evidence in the perceptual and non-perceptual regions. In these preliminary analyses, we correlated individuals' MVPA decoding evidence with their behavioral performance on STM. However, it should be noted that the small sample size may increase the false positive rate and give rise to the inflated effect size in the individual difference correlation analyses (Yarkoni 2009; Button et al. 2013), and thus any results should be treated with caution. Future studies with adequate sample size would be needed to replicate these results.)

Implications for Theories of the Neural Basis of Short-term Memory

We have argued that the current findings provide more support for a buffer than an embedded processes account of

phonological STM. Nonetheless, there are many unresolved issues with respect to the neural instantiation of both approaches and further theoretical and empirical work is needed to resolve these issues. One issue concerns the nature of the representations stored in the buffer, regarding whether they duplicate sensory information or represent a different format of stimulus information. These possibilities could be teased apart in future work by designing experiments that yield data that can be analyzed to differentiate representations within the speech and non-speech domains, for instance, by using representational similarity analysis (Kriegeskorte et al. 2008).

Moreover, at a theoretical level, it is unclear how persisting activity at any level can be used to support the retention of multiple items and their order. Even though our task involved recognition of a single probe item, in the speech condition, subjects had to maintain the order and grouping of phonemes in order to reject non-matching probes which contained the initial phonemes from one stimulus and the final phonemes from another. Computational cognitive models of phonological retention for a series of items often assume that it involves the convolution of item information (i.e., phonemic information) with a timing signal (Burgess and Hitch 1999, 2005; Botvinick and Plaut 2006). In neural terms, one might postulate that some region receives input from both content information (e.g., phonemes) and from a timing signal or other signal encoding order information. The maintenance of this convolution of information is critical for STM. At present, the suggestion that the SMG might play this role must remain quite speculative. While considerable evidence indicates that regions in the STG are connected to the SMG, it is unclear what the neural source of the timing or order code would be and whether the SMG receives the appropriate input from such a source. Recently, theta-band oscillations have been argued to reflect working memory capacity, and Roux and Uhlhaas (2014) have argued that theta activity occurs predominantly in tasks involving sequential encoding of multiple items in working memory. Lundqvist et al. (2016) have shown a critical role for the interplay between beta and gamma bursts in coding for dynamic and transient representations in working memory, making the provocative claim that sustained delay period activations are merely an artifact of averaging such bursts across trials (see also Sreenivasan et al. 2014; Siebenhühner et al. 2016). However, the links between the STG, SMG, IFG, and PreCG via patterns of functional connectivity and oscillatory dynamics in supporting phonological STM remain to be established. Clearly further work using neural signals with appropriate temporal and spatial resolution, multi-modal perturb-and-record methods, and dynamic causal modeling (Friston et al. in press) will be needed to flesh out this possibility.

Conclusion

In this study, we uncovered evidence more consistent with a buffer than an embedded processes approach to phonological STM. That is, we found no evidence of sustained activity and only limited evidence of MVPA decoding of stimulus representations in sensory cortex (STG), but did find such evidence in an inferior parietal region (SMG). Future work will be needed to uncover the nature of the representations maintained downstream from the sensory region in the buffer region as well as to uncover how multiple items and their order can be encoded in buffer regions.

Supplementary Material

Supplementary material is available at *Cerebral Cortex* online.

Funding

This work was supported by the T.L.L. Temple Foundation Neuroplasticity Laboratory award to Rice University, a Rice University Social Sciences Research Institute (SSRI) grant to R.C.M. and A.C.H., a Rice University SSRI Pre-dissertation Research grant and a Dissertation Fellowship from the Dingwall Foundation to Q.Y.

Notes

We thank Tristanne Mock for her help with creation of the stimuli and the Baylor College of Medicine Core for Advanced MRI (CAMRI) for assistance with data collection. We also thank Michael Beauchamp and Morgan Widhalm for their comments on an earlier version of this manuscript. *Conflict of Interest*: None declared.

References

- Allport DA. 1984. Auditory-verbal short-term memory and conduction aphasia. In: Bouma H., Bouwhuis D.G., editors. *Attention and performance X: control of language processes*. Hove, UK: Lawrence Erlbaum. p. 313–326.
- Aron AR, Robbins TW, Poldrack RA. 2014. Inhibition and the right inferior frontal cortex: one decade on. *Trends Cogn Sci*. 18:177–185.
- Arsenault JS, Buchsbaum BR. 2015. Distributed neural representations of phonological features during speech perception. *J Neurosci*. 35:634–642.
- Baddeley A, Gathercole S, Papagno C. 1998. The phonological loop as a language learning device. *Psychol Rev*. 105:158–173.
- Baddeley A, Lewis V, Vallar G. 1984. Exploring the articulatory loop. *Q J Exp Psychol A*. 36:233–252.
- Baldo JV, Katseff S, Dronkers NF. 2012. Brain regions underlying repetition and auditory-verbal short-term memory deficits in aphasia: Evidence from voxel-based lesion symptom mapping. *Aphasiology*. 26:338–354.
- Barbosa J. 2017. Working memories are maintained in a stable code. *J Neurosci*. 37:8309–8311.
- Belleville S, Caza N, Peretz I. 2003. A neuropsychological argument for a processing view of memory. *J Mem Lang*. 48:686–703.
- Bettencourt KC, Xu Y. 2016. Decoding the content of visual short-term memory under distraction in occipital and parietal areas. *Nat Neurosci*. 19:150–157.
- Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET. 2000. Human temporal lobe activation by speech and nonspeech sounds. *Cereb Cortex*. 10: 512–528.
- Bolkan SS, Stujenske JM, Parnaudeau S, Spellman TJ, Rauffenbart C, Abbas AI, Harris AZ, Gordon JA, Kellendonk C. 2017. Thalamic projections sustain prefrontal activity during working memory maintenance. *Nat Neurosci*. 20:987–996.
- Botvinick MM, Plaut DC. 2006. Short-term memory for serial order: a recurrent neural network model. *Psychol Rev*. 113: 201–233.
- Braver TS, Cohen JD, Nystrom LE, Jonides J, Smith EE, Noll DC. 1997. A parametric study of prefrontal cortex involvement in human working memory. *NeuroImage*. 5:49–62.
- Bray N. 2017. Working memory: persistence is key. *Nat Rev Neurosci*. 18:385–385.
- Buchsbaum BR, Hickok G, Humphries C. 2001. Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cogn Sci*. 25:663–678.

- Buchsbaum BR, Olsen RK, Koch P, Berman KF. 2005. Human dorsal and ventral auditory streams subserved rehearsal-based and echoic processes during verbal working memory. *Neuron*. 48:687–697.
- Burgess N, Hitch GJ. 1999. Memory for serial order: a network model of the phonological loop and its timing. *Psychol Rev*. 106:551–581.
- Burgess N, Hitch G. 2005. Computational models of working memory: putting long-term memory into context. *Trends Cogn Sci*. 9:535–541.
- Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, Munafò MR. 2013. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci*. 14:365–376.
- Chein JM, Fiez JA. 2001. Dissociation of verbal working memory system components using a delayed serial recall task. *Cereb Cortex*. 11:1003–1014.
- Christophel TB, Cichy RM, Hebart MN, Haynes J-D. 2015. Parietal and early visual cortices encode working memory content across mental transformations. *NeuroImage*. 106:198–206.
- Christophel TB, Hebart MN, Haynes J-D. 2012. Decoding the contents of visual short-term memory from human visual and parietal cortex. *J Neurosci*. 32:12983–12989.
- Cisler JM, Bush K, Steele JS. 2014. A comparison of statistical methods for detecting context-modulated functional connectivity in fMRI. *NeuroImage*. 84:1042–1052.
- Cohen JD, Perlstein WM, Braver TS, Nystrom LE, Noll DC, Jonides J, Smith EE. 1997. Temporal dynamics of brain activation during a working memory task. *Nature*. 386:604–608.
- Corbetta M, Shulman GL. 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci*. 3:201–215.
- Cowan N. 2001. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav Brain Sci*. 24:87–114.
- Cowan N, Li D, Moffitt A, Becker TM, Martin EA, Sauls JS, Christ SE. 2011. A neural region of abstract working memory. *J Cogn Neurosci*. 23:2852–2863.
- Cox RW. 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res*. 29:162–173.
- Cox RW, Chen G, Glen DR, Reynolds RC, Taylor PA. 2017. FMRI clustering in AFNI: false-positive rates redux. *Brain Connect*. 7:152–171.
- Curtis CE, D’Esposito M. 2003. Persistent activity in the prefrontal cortex during working memory. *Trends Cogn Sci*. 7:415–423.
- Dale AM, Fischl B, Sereno MI. 1999. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage*. 9:179–194.
- D’Esposito M, Postle BR. 2015. The cognitive neuroscience of working memory. *Annu Rev Psychol*. 66:115–142.
- Eklund A, Nichols TE, Knutsson H. 2016. Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. *Proc Natl Acad Sci USA*. 113:7900–7905.
- Emrich SM, Riggall AC, LaRocque JJ, Postle BR. 2013. Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory. *J Neurosci*. 33:6516–6523.
- Ester EF, Sprague TC, Serences JT. 2015. Parietal and frontal cortex encode stimulus-specific mnemonic representations during visual working memory. *Neuron*. 87:893–905.
- Fiebach CJ, Rissman J, D’Esposito M. 2006. Modulation of inferotemporal cortex activation during verbal working memory maintenance. *Neuron*. 51:251–261.
- Fischl B, Sereno MI, Dale AM. 1999. Cortical surface-based analysis. II: inflation, flattening, and a surface-based coordinate system. *NeuroImage*. 9:195–207.
- Formisano E, De Martino F, Bonte M, Goebel R. 2008. “Who” is saying “what?” Brain-based decoding of human voice and speech. *Science*. 322:970–973.
- Friston KJ, Preller KH, Mathys C, Cagnan H, Heinzle J, Razi A, Zeidman P. in press. Dynamic causal modelling revisited. *NeuroImage*. doi: 10.1016/j.neuroimage.2017.02.045
- Gathercole SE, Baddeley AD. 1989. Evaluation of the role of phonological STM in the development of vocabulary in children: a longitudinal study. *J Mem Lang*. 28:200–213.
- Gathercole SE, Hitch GJ, Service E, Martin AJ. 1997. Phonological short-term memory and new word learning in children. *Dev Psychol*. 33:966–979.
- Goldman-Rakic PS. 1995. Cellular basis of working memory. *Neuron*. 14:477–485.
- Groetswagers T, Wardle SG, Carlson TA. 2017. Decoding dynamic brain patterns from evoked responses: a tutorial on multivariate pattern analysis applied to time series neuroimaging data. *J Cogn Neurosci*. 29:677–697.
- Gupta P. 2003. Examining the relationship between word learning, nonword repetition, and immediate serial recall in adults. *Q J Exp Psychol A*. 56:1213–1236.
- Gupta P, Lipinski J, Abbs B, Lin P-H. 2005. Serial position effects in nonword repetition. *J Mem Lang*. 53:141–162.
- Harrison SA, Tong F. 2009. Decoding reveals the contents of visual working memory in early visual areas. *Nature*. 458:632–635.
- Haxby JV, Connolly AC, Guntupalli JS. 2014. Decoding neural representational spaces using multivariate pattern analysis. *Annu Rev Neurosci*. 37:435–456.
- Hickok G, Poeppel D. 2004. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*. 92:67–99.
- Jonides J, Lacey SC, Nee DE. 2005. Processes of working memory in mind and brain. *Curr Dir Psychol Sci*. 14:2–5.
- Jonides J, Schumacher EH, Smith EE, Koeppel RA, Awh E, Reuter-Lorenz PA, Marshuetz C, Willis CR. 1998. The role of parietal cortex in verbal working memory. *J Neurosci*. 18:5026–5034.
- Jonides J, Schumacher EH, Smith EE, Lauber EJ, Awh E, Minoshima S, Koeppel RA. 1997. Verbal working memory load affects regional brain activation as measured by PET. *J Cogn Neurosci*. 9:462–475.
- Kalm K, Norris D. 2014. The representation of order information in auditory-verbal short-term memory. *J Neurosci*. 34:6879–6886.
- Kamiński J, Sullivan S, Chung JM, Ross IB, Mamelak AN, Rutishauser U. 2017. Persistently active neurons in human medial frontal and medial temporal lobe support working memory. *Nat Neurosci*. 20:590–601.
- King J-R, Dehaene S. 2014. Characterizing the dynamics of mental representations: the temporal generalization method. *Trends Cogn Sci*. 18:203–210.
- King AJ, Nelken I. 2009. Unraveling the principles of auditory cortical processing: can we learn from the visual system? *Nat Neurosci*. 12:698–701.
- Kriegeskorte N, Mur M, Bandettini P. 2008. Representational similarity analysis—connecting the branches of systems neuroscience. *Front Syst Neurosci*. 2:4.

- Langel J, Hakun J, Zhu DC, Ravizza SM. 2014. Functional specialization of the left ventral parietal cortex in working memory. *Front Hum Neurosci.* 8:440.
- LaRocque JJ, Lewis-Peacock JA, Postle BR. 2014. Multiple neural states of representation in short-term memory? It's a matter of attention. *Front Hum Neurosci.* 8:5.
- LaRocque JJ, Riggall AC, Emrich SM, Postle BR. 2017. Within-category decoding of information in different attentional states in short-term memory. *Cereb Cortex.* 27:4881–4890.
- Leavitt ML, Mendoza-Halliday D, Martinez-Trujillo JC. 2017. Sustained activity encoding working memories: not fully distributed. *Trends Neurosci.* 40:328–346.
- Lee S-H, Baker CI. 2016. Multi-voxel decoding and the topography of maintained information during visual working memory. *Front Syst Neurosci.* 10:2.
- Lee S-H, Kravitz DJ, Baker CI. 2013. Goal-dependent dissociation of visual and prefrontal cortices during working memory. *Nat Neurosci.* 16:997–999.
- Leff AP, Schofield TM, Crinion JT, Seghier ML, Grogan A, Green DW, Price CJ. 2009. The left superior temporal gyrus is a shared substrate for auditory short-term memory and speech comprehension: evidence from 210 patients with stroke. *Brain.* 132:3401–3410.
- Lewis-Peacock JA, Drysdale AT, Oberauer K, Postle BR. 2012. Neural evidence for a distinction between short-term memory and the focus of attention. *J Cogn Neurosci.* 24:61–79.
- Linke AC, Cusack R. 2015. Flexible information coding in human auditory cortex during perception, imagery, and STM of complex sounds. *J Cogn Neurosci.* 27:1322–1333.
- Linke AC, Vicente-Grabovetsky A, Cusack R. 2011. Stimulus-specific suppression preserves information in auditory short-term memory. *Proc Natl Acad Sci USA.* 108:12961–12966.
- Lundqvist M, Rose J, Herman P, Brincat SL, Buschman TJ, Miller EK. 2016. Gamma and beta bursts underlie working memory. *Neuron.* 90:152–164.
- Majerus S. 2013. Language repetition and short-term memory: an integrative framework. *Front Hum Neurosci.* 7:357.
- Majerus S, Attout L, D'Argembeau A, Degueldre C, Fias W, Maquet P, Martinez Perez T, Stawarczyk D, Salmon E, Van der Linden M, et al. 2012. Attention supports verbal short-term memory via competition between dorsal and ventral attention networks. *Cereb Cortex.* 22:1086–1097.
- Martin RC. 2005. Components of short-term memory and their relation to language processing: evidence from neuropsychology and neuroimaging. *Curr Dir Psychol Sci.* 14:204–208.
- Martin RC, Breedin SD. 1992. Dissociations between speech perception and phonological short-term memory deficits. *Cogn Neuropsychol.* 9:509–534.
- Martin RC, Lesch MF, Bartha MC. 1999. Independence of input and output phonology in word processing and short-term memory. *J Mem Lang.* 41:3–29.
- Martin N, Saffran EM. 1997. Language and auditory-verbal short-term memory impairments: evidence for common underlying processes. *Cogn Neuropsychol.* 14:641–682.
- Martin RC, Wu D, Freedman M, Jackson EF, Lesch M. 2003. An event-related fMRI investigation of phonological versus semantic short-term memory. *J Neurolinguist.* 16:341–360.
- McLaren DG, Ries ML, Xu G, Johnson SC. 2012. A generalized form of context-dependent psychophysiological interactions (gPPI): a comparison to standard approaches. *NeuroImage.* 61:1277–1286.
- Mendoza-Halliday D, Torres S, Martinez-Trujillo JC. 2014. Sharp emergence of feature-selective sustained activity along the dorsal visual pathway. *Nat Neurosci.* 17:1255–1262.
- Mesgarani N, Cheung C, Johnson K, Chang EF. 2014. Phonetic feature encoding in human superior temporal gyrus. *Science.* 343:1006–1010.
- Mongillo G, Barak O, Tsodyks M. 2008. Synaptic theory of working memory. *Science.* 319:1543–1546.
- Myers NE, Stokes MG, Nobre AC. 2017. Prioritizing information during working memory: beyond sustained internal attention. *Trends Cogn Sci.* 21:449–461.
- Narayanan NS, Prabhakaran V, Bunge SA, Christoff K, Fine EM, Gabrieli JDE. 2005. The role of the prefrontal cortex in the maintenance of verbal working memory: an event-related fMRI analysis. *Neuropsychology.* 19:223–232.
- Nee DE, Brown JW, Askren MK, Berman MG, Demiralp E, Krawitz A, Jonides J. 2013. A meta-analysis of executive components of working memory. *Cereb Cortex.* 23:264–282.
- Norman KA, Polyn SM, Detre GJ, Haxby JV. 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci.* 10:424–430.
- Oberauer K, Lange EB. 2009. Activation and binding in verbal working memory: a dual-process model for the recognition of nonwords. *Cogn Psychol.* 58:102–136.
- Obleser J, Leaver AM, Vanmeter J, Rauschecker JP. 2010. Segregation of vowels and consonants in human auditory cortex: evidence for distributed hierarchical organization. *Front Psychol.* 1:232.
- Paulesu E, Frith CD, Frackowiak RS. 1993. The neural correlates of the verbal component of working memory. *Nature.* 362:342–345.
- Paulesu E, Shallice T, Danelli L, Sberna M, Frackowiak RSJ, Frith CD. 2017. Anatomical modularity of verbal working memory? functional anatomical evidence from a famous patient with short-term memory deficits. *Front Hum Neurosci.* 11:231.
- Postle BR. 2006. Working memory as an emergent property of the mind and brain. *Neuroscience.* 139:23–38.
- Postle BR, Berger JS, D'Esposito M. 1999. Functional neuroanatomical double dissociation of mnemonic and executive control processes contributing to working memory performance. *Proc Natl Acad Sci USA.* 96:12959–12964.
- Price CJ. 2010. The anatomy of language: a review of 100 fMRI studies published in 2009. *Ann N Y Acad Sci.* 1191:62–88.
- Price CJ. 2012. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage.* 62:816–847.
- Ravizza SM, Delgado MR, Chein JM, Becker JT, Fiez JA. 2004. Functional dissociations within the inferior parietal cortex in verbal working memory. *NeuroImage.* 22:562–573.
- Ravizza SM, Hazeltine E, Ruiz S, Zhu DC. 2011. Left TPJ activity in verbal working memory: implications for storage- and sensory-specific models of short term memory. *NeuroImage.* 55:1836–1846.
- Riggall AC, Postle BR. 2012. The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. *J Neurosci.* 32:12990–12998.
- Rissman J, Gazzaley A, D'Esposito M. 2008. Dynamic adjustments in prefrontal, hippocampal, and inferior temporal interactions with increasing visual working memory load. *Cereb Cortex.* 18:1618–1629.
- Romero L, Walsh V, Papagno C. 2006. The neural correlates of phonological short-term memory: a repetitive transcranial magnetic stimulation study. *J Cogn Neurosci.* 18:1147–1155.
- Rose NS, LaRocque JJ, Riggall AC, Gosseries O, Starrett MJ, Meyerling EE, Postle BR. 2016. Reactivation of latent working

- memories with transcranial magnetic stimulation. *Science*. 354:1136–1139.
- Roux F, Uhlhaas PJ. 2014. Working memory and neural oscillations: α - γ versus θ - γ codes for distinct WM information? *Trends Cogn Sci*. 18:16–25.
- Rypma B, D'Esposito M. 1999. The roles of prefrontal brain regions in components of working memory: effects of memory load and individual differences. *Proc Natl Acad Sci USA*. 96:6558–6563.
- Rypma B, Prabhakaran V, Desmond JE, Glover GH, Gabrieli JD. 1999. Load-dependent roles of frontal brain regions in the maintenance of working memory. *NeuroImage*. 9:216–226.
- Saad ZS, Reynolds RC. 2012. SUMA. *NeuroImage*. 62:768–773.
- Salamé P, Baddeley A. 1982. Disruption of short-term memory by unattended speech: implications for the structure of working memory. *J Verbal Learning Verbal Behav*. 21:150–164.
- Salmon E, Van der Linden M, Collette F, Delfiore G, Maquet P, Degueldre C, Luxen A, Franck G. 1996. Regional brain activity during working memory tasks. *Brain*. 119:1617–1625.
- Schneegans S, Bays PM. 2017. Restoration of fMRI decodability does not imply latent working memory states. *J Cogn Neurosci*. 29:1977–1994.
- Scott SK, Johnsrude IS. 2003. The neuroanatomical and functional organization of speech perception. *Trends Neurosci*. 26:100–107.
- Serences JT, Ester EF, Vogel EK, Awh E. 2009. Stimulus-specific delay activity in human primary visual cortex. *Psychol Sci*. 20:207–214.
- Shallice T, Vallar G. 1990. The impairment of auditory-verbal short-term storage. In: Shallice GVT, editor. *Neuropsychological impairments of short-term memory*. New York, NY, US: Cambridge University Press. p. 11–53.
- Shallice T, Warrington EK. 1977. Auditory-verbal short-term memory impairment and conduction aphasia. *Brain Lang*. 4:479–491.
- Siebenhühner F, Wang SH, Palva JM, Palva S. 2016. Cross-frequency synchronization connects networks of fast and slow oscillations during visual working memory maintenance. *eLife*. 5:e13451.
- Smith EE, Jonides J. 1998. Neuroimaging analyses of human working memory. *Proc Natl Acad Sci USA*. 95:12061–12068.
- Smith EE, Jonides J, Marshuetz C, Koeppel RA. 1998. Components of verbal working memory: evidence from neuroimaging. *Proc Natl Acad Sci USA*. 95:876–882.
- Sreenivasan KK, Curtis CE, D'Esposito M. 2014. Revisiting the role of persistent neural activity during working memory. *Trends Cogn Sci*. 18:82–89.
- Sternberg S. 1966. High-speed scanning in human memory. *Science*. 153:652–654.
- Stokes MG. 2015. “Activity-silent” working memory in prefrontal cortex: a dynamic coding framework. *Trends Cogn Sci*. 19:394–405.
- Talairach J, Tournoux P. 1988. Co-planar stereotaxic atlas of the human brain: 3-dimensional approach system: an approach to cerebral imaging. New York: Theime Medical.
- Todd JJ, Fougny D, Marois R. 2005. Visual short-term memory load suppresses temporo-parietal junction activity and induces inattentive blindness. *Psychol Sci*. 16:965–972.
- Turkeltaub PE, Coslett HB. 2010. Localization of sublexical speech perception components. *Brain Lang*. 114:1–15.
- Vallar G, Baddeley AD. 1984. Fractionation of working memory: neuropsychological evidence for a phonological short-term store. *Journal of Verbal Learning and Verbal Behavior*. 23:151–161.
- Vallar G, Papagno C. 1995. Neuropsychological impairments of short-term memory. In: Baddeley AD, Wilson BA, Watts FN, editors. *Handbook of memory disorders*. Oxford, England: John Wiley & Sons. p. 135–165.
- Vigneau M, Beaucois V, Hervé PY, Duffau H, Crivello F, Houdé O, Mazoyer B, Tzourio-Mazoyer N. 2006. Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *NeuroImage*. 30:1414–1432.
- Wager TD, Jonides J, Reading S. 2004. Neuroimaging studies of shifting attention: a meta-analysis. *NeuroImage*. 22:1679–1693.
- Warrington EK, Logue V, Pratt RT. 1971. The anatomical localization of selective impairment of auditory verbal short-term memory. *Neuropsychologia*. 9:377–387.
- Warrington EK, Shallice T. 1969. The selective impairment of auditory verbal short-term memory. *Brain*. 92:885–896.
- Watanabe K, Funahashi S. 2014. Neural mechanisms of dual-task interference and cognitive capacity limitation in the prefrontal cortex. *Nat Neurosci*. 17:601–611.
- Wolff MJ, Jochim J, Akyurek EG, Stokes MG. 2017. Dynamic hidden states underlying working-memory-guided behavior. *Nat Neurosci*. 20:864–871.
- Xu Y. 2017. Reevaluating the sensory account of visual working memory storage. *Trends Cogn Sci*. 21:794–815.
- Xu Y, Chun MM. 2006. Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*. 440:91–95.
- Yarkoni T. 2009. Big correlations in little studies: inflated fMRI correlations reflect low statistical power—commentary on Vul et al. (2009). *Perspect Psychol Sci*. 4:294–298.
- Zarahn E, Rakitin B, Abela D, Flynn J, Stern Y. 2005. Positive evidence against human hippocampal involvement in working memory maintenance of familiar stimuli. *Cereb Cortex*. 15:303–316.