UC Berkeley
UC Berkeley Previously Published Works

Title

The effect of rehearsal rate and memory load on verbal working memory

Permalink

Authors

Fegen, David
Buchsbaum, Bradley R
D'Esposito, Mark

Publication Date

DOI

Peer reviewed

The Effect of Rehearsal Rate and Memory Load on Verbal Working Memory

David Fegen[a,*], **Bradley R. Buchsbaum**[c], and **Mark D'Esposito**[a,b]

[a]Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, CA 94720, USA

[b]Department of Psychology, University of California, Berkeley, Berkeley, CA 94720, USA

[c]Rotman Research Institute, Baycrest Hospital, Toronto ON M6A 2E1, Canada

Abstract

While many neuroimaging studies have investigated verbal working memory (WM) by manipulating memory load, the subvocal rehearsal rate at these various memory loads has generally been left uncontrolled. Therefore, the goal of this study was to investigate how mnemonic load and the rate of subvocal rehearsal modulate patterns of activity in the core neural circuits underlying verbal working memory. Using fMRI in healthy subjects, we orthogonally manipulated subvocal rehearsal rate and memory load in a verbal WM task with long 45-second delay periods. We found that middle frontal gyrus (MFG) and superior parietal lobule (SPL) exhibited memory load effects primarily early in the delay period and did not exhibit rehearsal rate effects. In contrast, we found that inferior frontal gyrus (IFG), premotor cortex (PM) and Sylvian-parietal-temporal region (area Spt) exhibited approximately linear memory load and rehearsal rate effects, with rehearsal rate effects lasting through the entire delay period. These results indicate that IFG, PM and area Spt comprise the core articulatory rehearsal areas involved in verbal WM, while MFG and SPL are recruited in a general supervisory role once a memory load threshold in the core rehearsal network has been exceeded.

Keywords

memory load; subvocal rehearsal; fMRI; verbal working memory; middle frontal gyrus; Sylvian-parietal-temporal (area Spt)

1. Introduction

Working memory (WM) is what allows one to maintain and manipulate task-relevant information over short time periods and is critical for various cognitive tasks such as problem solving, reasoning and comprehension (Daneman & Merikle, 1996; Kyllonen & Christal, 1990; Logie et al., 1994). At the center of many theoretical models of WM is an attempt to explain how task-relevant information is maintained in an activated state over a delay period. For example, Baddeley and colleagues' classic WM model consists of domain-

*Corresponding author. Present address: NYU, 6 Washington Place, Room 863, New York, NY 10003, USA. Tel.: +1 212 998 8347. dfegen@gmail.com.

specific storage components for visuospatial and verbal information coupled with rehearsal processes that serve to update and refresh items currently held in memory (Baddeley, 1986). Current neuroscientific models of WM such as the "emergent property view" (Postle, 2006a) propose that information is effectively "stored" in memory by the repeated reactivation of the same cortical regions that were involved in the initial perception of the task-relevant information (Buchsbaum & D'Esposito, 2008; D'Esposito, 2007; Postle, 2006a). In these models rehearsal is one mechanism by which transient representations can be reactivated and is defined as the repeated selection of, or the repeated attention to, task-relevant mnemonic representations (Curtis & D'Esposito, 2003). However, despite the importance of strategic rehearsal processes to theoretical WM models, neuroscience investigations of WM typically allow subjects to freely choose the rate and manner in which they maintain information in working memory. A drawback of this naturalistic approach is that a subject's internal rehearsal strategy may change as a function of other experimentally manipulated variables, such as memory load (i.e. the number of items that must be retained in memory). Indeed, memory load manipulations are often used as a way of indexing working memory storage processes (Awh et al., 1996; Todd & Marois, 2005), and to the extent that these manipulations are used to make inferences about the informational capacity of a brain region or system, it is important to understand how rehearsal processes scale with memory load.

The primary question of the current study is whether the neural systems that vary as a function of rehearsal rate are modulated by changes in memory load, and vice versa. While it has been established in behavioral studies that verbal WM capacity is strongly correlated with a person's ability to rapidly produce speech (Cowan et al., 1998; Dasí et al., 2008; Hulme et al., 1984), the connection between rehearsal rate and memory load has never been examined to our knowledge in neuroscience studies of WM. Indeed, in the context of manipulations of memory load, a corresponding increase in the rate of subvocal rehearsal acts as a confounding variable. Moreover, the confounding of load and rate of rehearsal may partially account for the between-study variability in neural localization of load effects in previous neuroimaging studies (Postle et al., 1999; Rypma et al., 1999a,b; Rypma et al., 2002; Zarahn et al., 2005) contributing to what has already been described about the issue (Feredoes & Postle, 2007). Thus, because memory load manipulations are not particularly well-suited to dissociate domain-general executive processes (i.e. attentional, executive, internal monitoring) from rehearsal or reactivation processes (Buchsbaum & D'Esposito, 2008), previous verbal WM load studies have demonstrated activity across numerous cortical areas including middle frontal gyrus (MFG, BA 9/46), inferior frontal gyrus (IFG, BA 44/45), premotor cortex (PM, BA 6), Sylvian-parietal-temporal area (area Spt) and superior parietal lobule (SPL, BA 7). While MFG and SPL are generally thought of as domain-general executive regions and known to be activated by high memory loads (Cohen et al., 1997; Rypma et al., 1999a; Zarahn et al., 2005), the involvement of these regions in subvocal rehearsal is less clear. In particular, while it is known that MFG and SPL are not required for WM tasks with low memory loads (Barbey et al., 2013; D'Esposito & Postle, 1999a; Hamidi et al., 2008; Postle et al., 2006b) and thus not required for rehearsal of task-relevant items, it may be the case that these areas may be recruited at higher rates of rehearsal. On the other hand, in areas such as IFG, PM and area Spt, which are known to be involved in articulatory rehearsal and speech production more generally (Hickok et al.,

2003; Shergill et al., 2002; Wildgruber et al., 2001, 1999), it is not clear if memory load influences rehearsal processing in these regions independently of subvocal rehearsal rate. In summary, it is currently unknown whether domain-general executive regions (MFG, SPL) exhibit selective activity related to the rehearsal of task-relevant items, and if domain-specific nodes of the verbal rehearsal network (IFG, PM, Spt) display memory load effects. Understanding how activity in the nodes of the WM circuit are modulated by rehearsal rate and memory load will lead to a deeper understanding of the computations that each of these cortical regions are performing and is vital to our understanding of WM.

A second question regarding WM is how the representation of task-relevant information changes as it is being rehearsed at a constant rate over time. Behavioral studies have shown that WM tasks with long retention intervals involve an effortful first stage followed by an automatized and less effortful second stage (Aldridge et al., 1987; Greene, 1987; Naveh-Benjamin & Jonides, 1984; Phaf & Wolters, 1993). Several fMRI studies of WM maintenance have found decreasing activity as the delay period progressed in the cortical regions involved in maintaining task-relevant information (Chein & Fiez, 2001; Jha & McCarthy, 2000). The finding of decreasing activity over time may reflect a "sharpening" of task-relevant neural representations. With time and increasing number of rehearsals, neural activation associated with the coding of irrelevant features may begin to wane, a phenomenon that has been referred to as "repetition suppression" (Desimone, 1996; Wiggs & Martin, 1998). However, because these previous studies of WM maintenance did not directly control rehearsal it is not known if these cortical regions demonstrated decreasing activity over the delay period because participants slowed or stopped rehearsing before the delay period was over, or whether activity decreases might be genuinely attributed to a neural phenomenon such as repetition suppression. These alternatives can be better distinguished by explicitly controlling rehearsal rate and examining activity changes over the delay period.

A third question that we will examine is how activity in regions supporting subvocal rehearsal is modulated by rehearsal rate. While several fMRI studies investigating speech found a linear relationship between rehearsal rate and cortical activity (Riecker et al., 2006, 2005; Shergill et al., 2002; Wildgruber et al., 2001), it is unclear if this same pattern of activation holds for WM rehearsal as these speech studies did not contain a memory component and simply had subjects repeatedly rehearse single syllables like "ta". The attentional demands associated with a WM task may affect the neural systems involved in subvocal rehearsal. For example, top-down attention may lead to synaptic potentiation, a form of synaptic plasticity that may result in less activity as the rate of activation is increased. If there is synaptic efficiency at higher rehearsal rates then this would result in a nonlinear relationship between cortical activity and rehearsal rate with proportionately less cortical activity required.

In order to investigate how neural activity during the maintenance of task-relevant information changes with memory load, time, and rehearsal rate, we employed a novel WM paradigm that explicitly and directly controlled subvocal rehearsal rate as well as memory load over 45-second delay periods. We then addressed the above questions by investigating: 1) which cortical regions are involved in computations related to memory load, rehearsal

rate, or both and if these cortical areas can be dissociated on the basis of these factors (behaviorally, the relationship of memory load and rehearsal rate will be tested in a related behavioral task), 2) how neural activity changes through time while keeping rehearsal rate constant, and 3) how neural activity changes with different subvocal rehearsal rates, especially in the critical rehearsal regions PM and area Spt.

2. Methods

2.1 Subjects

Twenty-eight subjects gave informed written consent according to procedures approved by the University of California and participated in the study. All were right-handed, native English speakers, had normal or corrected-to-normal vision, and normal hearing. All subjects were healthy with no neurological or psychiatric disease. One subject was eliminated due to falling asleep in the scanner and three subjects were eliminated for failing to follow the instructions properly (subvocally rehearsing when not explicitly prompted by the task). Thus, a total of twenty-four subjects (13 females; age: 18–32, mean: 21.3) were included in the final analyses.

2.2 Experimental Stimuli

Letters were chosen pseudorandomly from a pool of 19 consonants (b,c,d,f,g,h,j,k,l,m,n,p,q,r,s,t,v,x,z) with the only constraint being that a letter could not be repeated within the same trial. Vowels (a,e,i,o,u) and the letter "y" were excluded to minimize chunking of letter sequences into words; and the letter "w" was excluded because it has three syllables. Letters were spoken by a female voice that was generated with text-to-speech software (Nuance Speechify, Burlington, MA).

2.3 Behavioral task performed prior to fMRI scanning

Before being informed of any of the details of the fMRI experiment, subjects performed a verbal WM task to determine the effect of memory load on rehearsal rate. Subjects were presented with 2, 4, 6, or 8 letters at a rate of one letter per second. Each letter was presented simultaneously in the visual and auditory modalities. Following the presentation of the final letter in the sequence there was a 1-second pause before a 500ms beep sounded informing subjects to begin overtly rehearsing the letter sequence over a 15-second delay period. Subjects were instructed to rehearse the letters one letter at a time, in the original order, at a normal speaking voice at whatever rate was comfortable for them. After the delay period subjects were prompted with a recall probe (green triangle that appeared in the center if screen) and given four seconds to recall in order as many of the letters as possible. Overt rehearsal and recall responses were recorded by a digital recorder and then manually transcribed and scored.

Each subject was given a total of five blocks of trials with two-minute breaks between blocks. The first block was a practice block that was not scored and the remaining four blocks were test blocks with the first trial of each block also counted as practice. Each block contained eight scored trials, two at each memory load, which were pseudorandomly

ordered for each subject. Therefore, each subject had a total of eight trials at each memory load.

2.4 Behavioral task performed during fMRI scanning

Memory load (2, 4, 6 letters) and rehearsal rate (0.8, 1.1, 1.4, 1.7, 2.0, 2.3, 2.6Hz) were independently manipulated resulting in a fully crossed 3×7 factorial design (Figure 1). During encoding subjects were presented with 2, 4, or 6 letters at a rate of one letter per second. As in the behavioral task, each letter was presented simultaneously in the auditory and visual modalities. During a 45-second delay period, subjects were required to subvocally rehearse the sequence one letter at a time, in the original order, in synchrony with a visual pacing cue (flashing gray circle). Each pacing cue was visually presented for 300 msec while the interval between pacing cues varied across trials. Specifically, subjects were paced to rehearse each letter at 0.8Hz (1 rehearsal/1250msec), 1.1Hz, 1.4Hz, 1.7Hz, 2.0Hz, 2.3Hz and 2.6Hz (1 rehearsal/385msec). Critically, to ensure that subjects were subvocally rehearsing one letter with each visual pacing cue for the entire delay period, at the end of each trial subjects were prompted with a probe (red circle) and given two seconds to verbally report the last letter they had subvocally rehearsed. Because the correct answer depended on the number of rehearsals over a fixed time interval, the correct answer varied trial-by-trial both as a function of how many letters subjects were supposed to rehearse as well as the pacing rate -- so subjects could not use a simple rule to answer the order probe correctly (e.g. always the second letter). Because the memory load and rehearsal rate were independently manipulated for every trial, the correct answer randomly varied among all possible letter positions. After the order probe, subjects were prompted with a recall probe (green triangle) and given four seconds to verbally report the full list of letters in order. Verbal answers for the order and recall probes were digitally recorded and manually scored after the experiment.

In a control task subjects heard time-reversed versions of the letters used in the rehearsal trials, followed by a 45-second period where they were instructed to observe the flashing gray circles without subvocally rehearsing. On these control trials, although subjects did not engage in subvocal rehearsal, the rate of the pacing cue was varied in the same manner as it was in the memory trials. At the end of the control task trials subjects were given four seconds to say "square" as soon as they saw the blue square probe appear. To ensure subjects remained attentive during these long periods, catch trials were included where the blue square appeared before the full 45 seconds so that subjects could not anticipate when the trial would end. Delay periods in catch trials were pseudorandomly chosen from a set of durations spanning from 25%–75% of the full 45-second delay periods, with the constraint that once a duration was used it could not be repeated. Full-length control and interrupted control ("catch") trials were randomly intermixed.

Each fMRI session consisted of ten scanning runs that lasted approximately eight minutes each. Each run contained a total of seven trials: four or five task trials, one or two control trials, and one or two catch trials. For all trial types, memory load and rehearsal rate were independently and pseudorandomly chosen with the constraint that all conditions had to occur before a condition could be repeated. In total, each fMRI session contained 42-task

trials, 14-control trials and 14-catch trials. Memory load and rehearsal rate were fully crossed in the task condition ($3 \times 7 = 21$), resulting in two trials per condition. Although there were a small number of trials per cell, none of our statistical analyses compared individual conditions with one another. Rather, for each factor, either the main effect (which collapses over the trials of the other factor) or linear trend (which includes all trials in each condition of the relevant factor) were statistically tested. Due to limited scanner time, the load and rate factors were not fully crossed within the control trials, however, each of the seven rates was presented twice in order to control over the full range of visual flashing cues in the absence of subvocal rehearsal.

2.5 Acquisition of Functional MRI Data

MR data were acquired with a Siemens TIM/Trio 3 Tesla scanner (Berlin/Munich, Germany). Functional data were obtained using a 32-channel radiofrequency head coil and a 1-shot T_2*-weighted echo-planar imaging (EPI) sequence sensitive to blood oxygenation level-dependent (BOLD) contrast (TR = 2000msec, TE = 25msec, 224mm field of view with a 72×72 matrix size, in-plane resolution 3.1mm $\times$ 3.1mm). Each functional volume contained 34-contiguous 3.3mm-thick axial slices separated by a 0.5mm interslice gap acquired in an interleaved fashion. High-resolution whole-brain MP Flash T_1-weighted scans were acquired for anatomical localization.

Presentation software (Neurobehavioral Systems, Albany, CA) was used for auditory and visual stimulus delivery. Auditory stimuli were delivered via MRI-compatible form-fitting foam insert earphones (Sensimetrics, MA). Visual stimuli were presented via a liquid crystal display projector (Avotec, FL), which displayed images on a screen located in the center of the scanner bore. Subjects viewed the screen by looking at a mirror mounted on the radiofrequency coil. Overt responses in the fMRI scanner were recorded with a dual-channel, noise cancelling fiber optical microphone system and noise reduction software (Optoacoustics Ltd., Or-Yehuda, Israel).

2.6 Pre-Processing of fMRI Data

MRI data were converted to NifTI format. Functional data were slice-time corrected, realigned to the first acquired volume using the AFNI (Cox, 1996) program 3dVolreg and spatially smoothed with a 5-mm full-width at half-maximum Gaussian kernel. No runs were removed for excessive head motion (>3mm within a single run). All subsequent statistical analyses were performed on these realigned and smoothed images. To view the functional MRI data on a representative normalized structural anatomical image, a study specific group template was created with the program Advanced Normalization Tools (ANTS: http://www.picsl.upenn.edu/ANTS/). In an iterative fashion, each subject's high-resolution anatomical scan was warped (symmetric normalization algorithm, SyN) into registration with one another, and a group mean image was generated. The study specific group template was then normalized to Montreal Neurological Institute (MNI) space with a 12-parameter affine transformation. Each subject's high-resolution anatomical scan was then registered to this study specific group template in MNI space utilizing the SyN algorithm (ANTS). These warping parameters were then applied to the native space EPI data/statistical maps as needed to transform them into normalized template space. The study-specific group template and

group functional data were converted to cortical surface maps for visualization purposes (SUMA: http://afni.nimh.nih.gov/afni/suma).

2.7. Statistical Analyses of fMRI data

Regression modeling at the single-subject level was performed with the AFNI program 3dDeconvolve. All trials whether behaviorally correct or incorrect (as defined by either the order or recall probe) were included in all fMRI analyses. Trials with order probe errors were included because this metric did not directly relate to the main question of this study (the rehearsal of task-relevant information), and instead was meant to ensure subjects had an incentive to perform the task as instructed. We expect that as rehearsal rates increase, order accuracy will decrease, as there is an increased chance subjects will miss a visual pacing cue somewhere within the long delay periods (e.g. if they happen to blink at the wrong time). Therefore, a decrease in accuracy for this probe does not necessarily indicate subjects are not rehearsing as they are instructed to. Trials with recall probe errors were included because they generally occurred on a small percentage of trials (group average accuracy: 95% correct) and tended to be slight confusions between similar phonemes (e.g. replacing "p" with "b"). Block regressors were generated by convolving a boxcar function with a hemodynamic response function (for further details please see AFNI documentation). For all trials, encoding periods were modeled as blocks with durations matching stimulus presentation lengths (two second block for 2-letter trial, four second block for 4-letter trial, six second block for 6-letter trial) and the delay period of every task, control and catch trial was also modeled as a block. Each delay period condition was modeled with a separate regressor ($LOAD_{[2L,4L,6L]} \times RATE_{[0.8,1.1,1.4,1.7,2.0,2.3,2.6\ Hz]}$). To reduce collinearity between encoding and delay period regressors, a two-second gap was introduced between these two regressors. Although encoding and delay regressors were only separated by two seconds, raising concern that they might be collinear, the computed correlations between encoding and delay regressors, for all levels of load, were near zero (between .018 and .002). Thus, for analyses involving the entire delay period, a 43-second block regressor was placed beginning two seconds after the onset of the delay period. For analyses involving different phases of the delay period, the delay period was divided into three separate 14-second segments. Therefore, the early delay period was modeled as a block 2–16 seconds into the delay period, the middle of the delay period as a block 16–30 seconds into the delay period, and the end of the delay period as a block 30–44 seconds into the delay period. Order probe and recall probe periods were also modeled. For each scanning run a set of nuisance regressors (constant term plus linear, quadratic and higher order polynomial terms) were included to model low frequency noise. Head movement was also modeled using six motion parameters estimated from the motion correction algorithm. Statistical contrasts at the single-subject level were carried out in native space and were computed as weighted sums of the estimated beta coefficients divided by an estimate of the standard error, resulting in a *t*-statistic. For linear trend analyses, beta coefficients were used in linear trend contrasts for the three equally spaced levels of LOAD (−1,0,1) and the seven equally spaced levels of RATE (−1,−0.667,−0.333,0,0.333,0.667,1) at the single-subject level and resulting *t*-statistics were used in further analyses. Each subject's *t*-statistic map was normalized to MNI space and random effects group analyses were computed on *t*-values. Beta values scaled by the variability (i.e. *t*-values) were utilized because *t*-values have been empirically

shown to be more normally distributed than raw beta values (Thirion et al., 2007). Left-hemisphere group results were corrected for multiple comparisons by thresholding to $q<0.05$ using the False Discovery rate (FDR) method (Genovese et al., 2002). Effect sizes were estimated with a parametric generalized eta squared (η_G^2) analysis.

2.8 Region of Interest Definition

To further investigate the effects of memory load and rehearsal rate in cortical areas commonly implicated in verbal working memory, region of interest (ROI) analyses were performed within MFG, SPL, IFG, PM and area Spt. IFG was subdivided into a more posterior (IFG pars operculum, IFGpo, BA 44) and more anterior region (IFG pars triangularis, IFGpt, BA 45) because of evidence that these two subdivisions have different functional (Badre & Wagner, 2007; Wagner et al., 2001) and anatomical connectivity (Frey et al., 2008).

ROIs were created using a multi-stage process: first defining a "search space" mask in group normalized space, reverse normalizing these masks to native space, then finding the top-10 statistically significant contiguous voxels within resulting search space masks for each subject. Search space masks were defined in MNI space using both anatomical and functional methods. Anatomical regions were defined using unthresholded masks from the Harvard-Oxford Cortical and Subcortical Structural Atlas (included in FSL: http://www.fmrib.ox.ac.uk). Functional activation for search spaces was assessed with statistical contrasts computed at the single-subject level for each voxel as a t-statistic for the contrast [$TASK_{ALL} > CONTROL_{ALL}$] over 43 seconds of the delay period (not including the first two seconds). Therefore, these statistical contrasts were not biased to a particular time phase of the delay period (since the whole delay period was used) or biased to a particular task condition (since all task conditions were included). These single subject t-statistic maps were spatially normalized and then tested against zero with a one-sample t-test.

Search space masks were defined slightly differently based upon the functional and anatomic details of each region. Anatomical definitions (Harvard-Oxford atlas) were used for IFGpo and SPL. However, for areas demonstrating spatially heterogeneous activity peaks in MNI space, a combined anatomical and functional approach was used in order to choose, in a principled manner, a single anatomical location for subsequent analyses. Specifically, within each anatomical region the most significant voxel was identified in MNI space and a 10mm sphere was placed around this top-voxel before being re-intersected with the original group normalized anatomical mask. This approach was used to define IFGpt, MFG and PM search space masks. Lastly, because area Spt is functionally (as opposed to anatomically) defined (Buchsbaum et al., 2005), at the group level the voxel with the highest t-value located near the Sylvian fissure at the parietal-temporal boundary was located and a 10mm sphere was placed around this top-voxel.

These group normalized search space masks were reverse normalized (using each subject's inverse MNI to native EPI space transformation) and then searched within for the voxel with the highest t-value on a subject-by-subject basis. Once the highest t-value was located (using

the statistical contrast detailed above), the nine highest contiguous voxels were found, giving a top-10 voxel ROI for each subject in native space.

2.9 Analyses of Behavioral Data

Behavioral data collected before fMRI scanning were scored for rehearsal rate over the delay as well as for correct recall of letters at the end of the delay period. One subject was eliminated due to whispering rehearsals that could not be scored (subject was not removed from fMRI analyses). Trials in which subjects used elaborative rehearsal (e.g. if the letters were "sdnz" and the subject began rehearsing, "San Diego, New Zealand") were eliminated for failing to follow instructions (instructed only to rehearse the given letters). Overt rehearsals were manually recorded by counting the number of letters spoken over the full delay period (15 seconds) as well as over the first half (7.5 seconds) and the second half (7.5 seconds). To determine the effect of rehearsal rate on accuracy, trials were classified as "all correct" (all letters in trial correct and in correct position) or as "error" (any letter in trial incorrect or in incorrect position). For both rate and recall, the mean value of each condition was computed for each subject and then used to compute the group mean.

Behavioral data collected during fMRI scanning were scored for the order probe (red circle) and recall probe (green triangle). Order probe was classified correct if the single correct letter was reported. Recall data were scored on a letter-by-letter basis requiring both letter identity and letter position to be correct to be classified as correct. The mean number correct for each condition was computed for each subject and then used to compute the group mean. Non-parametric permutation-based ANOVA tests were performed with the R package "lmPerm" (http://cran.r-project.org/web/packages/lmPerm/index.html), while two-sample permutation tests were performed with the R package "DAAG" (http://cran.r-project.org/web/packages/DAAG/index.html). Permutation ANOVA tests employed an "exact" test in which all permutations were computed, while two-sample permutation tests employed 1,000,000 randomly selected permutations. While we did not detect any irregularities in these behavioral data relative to other similar data, non-parametric permutation tests were employed to avoid making any assumptions about the distribution of behavioral data which may be incorrect. For data analyzed by ANOVA tests, effect sizes were estimated by computing parametric generalized eta squared (η_G^2) values, while for data analyzed by two sample tests, Cohen's d (d) values were computed.

3. Results

3.1 Behavioral data prior to fMRI scanning

To determine the relation between memory load and rehearsal rate, subjects performed an overt free rehearsal WM task. A permutation repeated-measures ANOVA was performed on recall accuracy with LOAD modeled as a within-subject factor and SUBJECT modeled as a random effect. The main effect of LOAD was significant ($F(3,66)=141$, $MSE=3.3$, $p<0.0001$; $\eta_G^2=0.86$). Post-hoc linear contrasts on the permutation ANOVA results revealed a significant decreasing trend for LOAD (linear coefficient$=-0.6$, $t(66)=-19.4$, $SE=0.03$, $p<0.0001$), confirming that recall accuracy decreased with increasing memory load.

To investigate the effect of load on rehearsal rate, a permutation repeated-measures ANOVA was performed on free-rehearsal rate with LOAD modeled as a within-subject factor and SUBJECT modeled as a random effect (Figure 2a). The main effect of LOAD was significant ($F(3,66)$=26.6, MSE=2.9, p<0.0001; η_G^2=0.55). Post-hoc paired two-sample permutation tests revealed that rehearsal rate significantly increased with memory load until the 6-letter condition, at which point rehearsal rate significantly decreased: 2-letter versus 4-letter trials ($t(22)$=−8.8, p<0.0001; d=1.84), 4-letter versus 6-letter trials ($t(22)$=−3.0, p<0.01; d=0.62) and 6-letter versus 8-letter trials ($t(22)$=4.0, p<0.001; d=0.83). Thus, subjects displayed an inverted U-shaped function with increases in rehearsal rate across increases in low memory loads, but with a decrease in rehearsal rate across an increase in high memory loads. When trials were divided into correct and error trials and compared across conditions, two-sample permutation tests revealed there were significant differences between 6-letter ($t(40)$=2.8, p<0.01; d=0.83) and 8-letter trials ($t(18)$=3.3, p<0.001; d=1.28). Specifically, subjects rehearsed faster in all correct trials and slower in error trials in both 6 and 8-letter trials.

To investigate if rehearsal rate was constant or varied across the delay period, the rehearsal rate in the first half (first 7.5 seconds) and second half (last 7.5 seconds) of the delay period was compared. A two-way permutation repeated-measure ANOVA was performed with LOAD and TIME modeled as within-subject factors and SUBJECT modeled as a random effect (Figure 2b). There was a significant LOAD×TIME interaction ($F(3,66)$=8.6, MSE=0.2, p<0.001; η_G^2=0.03) as rehearsal rate decreased over the delay period for low memory loads (2 and 4 letters) but increased over the delay period for high memory loads (6 and 8 letters). Specifically, post-hoc paired two-sample permutation tests across TIME revealed a significant difference for 2-letter trials ($t(22)$=2.8, p<0.05; d=0.58), a marginally significant difference for 4-letter trials ($t(22)$=2.0, p=0.06; d=0.42), no significant difference for 6-letter trials ($t(22)$= −1.6, p=0.11; d=0.34) and a significant difference for 8-letter trials ($t(22)$=−2.4, p<0.05; d=0.49).

3.2 Behavioral data during fMRI scanning

A permutation two-way repeated-measures ANOVA was performed on the order probe accuracy with LOAD and RATE modeled as within-subject factors and SUBJECT modeled as a random effect (Figure 3, left). The main effect of LOAD was significant ($F(2,46)$=9.0, MSE=8994, p<0.0001; η_G^2=0.04) and RATE was significant ($F(6,138)$=24.1, MSE=23237, p<0.0001; η_G^2=0.25), while the interaction of LOAD×RATE was not significant ($F(12,276)$=1.4, MSE=1196, p=0.24; η_G^2=0.03). Post-hoc tests on the permutation ANOVA results revealed a significant decreasing linear trend for LOAD (linear coefficient=−10.3, $t(46)$=−4.3, SE=2.4, p<0.0001) and RATE (linear coefficient=−42.3, $t(138)$=−11.6, SE=3.7, p<0.0001) on order probe accuracy.

A permutation two-way repeated-measures ANOVA was performed on the recall probe accuracy with LOAD and RATE modeled as within-subject factors and SUBJECT modeled as a random effect (Figure 3, right). The main effect of LOAD was significant ($F(2,46)$=31.6, MSE=7514, p<0.0001; η_G^2=0.25) while the main effect of RATE

$(F(6,138)=0.4$, $MSE=33$, $p=0.77$; $\eta^2_G=0$) and the interaction of LOAD×RATE were not significant $(F(12,276)=0.5$, $MSE=49$, $p=0.85$; $\eta^2_G=0.01$). Post-hoc tests on the permutation ANOVA results revealed a significant decreasing linear trend for LOAD (linear coefficient= −8.5, $t(46)=-7.1$, $SE=1.2$, $p<0.0001$) on recall probe accuracy, but not for RATE (linear coefficient=−0.1, $t(138)=-0.1$, $SE=1.0$, $p=1.0$).

3.3 Whole brain fMRI Analyses

To identify cortical regions sensitive to rehearsal rate and memory load, a 43-second delay period activity estimate (*t*-value) for each of the 21 conditions (LOAD$_3$ ×RATE$_7$) for every subject was entered into a parametric two-way repeated-measures ANOVA with LOAD and RATE modeled as within-subject factors and SUBJECT modeled as a random effect (Figure 4). Resulting parametric ANOVA non-directional *F* values for the effects of LOAD and RATE were assigned direction based on parametric ANOVA linear trend analyses. There was a main effect of LOAD in many areas commonly implicated in verbal working memory including MFG, IFG, PM, SPL and area Spt. Additionally, superior frontal gyrus, inferior parietal lobule, and anterior temporal lobe areas demonstrated load effects. A main effect of RATE was found in IFG, PM, area Spt and occipital lobe. No cortical areas revealed a significant LOAD×RATE interaction.

As shown in Figure 4, the occipital lobe demonstrated a main effect of RATE presumably due to the changing frequency of visual pacing cues. To ensure that IFG, PM and area Spt were not also simply responding to visual sensory input, a parametric repeated measures ANOVA with RATE modeled as a within-subject factor and SUBJECT modeled as a random effect was performed on the control conditions which included all seven rehearsal rates (but without the requirement to subvocally rehearse). The occipital area was the only region that demonstrated a main effect of RATE confirming that IFG, PM and area Spt were not activated simply by the visual pacing cues (results not shown).

To investigate how rate and load effects changed across the delay period, activity estimates (*t*-values) for the early delay (2–16 seconds), mid delay (16–30 seconds) and late delay period (30–44 seconds) for each condition were computed. The estimates from each phase were entered into separate two-way parametric repeated measures ANOVAs with LOAD and RATE modeled as within-subject factors and SUBJECT modeled as a random effect (Figure 5). The main effect of LOAD during the early delay period revealed a similar pattern of activation as the main effect of LOAD when averaged across the entire delay period. However, the main effect of LOAD diminished through the delay period and by the late delay period only IFG and PM exhibited significant LOAD effects. While it might be assumed the decrease in memory load over time is due to less activity in higher memory loads through the delay period, further analysis detailed below demonstrates that the activity in high memory loads is actually constant and instead the activity in low memory loads increases over time (Figure 9). In contrast to these memory load effects, areas that exhibited a main effect of RATE during the early delay period remained active in the later delay periods. There were no cortical regions that exhibited a significant LOAD×RATE interaction in any phase of the delay period.

To determine which regions were sensitive to load and rate, as well as those sensitive only to load, two separate conjunction analyses were performed (Figure 6). One conjunction analysis was carried out for voxels showing both a significant parametric linear trend in LOAD and RATE, [LOAD$_{(FDR\ q<0.05)}$] ∩ [RATE$_{(FDR\ q<0.05)}$], and a second conjunction analysis for voxels showing a significant parametric linear trend in LOAD but *not* RATE (i.e. *absence of evidence* of a linear RATE effect at an uncorrected p < .05 threshold) [LOAD$_{(FDR\ q<0.05)}$] ∩ [RATE$_{(p>0.95)}$]. Results demonstrated that IFG, PM and area Spt were significant for both LOAD and RATE, while MFG and SPL were significant for LOAD but not RATE. A conjunction analysis for RATE but *not* LOAD only demonstrated overlap in occipital cortex and therefore was not included. Note that the conjunction analyses of the form [A and *not* B] are not intended as a formal test the hypothesis that A > 0 and B = 0, but rather as an informal approximation of such a test.

3.4 Region of Interest fMRI Analyses

To examine the precise pattern of rehearsal rate and memory load effects across cortical regions commonly recruited during verbal WM, ROI analyses were performed. ROI selection, as described in Methods, was based on identifying areas showing elevated delay period activation and subject to further additional anatomical constraints; it was therefore unbiased with respect to the experimental effects of load and rate. For each ROI, a two-way parametric repeated-measures ANOVA was carried out on the extracted *t*-values with LOAD and RATE modeled as within-subject factors and SUBJECT modeled as a random effect (Figure 7). Two patterns of results emerged. MFG and SPL revealed a main effect of LOAD with no main effect of RATE, while IFGpt, IFGpo, PM and area Spt revealed a main effect of both LOAD and RATE. Furthermore, MFG and SPL exhibited a non-linear or threshold pattern of activity for memory load (activity levels were similar in 2 and 4-letter conditions but significantly increased in the 6-letter condition), while PM and area Spt exhibited an approximately linear increase in activity across both load and rate. No region revealed a LOAD×RATE interaction. Effect sizes (parametric generalized eta squared values, η_G^2) and parametric statistical significance for these data are displayed in Table 1.

To investigate how these effects changed through the delay, the delay period was divided into early (2–16 seconds), middle (16–30 seconds) and late (30–44 seconds) periods and the ROI analysis was repeated. For each ROI, a three-way parametric repeated-measures ANOVA was performed on the ROI extracted fMRI *t*-values with LOAD, RATE and PHASE modeled as within-subject factors and SUBJECT modeled as a random effect (Figure 8). All ROIs showed a PHASE×LOAD interaction (all ROIs: *p*<0.001), while no ROI showed a PHASE×RATE interaction (all ROIs: *p*>0.10). To further examine the PHASE×LOAD interaction, the activity estimates were collapsed over rehearsal rate and ROI. As illustrated in Figure 9, whereas activity in the 2 and 4-letter conditions increased over the delay period, activity in the 6-letter condition was constant across the delay.

4. Discussion

In this fMRI study we investigated three main questions by employing a novel WM paradigm that orthogonally manipulated rehearsal rate and memory load. The primary

question examined was which cortical regions are involved in computations related to memory load, rehearsal rate, or both, and if these cortical areas can be dissociated on the basis of these factors (addressed in sections 4.1, 4.2 and 4.3 below).

4.1 Domain-general areas: MFG, SPL

While it has been previously shown that the domain-general executive regions MFG and SPL are not required for WM tasks with low memory loads (Barbey et al., 2013; D'Esposito & Postle, 1999a; Hamidi et al., 2008; Postle et al., 2006b), the current study asked whether these regions are engaged when low memory loads are rehearsed at a fast rate. Since these regions have been generally thought of as providing a top-down bias or attention signal to more posterior cortical regions (Miller & Cohen, 2001; Miller & D'Esposito, 2005), it is possible that the added control demand of rapid rehearsal might require increased engagement from the MFG and SPL even under low loads. However, our results indicate that MFG and SPL are largely insensitive to the rate at which the task-relevant representations are reactivated with subvocal rehearsal.

As in previous studies (Cohen et al., 1997; Jonides et al., 1997; Leung et al., 2002; Rypma et al., 1999b), we observed a memory load non-linearity or threshold effect (similar activity for 2 and 4-letter loads with significantly increased activity at the highest memory load, 6-letters), consistent with the behavioral evidence showing that working capacity is limited to approximately 3 or 4 items (Cowan, 2001). Therefore, it seems when this capacity is exceeded higher-level cognitive control processes are required to maintain the full set of items in memory. This may be achieved by alternative memory reactivation mechanisms that do not depend on articulatory rehearsal, but instead rely on general reactivation processes such as "attentional refreshing" (Camos et al., 2009; Raye et al., 2002). Overall, the activity pattern observed in MFG and SPL is consistent with these regions playing a key role in cognitive control processes such as the manipulation (Champod & Petrides, 2010, 2007; D'Esposito et al., 1999b; Postle et al., 1999), attentional refreshing, or strategic reorganization (Bor et al., 2003; Druzgal & D'Esposito, 2003; Rypma et al., 1999b) of information in working memory.

4.2 Verbal rehearsal areas: IFG, PM, Spt

We also asked whether memory load effects are evident in a network of areas—the IFG, PM, and Spt—associated with articulatory rehearsal after controlling for rehearsal rate. Our results for the first time indicate that when rehearsal rate is rigorously controlled, memory load effects still exist in IFG, PM and area Spt. The load-related effects in these regions are roughly linear, however, and therefore differ from the threshold-like pattern observed in the MFG and SPL. These findings are consistent with the proposed roles for these regions, with IFG (BA 44) thought to be involved in articulatory motor planning (Awh et al., 1996; Paulesu et al., 1993), PM (BA 6) thought to play an important role in the selection of phoneme- and syllable-level representations during speech production (Guenther et al., 2006) and area Spt thought act as an auditory-motor interface (Buchsbaum et al., 2005). The effects of rehearsal rate in these regions are further explored below (in sections 4.4 and 4.5).

4.3 Memory Load & Rehearsal Rate

Several previous studies have demonstrated that memory load manipulations non-specifically activate a network of cortical areas (Rypma et al., 1999b; Zarahn et al., 2005), but it has not been clear which of these areas are involved in domain-general executive process versus rehearsal, reactivation processes, or some combination thereof. We have shown that by controlling memory load and rehearsal rate, these cortical areas can be dissociated based on their activation profiles as a function of these two variables. Specifically, we found that MFG and SPL are not sensitive to rehearsal rate but are activated at high memory loads, while IFG, PM and area Spt exhibit a roughly linear increase in activity with increasing rehearsal rate, irrespective of the memory load or how long the rehearsal has been occurring. Taken together, these results suggest that IFG, PM and area Spt comprise the core areas involved in rehearsal and that once a certain memory load threshold is exceeded MFG and SPL are also involved in the maintenance of task-relevant information.

Our findings demonstrate that memory load manipulations cannot be interpreted without also accounting for potential changes in strategic maintenance factors such as the rate of subvocal rehearsal. This may explain why previous verbal WM studies report discrepant findings of memory load effects in various cortical areas including prefrontal, inferior parietal, temporal, and motor areas (Buchsbaum et al., 2005; Narayanan et al., 2005; Postle et al., 1999; Rypma et al., 1999b; Zarahn et al., 2005). While there are many different sources of variability that may help explain these findings such as differences in the task (recall versus recognition), number of items (below, at, or above span), analysis (whole-brain versus region of interest; how delay period modeled), our results demonstrate that memory load and rehearsal rate have both overlapping and non-overlapping neural effects and therefore future WM studies involving a manipulation of memory load should consider the effects of rehearsal rate. Moreover, this finding does not only pertain to verbal WM tasks, but may apply to object and spatial visual working memory studies where the rate of internal rehearsal, or the rate of internal shifts of visual attention (Awh et al., 2006), may relate to load-sensitive activity in non-verbal domains.

4.4 Temporal effects of sustained WM rehearsal

The second question investigated was whether and how task-relevant WM representations change over time when subvocal rehearsal is maintained at a constant rate. Behavioral studies of verbal WM have provided evidence for an effortful first stage followed by a second relatively automatized stage that is less attention-demanding (Aldridge et al., 1987; Greene, 1987; Naveh-Benjamin & Jonides, 1984; Phaf & Wolters, 1993) and neuroimaging studies have correspondingly found decreasing activity over the delay period in areas maintaining task-relevant information (Chein & Fiez, 2001; Jha & McCarthy, 2000). This may suggest that rehearsed representations are "sharpened" as neurons coding for irrelevant features respond less, similar to what is believed to occur in the phenomenon of repetition suppression (Desimone, 1996; Wiggs & Martin, 1998). Surprisingly, we did not find evidence of declining activity in the core nodes of the WM network. Rather, activity in the 2 and 4-letter conditions increased over time until it was approximately equal to the constant 6-letter activity (which led to the disappearing load effects over time). These results diverge

from what Jha and McCarthy (2000) found in a study employing long delay periods where diminishing load effects were due to decreases in high memory loads. One reason for the discrepancy might be controlling rehearsal rate. In the Jha and McCarthy study participants may have slowed or stopped their rehearsal of the high memory load information as the long delay periods progressed, bringing activity down to the level of low memory loads.

Our results do not demonstrate that activity in low memory loads remained constant, but that activity increased over the delay period. It may be that as the time into the delay period progresses subjects need to increase attention to the task as it becomes more difficult due to fatigue or other factors. These attentional effects may increase activity levels in the 2 and 4-letter conditions but not in the 6-letter condition, where activity levels may already be at a neuronal or hemodynamic ceiling. Relative to other published studies we are unaware of any WM study employing long delay periods (so that the relatively slow fMRI BOLD signal can show reliable increases or decreases over time) that demonstrate an increase in activity during sustained subvocal rehearsal. Therefore, when ensuring a constant rate of rehearsal across a memory delay period, we found no evidence for decreasing levels of activity and therefore no evidence for neural sharpening.

4.5 Effect of Rehearsal Rate

The third question was how activity in regions involved in the maintenance of task-relevant information changes as information is rehearsed or reactivated at increasing rates. It is conceivable that faster rehearsal rates could lead to synaptic facilitation, which is thought to be the result of repeated neural firing leading to an increased accumulation of residual calcium at the presynaptic terminal, that would then increase the probability of subsequent neurotransmitter release (Zucker & Regehr, 2002). Accordingly, WM modeling studies have found that models incorporating synaptic facilitation were able to maintain more items (Rolls et al., 2013) and better remember two items within the same neurons (Deco et al., 2010) than models that did not include synaptic facilitation. Our results demonstrate that activity in the critical rehearsal areas PM and area Spt was approximately linear, indicating a fixed cost for each rehearsed item, irrespective of the overall rate of rehearsal. Thus, there does not seem to be an increase in neural efficiency as more items are rehearsed per unit time. Of course, because our paradigm mandates a certain rate of rehearsal it may impede the acquisition of chunked representations that may form the basis of neural efficiency gains associated with short-term synaptic plasticity (Wickelgren, 1979).

We should also note that although we did not find any evidence for synaptic facilitation at high rehearsal rates (which would be demonstrated by a non-linear activity increase), it is not clear how this phenomenon would manifest in the fMRI BOLD signal, which seems to reflect both spiking activity and local field potentials (Logothetis et al., 2001). Finally, our findings are similar to results from studies of overt and covert speech which varied the rehearsal rate anywhere from 1–6Hz and found various motor/premotor areas that scaled their BOLD activity linearly with respect to rehearsal rate (Riecker et al., 2006, 2005; Shergill et al., 2002; Wildgruber et al., 2001). While these speech production studies did not involve attention or memory (subjects simply say "pa" or "ta" at different rates), the fact that these studies and the current study both found linear increases in activity with increasing

rehearsal rate suggest the same neural mechanisms might be engaged in speech production and verbal WM (Acheson et al., 2011; Acheson & MacDonald, 2009).

Acknowledgements

References

Acheson DJ, Hamidi M, Binder JR, Postle BR. A common neural substrate for language production and verbal working memory. Journal of Cognitive Neuroscience. 2011; 23(6):1358–1367. [PubMed: 20617889]

Acheson DJ, MacDonald MC. Verbal working memory and language production: common approaches to the serial ordering of verbal information. Psychological Bulletin. 2009; 135(1):50–68. [PubMed: 19210053]

Aldridge JW, Garcia HR, Mena G. Habituation as a necessary condition for maintenance rehearsal. Journal of Memory and Language. 1987; 26(6):632–637.

Awh E, Armstrong KM, Moore T. Visual and oculomotor selection: links, causes and implications for spatial attention. Trends in Cognitive Sciences. 2006; 10(3):124–130. [PubMed: 16469523]

Awh E, Jonides J, Smith EE, Schumacher EH, Koeppe RA, Katz S. Dissociation of storage and rehearsal in verbal working memory: Evidence from positron emission tomography. Psychological Science. 1996; 7:25–31.

Baddeley, AD. Working memory. Oxford, UK: Oxford University Press; 1986.

Badre D, Wagner AD. Left ventrolateral prefrontal cortex and the cognitive control of memory. Neuropsychologia. 2007; 45(13):2883–2901. [PubMed: 17675110]

Barbey AK, Koenigs M, Grafman J. Dorsolateral prefrontal contributions to human working memory. Cortex. 2013; 49(5):1195–1205. [PubMed: 22789779]

Bor D, Duncan J, Wiseman RJ, Owen AM. Encoding strategies dissociate prefrontal activity from working memory demand. Neuron. 2003; 37(2):361–367. [PubMed: 12546829]

Buchsbaum B, D'Esposito M. The search for the phonological store: from loop to convolution. Journal of Cognitive Neuroscience. 2008; 20(5):762–778. [PubMed: 18201133]

Buchsbaum B, Olsen R, Koch P, Berman K. Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. Neuron. 2005; 48(4):687–697. [PubMed: 16301183]

Camos V, Lagner P, Barrouillet P. Two maintenance mechanisms of verbal information in working memory. Journal of Memory and Language. 2009; 61(3):457–469.

Champod AS, Petrides M. Dissociation within the frontoparietal network in verbal working memory: A parametric functional magnetic resonance imaging study. Journal of Neuroscience. 2010; 30(10):3849–3856. [PubMed: 20220020]

Champod AS, Petrides M. Dissociable roles of the posterior parietal and the prefrontal cortex in manipulation and monitoring processes. Proc Natl Acad Sci USA. 2007; 104(37):14837–14842. [PubMed: 17804811]

Chein J, Fiez J. Dissociation of verbal working memory system components using a delayed serial recall task. Cerebral Cortex. 2001; 11(11):1003–1014. [PubMed: 11590110]

Cohen JD, Perlstein WM, Braver TS, Nystrom LE, Noll DC, Jonides J, Smith EE. Temporal dynamics of brain activation during a working memory task. Nature. 1997; 386(6625):604–608. [PubMed: 9121583]

Cowan N. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. Behav Brain Sci. 2001; 24(1):87–114. discussion 114–185. [PubMed: 11515286]

Cowan N, Wood NL, Wood PK, Keller TA, Nugent LD, Keller CV. Two separate verbal processing rates contributing to short-term memory span. Journal of Experimental Psychology General. 1998; 127(2):141–160. [PubMed: 9622911]

Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Computers and Biomedical Research. 1996; 29(3):162–173. [PubMed: 8812068]

Curtis CE, D'Esposito M. Persistent activity in the prefrontal cortex during working memory. Trends in Cognitive Sciences. 2003; 7(9):415–423. [PubMed: 12963473]

D'Esposito M. From cognitive to neural models of working memory. Philosophical Transactions of the Royal Society B: Biological Sciences. 2007; 362(1481):761.

D'Esposito M, Postle BR. The dependence of span and delayed-response performance on prefrontal cortex. Neuropsychologia. 1999a; 37(11):1303–1315. [PubMed: 10530730]

D'Esposito M, Postle BR, Ballard D, Lease J. Maintenance versus manipulation of information held in working memory: an event-related fMRI study. Brain Cogn. 1999b; 41(1):66–86. [PubMed: 10536086]

Daneman M, Merikle PM. Working memory and language comprehension: a meta-analysis. Psychonomic Bulletin & Review. 1996; 3(4):422–433. [PubMed: 24213976]

Dasí C, Soler M, Cervera T, Ruiz J. Influence of articulation rate on two memory tasks in young and older adults. Percept Mot Skills. 2008; 106(2):579–589. [PubMed: 18556911]

Deco G, Rolls ET, Romo R. Synaptic dynamics and decision making. Proc Natl Acad Sci USA. 2010; 107(16):7545–7549. [PubMed: 20360555]

Desimone R. Neural mechanisms for visual memory and their role in attention. Proc Natl Acad Sci USA. 1996; 93(24):13494–13499. [PubMed: 8942962]

Druzgal TJ, D'Esposito M. Dissecting contributions of prefrontal cortex and fusiform face area to face working memory. Journal of Cognitive Neuroscience. 2003; 15(6):771–784. [PubMed: 14511531]

Feredoes E, Postle BR. Localization of load sensitivity or working memory storage: quantitatively and qualitatively discrepant results yielded by single-subject and group-averaged approaches to fMRI group analysis. Neuroimage. 2007; 35(2):881–903. [PubMed: 17296315]

Frey S, Campbell JSW, Pike GB, Petrides M. Dissociating the human language pathways with high angular resolution diffusion fiber tractography. Journal of Neuroscience. 2008; 28(45):11435–11444. [PubMed: 18987180]

Genovese CR, Lazar NA, Nichols T. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage. 2002; 15(4):870–878. [PubMed: 11906227]

Greene RL. Effects of maintenance rehearsal on human memory. Psychological Bulletin. 1987; 102(3):403–413.

Guenther FH, Ghosh SS, Tourville JA. Neural modeling and imaging of the cortical interactions underlying syllable production. Brain Lang. 2006; 96(3):280–301. [PubMed: 16040108]

Hamidi M, Tononi G, Postle BR. Evaluating frontal and parietal contributions to spatial working memory with repetitive transcranial magnetic stimulation. Brain Res. 2008; 1230:202–210. [PubMed: 18662678]

Hickok G, Buchsbaum B, Humphries C, Muftuler T. Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. Journal of Cognitive Neuroscience. 2003; 15(5):673–82. [PubMed: 12965041]

Hulme C, Thomson N, Muir C, Lawrence A. Speech rate and the development of short-term memory span. Journal of Experimental Child Psychology. 1984; 38(2):241–253.

Jha AP, McCarthy G. The influence of memory load upon delay-interval activity in a working-memory task: an event-related functional MRI study. Journal of Cognitive Neuroscience. 2000; 12(S2):90–105. [PubMed: 11506650]

Jonides J, Schumacher EH, Smith EE, Lauber EJ, Awh E, Minoshima S, Koeppe RA. Verbal working memory load affects regional brain activation as measured by PET. Journal of Cognitive Neuroscience. 1997; 9(4):462–475. [PubMed: 23968211]

Kyllonen PC, Christal RE. Reasoning ability is (little more than) working-memory capacity?! Intelligence. 1990; 14(4):389–433.

Leung H-C, Gore JC, Goldman-Rakic PS. Sustained mnemonic response in the human middle frontal gyrus during on-line storage of spatial memoranda. Journal of Cognitive Neuroscience. 2002; 14(4):659–671. [PubMed: 12126506]

Logie RH, Gilhooly KJ, Wynn V. Counting on working memory in arithmetic problem solving. Mem Cognit. 1994; 22(4):395–410.

Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A. Neurophysiological investigation of the basis of the fMRI signal. Nature. 2001; 412(6843):150–157. [PubMed: 11449264]

Miller E, Cohen J. An integrative theory of prefrontal cortex function. Annual Review of Neuroscience. 2001; 24:167–202.

Miller BT, D'Esposito M. Searching for "the top" in top-down control. Neuron. 2005; 48(4):535–538. [PubMed: 16301170]

Narayanan NS, Prabhakaran V, Bunge SA, Christoff K, Fine EM, Gabrieli JDE. The role of the prefrontal cortex in the maintenance of verbal working memory: an event-related fMRI analysis. Neuropsychology. 2005; 19(2):223–232. [PubMed: 15769206]

Naveh-Benjamin M, Jonides J. Maintenance rehearsal: A two-component analysis. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1984; 10(3):369–385.

Paulesu E, Frith CD, Frackowiak RS. The neural correlates of the verbal component of working memory. Nature. 1993; 362(6418):342–345. [PubMed: 8455719]

Phaf RH, Wolters G. Attentional shifts in maintenance rehearsal. The American Journal of Psychology. 1993; 106(3):353–382.

Postle BR. Working memory as an emergent property of the mind and brain. Neuroscience. 2006a; 139(1):23–38. [PubMed: 16324795]

Postle BR, Berger JS, D'Esposito M. Functional neuroanatomical double dissociation of mnemonic and executive control processes contributing to working memory performance. Proc Natl Acad Sci USA. 1999; 96(22):12959–12964. [PubMed: 10536030]

Postle BR, Ferrarelli F, Hamidi M, Feredoes E, Massimini M, Peterson M, Alexander A, Tononi G. Repetitive transcranial magnetic stimulation dissociates working memory manipulation from retention functions in the prefrontal, but not posterior parietal, cortex. Journal of Cognitive Neuroscience. 2006b; 18(10):1712–1722. [PubMed: 17014375]

Raye C, Johnson M, Mitchell K, Reeder J. Neuroimaging a single thought: dorsolateral PFC activity associated with refreshing just-activated information. Neuroimage. 2002; 15(2):447–453. [PubMed: 11798278]

Riecker A, Kassubek J, Gröschel K, Grodd W, Ackermann H. The cerebral control of speech tempo: opposite relationship between speaking rate and BOLD signal changes at striatal and cerebellar structures. Neuroimage. 2006; 29(1):46–53. [PubMed: 16085428]

Riecker A, Mathiak K, Wildgruber D, Erb M, Hertrich I, Grodd W, Ackermann H. fMRI reveals two distinct cerebral networks subserving speech motor control. Neurology. 2005; 64(4):700–706. [PubMed: 15728295]

Rolls ET, Dempere-Marco L, Deco G. Holding multiple items in short term memory: a neural mechanism. PLoS ONE. 2013; 8(4)

Rypma B, D'Esposito M. The roles of prefrontal brain regions in components of working memory: effects of memory load and individual differences. Proc Natl Acad Sci USA. 1999a; 96(11):6558–6563. [PubMed: 10339627]

Rypma B, Prabhakaran V, Desmond JE, Glover GH, Gabrieli JD. Load-dependent roles of frontal brain regions in the maintenance of working memory. Neuroimage. 1999b; 9(2):216–226. [PubMed: 9927550]

Rypma B, Berger JS, D'Esposito M. The influence of working-memory demand and subject performance on prefrontal cortical activity. Journal of Cognitive Neuroscience. 2002; 14(5):721–731. [PubMed: 12167257]

Shergill SS, Brammer MJ, Fukuda R, Bullmore E, Amaro E, Murray RM, McGuire PK. Modulation of activity in temporal cortex during generation of inner speech. Hum Brain Mapp. 2002; 16(4):219–227. [PubMed: 12112764]

Thirion B, Pinel P, Mériaux S, Roche A, Dehaene S, Poline JB. Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. Neuroimage. 2007; 35(1):105–120. [PubMed: 17239619]

Todd JJ, Marois R. Posterior parietal cortex activity predicts individual differences in visual short-term memory capacity. Cogn Affect Behav Neurosci. 2005; 5(2):144–155. [PubMed: 16180621]

Wagner AD, Maril A, Bjork RA, Schacter DL. Prefrontal contributions to executive control: fMRI evidence for functional distinctions within lateral prefrontal cortex. Neuroimage. 2001; 14(6): 1337–1347. [PubMed: 11707089]

Wickelgren WA. Chunking and consolidation: a theoretical synthesis of semantic networks, configuring in conditioning, S–R versus cognitive learning, normal forgetting, the amnesic syndrome, and the hippocampal arousal system. Psychol Rev. 1979; 86(1):44–60. [PubMed: 451113]

Wiggs CL, Martin A. Properties and mechanisms of perceptual priming. Current Opinion in Neurobiology. 1998; 8(2):227–233. [PubMed: 9635206]

Wildgruber D, Ackermann H, Grodd W. Differential contributions of motor cortex, basal ganglia, and cerebellum to speech motor control: effects of syllable repetition rate evaluated by fMRI. Neuroimage. 2001; 13(1):101–109. [PubMed: 11133313]

Wildgruber D, Kischka U, Ackermann H, Klose U, Grodd W. Dynamic pattern of brain activation during sequencing of word strings evaluated by fMRI. Brain Res Cogn Brain Res. 1999; 7(3):285–294. [PubMed: 9838166]

Zarahn E, Rakitin B, Abela D, Flynn J, Stern Y. Positive evidence against human hippocampal involvement in working memory maintenance of familiar stimuli. Cereb Cortex. 2005; 15(3):303–316. [PubMed: 15342440]

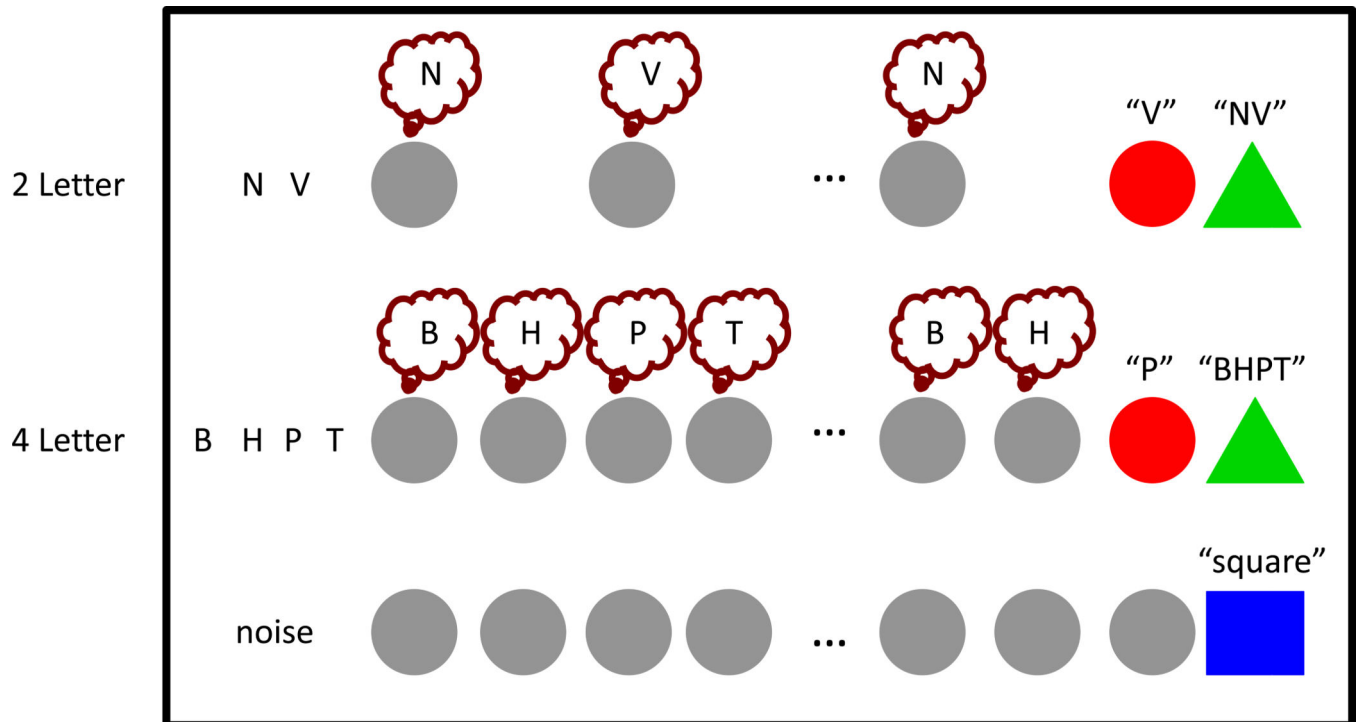Zucker RS, Regehr WG. Short-term synaptic plasticity. Annu. Rev. Physiol. 2002; 64:355–405. [PubMed: 11826273]

Figure 1. fMRI task design
Subjects (*N*=24) were presented with either 2, 4, or 6 letters that they were instructed to maintain over a 45-second delay period. During the delay period a repeating visual pacing cue prompted subjects to subvocally rehearse one letter at one of seven rates: 0.8, 1.1, 1.4, 1.7, 2.0, 2.3 or 2.6Hz. After the delay period subjects were first prompted by an order probe (red circle) to verbally report which letter they were on, followed by a recall probe (green triangle), to verbally report the original list of letters, in order. For control trials, subjects were instructed to listen to noise (time-reversed letters) during encoding before attending to the visual pacing cues without subvocally rehearsing, and verbally saying "square" as soon as they saw the blue square probe appear.
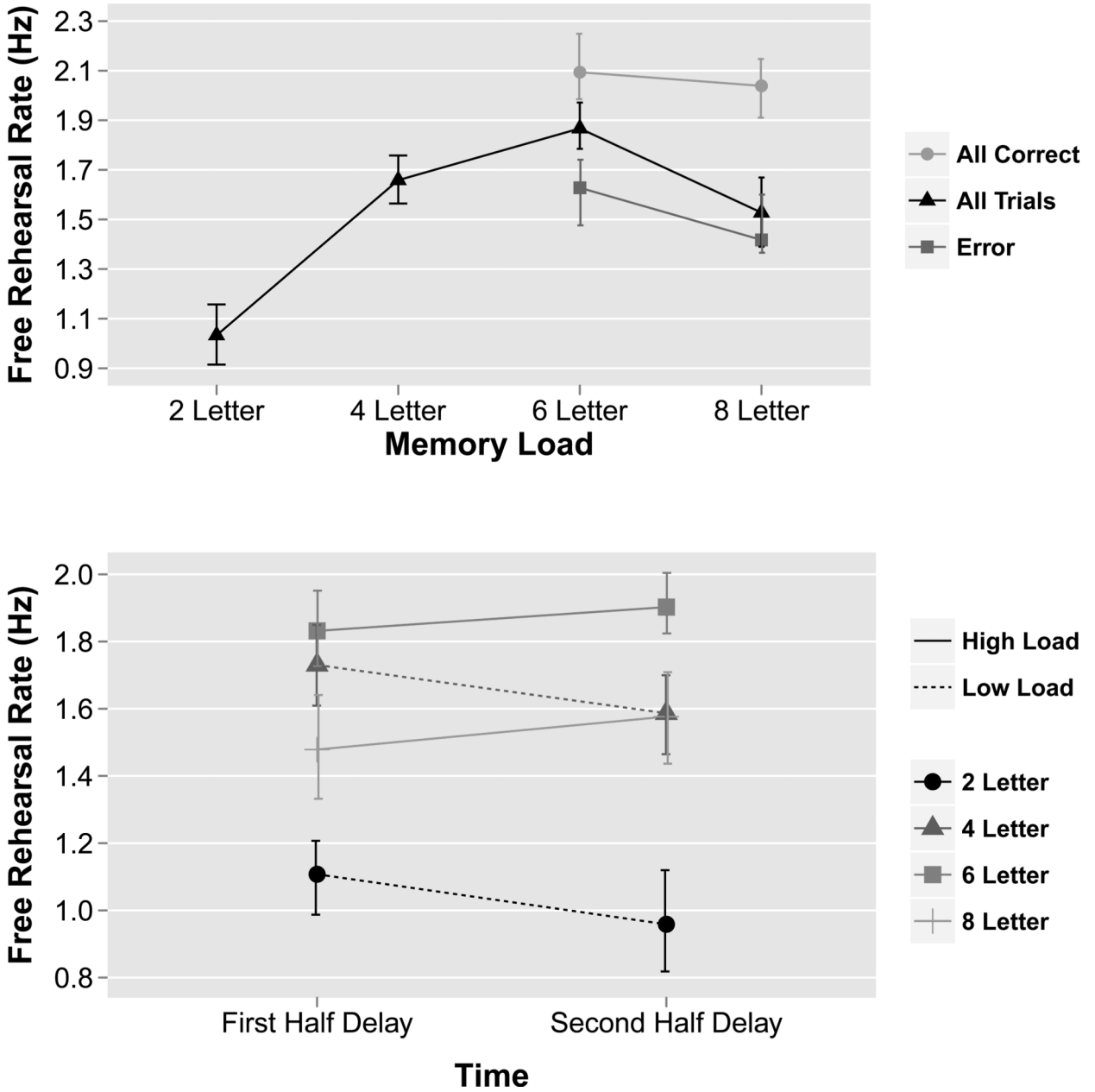
Figure 2. Overt free rehearsal behavioral data

Prior to the fMRI experiment, subjects overtly rehearsed either 2, 4, 6, or 8 letters over a 15-second delay period in a working memory task. (a) For each memory load, mean rehearsal rate (number of letters/second) is plotted. In general, as the memory load increased subjects rehearsed at a faster rate until the memory load exceeded 6 letters, at which point rehearsal rate decreased. (b) Rehearsal rate in the first half of the delay period versus the second half of the delay period for each memory load. Results were significant for a LOAD×TIME interaction (permutation ANOVA: $p<0.001$). In general, subjects rehearsed faster in the first half of the delay period for low memory loads (2 and 4 letters), but faster in the second half of the delay period for high memory loads (6 and 8 letters). Data plotted are group means.

Error bars represent bootstrapped within-subject 95% confidence intervals. Full results of statistical tests are included in the main text (section 3.1).
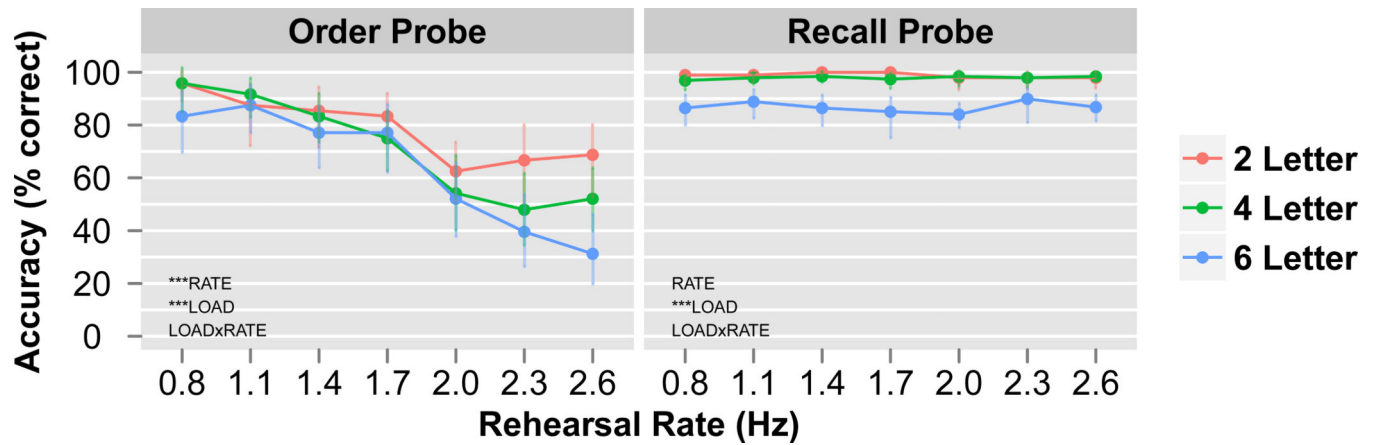
Figure 3. fMRI behavioral results

Order probe (red circle) accuracy on left, recall probe (green triangle) accuracy on right. Permutation ANOVA results demonstrate that the order probe was significant for RATE, with a large decrease in accuracy between 1.7 and 2.0 Hz (left). Conversely, the recall probe was not significant for RATE (right). However, both probes showed a memory LOAD effect as accuracy decreased from 2, to 4, to 6 letters. There was no LOAD×RATE interaction for either probe. Data plotted are group means. Error bars represent bootstrapped within-subject 95% confidence intervals. ***$p<0.0001$.
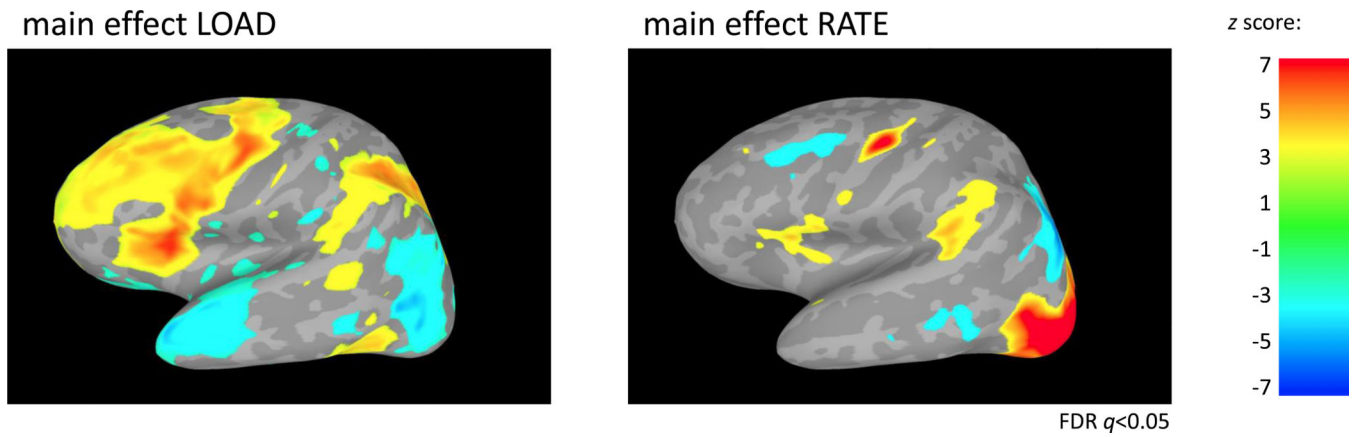
Figure 4. fMRI ANOVA results

Activity estimates (*t*-values) for each condition for 43 seconds of the delay period were entered into a two-way parametric repeated measures ANOVA (LOAD, RATE). A positive main effect of memory LOAD can be visualized across many regions implicated in verbal working memory including MFG, IFG, PM, SPL and area Spt. In contrast, a positive main effect of rehearsal RATE is mainly restricted to IFG, PM, area Spt and occipital lobe regions. No region demonstrated a LOAD×RATE interaction. Data from left hemisphere; data shown are *z*-scores thresholded at FDR *q*<0.05.
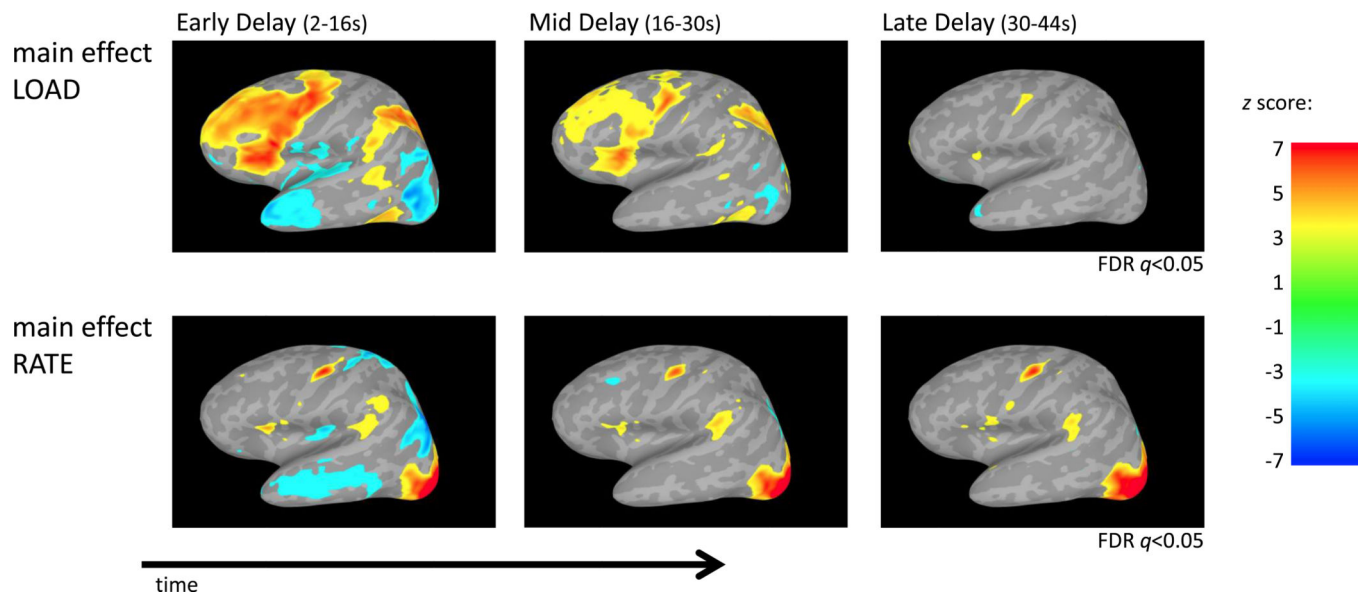
Figure 5. fMRI ANOVA results across time

Activity estimates (*t*-values) for each condition for the early, mid and late delay period time phases were separately entered into a two-way parametric repeated measures ANOVA (LOAD, RATE). Results show that the main effect of memory LOAD declined through time. Subsequent analysis reveals this is the result of activity in the high memory load remaining constant, while activity in the low memory loads increases over time (Figure 9). In contrast, the main effect of RATE was primarily restricted to IFG, PM and area Spt (and occipital lobe) and were constant through the delay period without diminishing. No region demonstrated a LOAD×RATE interaction during any part of the delay period. Data from left hemisphere; data shown are *z*-scores thresholded at FDR *q*<0.05 for the three time periods in load and rate separately in order to compare changes across time.
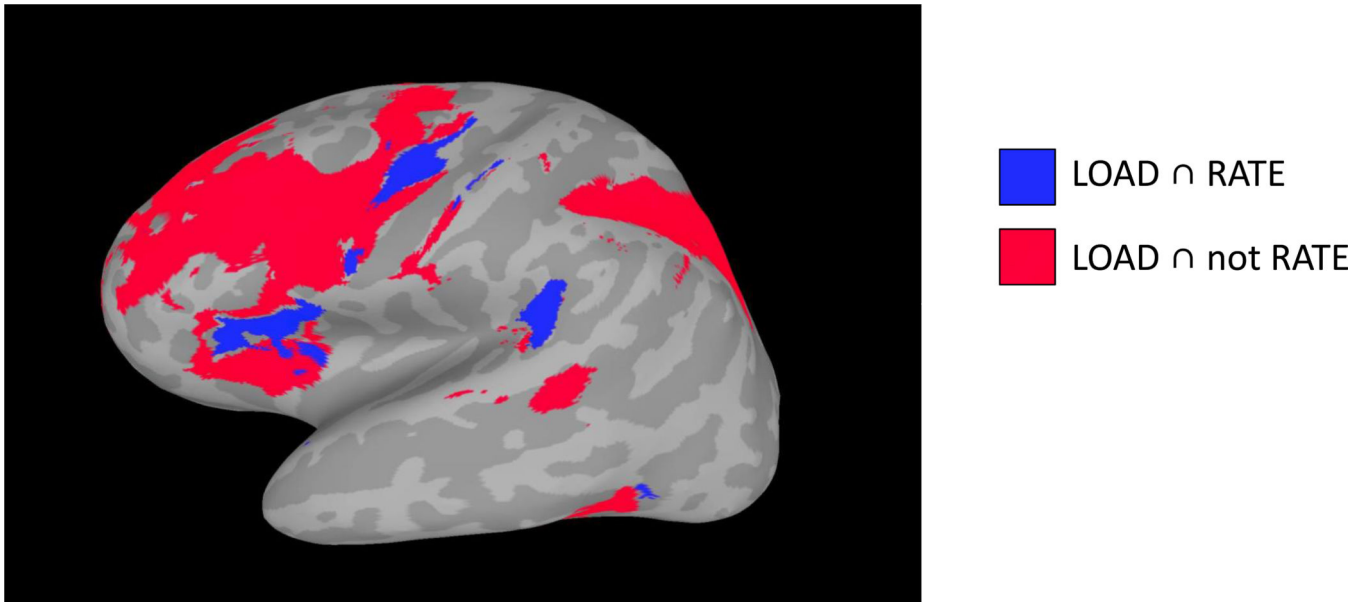
Figure 6. Conjunction analysis
Parametric linear trend analysis for LOAD and RATE were independently thresholded at the group level and two separate conjunction analyses were performed: [LOAD$_{(FDR\ q<0.05)}$] ∩ [RATE$_{(FDR\ q<0.05)}$], and [LOAD$_{(FDR\ q<0.05)}$] ∩ [RATE$_{(p>0.95)}$]. Separate conjunction maps (blue, red) were then combined into one figure for visualization (as shown above). Results show that IFG, PM and area Spt were sensitive to LOAD and RATE, while primarily MFG and SPL were responsive to LOAD but not RATE. Data from left hemisphere.
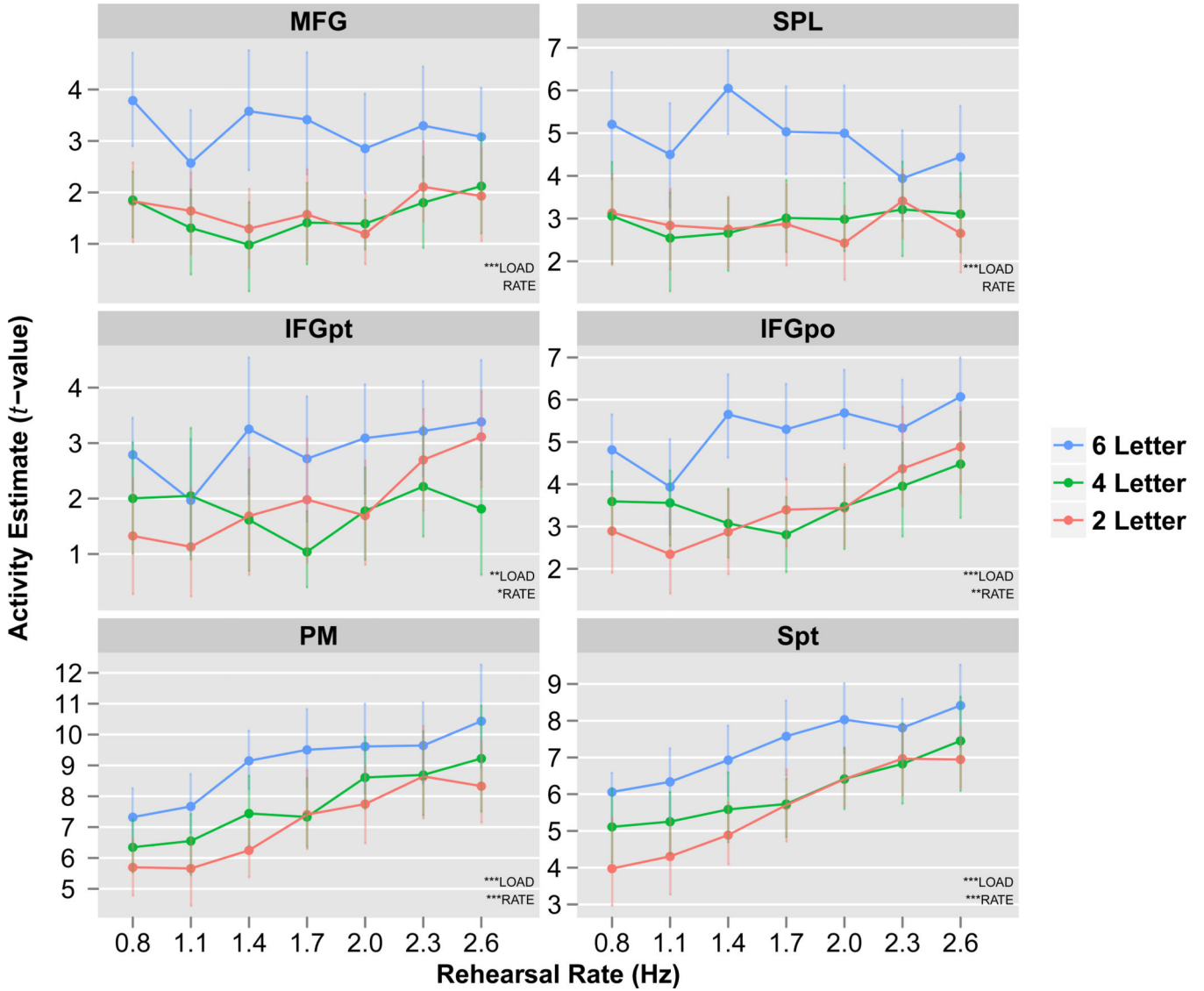
Figure 7. ROI analysis of key verbal working memory regions
Activity estimates (*t*-values) plotted for 43 seconds of the delay period for each rehearsal rate (x-axis) and memory load (different colors). Parametric ANOVA results show that MFG and SPL are significant for LOAD but not RATE, and that memory load effects in these regions appear non-linear. In contrast, IFGpt, IFGpo, PM and area Spt are all significant for LOAD and RATE. Furthermore, especially within PM and area Spt, the effects of load and rate appear to be approximately linear. There was no significant LOAD×RATE interaction within any region. Data plotted are group means. Error bars represent bootstrapped within-subject 95% confidence intervals. ***$p<0.0001$, **$p<0.001$, *$p<0.05$. See Table 1 for corresponding effect sizes.
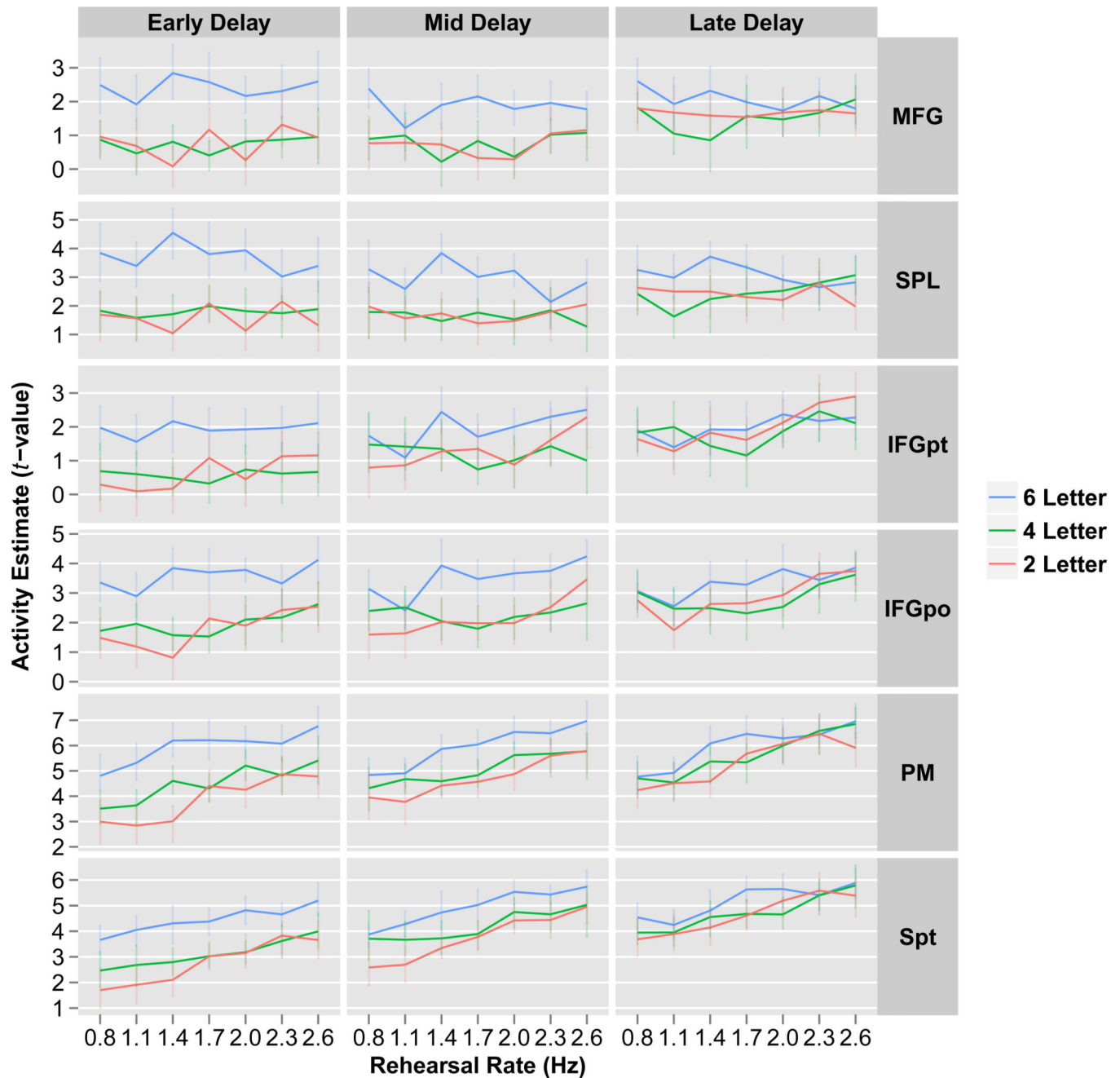
Figure 8. ROI analysis of key verbal working memory regions through time
Activity estimates (*t*-values) plotted for the early, mid and late delay period time phases (columns) for each rehearsal rate (x-axis) and memory load (different colors). Results demonstrate a reduction in load effects through time as all regions revealed a PHASE×LOAD interaction (all ROIs, parametric ANOVA: *p*<0.001). Conversely, rehearsal rate effects remained constant through time as no region demonstrated a PHASE×RATE interaction (all ROIs, parametric ANOVA: *p*>0.10). There was no LOAD×RATE interaction. Data plotted are group means. Error bars represent bootstrapped within-subject 95% confidence intervals.
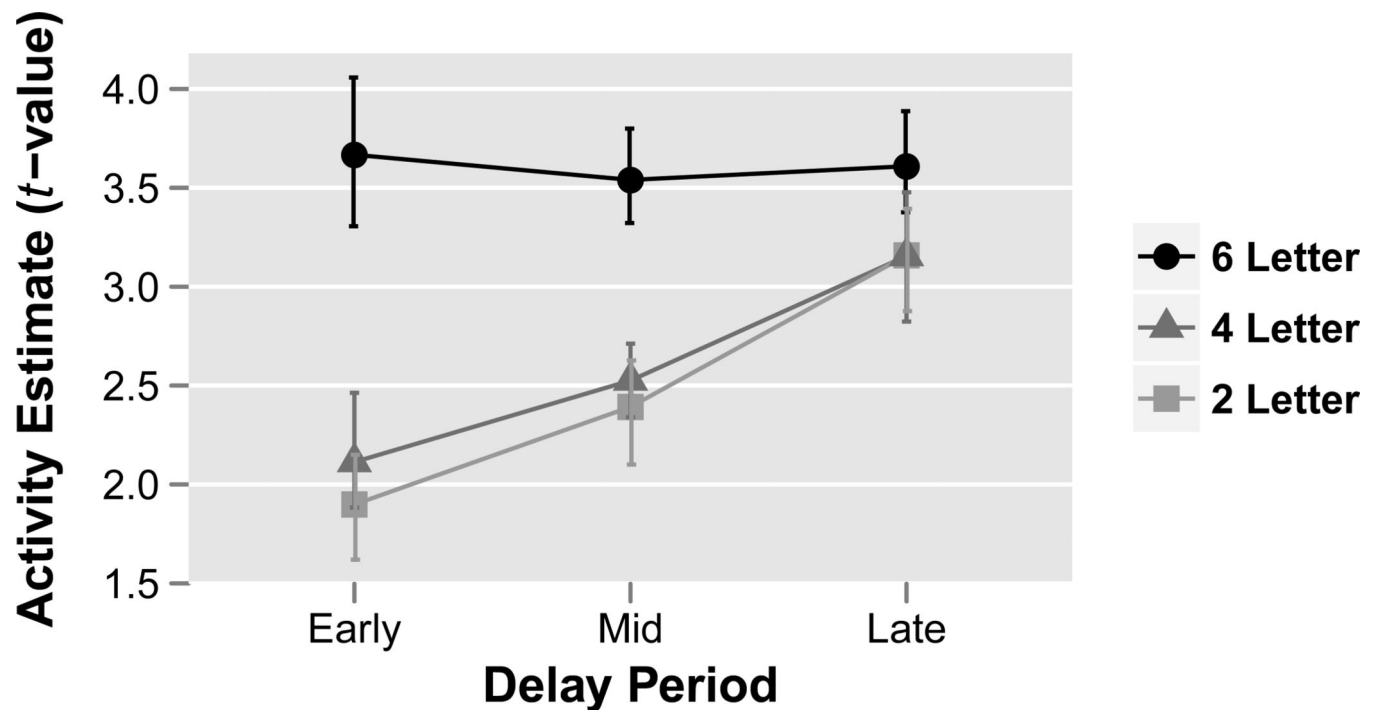
Figure 9. Average activity across all ROIs through time
Activity estimates (*t*-values) from Fig. 8 were averaged over rehearsal rate and ROI to investigate the PHASE×LOAD interaction across early, mid and late delay period time phases. Results reveal the dissipating load effects are due to activity in 6-letter condition remaining constant through the delay period while activity in the 2 and 4-letter conditions increase through time. Data plotted are group means. Error bars represent bootstrapped within-subject 95% confidence intervals.

Table 1

Effect sizes (parametric generalized eta squared values, η_G^2) and parametric statistical significance from ROI analysis presented in Fig. 7.

Condition	Region of Interest					
	MFG	SPL	IFGpt	IFGpo	PM	Spt
LOAD$_{(df=2,46)}$	.145***	.158***	.059**	.133***	.129***	.131***
RATE$_{(df=6,138)}$	.019	.007	.032*	.067**	.198***	.191***
LOAD×RATE$_{(df=12,276)}$	.016	.030	.030	.030	.016	.016

*
$p<0.05,$

**
$p<0.001,$

***
$p<0.0001$