A Common Neural Substrate for Language Production and Verbal Working Memory

Daniel J. Acheson[1], **Massihullah Hamidi**[1], **Jeffrey R. Binder**[2], and **Bradley R. Postle**[1]
[1]University of Wisconsin-Madison

[2]Medical College of Wisconsin

Abstract

Verbal working memory (VWM), the ability to maintain and manipulate representations of speech sounds over short periods, is held by some influential models to be independent from the systems responsible for language production and comprehension [e.g., Baddeley, A. D. *Working memory, thought, and action.* New York, NY: Oxford University Press, 2007]. We explore the alternative hypothesis that maintenance in VWM is subserved by temporary activation of the language production system [Acheson, D. J., & MacDonald, M. C. Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin, 135*, 50–68, 2009b]. Specifically, we hypothesized that for stimuli lacking a semantic representation (e.g., nonwords such as *mun*), maintenance in VWM can be achieved by cycling information back and forth between the stages of phonological encoding and articulatory planning. First, fMRI was used to identify regions associated with two different stages of language production planning: the posterior superior temporal gyrus (pSTG) for phonological encoding (critical for VWM of nonwords) and the middle temporal gyrus (MTG) for lexical–semantic retrieval (not critical for VWM of nonwords). Next, in the same subjects, these regions were targeted with repetitive transcranial magnetic stimulation (rTMS) during language production and VWM task performance. Results showed that rTMS to the pSTG, but not the MTG, increased error rates on paced reading (a language production task) and on delayed serial recall of nonwords (a test of VWM). Performance on a lexical–semantic retrieval task (picture naming), in contrast, was significantly sensitive to rTMS of the MTG. Because rTMS was guided by language production-related activity, these results provide the first causal evidence that maintenance in VWM directly depends on the long-term representations and processes used in speech production.

Introduction

The construct of working memory (WM) is one of the most extensively studied in cognitive psychology, in large part because performance on WM tasks correlates with many complex behaviors, such as language comprehension and development, reasoning and problem solving, and performance on intelligence tests (Baddeley, 2007). The modal view of WM is that maintenance is achieved via specialized storage buffers that are independent of long-term memory (LTM; Atkinson & Shiffrin, 1968). This perspective has its roots in the mind-as-computer metaphor, with short-term buffers akin to RAM and long-term storage to the hard drives of modern computers. The multicomponent model, arguably the most influential cognitive model of WM for the past 35 years, holds that verbal information is maintained for short periods by a specialized phonological store whose contents decay over time unless

refreshed by a process of subvocal articulation (i.e., the phonological loop; Baddeley, 2007). The distinction between short-term memory and LTM is maintained within this model, in that representations within the phonological store are thought to be independent of those responsible for language comprehension and production. The influence of this model is such that some of the earliest cognitive neuroimaging studies were specifically designed to find the neural basis of this WM buffer (e.g., Awh & Jonides, 2001; Paulesu, Frith, & Frackowiak, 1993), and subsequently, a growing consensus has developed that inferior parietal cortex in the left hemisphere is the neural basis of the phonological store (reviewed in Buchsbaum & D'Esposito, 2008). More recently, however, these cognitive neuroscience extensions of the multicomponent model have been questioned, as have some of the core assumptions of the model itself.

The association of inferior parietal cortex with the phonological store has been challenged on the basis that it does not demonstrate one of the most fundamental properties of the phonological loop model. Specifically, auditorily presented information is said to have direct access to the phonological store, whereas visually presented information must first be recoded into a phonological form (Baddeley, 2007). The inferior parietal regions identified by most neuroimaging studies of verbal working memory (VWM), however, are located superior to the Sylvian fissure, and are not activated by passive listening (Buchsbaum & D'Esposito, 2008), nor are they sensitive to phonological similarity among to-be-remembered items (Chein & Fiez, 2001), a manipulation that is thought to induce interference within the phonological store. From a broader perspective, many neuroimaging studies of VWM are susceptible to the criticism that they amount to an enterprise in which the reality of a cognitive construct (in this case, a WM buffer called the phonological store) is assumed to be true, and the only goal of the experiment is to identity where in the brain this cognitive construct is implemented (e.g., van Eijsden, Hyder, Rothman, & Shulman, 2009; Buzsaki, 2006).

With regard to the multicomponent model itself, one of the core assumptions of the model is the independence of short- and long-term representations. Within the verbal domain, however, many studies have shown that long-term, linguistic representation affects VWM performance (Walker & Hulme, 1999; Roodenrys, Hulme, Alban, Ellis, & Brown, 1994; Tehan & Humphreys, 1988; Watkins & Watkins, 1977; Crowder, 1976). Furthermore, patient populations with damage to language comprehension and production often exhibit deficits in VWM task performance (Martin & Saffran, 1997). There is thus considerable evidence from experimental psychology, psycholinguistics, computational modeling, and neuropsychology that is consistent with the idea that interactions between multiple levels of linguistic representation support VWM function (Acheson & MacDonald, 2009b; Martin & Saffran, 1997).

An alternative to the idea that WM is supported by specialized storage buffers is the idea that WM functions might emerge from the temporary activation of neural systems that have evolved to support functions of perception, action, and representation that are not specific to WM (Postle, 2006; Ruchkin, Grafman, Cameron, & Berndt, 2003). This hypothesis is consistent with cognitive models of WM in which maintenance is achieved via the focusing of attention onto different components of LTM (e.g., Oberauer, 2002; Cowan, 1995), and also with the idea that short-term maintenance is achieved via "reverberatory activity" with neural circuits whose synapses also encode and represent LTM (Hebb, 1949). Indeed, many neuroimaging studies have indicated that the same brain regions associated with long-term phonological storage also support maintenance in VWM task performance (Leff et al., 2009; Buchsbaum & D'Esposito, 2008).

Although it has long been acknowledged that phonological representations are maintained in VWM, the present study was motivated by a more specific hypothesis: that the short-term retention of verbal information ("VWM") is accomplished via temporary activation of representations within the language production architecture. Language production is a skill acquired over a lifetime of experience, hence, the representations and processes involved necessarily reflect long-term learning (i.e., LTM). Language production occurs via a series of stages, beginning with the formulation of a message, followed by lexical retrieval (mapping an abstract, conceptual message onto words), phonological encoding (specifying the speech sounds that comprise the words), articulatory planning (formulation of a sequential motor plan for speech output), and articulation (Indefrey & Levelt, 2004). Our model holds that VWM can be accomplished by cycling information back and forth across these levels of production planning. For stimuli devoid of semantic content (in our case, nonwords), it holds that VWM can be accomplished via sustained interaction between the processes of phonological encoding and articulatory planning. In essence, this account of VWM suggests that the two major components responsible for maintaining verbal information within the multicomponent model (e.g., the phonological store and articulatory rehearsal) may simply reflect a relabeling of two of the stages of language production planning (i.e., phonological encoding and articulatory planning).

In this study, we hypothesized that repetitive transcranial magnetic stimulation (rTMS) targeting a brain region that supports the process of phonological encoding would have qualitatively similar effects on nonword reading (i.e., speech production) and the short-term retention of nonwords (i.e., VWM). Confirmation of this hypothesis would overturn a core feature of the multicomponent and related models, which is the independence of the phonological store from long-term, phonological representations (i.e., the independence of WM from LTM). The reason for this is that phonological encoding necessarily entails use of LTM in that the representations and processes involved reflect the outcome of many years of language production experience and learning. Previously, when considering evidence of LTM influences on VWM such as word frequency (Roodenrys et al., 1994) and word concreteness (Walker & Hulme, 1999), proponents of the multicomponent model have suggested that LTM influences only occurs when degraded short-term traces are compared to LTM at the time of retrieval (i.e., trace redintegration; Hulme, Roodenrys, Schweickert, & Brown, 1997; Schweickert, 1993). Thus, the theoretical claim of our model—that maintenance in VWM entails the activation of long-term, speech-based representations— invalidates a central tenet of the multicomponent model, and would represent more evidence in support of the idea that WM reflects the temporary activation of LTM representations (Lewis-Peacock & Postle, 2008; Postle, 2006; Ruchkin et al., 2003; Oberauer, 2002; Cowan, 1995).

Our approach took advantage of research showing that phonological encoding is dissociable behaviorally, temporally and anatomically from lexical–semantic retrieval (Wilson, Isenberg, & Hickok, 2009; Indefrey & Levelt, 2004). The experiment proceeded in two stages. First, regions of the brain associated with lexical–semantic retrieval and with phonological encoding were identified by having subjects perform overt picture naming and nonword reading during functional magnetic resonance imaging (fMRI). Nonwords represent legal combinations of speech sounds without an associated meaning (e.g., *mun*). Although many brain areas would be expected to show differential activation during these behaviors, we planned a priori to select voxels in the posterior superior temporal gyrus (pSTG) and in the middle temporal gyrus (MTG), on the basis of their previous association with the phonological and lexical–semantic retrieval of words, respectively (Binder, Desai, Graves, & Conant, 2009; Wilson et al., 2009; Indefrey & Levelt, 2004). Second, these brain regions were targeted with rTMS while subjects performed speech production tasks that either did or did not require accessing lexical–semantic representations (picture naming and

paced reading of nonwords, respectively), as well as a test of VWM (delayed serial recall of nonwords).

In the context of the present investigation, the MTG served as a control region to establish the specificity of the predicted effect of rTMS of the pSTG on nonword reading and delayed serial recall. rTMS of the MTG was expected to produce comparable peripheral effects (e.g., jaw movement), but to produce neural effects that would emphasize a stage of language production planning (lexical retrieval) that was not hypothesized to be important for paced reading and delayed serial recall of nonwords. Thus, using the MTG as a control region could narrow the candidate stages of production planning that could account for the results of rTMS of the pSTG. The task that we used to operationalize lexical–semantic retrieval—picture naming—also necessarily engages phonological encoding and articulatory processes (to accomplish the overt naming component of the task). Therefore, we sought to target the earlier stage of production by controlling the timing of rTMS so as to primarily affect the lexical–semantic retrieval stage, and not the phonological encoding stage of the task (see Methods for details).

Methods

Subjects

A total of 14 individuals (7 women) with no history of psychiatric or neurological disorder participated in the study and were compensated at \$20/hour. The mean age was 24.5 (*SD* = 4.2). Subjects gave informed consent, and the experiment was approved by the University of Wisconsin institutional review board. Two subjects were excluded from analysis due to discomfort from rTMS which did not allow them to complete the TMS portion of the experiment.

fMRI

Acquisition—Magnetic resonance images were acquired on a 3-T scanner (GE Signa VH/ 1). Two sets of T1-weighted images were collected: 30 axial slices ($0.9375 \times 0.9375 \times 4$ mm) coplanar with the functional images and 248 axial slices (0.5 mm $\times 0.5$ mm $\times 0.8$ mm) that were later reconstructed into a three-dimensional image for use in targeting with rTMS. Functional images were collected using gradient-echo, echo-planar sequences (TR = 2000 msec, TE = 30 msec) to acquire data sensitive to blood oxygen level dependent (BOLD) signal within a 64×64 matrix (30 axial slices; $3.75 \times 3.75 \times 4$ mm). Each functional run lasted 7:20, including an initial 20 sec of discarded acquisitions to achieve a steady state of tissue magnetization.

Design—Stimuli were presented within a rapid event-related design, with random interstimulus intervals between 4 and 8 sec (mean of 6 sec). Each functional run was divided between two tasks, picture naming and nonword reading. Instructions were displayed for 10 sec before each task. Stimuli were presented for 2 sec, and responses were monitored using an MR-compatible microphone and through button presses indicating whether stimuli were nameable or not. This latter step was taken as there were times when subject responses were difficult to monitor due to scanner noise. A fixation cross remained on the display at all other times.

Tasks

Picture naming: Subjects saw a total of 200 black and white pictures, half of which were nameable objects (Rossion & Pourtois, 2004) and half visual control images that were scrambled versions of the nameable picture. Subjects were instructed to overtly name the object in the picture if they could, or say the word "nothing" if they could not. As an

additional means of monitoring performance, subjects pressed a button after presentation of each stimulus to indicate whether they had been able to generate a name or not. Each functional run consisted of 50 items, half of which were pictures.

Nonword reading: Subjects saw a total of 100 non-words drawn from the English Lexicon Project (Balota et al., 2007), randomly intermixed with 100 consonant strings that served as a non-nameable orthographic control. Task instructions were the same as the picture naming task, and each functional run consisted of 50 stimuli, half of which were nameable.

fMRI Data Analysis—Functional analysis was carried out using the AFNI software package (Cox, 1996). Preprocessing steps included (in order) correction for slice time acquisition, rigid-body realignment to the first volume (3dvolreg) and correction for magnetic field inhomogeneities (using in-house software). Spatial smoothing was not imposed. Functional data were analyzed using linear regression models (3dDeconvolve) using a zero-parameter gamma variate to estimate the BOLD response for each voxel. Trials in which participants indicated they could not name the picture or in which no response was given were excluded from the analysis. Activity associated with lexical-semantic retrieval was elicited with the contrast [picture naming–scrambled picture]– [nonword–consonants]; phonological encoding with [nonword reading–consonant string]. Targets for rTMS were selected from among activated voxels based on their proximity to the two anatomical regions associated with these two stages of production planning: pSTG and MTG (Indefrey & Levelt, 2004). In cases where multiple clusters of voxels were activated within these anatomical regions of interest, the center of the largest cluster was chosen for targeting with rTMS. Finally, the statistical maps containing the selected target voxels were coregistered and merged with the high-resolution T1 images, such that these activated regions appeared on the 3-D reconstructed brain images (see Figure 1 for the stimulation locations for all subjects).

rTMS

Stimulation Parameters—rTMS was delivered with a Magstim Super Rapid magnetic stimulator fit with a 70-mm figure-of-eight air-cooled coil (Magstim, Whitland, UK). Position of the stimulating coil was guided by infrared-based frameless stereotaxy (eXimia Navigated Brain Stimulation, Helsinki, Finland), so as to target the regions selected for each subject from their functional scans. The coil was oriented along the anterior– posterior axis of the temporal-lobe gyrus being stimulated, with the handle facing toward the back of the subject's head (approximately 30° from horizontal). In order to minimize discomfort, the angle of the coil handle was rotated toward a more vertical direction (approximately 45° from horizontal) for three subjects.

Stimulation intensity was set at 110% of resting motor threshold and was corrected for scalp-to-cortex distance of the target (Stokes et al., 2005). Average scalp-to-cortex distance for the pSTG was 16.2 mm (range = 11.2–20.9 mm), and 15.3 mm for the MTG (range = 11.4–20.8 mm). Average resting motor threshold across subjects was 58% of stimulator output (range = 46%–66%). After correcting for scalp-to-cortex distance, this resulted in an average stimulation of 61% (range = 48%–73%) of stimulator output for the pSTG, and 58% (range 47%–68%) of stimulator output for the MTG. Although the depth of stimulation did not vary between the two regions [$t(11) = 1.63$, $p > .1$], there was a trend toward differences in stimulator output intensity, as the pSTG received slightly higher intensity than the MTG across subjects [$t(11) = 2.14$, $p = .056$]. rTMS at 10 Hz was delivered unpredictably on half of the trials, and the order in which regions were stimulated, as well as which items were presented during rTMS$_{present}$ trials, was counterbalanced across subjects. These rTMS

stimulation parameters were within international safety standards for maximum train duration and minimum intertrain interval was followed (Wassermann, 1998).

Tasks—Subjects performed two blocks of three tasks per region (see Figure 2) in the following order: paced reading, picture naming, delayed serial recall, and picture naming. Subjects took a break of approximately 20 min in between blocks. Task instructions were given for 10 sec before each task, followed by 20 sec of fixation. Stimuli were presented and responses recorded using E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA). Subject responses were recorded with a lapel microphone, which provided recording quality sufficient to complete the scoring procedures described below.

<u>Paced reading:</u> Subjects read a total of 80 lists of phonologically similar nonwords, 40 per stimulated brain region. Phonological similarity was defined as sharing a common rhyme unit (e.g., pof, rof, nof, etc.). This manipulation was chosen to maximize the likelihood of inducing errors (Acheson & MacDonald, 2009b). Each list comprised five items. On each trial, subjects first read nonwords individually at a rate of 1/sec, followed by silent viewing of the entire list for 2 sec, followed by paced reading, which was initiated by a 200-msec tone. During the paced-reading epoch, the list was read through twice at a rate of 300 msec/ nonword (total duration of epoch = 10 sec), paced by sequentially changing the color of each nonword from black to red. On half of the trials, unpredictably, a 3-sec train of rTMS (30 pulses) was initiated 200 msec prior to the onset of paced reading to capture ongoing phonological encoding prior to articulation (Indefrey & Levelt, 2004). Mean ITI was 12 sec.

<u>Delayed serial recall:</u> Subjects read and recalled a total of 80 lists of nonwords, 40 per region. Nonword lists were the same as in paced reading, but occurred in a different order. Stimuli were selected such that the stimuli used for paced reading in one region differed from those used in serial recall in the same region. On each trial (initiated by the subject with a button press), five nonwords were presented at a rate of 1/sec, followed by a 3-sec delay, followed by a red question mark cueing spoken recall of the list in order. The rTMS train began with the offset of the final item of the presented list and lasted throughout the delay period (30 pulses). Mean ITI was 12 sec.

<u>Picture naming:</u> Subjects named a total of 160 colored pictures of common objects (Rossion & Pourtois, 2004), 80 pictures per region. Pictures were presented for 2 sec, the onset of which was accompanied by a 200-msec tone. Target responses to each picture were a mixture of single and multisyllabic words, and the average number of syllables per target was matched across region and rTMS conditions. Across subjects, each picture was presented equally often in each combination of rTMS condition and region (e.g., pSTG, $rTMS_{present}$; pSTG, $rTMS_{absent}$, etc.). Subjects viewed a fixation cross when pictures were not present. On rTMS trials, to maximize the extent to which rTMS would affect lexical–semantic retrieval rather than phonological encoding, a train of four rTMS pulses was delivered beginning 100 msec prior to stimulus presentation. The existing literature on TMS of picture naming is small, and provides contradictory guidance with regard to optimal timing of TMS for this task (Mottaghy, Sparing, & Töpper, 2006; Mottaghy et al., 1999; Topper, Mottaghy, Brugmann, Noth, & Huber, 1998). Therefore, we based our rTMS procedure on a model of the timing of the different stages of language production arising from a recent meta-analysis of word production (Indefrey & Levelt, 2004). Specifically, this model posits that, relative to picture onset, conceptual (i.e., semantic) and lexical retrieval occurs with a latency of 0–250 msec, phonological encoding with a latency of 200–400 msec, and articulatory planning with a latency of 400–600 msec. Thus, we assumed that rTMS spanning from −100 to 200 msec relative to picture onset would primarily affect lexical retrieval. Mean ITI was 5 sec.

Data Scoring—For all three tasks, verbal responses were digitized and scored off-line using Praat software (Boersma, 2001). Note that there was no difficulty with speech being obscured by the clicks from TMS coil discharge or noise from the air cooling mechanism because the lapel microphone was placed very close to the subject's throat. Also, precise measurements of speech timing were possible because raters had both acoustic and visual cues from the speech spectrogram which marked the onset of each trial, as well as the onset and offset of speech (see below).

<u>**Paced reading and delayed serial recall:**</u> Subject responses were coded for three types of speech errors (additions, omissions, and substitutions) across both whole items and individual phonemes. Additions occurred when a phoneme was added to an item or when a whole item was added to a list. Omissions occurred when a phoneme was left out of an item or an item itself was left out of a list. Substitutions occurred when a phoneme was substituted for another or when an item was substituted for another. A substituted item was scored as contextual if it was from the current list, and noncontextual if it was not. Data scoring was conducted using previously established procedures (Acheson & MacDonald, 2009a). Subject responses were phonetically transcribed by two trained individuals (interrater reliability = .92), and error scoring was automated with an in-house perl script. Only two types of errors showed an effect of rTMS, and are therefore the only ones reported here: item contextual substitutions (hereafter, item ordering errors) and item omissions.

<u>**Picture naming:**</u> Accuracy of response was coded based on the norms for each picture (Rossion & Pourtois, 2004). In order to avoid difficulties associated with voice keys in noisy environments, speaking times were manually scored by listening to subject responses and visually examining the speech spectrogram using Praat software (Boersma, 2001). Speech initiation latency was coded as the time from the onset of the trial (marked by the tone) until speech began. Total speech duration was coded as the time from the onset of the trial until speaking had stopped. Only correctly named pictures were included in the analysis of speaking times. Two individuals coded speech times (interrater reliabilities = .94). Overall naming accuracies were very high (96%), and did not show an effect of rTMS ($Fs < 1$), hence, only 4% of the naming data was removed prior to conducting the naming latency analyses.

Results

Paced Reading

Results for paced reading are summarized in Figure 3. A 2 (region) $\times$ 2 (TMS) repeated measures ANOVA on the proportion of item ordering errors revealed a TMS $\times$ Region interaction [$F(1, 11) = 7.46$, $p = .02$], but no main effect of region [$F(1, 11) = 0.848$, $p = .38$] or TMS [$F(1, 11) = 2.69$, $p = .13$]. Pairwise tests confirmed that, consistent with our prediction, subjects made more item ordering errors when rTMS was delivered to the pSTG [$t(11) = 2.32$, $p = .04$; $\mu D = 0.016$, $SD = 0.024$], but not to the MTG [$t(11) = 0.297$, $p = .77$; $\mu D = 0.0014$, $SD = 0.016$]. All mean differences reported here and below are rTMS$_{present}$ − rTMS$_{absent}$, thus negative numbers indicate either more errors or faster speaking times for the rTMS$_{absent}$ condition. No effect of TMS was observed for item omissions [TMS: $F(1, 11) = 1.01$, $p = .34$; Region: $F(1, 11) = 0.015$, $p = .905$; TMS $\times$ Region: $F(1, 11) = 1.23$, $p = .29$]; this result is not surprising given that individuals could view the items they were reading throughout task performance. Because rTMS effects were specific to the pSTG, these results confirm that one of the roles of the pSTG is the representation of the phonological form of words (Wilson et al., 2009; Graves, Grabowski, Mehta, & Gupta, 2008).

Delayed Serial Recall

The pattern of results on the WM task (Figure 4) was similar to that observed for paced reading, except the selective deficit manifested itself in item omissions rather than item ordering errors. An ANOVA revealed a significant TMS × Region interaction [$F(1, 11)$ = 5.64, p = .04], but no main effect of either TMS [$F(1, 11)$ = 4.60, p = .06] or region [$F(1, 11)$ = 3.29, p = .10]. Pairwise tests confirmed that subjects were more likely to omit entire items when rTMS was applied to the pSTG [$t(11)$ = 3.09, p = .01; µD = 0.052, SD = 0.06], but not to the MTG [$t(11)$ = 0.72, p = .49; µD = 0.01, SD = 0.06]. The marginal main effects of TMS and region are explained by the fact that subjects were overall more likely to omit for the pSTG relative to the MTG (µD = 0.02, SD = 0.09), and were more likely to omit items when rTMS was present relative to when it was absent (µD = 0.03, SD = 0.09). No effects were observed for item ordering errors [region: $F(1, 11)$ = 1.95, p = .19; TMS: $F(1, 11)$ = 0.61, p = .45; TMS × Region: $F(1, 11)$ = 0.45, p = .52]. Thus, this pattern of results indicates that nonword reading and VWM performance were similarly sensitive to rTMS of the pSTG, and not sensitive to rTMS of the MTG.

Picture Naming

Given the timing of rTMS with this task (100 msec prior to up to 200 msec after picture onset), which was intended to target semantic and lexical retrieval stages of the naming process, we predicted that rTMS of the MTG would show a larger effect on picture naming times than rTMS of the pSTG. Results were in line with this prediction, although TMS × Region interaction failed to reach significance. Similar to previous research that has used TMS to investigate picture naming (e.g., Mottaghy et al., 2006), rTMS decreased picture naming times (Figure 5). Speech initiation latencies showed a main effect of TMS [$F(1, 11)$ = 6.65, p < .05], but no main effect of region [$F(1, 11)$ = 0.18, 0.68] and no interaction [$F(1, 11)$ = 0.91, p = .36]. Total speech durations showed a main effect of TMS [$F(1, 11)$ = 6.36, p = .03] and a trend toward a TMS × Region interaction [$F(1, 11)$ = 3.56, p = .08], but no main effect of region [$F(1, 11)$ = 0.078, p = .79].

Planned comparisons for an effect of TMS at each region showed that the decrease in picture naming latencies after stimulation of the MTG was numerically larger than the effect observed for the pSTG. For the MTG, the difference in naming times between $rTMS_{present}$ and $rTMS_{absent}$ conditions was significant for both speech initiation latency [$t(11)$ = 2.21, p = .04; µD = −0.037, SD = 0.05] and total speech duration [$t(11)$ = 2.43, p = .03; µD = −0.053, SD = 0.08]. No such effect was observed for the pSTG for either onset latencies [$t(11)$ = 0.83, p = .43; µD = −0.012, SD = 0.10] or total speaking duration [$t(11)$ = 0.17, p = .87; µD = −0.002, SD = 0.14]. Although these results are thus consistent with a well-established role of the MTG in lexical–semantic retrieval (Binder et al., 2009; Wilson et al., 2009; Indefrey & Levelt, 2004), they cannot be taken as definitive evidence for a greater role for this region than for the pSTG in this function. We return to this point in the Discussion section below.

Discussion

The results of the present study demonstrate that maintenance in VWM is critically dependent on representations in LTM. Specifically, the ability to maintain a sequence of speech sounds over short periods of time depends on long-term representations within the language production architecture, in this case, the long-term phonological representations that are used to assemble an articulatory gesture. Although similar conclusions have been provided by neuropsychological (Martin & Saffran, 1997) and neuroimaging (Leff et al., 2009; Buchsbaum & D'Esposito, 2008) investigations, the former is inferentially limited by the fact that brain damage in patient populations is rarely localized to a single brain area, and

the latter by the fact that the brain–behavior relationship inferred by neuroimaging studies is correlational (although see Leff et al., 2009 for a combination of these two approaches). In the present study, the focal modulation of brain activity with rTMS provides compelling evidence for a causal link between language production and VWM.

With regard to our theoretical motivation, our results are consistent with the idea that the short-term retention of verbal information (i.e., VWM) is accomplished via the phonological encoding component of speech production (Acheson & MacDonald, 2009b; Martin & Saffran, 1997). The inferential basis for this interpretation comes from two important aspects of the design. First, the pSTG region targeted with rTMS was not defined according to its activity in VWM tasks, but rather to its activity in a language production task. Second, the specificity of the stage of language production implicated in VWM, phonological encoding, was established by differential effects of rTMS of the pSTG versus the MTG. The insensitivity of nonword reading and delayed serial recall to rTMS of the MTG indicates that it is not the case that these tasks are simply sensitive to rTMS in general, even when rTMS targets an anatomically adjacent control region that is also implicated in speech production. And because rTMS of the MTG does significantly affect a different speech production task, picture naming, it cannot be the case that this control area is simply less susceptible to the effects of rTMS.

The results of the present study add to a growing body of research showing that rTMS of peri-sylvian regions can influence VWM performance (Feredoes & Postle, 2007a, 2007b; Kirschen, Davis-Ratner, Jerde, Schraedley-Desmond, & Desmond, 2006; Romero, Walsh, & Papagno, 2006; Mottaghy, Doring, Muller-Gartner, Topper, & Krause, 2002; Duzel, Hufnagel, Helmstaedter, & Elger, 1996; Grafman et al., 1994). However, there exists considerable variability across these studies in the tasks used to test WM performance, the timing of TMS, the regions stimulated, hence, the results obtained. For instance, whereas some studies that have targeted inferior parietal regions have shown rTMS-related decrements in performance on the *n*-back and digit-span tasks (Romero et al., 2006; Mottaghy et al., 2002), another targeting this region has reported an rTMS-related improvement in serial recall of phonologically similar nonwords (Kirschen et al., 2006). Similarly, whereas one study has produced disruption of free recall performance when rTMS was applied to mid-temporal regions (Grafman et al., 1994), another has produced rTMS-related improvements in the recency portion of the digit-span task (Duzel et al., 1996). Thus, an important area for future research will be to disentangle which TMS parameters (i.e., when, where and how much) are likely to positively or negatively affect performance. It is noteworthy, however, that the single study that used words as stimuli (Grafman et al., 1994) found that rTMS of mid-temporal regions that were similar to the control region in the present study deleteriously affected recall performance. This is precisely what would be predicted if VWM involves activating different levels of production planning, including lexical–semantic representations when available.

Apart from our principal hypothesis, our results also make contact with several areas of language production research. With regard to phonological encoding, the fact that rTMS of the pSTG but not the MTG affected paced reading of nonwords provides evidence supporting the critical role of the pSTG in retrieving the speech sounds used for language production (Graves et al., 2008). With regard to lexical–semantic retrieval, our picture naming results add to a small literature showing that both single-pulse and rTMS of posterior temporal regions facilitates picture naming (e.g., Mottaghy et al., 1999, 2006; Topper et al., 1998). The previous studies have reported facilitation of picture naming following TMS of the pSTG either using an off-line repetitive protocol (i.e., 20 Hz trains followed by picture naming; Mottaghy et al., 1999), or a single-pulse protocol, with which facilitation only occurred when TMS was delivered either 1000 or 500 msec prior to picture

naming (Mottaghy et al., 2006; Topper et al., 1998). In our study, although rTMS of the pSTG produced a speeding of speech initiation latencies, this effect was not reliable. rTMS of the MTG, however, did produce reliable facilitation. One possible reason for the difference between the present versus the earlier results may be methodological. Whereas we targeted regions that showed functional activation for the task at the level of the individual subject, the studies by Mottaghy and colleagues guided rTMS anatomically, as inferred from the 10–20 navigation system. This latter procedure is likely to produce greater variability in the functional regions targeted with TMS. With regard to timing, our results are generally consistent with those from earlier studies that used single-pulse TMS during picture naming, in which pulses delivered 100 msec prior to up to 300 msec post picture onset showed some evidence of facilitation, although this failed to reach statistical significance (see Mottaghy et al., 2006). In our study, subjects were also numerically faster to initiate picture naming when TMS was applied to the pSTG 100 msec prior to up to 200 msec post picture onset.

Our prediction that picture naming performance would be more sensitive to rTMS of the MTG than of the pSTG was derived from a recent meta-analysis of the language production literature (Indefrey & Levelt, 2004). The fact that our results failed to show the clear anatomical dissociation for picture naming results, however, suggests that further prospective investigation of the model derived from this meta-analysis is warranted. It may prove to be the case, however, that observing such a dissociation within two such anatomically and functionally adjacent areas may be difficult. With regards to anatomy, there may prove to be inherent limitations on the spatial resolution of rTMS and the spread of activation that it induces in the cortex. In terms of the functional architecture of language production, it may be difficult to dissociate phonological and lexical– semantic retrieval given that phonological representations feedback and influence the activation of lexical representations (e.g., within interactive activation accounts; Dell, 1986). This account may explain, for instance, why some speeding of picture naming was observed when rTMS was applied to the pSTG. Such a possibility could plausibly be tested in future experimentations through use of a functionally guided, single-pulse TMS paradigm in which the spatial and temporal parameters of stimulation are independently manipulated.

Finally, although we have emphasized that the long-term, phonological representations targeted in this experiment are critical to the process of phonological encoding in speech production, it remains to be determined whether these same representations might also be used in the service of language comprehension. For instance, a recent voxel-based analysis of a large sample of stroke patients provided support for the left pSTG as a common substrate for VWM and language comprehension (Leff et al., 2009). An ongoing question for many language researchers is the extent to which phonological and other linguistic representations are shared between language production and language comprehension (e.g., Pickering & Garrod, 2007; Heim, Opitz, Muller, & Friederici, 2003; Watkins, Strafella, & Paus, 2003; Martin, Lesch, & Bartha, 1999). Although the present results in concert with those of Leff et al. (2009) and others (e.g., Buchsbaum & D'Esposito, 2008; Graves et al., 2008) suggest a common neural substrate for phonological representations in the service of language production and comprehension, an important area for future research will be to test this hypothesis directly through the causal inferences afforded by TMS.

Conclusion

Although some theoretical accounts hold the representations maintained in VWM to be independent of the long-term representations responsible for language comprehension and production, our results support the view that this distinction may not be necessary. Instead, it may be that one of the properties of the language system is the ability to maintain a production plan over an extended period of time via repeated interaction across multiple

levels of linguistic representation. A strong version of this view is that the "phonological store" and "articulatory loop" invoked by memory researchers may correspond to the same entities that speech production researchers call "phonological encoding" and "articulatory planning." This is consistent with a growing body of evidence from both the behavioral (Acheson & MacDonald, 2009a; Page, Madge, Cumming, & Norris, 2007) and neuroscience (Buchsbaum & D'Esposito, 2008) literatures that link language production and VWM processes. The present results provide the first direct, causal evidence that regions of the brain responsible for phonological encoding in language production are also responsible for the short-term retention of speech sounds in a test of VWM. More broadly, they speak against the long-held assumption that short-term memory and LTM representations are independent, and instead suggest that the representations maintained in WM are the same representations coded in long-term memory.

Acknowledgments

References

Acheson DJ, MacDonald MC. Twisting tongues and memories: Explorations into the relationship between language production and verbal working memory. Journal of Memory and Language. 2009a; 60:329–350. [PubMed: 21165150]

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

Atkinson, RC.; Shiffrin, RM. Human memory: A proposed system and its control processes. In: Spence, KW.; Spence, JT., editors. The psychology of learning and motivation: II. New York: Academic Press; 1968. p. 89-195.

Awh E, Jonides J. Overlapping mechanisms of attention and spatial working memory. Trends in Cognitive Sciences. 2001; 5:119–126. [PubMed: 11239812]

Baddeley, AD. Working memory, thought, and action. New York NY: Oxford University Press; 2007.

Balota DA, Yap MJ, Cortese MJ, Hutchison KA, Kessler B, Loftis B, et al. The English Lexicon project. Behavior Research Methods. 2007; 39:445–459. [PubMed: 17958156]

Binder JR, Desai RH, Graves WW, Conant LL. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cerebral Cortex. 2009; 19:2767–2796. [PubMed: 19329570]

Boersma P. Praat, a system for doing phonetic by computer. Glot International. 2001; 5:341–345.

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

Buzsaki, G. Rhythms of the brain. New York, NY: Oxford University Press; 2006.

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

Cowan, N. Attention and memory: An integrated framework. New York, NY: Oxford University Press; 1995.

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

Crowder RG. The locus of the lexicality effect in short-term memory for phonologically identical lists. Bulletin of the Psychonomic Society. 1976; 7:361–363.

Dell GS. A spreading-activation theory of retrieval in sentence production. Psychological Review. 1986; 93:283–321. [PubMed: 3749399]

Duzel E, Hufnagel A, Helmstaedter C, Elger C. Verbal working memory components can be selectively influenced by transcranial magnetic stimulation in patients with left temporal lobe epilepsy. Neuropsychologia. 1996; 34:775–783. [PubMed: 8817507]

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

Feredoes E, Postle BR. The neural basis of the short-term storage of verbal information are anatomically variable across individuals. Journal of Neuroscience. 2007b; 27:11003–11008. [PubMed: 17928441]

Grafman J, Pascual-Leone A, Alway D, Nichelli P, Gomez-Tortosa E, Hallett M. Induction of a recall deficity by rapid-rate transcranial magnetic stimulation. NeuroReport. 1994; 5:1157–1160. [PubMed: 8080978]

Graves W, Grabowski T, Mehta S, Gupta P. Left posterior superior temporal gyrus participates specifically in accessing lexical phonology. Journal of Cognitive Neuroscience. 2008; 20:1698–1710. [PubMed: 18345989]

Hebb, DO. The organization of behavior. New York, NY: John Wiley; 1949.

Heim S, Opitz B, Muller K, Friederici AD. Phonological processing during language production: fMRI evidence for a shared production–comprehension network. Brain Research, Cognitive Brain Research. 2003; 16:285–296. [PubMed: 12668238]

Hulme C, Roodenrys S, Schweickert R, Brown GDA. Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1997; 23:1217–1232.

Indefrey P, Levelt WJM. The spatial and temporal signatures of word production components. Cognition. 2004; 92:101–144. [PubMed: 15037128]

Kirschen MP, Davis-Ratner MS, Jerde TE, Schraedley-Desmond P, Desmond JE. Enhancement of phonological memory following transcranial magnetic stimulation (TMS). Behavioural Neurology. 2006; 17:187–194. [PubMed: 17148839]

Leff AP, Schofield TM, Crinion JT, Seghier ML, Grogan A, Green DW, et al. The left superior temporal gyrus is a shared substrate for auditory short-term memory and speech comprehension: Evidence from 210 patients with stroke. Brain. 2009; 132:3401–3410. [PubMed: 19892765]

Lewis-Peacock JA, Postle BR. Temporary activation of long-term memory supports working memory. Journal of Neuroscience. 2008; 28:8765–8771. [PubMed: 18753378]

Martin N, Saffran EM. Language and auditory–verbal short-term memory impairments: Evidence for common underlying processes. Cognitive Neuropsychology. 1997; 14:641–682.

Martin RC, Lesch MF, Bartha MC. Independence of input and output phonology in word processing and short-term memory. Journal of Memory and Language. 1999; 41:3–29.

Mottaghy FM, Doring T, Muller-Gartner HW, Topper R, Krause BJ. Bilateral parieto-frontal network for verbal working memory: An interference approach using repetitive transcranial magnetic stimulation (rTMS). European Journal of Neuroscience. 2002; 16:1627–1632. [PubMed: 12405977]

Mottaghy FM, Hungs M, Brugmann M, Sparing R, Boroojerdi B, Foltys H, et al. Facilitation of picture naming after repetitive transcranial magnetic stimulation. Neurology. 1999; 53:1806–1812. [PubMed: 10563632]

Mottaghy FM, Sparing R, Töpper R. Enhancing picture naming with transcranial magnetic stimulation. Behavioural Neurology. 2006; 17:177–186. [PubMed: 17148838]

Oberauer K. Access to information in working memory: Exploring the focus of attention. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2002; 28:411–421.

Page MPA, Madge A, Cumming N, Norris DG. Speech errors and the phonological similarity effect in short-term memory: Evidence suggesting a common locus. Journal of Memory and Language. 2007; 56:49–64.

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

Pickering MJ, Garrod S. Do people use language production to make predictions during comprehension? Trends in Cognitive Sciences. 2007; 11:105–110. [PubMed: 17254833]

Postle BR. Working memory as an emergent property of mind and brain. Neuroimage. 2006; 30:950–962. [PubMed: 16413797]

Romero L, Walsh V, Papagno C. The neural correlates of phonological short-term memory: A repetitive transcranial magnetic stimulation study. Journal of Cognitive Neuroscience. 2006; 18:1147–1155. [PubMed: 16839288]

Roodenrys S, Hulme C, Alban J, Ellis AW, Brown GDA. Effects of word frequency and age of acquisition on short-term memory span. Memory & Cognition. 1994; 22:695–701.

Rossion B, Pourtois G. Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. Perception. 2004; 33:217–236. [PubMed: 15109163]

Ruchkin DS, Grafman J, Cameron K, Berndt RS. Working memory retention systems: A state of activated long-term memory. Behavioral and Brain Sciences. 2003; 26:709–777. [PubMed: 15377128]

Schweickert R. A multinomial processing tree model for degradation and redintegration in immediate recall. Memory & Cognition. 1993; 21:168–175.

Stokes MG, Chambers CD, Gould IC, Henderson TR, Janko NE, Allen NB, et al. Simple metric for scaling motor threshold based on scalp–cortex distance: Application to studies using transcranial magnetic stimulation. Journal of Neurophysiology. 2005; 94:4520–4527. [PubMed: 16135552]

Tehan G, Humphreys MS. Articulatory loop explanations of memory span and pronunciation rate correspondences: A cautionary note. Bulletin of the Psychonomic Society. 1988; 26:293–296.

Topper R, Mottaghy FM, Brugmann M, Noth J, Huber W. Facilitation of picture naming by focal transcranial magnetic stimulation of Wernicke's area. Experimental Brain Research. 1998; 121:371–378.

van Eijsden P, Hyder F, Rothman DL, Shulman RG. Neurophysiology of functional imaging. Neuroimage. 2009; 45:1047–1054. [PubMed: 18801442]

Walker I, Hulme C. Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1999; 25:1256–1271.

Wassermann EM. Risk and safety of repetitive transcranial magnetic stimulation: Report and suggested guidelines from the International Workshop on the Safety of Repetitive Transcranial Magnetic Stimulation, June 5–8, 1996. Electroencephalography and Clinical Neurophysiology. 1998; 108:1–16. [PubMed: 9474057]

Watkins KE, Strafella AP, Paus T. Seeing and hearing speech excites the motor system involved in speech production. Neuropsychologia. 2003; 41:989–994. [PubMed: 12667534]

Watkins OC, Watkins MJ. Serial recall and the modality effect: Effects of word frequency. Journal of Experimental Psychology: Human Learning and Memory. 1977; 3:712–718.

Wilson SM, Isenberg AK, Hickok G. Neural correlates of word production stages delineated by parametric modulation of psycholinguistic variables. Human Brain Mapping. 2009; 30:3596–3608. [PubMed: 19365800]
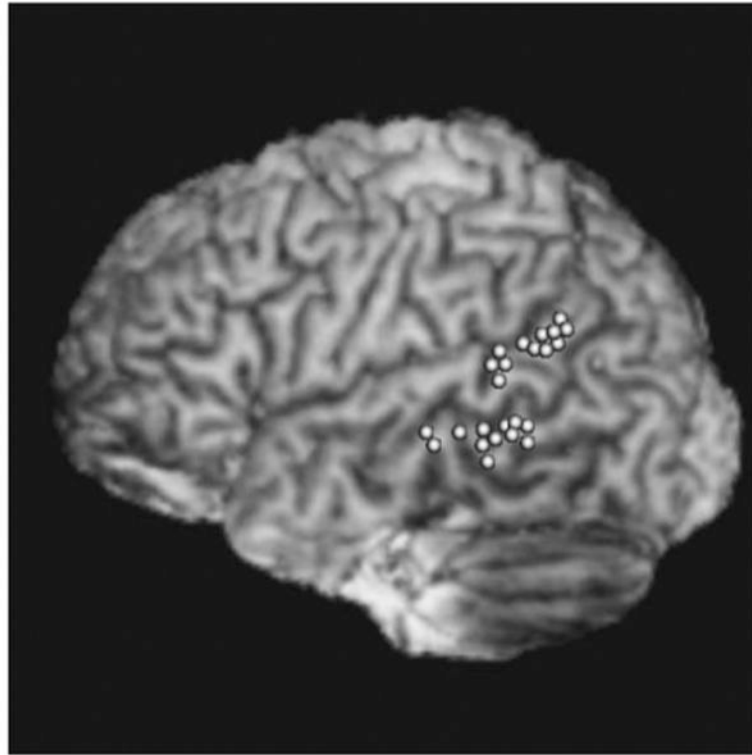
Figure 1.
Stimulation locations for each subject in this study, transformed to a common atlas space (ICBM 452) for display purposes. Average MNI coordinates for the pSTG stimulation were −64, −38, −13, and for the MTG, −69, −39, −2.
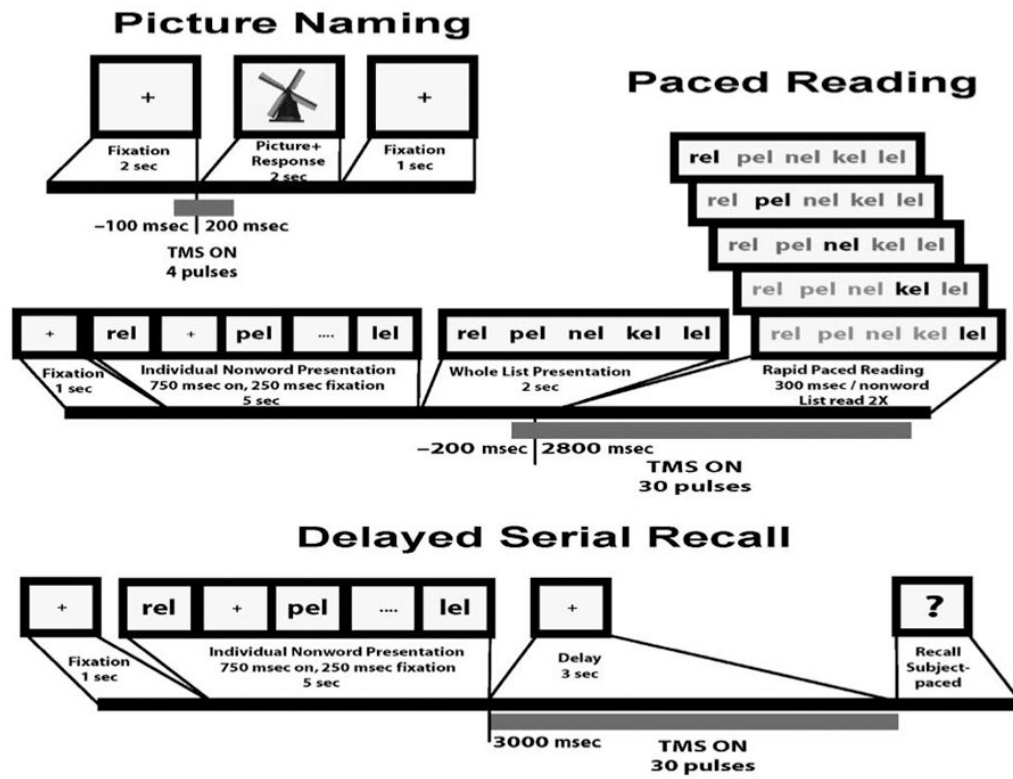
Figure 2.
Pictorial depiction of the picture naming, paced reading, and delayed serial recall tasks used in concert with rTMS. Each figure contains the timeline over which the stimuli and rTMS pulses were delivered for each task.
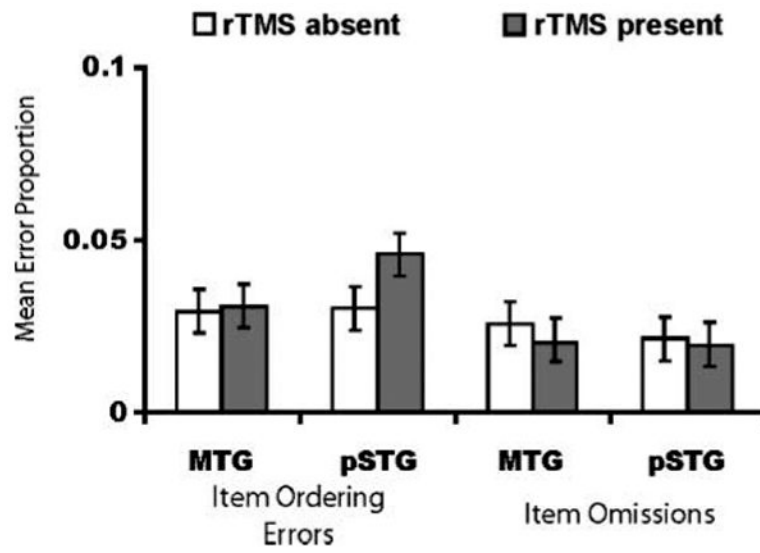
Figure 3.
Mean error proportions (95% CI) for paced reading as a function of region and rTMS condition. rTMS of the pSTG, but not the MTG, increased item ordering errors; no effect was observed for item omissions.
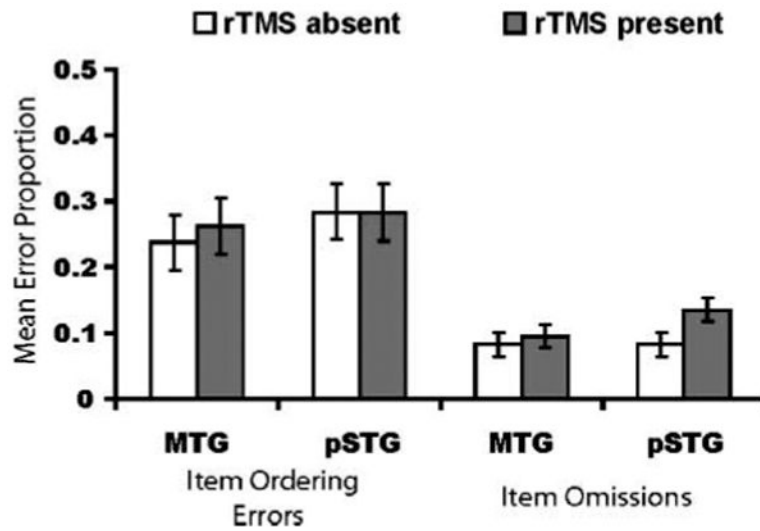
Figure 4.
Mean error proportions (95% CI) for delayed serial recall as a function of region and rTMS condition. rTMS of the pSTG, but not the MTG, led to an increase in item omission errors; no effect was observed for item ordering errors.
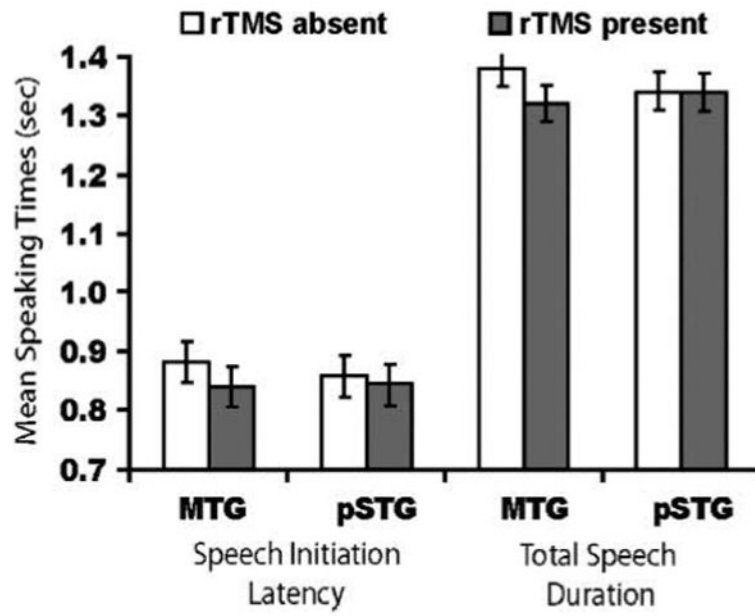
Figure 5.
Mean picture naming latency and speaking duration (95% CI) as a function of region and rTMS condition. rTMS speeded both measures, and the magnitude of this effect was numerically larger for MTG than the pSTG.