Experimental Brain Research (2020) 238:2307–2321
https://doi.org/10.1007/s00221-020-05886-9

RESEARCH ARTICLE

The phonological loop: is speech special?

Isabelle Deschamps[1,2] · Melody Courson[1,2] · Anthony Steven Dick[3] · Pascale Tremblay[1,2]

Received: 22 February 2020 / Accepted: 16 July 2020 / Published online: 30 July 2020

Abstract

It has been proposed that the maintenance of phonological information in verbal working memory (vWM) is carried by a domain-specific short-term storage center—the phonological loop—which is composed of a phonological store and an articulatory rehearsal system. Several brain regions including the left posterior inferior frontal gyrus (pIFG) and anterior supramarginal gyri (aSMG) are thought to support these processes. However, recent behavioral evidence suggests that verbal and non-verbal auditory information may be processed as part of a unique domain general short-term storage center instead of through specialized subsystems such as the phonological loop. In the current study, we used a single-pulse transcranial magnetic stimulation (TMS)-delayed priming paradigm with speech (syllables) and acoustically complex non-speech sounds (bird songs) to examine whether the pIFG and aSMG are involved in the processing of verbal information or, alternatively, in the processing of any complex auditory information. Our results demonstrate that TMS delivered to both regions had an effect on performance for speech and non-speech stimuli, but the nature of the effect was different. That is, priming was reduced for the speech sounds because TMS facilitated the detection of different but not identical stimuli, and accuracy was decreased for non-speech sounds. Since TMS interfered with both speech and non-speech sounds, these findings support the existence of an auditory short-term storage center located within the dorsal auditory stream.

Keywords Transcranial magnetic stimulation · Auditory working memory · Speech perception · Auditory discrimination · Inferior frontal gyrus · Supramarginal gyrus

Introduction

The process of spoken language comprehension is complex, involving perceptual, linguistic as well as cognitive and executive functions such as working memory (WM) and attention. According to most models of speech perception/recognition, to allow for meaning to be extracted from the auditory stream, the auditory speech signal must be processed acoustically and decomposed into phonological units (Gaskell and Marslen-Wilson 1997, 2002; McClelland and Elman 1986; Norris 1994; Stevens 1972). These phonological units are maintained in transient storage through WM processes (Jacquemot and Scott 2006) allowing for words to be recognized and meaning to be accessed.

While most theoretical models converge on the notion that WM involves several processes related to the encoding, maintenance (i.e., storage), and retrieval of task-relevant information, questions arise regarding the scope of these mechanisms: are they domain specific (i.e., verbal, spatial, auditory) or domain general? In domain-specific WM models (e.g. Baddeley and Hitch 1974; Martin 2005), different subsystems are dedicated to the transient storage of specific types of information (examples of dissociation between phonological WM and semantic WM can be found in Martin (2005); Martin et al. (1994); Martin et al. (2003)); whereas in domain-general WM models (e.g. Atkinson and Shiffrin 1968; Barrouillet et al. 2004; Cowan 2001), a single system receives and stores input from different types of information.

Communicated by Melvyn A. Goodale.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00221-020-05886-9) contains supplementary material, which is available to authorized users.

✉ Pascale Tremblay
Pascale.Tremblay@fmed.ulaval.ca

[1] Département de Réadaptation, Faculté de Médecine, Université Laval, 1050 Avenue de La Médecine, Office 4109, Quebec City, QC G1V 0A6, Canada

[2] CERVO Brain Research Centre, Quebec City, QC, Canada

[3] Florida International University, Miami, FL, USA

One of the most influential domain-specific models of WM (Baddeley and Hitch 1974) proposes four subsystems: an episodic buffer, a central executive system, a visuo-spatial sketch pad and a phonological loop. Whereas the visuo-spatial sketchpad is dedicated to the temporary storage of visual or spatial information, the phonological loop is involved with the transient storage of verbal information. The phonological loop is composed of a phonological store and an articulatory rehearsal system (Baddeley 1975, 1986, 1992, 2003; Hitch and Baddeley 1976). The phonological store holds memory traces for approximately 1–2 s and the articulatory rehearsal system refreshes those memory traces by way of subvocal speech allowing them to stay in WM for longer periods of time. In this model, auditory verbal material has direct access to the phonological store, whereas the visual verbal material (i.e., print) has indirect access; visually presented verbal stimuli are first coded in the visuo-spatial sketchpad and then converted into phonological representations (i.e., sound-based representations) through the subvocal rehearsal system (subvocal speech).[1] While the phonological loop was initially assumed to hold phonological information *only* (i.e., sound-based representations of verbal material), there is some evidence to suggest that it may also process non-speech sounds. For instance, Salamé and Baddeley (1982, 1986, 1989) demonstrated that during the serial recall of verbal material, the presentation of irrelevant speech material, as well as vocal and instrumental music, had a detrimental effect on performance. Similarly, Jones and Macken (1993) have shown that the presentation of auditory tones can disrupt the serial recall of letters, and Pechmann and Mohr (1992) demonstrated that the presentation of spoken verbal material can disrupt the serial recall of auditory non-verbal material. These findings suggest that auditory verbal and non-verbal information may be processed within the same transient storage systems during WM tasks (Baddeley 2012). However, the extent to which components of the phonological loop are used to remember non-speech sounds remains unclear.

The functional architecture of the phonological loop has been investigated extensively in the context of verbal WM via fMRI studies (Awh et al. 1996; Buchsbaum et al. 2001; Chen and Desmond 2005; Fiez et al. 1996; Gruber 2000; Hartwigsen et al. 2016; Henson et al. 2000; Herwig et al. 2003; Kirschen et al. 2010; Marvel and Desmond 2012; Paulesu et al. 1993; Rodriguez-Jimenez et al. 2009; Shen et al. 2015). The results of these studies suggest that the articulatory rehearsal system engages a network of regions including the posterior part of the inferior frontal gyrus (pIFG), the supplementary motor area (SMA), the cerebellum as well as the ventral premotor cortex (PM). In contrast, the phonological store engages parietal regions such as the supramarginal gyrus [SMG], the intraparietal sulcus [IPS], and the parietal operculum as well as area Spt (Sylvian parietal temporal area). However, depending on task demands and the modality of stimulus presentation (auditory, visual), different regions within the inferior and superior parietal lobules as well as the posterior superior temporal gyrus have been reported. Indeed, some have argued that no single brain region has a functional profile matching perfectly with the phonological store, and it may be better to ascribe phonological short-term memory as something that emerges from the "integrated neural processes that underlie the perception and production of speech" (Buchsbaum and D'Esposito 2008, p. 762). Interestingly, studies investigating auditory non-verbal WM (Gaab et al. 2003; Kumar et al. 2016; Martinkauppi et al. 2000; Zatorre et al. 1994) and the overlap between auditory verbal and non-verbal WM (Hickok et al. 2003; Koelsch et al. 2009; Schulze et al. 2011) also report a similar network of regions. These results suggest that the maintenance of auditory verbal and non-verbal information involves similar WM components.

Several TMS studies using verbal stimuli have confirmed an implication for the pIFG in the articulatory system of the phonological loop and an implication for the SMG in the phonological store component (Deschamps et al. 2014; Gough et al. 2005; Hartwigsen et al. 2010a, b; Hartwigsen et al. 2016; Herwig et al. 2003; Kirschen et al. 2006; Nixon et al. 2004; Romero et al. 2006). Interestingly, a few studies using repetitive TMS (rTMS) or transcranial direct stimulation (tDCS) have shown that the left SMG is involved in the storage of non-verbal information. These studies show that left SMG stimulation is detrimental when applied during the interval between the first and second tone sequences during a pitch memory task but not when applied during the presentation of the first sequence (Schaal et al. 2013, 2015a, b; Vines et al. 2006). The authors interpreted these findings as indicating that the left SMG is involved in the maintenance of pitch information, which suggests that the phonological store holds non-verbal auditory information in transient storage. Whilst results from TMS studies investigating WM with verbal (print or speech) and non-verbal auditory material suggest a role for the left SMG in the transient storage of auditory verbal and non-verbal information, to our knowledge, no study has addressed the role of the pIFG in articulatory rehearsal (i.e., subvocal speech) during WM using both auditory verbal and non-verbal materials. Given that the pIFG is a core region for articulatory rehearsal, determining whether it is recruited during non-verbal auditory WM task could provide valuable insights into the mechanisms underlying the phonological loop components. More

---

[1] In Baddeley's model of vWM, since print is converted into sound-based representations before accessing the phonological loop, the assumption is that that the phonological loop holds sound-based representations of print material.

generally, mapping the scope of the different phonological loop processes is key to advance current understanding of the neurobiology of WM.

In the current study, we sought to investigate, using a TMS-delayed priming paradigm, whether the left pIFG and aSMG are involved in the storage of auditory verbal and non-verbal information. To our knowledge, this is the first TMS study to investigate the role of these two regions using both auditory verbal (i.e., speech) and non-verbal (i.e., bird sounds) stimuli. To test this hypothesis, we used a delayed auditory discrimination task in which identical and different pairs of sounds were presented. This allowed us to take advantage of the well-established phenomenon of repetition priming—the facilitation of a stimulus that has previously been processed—to evaluate the effect of TMS during WM processes. To target WM processes, that is, the phonological store (maintenance of information) and articulatory rehearsal mechanisms, we introduced a 750-ms delay between the two stimuli. This delay was based on a prior electroencephalography (EEG) study investigating the time course of verbal working memory processes (maintenance and rehearsal processes) (Ruchkin et al. 1997). Repetition priming manifests itself behaviorally as a decrease in reaction times (RTs) for the processing of a repeated stimulus and at the neural level as a reduction in activation (Henson et al. 2000b; Wig et al. 2005). Repetition priming has been reported for the processing of verbal (Kouider and Dupoux 2005; Light et al. 1995; Rueckl 1990) and non-verbal stimuli (Bergerbest et al. 2004; Hutchins and Palmer 2008). Of particular interest is the fact that repetition priming has been observed with both short (2 s) and long delay (4–12 s) intervals between the presentation of the first stimulus and repeated stimulus (Chao et al. 1995). This allowed us to investigate, using a TMS-priming paradigm, whether stimulation to the left aSMG would interfere with the maintenance in memory of a stimulus. TMS-priming paradigm has been used to investigate the tuning of neural populations to specific features contained in a stimulus that has been previously processed (Cattaneo et al. 2008; Romei et al. 2016; Silvanto and Pascual-Leone 2008). TMS-priming paradigms are based on the concept that the effects of TMS pulses are "state dependent". In other words, the effects of TMS on specific populations of neurons are dependent on their firing history (Silvanto and Pascual-Leone 2008). For instance, Cattaneo (2010) demonstrated, using a TMS-priming paradigm, that neurons within the ventral PM were selectively tuned to different types of grasping actions (i.e., whole-hand grasp vs. precision grip).

In the present study, we expected TMS to reduce priming (by increasing the RTs of identical sequences) and sensitivity, reflecting a decrease in the perceptual advantage of the repeated stimuli. To examine the temporal unfolding of the involvement of the left pIFG and aSMG throughout the maintenance phase, TMS was delivered at two different time points, early (250 ms—storage phase) or late (500 ms—rehearsal phase), during the interstimulus interval. Our main hypothesis was that TMS to the left aSMG during the "storage phase" (250 ms into the delay) would disrupt the performance for both auditory speech and non-speech stimuli reflecting a role for this region in the *maintenance* of auditory information in the phonological store, a domain-general mechanism. In contrast, we expected TMS to the left pIFG during the "rehearsal phase" (500 ms into the delay) to disrupt performance only for the speech stimuli reflecting a role for this region in the *articulatory rehearsal processes*, a domain-specific mechanism.

Method

Participants

Nineteen healthy adults were recruited through emails sent to Université Laval's mailing lists as well as from the Lab participant databank (BACH, #2014-369). We report the data from 18 participants (mean age: 25.20 ± 3.91; 10 females), as one participant was unable to complete the TMS session due to a headache. All participants were native speakers of Quebec French, had normal or corrected-to-normal vision and no self-reported speech, voice, language, swallowing, psychological, neurological or neurodegenerative disorder, and no severe respiratory disorder. All participants were right-handed as assessed by the Edinburgh Handedness Inventory (mean 18.66 ± 1.76/20) (Oldfield 1971). Participants' cognitive level was assessed using the Montreal Cognitive Assessment scale (MOCA) (mean 28.79 ± 1.13/30). Hearing was assessed using a standard pure tone audiometry procedure using a clinical audiometer (AC40, Interacoustic). A pure tone average was computed (at frequencies: .5, 1 and 2 kHz) for each ear. This evaluation confirmed that all participants had normal hearing (right ear: −5.09 ± 3.74; left ear: −2.77 ± 4.35 dB HL). Prior to the experimental session, all participants were screened for any relative or absolute contraindication to TMS (Rossi et al. 2009; Wasserman 1998). The study was approved by the Institutional Ethical Committee of the Institut Universitaire en Santé Mentale de Quebec (#351-2013).

Materials and procedure

The experiment entailed two visits on two different days. During the first visit, participants underwent structural magnetic resonance imaging (MRI). For 8 of the 19 participants, the MRI images were acquired as part of previous projects and retrieved from the Lab databank BACH (#369-2014). During the second visit, participants completed a delayed

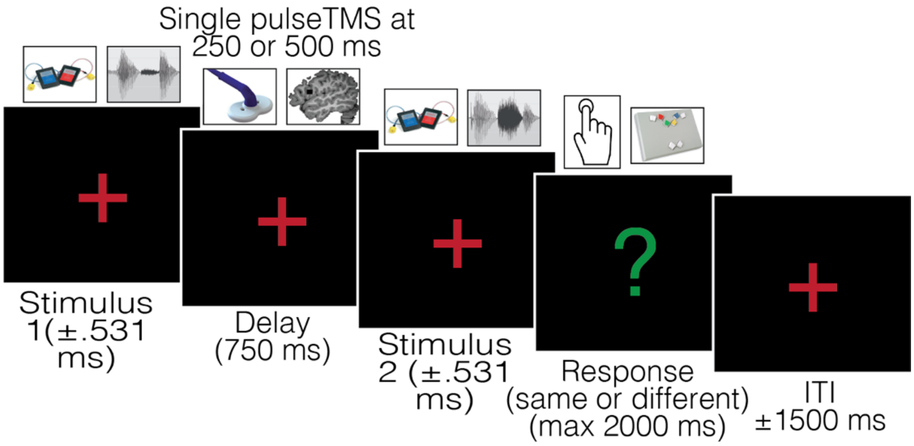

Fig. 1 Experimental design and trial structure. During the 750-ms delay between the two stimuli, a single TMS pulse was delivered either at 250 ms or 500 ms. Following the offset of the second stimuli, a green question mark appeared on the screen indicating to participants to respond (same/different) using a response pad. Participants had 2000 ms to respond before the trial ended

auditory discrimination task and two visual vWM tasks (n-back, Sternberg scanning, see Online Resource 1).

Task

A delayed auditory discrimination task was used to investigating the *maintenance* of auditory verbal and non-verbal information (i.e., transient storage). Unlike other WM tasks, such as the n-back task, discrimination tasks do not require manipulation and updating of information, thereby isolating maintenance mechanisms from other WM processes.

Participants were comfortably seated in a sound-attenuated room wearing insert earphones (TMS compatible, non-metallic, Etimotic Research INC, ER·1, Illinois, USA). The auditory stimuli were presented through via a computer controlled by the Presentation® software (version 18.1, http://www.neurobs.com). For each participant, the volume was adjusted using an amplifier (HP4, Presonus, Baton Rouge, Louisiana, US) prior to beginning the experiment. Participants responded using a response pad (RB-840 model, Cedrus, San Pedro, California, US).

The task included a total of 384 trials. Each trial began with the presentation of a red fixation cross, followed by the presentation of two auditory stimuli (Fig. 1). The offset of the first stimulus and the onset of the second stimulus were separated by a 750-ms delay. This delay was sufficiently long to target the phonological store and articulatory rehearsal processes while minimizing difficulty associated with very long delays. It has been shown, using event-related brain potentials during a delayed matched to sample task, that brain waves sensitive to memory load emerge immediately following presentation of an auditory stimulus (a negative frontal and a positive centro-posterior) (Ruchkin et al. 1997), suggesting that such mechanisms were at play during our 750-ms delay.

On half of the trials ($N=192$), a single TMS pulse was delivered either at 250 ms or 500 ms during the delay. On the other half of the trials ($N=192$), no stimulation was delivered (no-TMS). Following the offset of the second stimulus, a green question mark appeared on the screen telling participants that they should indicate whether the second stimulus was identical to the first or different by pressing a button on a response box. When the two stimuli in a pair were identical, participants pressed a green button with their right hand and when the stimuli differed, they pressed a red button. The position of the green and red buttons on the response pad was counterbalanced across participants. Participants were given 2000 ms after the presentation of the question mark to provide an answer. The trial ended with either the button press or, if the participant failed to respond, the trial ended after 2000 ms. The average inter-trial interval was 1516.75 ± 280.61 ms, ranging from 1000 to 2000 ms.

Stimuli

Two types of auditory stimuli were used: speech and non-speech sounds. We refer to this as the “Stimuli” factor. The speech sounds were 96[2] pairs of bisyllabic nonwords and the non-speech sounds were 96 pairs of edited bird sounds. Half the stimulus pairs were identical and the other half were different. In addition, half of the pairs was produced by the same speaker or same bird (i.e., one speaker or one bird produced all of the sounds within a sequence); whereas, the other half was produced by different speakers or birds (i.e., two different speakers or birds produced the sounds within one sequence). This was done to make the task more challenging and more naturalistic by increasing the acoustical variability of the stimuli.[3] Participants were asked to focus on the nature of the sound ignoring the speaker/bird.

---

[2] Each of the 96 speech and 96 non-speech stimuli were presented twice, once with each stimulation site.

[3] This factor was not included in the analyses as it was not a factor of interest.

The 96 non-speech sounds were created from 68 unique sound files featuring thirty-seven different birds (e.g., parrot, falcon, starling). These recordings are commercially available on iTunes (The Ultimate Sound Effects Collection: Birds, 2010 by HDsoundFX). The bird sounds were recorded at 44 kHz. Bird sounds were chosen, because they are, like speech, time-varying and acoustically complex (Tremblay et al. 2012). All sounds were edited for duration using Praat software (Boersma and Weenink 2011). The resulting stimuli were normalized to a mean intensity of 85 dB HL using Praat. A complete description of these stimuli (including their spectro-temporal characteristics) can be found in Tremblay et al. (2012).

The 96 speech pairs were composed of Quebec French bisyllabic nonwords not forming real words (a full list of stimuli is supplied in Online Resource 2). The nonwords were created from a set of 48 different consonant vowels (CV) syllables produced by three different native adult male French speakers from Quebec City (total of 144 tokens). The syllables were composed of combinations of 12 consonants (/f, s, ʃ, v, z, ʒ, p, t, k, b, d, g/) and 4 vowels (/e, ə, ɛ, ø/). The syllabes (but not the pairs) have been used in a prior study from our group (Tremblay et al. 2018). Each syllable was recorded at least three times, as part of a carrier sentence (i.e., Now I say _____). The recordings were made in a double-walled soundproof room using a high-quality headset microphone (Shure Microflex Beta 53) connected to a Quartet USB audio interface (Apogee Electronics, Santa Monica, CA 90,404, USA) and an iMac computer (OS X 10.9.4). All syllables were recorded using Sound Studio 4 software (Felt Tip inc., NYC, USA) at a sampling signal of 44.1 kHz and 24 bits of quantization. The sound files were edited offline using Praat software (Boersma and Weenink 2011). Each syllable was segmented, and for each speaker, the best exemplar of each syllable was selected. The resulting stimuli were normalized to a mean intensity of 85 dB HL using Praat. The mean duration of the individual syllables was 299 ± 72.55 ms; whereas, the mean duration of the individual bird sounds was 233 ± 20.45 ms. The mean duration of the individual syllables was significantly longer than the mean duration of the individual bird sounds ($t_{17} = 10.16$, $p < .001$, Cohen's $d = 1.25$). Since the onset of the delay and stimulation time were based on the offset of the first sequence, the first sequences with shorter sound duration were not kept longer in memory, as they would have been with a fixed onset time for the delay and the stimulation for all sequences. All stimuli and experiment files are available online on the Scholar Portal Dataverse: https://doi.org/10.5683/SP2/9YLXDF.

MRI acquisition and co-registration

A high-resolution T1-weighted anatomical MRI scan was obtained for all participants on a 3T Philips Achieva TX MRI scanner at the Clinic IRM Quebec-Mailloux in Quebec City (matrix 256 mm × 256 mm, 180 slices, 1 mm × 1 mm × 1 mm, no gap). Prior to the TMS session, the anatomical MRI was incorporated into Brainsight 2 (Rogue Research, Montreal, Canada) to guide coil placement and identify the stimulation targets. For each participant, four anatomical landmarks (tip of the nose, bridge of noise and intertragic notch) were identified on their T1 image to guide MRI-to-head co-registration using an infrared tracking system (Polaris, Northern Digital, Waterloo, Canada). Throughout the TMS session, infrared tracking was used to monitor the position of the coil with respect to the participant's brain.

TMS

Participants were seated in a padded TMS chair with their head comfortably held in place by a headrest (Rogue Research, Montreal, Quebec, Canada). Stimulation was performed with a figure of eight D70^2 coil combined to a Magstim Rapid^2 stimulator (Magstim Company, Dyfed, UK). To establish the resting motor threshold (RMT) of each participant, the TMS coil was placed over the hand area of the participant's left primary motor cortex, previously identified on the participant's MRI scan. The coil was held tangentially to the skull with the handle pointing posteriorly and inferiorly. Single pulses were delivered to the hand motor cortex and the intensity of the stimulation was adjusted until the motor evoked potential (MEP) in the right FDI (EMG Isolation Unit, Brainsight 2, Rogue Research, Montreal, QC, Canada) was observed in 5 of the 10 trials with an amplitude of at least 50 µV (Rossini et al. 1994). Adjustments to the location of the stimulation were made to locate the maximum excitable hand area. The stimulation intensity was set to 110%[4] of the participant's RMT, which ranged from 45 to 73% (mean 60.5 ± 8.3) of the output capacity of the stimulator. Importantly, the stimulation parameters used were well within TMS safety guidelines (Rossi et al. 2009; Wasserman 1998). Two cortical sites were selected for stimulation (Fig. 2): the left pIFG[5] and the left aSMG.[6] For each participant, the coordinates of the stimulation sites

---

[4] We selected this parameter based on previous studies that stimulated similar regions within the inferior frontal cortex or the supramarginal gyrus (Devlin et al. 2003; Romero et al. 2006).

[5] We selected this region based on studies investigating phonological processes (Gitelman et al. 2005; McDermott et al. 2003).

[6] We selected this region based on a previous study from our group (Deschamps et al. 2014).

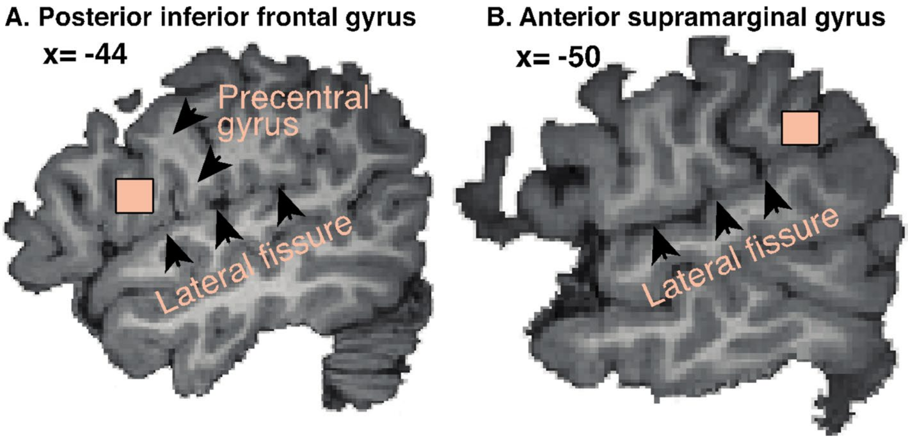

Fig. 2 Target region for TMS stimulation. Sagittal views of one of the participants' brain illustrating the target regions (pIFG and aSMG)

(left pIFG and left aSMG) were determined based on their own macrostructural gyral anatomy using Brainsight 2 software, a computerized frameless stereotaxic system (Rogue Research, Montreal, Canada). For the left aSMG, we first identified the end point of the lateral fissure. The stimulation site was then set by selecting a point that was 5 mm anterior to the endpoint of the posterior ramus of the lateral fissure and 5 mm inferior to the anterior point. The mean coordinates in Talairach space for the left aSMG as defined by the MNI-152 template in BrainSight 2 software (Rogue Research, Montreal, Canada) were − 50 to 38 26. For the left pIFG, we first identified the junction between the precentral gyrus (PrCG) and the inferior frontal sulcus (IFS). The stimulation site was then set by selecting a point that was 5 mm anterior to the PrCG/IFS junction and halfway between the IFS and the lateral fissure. The mean coordinates in Talairach space as defined by the MNI-152 template in BrainSight 2 software (Rogue Research, Montreal, Canada) for the left pIFG were − 44 10 14.

Each participant underwent four runs of single-pulse TMS: two runs of pIFG stimulation (including one run of speech and one run of non-speech stimuli) and two runs of aSMG stimulation (one run of speech and one run of non-speech stimuli). At the beginning of each run, the coil was placed either on the left pIFG or the left aSMG. The accuracy of the coil placement was monitored continuously using Brainsight 2. Throughout the experiment, the position was maintained within .22 ± .18 mm of the target on average.

Stimulation was delivered on half of the trials ($N = 192$), while no stimulation was applied during the other half (no-TMS) ($N = 192$). On half of the TMS trials ($N = 48$), a pulse was delivered early during the delay (i.e., 250 ms after delay onset, "TMS@250"); while on the other half of the trials ($N = 48$), a pulse was delivered later during the delay (i.e., 500 ms after delay onset, "TMS@500″; Fig. 1). The order of the runs and target regions was counterbalanced across participants. Trial order was randomized within each run. This experimental set-up resulted in 384 trials distributed across 16 conditions: 4 stimulation levels (no-TMS@250, no-TMS@500, TMS@250 ms, TMS@500 ms) × 2 target areas (pIFG, aSMG) × 2 stimuli (speech, non-speech).

Data analysis

All data were analyzed using SPSS 22 (IBM, Armonk, NY). Response accuracy was analyzed within the framework of signal detection theory (Macmillan and Creelman 2004). Specifically, *d*-prime (d') and criterion (c) were calculated. D-prime (d') is defined as the ability to accurately discriminate between identical syllable pairs (target) and non-identical syllable pairs (non-target trials). Criterion (c) is defined as the tendency to select one response (i.e., same or different) over the other (Macmillan and Creelman 1990). *C* is, thus, a measure of response bias; when $c = 0$, participants do not have a bias towards one response over the other. For RTs, a priming score (i.e., the difference in RTs between the processing of non-identical and identical stimuli) was calculated for each condition. RT was defined as the time from the offset of the second sound in the second sequence, which corresponds to the appearance of the question mark (Fig. 1), to the onset of participants' response. Priming scores were calculated after removing incorrect trials and removing RTs that were three times above or below the interquartile range for each participant in each condition. Trials containing errors represented 4.5% of all trials.

For each dependent measure (d', c, and priming), we calculated two difference scores ([TMS@250—no-TMS@250]; [TMS@500—no-TMS@500]). Each of these scores was entered in a separate ANOVA with three repeated factors: Region (pIFG, aSMG), Stimuli (speech, non-speech), and Time (TMS@250, TMS@500). Measures of effect sizes are provided in the form of partial eta squared (η_p^2), which are reported for all main effects and interactions. Significant interactions were decomposed into simple effects. For all significant group results, we show corresponding individual data (Weissgerber et al. 2015).

Table 1 Discriminability indices: *d′* and *c* for each condition

Stimuli	pIFG		aSMG	
	TMS@250—no-TMS@250	TMS@500—no-TMS@500	TMS@250—no-TMS@250	TMS@500—no-TMS@500
Speech *d′*	.06 (.07)	− .02 (.06)	0 (.04)	.04 (.07)
Non-Speech *d′*	− .02 (.09)	− .44 (.13)	.11 (.14)	− .31 (.12)
Speech *c*	.05 (.03)	.01 (.02)	.2 (.03)	.02 (.03)
Non-speech *c*	.09 (.07)	.02 (.06)	.07 (.06)	− .14 (.07)

Mean (SEM)

SEM standard error of the mean

Results

Discrimination performance

First, we examined whether non-speech stimuli were harder to discriminate (using *d′*) than speech stimuli in the absence of TMS. A paired-sample *t* test was computed. The results demonstrate that both sound categories were highly discriminable (speech: mean $d' = 3.35$, SD = .13; non-speech: mean $d' = 2.93$ SD = .34), though the non-speech sounds were significantly harder to discriminate than the speech sounds ($t_{17} = -5.37$, $p = .00005$, 95% CI [.25, .57], Cohen's $d = 1.31$).

Next, we examined the difference between TMS and no-TMS trial for each discriminability index (*d′* and *c*) for each condition (Table 1; Online Resources 3 and 4, respectively). The results of the repeated measures ANOVA for *d′* and *c* are detailed in Table 2A, B, respectively. Only significant effects are reported in the text. For *d′*, a main effect of Stimuli was found (F (1, 17) = 7.76, $p = .014$, $n_{\mathrm{p}}^2 = .31$), with participants being better at discriminating between speech than non-speech sequences. A main effect of Time was also found (F (1, 17) = 8.35, $p = .010$, $n_{\mathrm{p}}^2 = .33$), with participants being worse at discriminating between sequences following stimulation at TMS@500 than TMS@250. The Stimuli*Time interaction was also significant (F (1, 17) = 6.76, $p = .019$, $n_{\mathrm{p}}^2 = .28$). A simple-effect analysis revealed a significant difference between the two stimulation times for the non-speech stimuli (F (1, 17) = 8.54, $p = .01$, $n_{\mathrm{p}}^2 = .33$), with participants being worse at discriminating between sequences following stimulation at TMS@500 than TMS@250. No significant difference between the two stimulation times was found for the speech stimuli (F (1, 17) = .11, $p = .8$, $n_{\mathrm{p}}^2 = .006$). To determine the source of difference for the non-speech stimuli (i.e. [TMS@250—no-TMS@250] ≠ [TMS@500—no-TMS@500]), a series of Bonferroni-corrected paired-sample t-tests were conducted between TMS trials and the no-TMS trials at each time point for the non-speech sounds. To ensure that the interaction was not driven by differences between no-TMS trials, additional paired-sample *t* tests were also conducted to compare TMS@250 and TMS@500 as well as no-TMS@250 and no-TMS@500 trial (level of significant needed is .0125, .05/4 comparisons). As illustrated in Fig. 3, a strong significant difference was found at TMS@500 ($t_{17} = 4.06$, $p = .001$, 95% CI [.18, .58], Cohen's $d = .97$). A significant difference was also found between TMS trials (TMS@250 — TMS@500) ($t_{17} = -3.07$, $p = .005$, 95% CI [.09, .46], Cohen's $d = .74$). No other comparison reached significance. Thus, TMS interfered with participants' ability to discriminate between sequences when delivered 500 ms into the delay. For *c*, no significant main effects or interactions were found.

Table 2 Results of the repeated-measures ANOVA for *d′* and *c*

Effects	Type II sum of squares	*f* value	*df*	*p* value	n_{p}^2
A. Results for *d′*					
Region	.16	1.62	1,17	.22	.09
Stimuli	1.26	7.59	1,17	**.01**	.31
Time	1.75	8.39	1.17	**.01**	.33
Region*Stimuli	.16	.76	1,17	.40	.04
Region*Time	.03	.13	1,17	.73	.007
Stimuli*Time	1.49	6.76	1,17	**.02**	.28
Region*Stimuli*Time	.04	.45	1,17	.51	.03
B. Results for *c*					
Region	.09	4.03	1,17	.06	.19
Stimuli	.005	.08	1,17	.79	.004
Time	.23	3.57	1,17	.08	.17
Region*Stimuli	.06	1.53	1,17	.23	.08
Region*Time	.02	.38	1,17	.55	.02
Stimuli*Time	.14	3.16	1,17	.09	.16
Region*Stimuli*Time	.07	1.16	1,17	.30	.06

Significant effects are in bold. n_{p}^2 partial eta square, *df* degrees of freedom

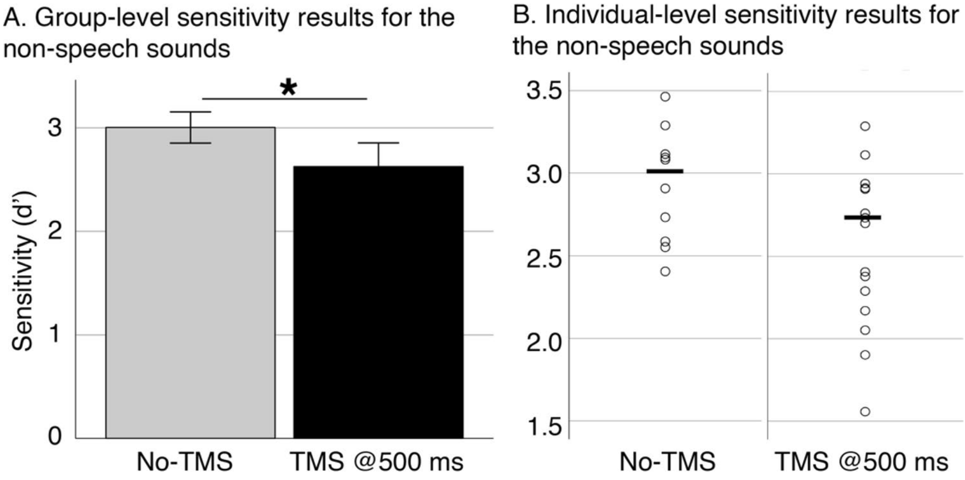

Fig. 3 TMS-induced changes in sensitivity for the non-speech sounds. **a** The bar charts illustrate the decomposition of the Stimuli*Stimulation time interaction for *d*′. The *y* axis represents sensitivity measured as *d*′. Error bars represent the standard error of the mean. Asterisks indicate a significant difference. **b** The individual value plots show the *d*′ values for each subject. Positive values indicate better sensitivity. The median sensitivity is represented as a thick black line

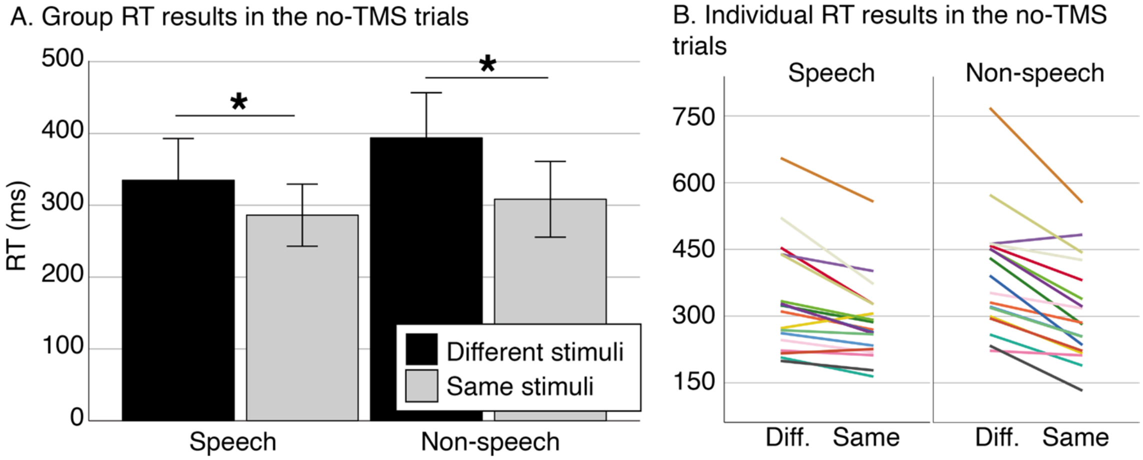

Fig. 4 Priming effect during the no-TMS trials. **a** The bar graphs representing reaction times (RT) for identical and different stimuli during the no-TMS trials for speech and non-speech stimuli. Lower values indicate a reduced processing time for repeated stimuli (priming). Error bars represent the standard error of the mean. Asterisks indicate a significant difference. **b** The line charts illustrate individual RT data for the different and same stimulus trials. Each line represents one subject

Table 3 Priming expressed as the difference in $\log_{\text{base } 10}$ transformed RT between non-identical and identical stimuli during no-TMS

Stimuli	Mean	SEM	95% CI	*t* value	*df*	*p* value	Cohen's *d*
Speech	.06	.01	[.04, .09]	5.04	17	< .001	1.33
Non-speech	.11	.02	[.08, .15]	6.98	17	< .0001	1.55

MS milliseconds, *SEM* standard error of the mean, *CI* confidence interval, *df* degrees of freedom

Priming

First, we validated that priming was indeed observed on RT in the absence of TMS. To do so, the RTs were first transformed to the log of 10 because they were not normally distributed. Following the transformation, we compared the RT of non-identical stimuli to the RT of identical stimuli using paired-sample *t* tests, separately for the speech and non-speech trials. As illustrated in Fig. 4 and detailed in Table 3, the results demonstrated that non-identical stimuli took longer to process than identical stimuli regardless of Stimuli (speech: $t_{17} = 5.04$, $p < .001$, 95% CI [.04, .09], Cohen's $d = 1.33$; non-speech: $t_{17} = 6.98$, $p < .0001$, 95% CI [.08, .15], Cohen's $d = 1.55$). Next, we examined whether non-speech stimuli regardless of whether the sequences were identical or different took longer to process than speech stimuli in the absence of TMS. While both speech sounds and non-speech sounds were processed relatively fast (speech sounds mean:

Table 4 Mean (SEM) priming (difference in RT between non-identical and identical stimuli)

Stimuli	pIFG		aSMG	
	TMS@250—no-TMS@250	TMS@500—no-TMS@500	TMS@250—no-TMS@250	TMS@500—no-TMS@500
Speech	65.26 (21.5) ms	−70.3 (22.6) ms	60.3 (29.9) ms	−52.1 (14.9) ms
Non-speech	−6.4 (38.6) ms	−34 (22.9) ms	−3.1 (21.5) ms	−41.3 (22.8) ms

SEM standard error from the mean

Table 5 Results of the repeated measures ANOVA for priming

Effects	Type II sum of squares	*f* value	*df*	*p* value	n_p^2
Region	19,655.39	.07	1,17	.8	.004
Stimuli	1,739,627.93	1.53	1,17	.23	.08
Time	22,134,226	17.41	1,17	**.001**	.51
Region*Stimuli	67,380.30	.06	1,17	.80	.004
Region*Time	36,129.22	.03	1,17	.88	.001
Stimuli*Time	7,472,528.33	5.65	1,17	**.03**	.25
Region*Stimuli*Time	256,813.08	.20	1,17	.66	.01

Significant effects are in bold. n_p^2 partial eta square, d*f* degrees of freedom

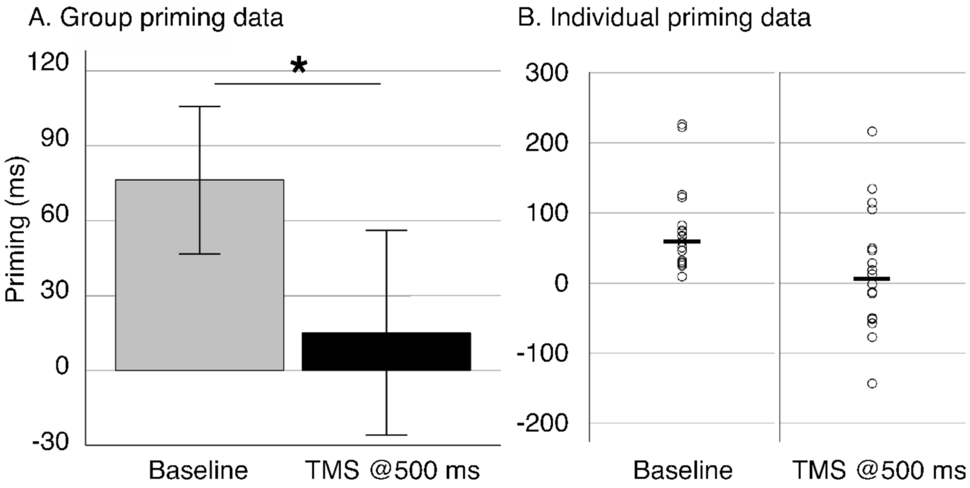

Fig. 5 TMS-induced changes in priming. **a** The bar charts illustrate the decomposition of the Stimuli*Time interaction for priming. The *y* axis represents priming in milliseconds (different–same stimuli) for speech sounds. Error bars represent the standard error of the mean. Asterisks indicate a significant difference. **b** The individual value plots show priming in ms for each subject. Positive values indicate a stronger priming effect. The median priming values are represented as thick black lines

RT = 310.57 ms, SD = 105.8 ms, non-speech sounds mean: 351.14, SD = 110.76), a paired-sample *t* test demonstrated that the non-speech sounds took significantly more time to process than the speech sounds ($t_{17} = 4.24$, $p = .00005$, 95% CI −60.76, −20.4], Cohen's $d = .90$).

The priming scores representing the difference between identical and different stimuli for each condition are listed in Table 4. The result of the repeated-measures ANOVA for priming is summarized in Table 5 and Online Resource 5. A main effect of Time was found (F (1, 17) = 17.41, $p = .001$, $n_p^2 = .51$), with participants exhibiting less priming following stimulation at TMS@500. The Stimuli*Time interaction was significant (F (1, 17) = 5.65, $p = .03$, $n_p^2 = .25$). A simple-effect analysis revealed a significant difference between the two stimulation times only for the speech stimuli (F (1, 17) = 23.95, $p < .0001$, $n_p^2 = .58$), with participants exhibiting reduced priming following stimulation at TMS@500 (Fig. 5). No significant difference between the two stimulation times was found for the non-speech stimuli (F (1, 17) = 1.35, $p = .26$, $n_p^2 = .074$). To gain a better understanding of these results, a series of Bonferroni-corrected paired-sample *t* tests was conducted for speech sounds in TMS and no-TMS trials at each time point to determine whether TMS trials were significantly different from no-TMS trials. To ensure that the effect was not driven by differences between no-TMS trials, additional paired-sample *t* tests were also conducted to compare TMS@250 and TMS@500 as well as no-TMS@250 and no-TMS@500 trial (level of significant needed is .0125, .05/4 comparisons). A strong and significant difference was found for speech sounds at

TMS@500 ($t_{17} = 4.42$, $p = .0004$, 95% CI [31.93, 90.44], Cohen's $d = 1.04$) and between TMS trials (TMS@250 — TMS@500) ($t_{17} = -3.15$, $p = .005$, 95% CI [226.21, 1145.98], Cohen's $d = .74$). No other comparisons reached significance. These results demonstrate that the source of the effect observed for speech sounds is due to the application of TMS at 500 ms into the delay (TMS@500). To investigate the nature of the TMS effect, that is, whether the reduction in priming that was observed was caused by an inhibitory effect of TMS (i.e., an increase in RTs for identical sequences during TMS trials) or a facilitatory effect (i.e., a decrease in RTs for different sequences during TMS trials), two paired-sample *t* tests were conducted between no-TMS@500 and TMS@500 trials for both identical and different sequences. A significant difference was found for the different sequences ($t_{17} = 3.33$, $p = .004$, 95% CI [17.76, 79.19], Cohen's $d = .78$) but not for the identical sequences ($t_{17} = .42$, $p = .68$, 95% CI [−73.98, 48.56], Cohen's $d = -.42$). Thus, TMS@500 facilitated the processing of non-identical sequences, thereby reducing the priming effect.

Discussion

The present experiment aimed to investigate the scope (domain specific vs. domain general) of auditory WM mechanisms during the processing of auditory verbal and non-verbal information. Based on previous studies, we selected two core regions of the phonological loop, the left aSMG and the left pIFG. We hypothesized that the left aSMG would exhibit domain-general sensitivity because of its involvement in the temporary storage of auditory verbal and non-verbal information, and that the pIFG would exhibit domain-specific sensitivity because of its involvement in articulatory rehearsal. We also expected that TMS applied to the aSMG during the "storage phase" (TMS@250) would disrupt performance for both speech and non-speech sounds; whereas TMS applied to the pIFG during the "rehearsal phase" (TMS@500) would disrupt only speech sounds. Our hypotheses were partially verified since TMS to both regions disrupted performance for the speech and non-speech stimuli. There were three main findings: (1) TMS during the late part of the delay disrupted performance for both speech and non-speech independently of the target; these effects were strong (sensitivity: Cohen's $d = .97$; priming Cohen's $d = 1.04$); (2) specifically, TMS led to a decline in discrimination of *non-speech sounds* but (3) to reduced priming for the *speech sounds* (due to a facilitation effect).

The finding that TMS only disrupted performance during the later part of the delay is at odds with our initial hypothesis. The lack of TMS effect during the early phase of the delay (i.e., TMS@250) for both regions could indicate that, early into the delay, maintenance/articulatory rehearsal mechanisms may not be deployed yet, though this is in contrast with prior ERP evidence (Ruchkin et al. 1997). It is possible that encoding mechanisms may still be at play within the first 500 ms of the delay (Schiller et al. 2003). The results of the current experiment support this hypothesis as an effect of TMS was only observed during the late phase of the delay for non-speech and speech stimuli. These findings are congruent with the results of a TMS experiment that reported TMS effects 300 ms post delay onset during an auditory verbal and non-verbal n-back tasks (Imm et al. 2008). Future single-pulse TMS studies using longer intervals between stimuli (> 1 s) as well as a broader set of intervals for stimulation within the delay (for example, every 100 ms for a second) are needed to map the temporal unfolding of WM mechanisms and better understand the neural architecture of WM.

The findings that TMS has a different impact on sensitivity and RT as a function of stimulus type (speech, non-speech) and that TMS reduced priming by facilitating the detection of different sequences were also unexpected. Our original hypothesis was that TMS would disrupt performance by reducing or abolishing priming for both speech and non-speech sounds. It is possible that a difference in the vulnerability of the mechanisms involved in processing/storing speech and non-speech within the phonological loop may account for this different impact. Moreover, we found that non-speech sounds were harder to process than speech sounds, as indicated by globally slower RTs (speech sounds mean: RT = 310.57 ms, SD = 105.8 ms, non-speech sounds mean: 351.14, SD = 110.76, a 13% increase) and lower discrimination scores (speech: mean $d' = 3.35$, SD = .13; non-speech: mean $d' = 2.93$ SD = .34, a 12.5% decrease) It is possible that non-speech sounds were processed too slowly to be affected by TMS in terms of RT. The fact that discriminability was affected only for the non-speech sounds may reflect a harder processing difficulty level. Future TMS studies using different types of non-speech sounds (e.g., tones, whistling, humming) are needed to investigate the source of this difference and to reveal underlying mechanisms. Though discrimination was high in both conditions, matching performance between verbal and non-verbal material may help dissociate difficulty from processing-related differences.

As for the facilitatory effect of TMS observed in the current study, state-dependent TMS studies have shown that TMS can reduce priming by selectively enhancing the detection of non-primed targets. It has been suggested that in state-dependent priming paradigms, neurons that are tuned to the prime are more active at the time of the TMS application than neurons tuned to the non-prime target. The application of TMS enhances the processing of the less active neurons prior to the presentation of the target stimulus, thereby facilitating the detection of non-primed targets and leaving primed targets unaffected (Cattaneo et al. 2008; Silvanto et al. 2007; Silvanto and Pascual-Leone 2008). Although these studies are based on state-dependent priming paradigms, it is plausible that in the current study, TMS introduced "neuronal noise, a term that describes the ''random'' activity of neurons that is not associated with the encoding of behaviourally relevant variables." (Miniussi et al. 2010, p. 129). In non-linear system, such as the brain, the addition of noise can increase the distinctiveness between the prime and non-primed target, akin to stochastic resonance (Miniussi et al. 2010). It is also congruent with a study on auditory working memory, which found that the addition of white noise can improve performance on a recall task because the items become more dissimilar (Othman et al. 2019).

Parallel processing within the phonological loop?

Our finding that TMS to the left aSMG and left pIFG during the late part of the delay (TMS@500) affected both speech and non-speech sounds, suggests that these regions may operate in parallel, at least during the first 500-ms window and that damage to either part of the system leads to a breakdown. The hypothesis that WM regions operate in parallel is consistent with a recent fMRI study by Fegen et al. (2015), in which the authors demonstrated that the pIFG and area Spt were sensitive to memory load (i.e., number of letters recalled) and rehearsal rates during a vWM task, suggesting that these regions are involved in the maintenance/articulatory rehearsal of information. Importantly, the authors used a long delay (44 s), which was divided into three equal time bins. The comparison of pIFG and Spt activation levels across these bins revealed no significant difference. The author interpreted these findings as suggesting that maintenance/articulatory rehearsal mechanisms may operate in parallel. Here, it is important to note that area Spt, not the aSMG was identified as the anatomical locus of the phonological store. This discrepancy will be addressed in the next section (i.e., auditory verbal and non-verbal information and the phonological loop).

It is worth noting that other WM components have been dissociated in time using single-pulse TMS. Specifically, a single pulse TMS study by Mottaghy et al. (2003) documented a detrimental effect of TMS on accuracy 260 ms post-stimulus onset during a visuo-verbal n-back task when TMS was applied to the left middle frontal gyrus (MFG) and 220 ms post-stimulus onset when TMS was applied to the left IPS. These results suggest that the left MFG and left IPS are involved in different mechanisms during visual verbal n-back tasks, or at least that they are involved in processes that operate sequentially within the first 260 ms post-stimulus onset. Mottaghy and colleagues suggested that the left MFG may be involved with processes associated with the central executive system, while the IPS may be involved with storage processes. Thus, it is possible that WM mechanisms associated with different components (e.g., central executive vs. phonological loop) are easier to dissociate in time compared to WM mechanisms that are associated with the same component (i.e., phonological loop), which may occur in parallel or closer together.

Another possibility is that, in the present study, WM mechanisms and auditory sensory memory mechanisms were involved. Auditory sensory memory (ASM) has been defined as a modality-specific, passive store not involving active manipulation or rehearsal, unlike WM, that can be assessed during a task to compare sounds (Nees 2016). It was recently suggested that ASM operates in parallel with WM mechanisms, and that ASM mechanisms can be observed up to 20 s after the presentation of a stimulus (Nees 2016). The short delay used in the current study does not allow us to separate ASM from WM mechanisms. As such, future single-pulse TMS studies using either longer delays (> 20 s) or the presentation of an irrelevant auditory stimulus just after the target sound to overwrite the ASM trace are necessary to determine whether WM mechanisms can be dissociated from ASM mechanisms.

Auditory verbal and non-verbal information and the phonological loop

The question of the role that the pIFG and the aSMG play during auditory verbal and non-verbal tasks remains unanswered. Are they part of the phonological loop or do they support other functions involved in auditory WM tasks (e.g., ASM, attentional processes, sensorimotor processes)? If these two regions support functions involved in WM tasks, perhaps, as suggested by some authors, the phonological store should not be viewed as a distinct construct localizable to a specific anatomical area, but rather as the by-product of regions that collaborate during WM tasks and support speech perception and production (Buchsbaum and D'Esposito 2008, 2019; Macken and Jones 2003). Each of these possibilities will be discussed in the following paragraphs.

In a recent TMS study conducted by Hartwigsen et al. (2016), it was found that the combination of offline aSMG TMS prior to online pIFG rTMS leads to the same detrimental effect on performance during a phonological decision task (i.e., counting syllables) than the one observed when stimulating only one of these two regions while performing the same phonological task. This finding suggests that each region makes a unique contribution to phonological decision tasks and that stimulation to one of these two regions is sufficient to disrupt performance. The authors interpreted these findings as suggesting that these regions (aSMG and pIFG) might play a role in working memory. Based on previous TMS studies, the authors suggested that the aSMG might be involved in the maintenance of information in the phonological store and that the pIFG might be involved in subarticulatory rehearsal. Of particular interest to the current study is the finding that the syllable counting task used by Hartwigsen et al. (2016), in addition to being a phonological task, also required WM and ASM. That is, in order to be able to count the number of syllables in a word, participants had to rely on the short-term storage of phonological information as well as rehearsal processes. Similar to the current study, their task also recruited WM and ASM processes. It is, therefore, possible that the aSMG and the pIFG are involved in parallel processes; the aSMG in ASM processes and the pIFG in WM processes (i.e., articulatory rehearsal). This also raises the possibility that perhaps the "phonological store" as described by Baddeley and Hitch is not a distinct construct, but rather the by-product of regions within sensorimotor and cognitive networks (WM, attentional) that collaborates during WM tasks. Stimulation to either of these regions would affect performance on WM tasks. This is the notion of the phonological store presented by Buchsbaum and D'Esposito (2008, 2019), and others (Jones et al. 2007; Macken and Jones 2003).

In the current study, stimulation to the aSMG during the late phase of the delay disrupted the performance for both speech and non-speech sounds (Cohen's $d = .97$; priming Cohen's $d = 1.04$). These results are consistent with previous TMS studies that have documented an effect of TMS and tDCS during verbal (Deschamps et al. 2014; Kirschen et al. 2006; Romero et al. 2006) and non-verbal auditory WM tasks (Schaal et al. 2015a, b; Vines et al. 2006). The finding that TMS also disrupted the performance for non-speech sounds raises the possibility that the aSMG is involved in the short-term maintenance of auditory information, a domain-general mechanism. It should be noted that based on the paradigm used in this study, it is not possible to dissociate whether the short-term maintenance of auditory information is part of the phonological loop or ASM. In the current study, to determine whether two sequences were identical or not, regardless of content (speech vs. non-speech), participants could either rely on the lingering ASM trace or the rehearsed WM trace, as both memory traces exist in parallel (Nees 2016). To be able to dissociate ASM from WM, a longer delay (> 20 s) or overwriting the auditory trace by presenting an irrelevant auditory stimulus just after the first stimuli is needed.

Another possibility that cannot be discarded is that TMS may not be able to dissociate processes related to the maintenance of speech and non-speech sounds in memory if they occur within adjacent subregions of the aSMG. Given that cytoarchitectonic and receptor architectonics analyses of the inferior parietal lobule have identified four subregions (area PF, PFop, PFt and PFcm) within the aSMG (Caspers et al. 2008, 2012) and that the resolution of TMS ranges between .5 and 1 cm (Sliwinska et al. 2014), if the stimulation site used in the current study overlapped two subregions, it might have affected the maintenance of speech and non-speech sounds simultaneously.

We also observed a disruption in the performance following stimulation to the left pIFG during the later phase of the delay for speech and non-speech sounds (Cohen's $d = .97$; priming Cohen's $d = 1.04$). While our results converge with previous TMS studies that have implicated the pIFG in articulatory rehearsal mechanisms during vWM tasks (Hartwigsen et al. 2010a, b; Nixon et al. 2004; Romero et al. 2006) and phonological judgments (Burton et al. 2000, 2005), the finding of an effect of TMS on non-speech stimuli is novel. One possibility is that participants tried to rehearse the sound they heard, as a strategy to facilitate discrimination. This hypothesis is congruent with a recent TMS study by Liao et al. (2014), which demonstrated that the left primary motor cortex (M1) was involved in motor-based strategies for rehearsal of verbal (i.e., words and pseudowords) but also non-verbal stimuli (i.e., foreign Chinese characters). Thus, the finding that verbal and non-verbal stimuli engage similar rehearsal mechanisms supports the hypothesis that WM mechanisms are domain general. This finding is also congruent with the idea that the phonological loop in humans evolved from an auditory–vocal interface also found in non-human primates to support language learning capacities. According to this evolutionary perspective, the evolution in humans of the auditory–vocal interface into the phonological loop also resulted in enhanced auditory working memory to support speech as we know it today. Most importantly, since the auditory–vocal interface interacted with other sensory modalities, it supports the notion that the phonological loop is also involved in non-speech processes (Aboitiz 2018; Michon et al. 2019).

Domain general models of memory (e.g. Atkinson and Shiffrin 1968; Barrouillet et al. 2004; Cowan 2001) propose that the same general memory principles (i.e., encoding, recall, etc.) apply to various stimulus modalities. Along the same conceptual line, it has recently been argued that language studies tend to focus on language-specific interpretations of results; whereas more basic, domain general computations and neurobiological mechanisms may provide valid alternative interpretations for language-related findings (Hasson, Egidi, Marelli, and Willems 2018). We suggest that the finding of a general contribution of the pIFG in rehearsal of auditory stimuli should be interpreted as evidence of domain-general mechanisms underlying the rehearsal of auditory information.

Conclusion

Our results demonstrate that TMS delivered to two important WM regions, the left pIFG and the left aSMG, impairs performance during the processing of speech and non-speech stimuli. These findings suggest the existence of domain-general auditory WM maintenance mechanisms that process both speech and non-speech stimuli. Yet, the nature of the disruptions found for each type of sound was different, with a reduction in RT for speech sounds and a decrease in accuracy for non-speech sounds, which suggests different computations being performed on these different inputs or, alternatively, different time courses. Future TMS studies are needed to replicate these results and explore the time course of auditory WM processes to gain a more thorough understanding the underlying mechanisms, as well as examine whether other WM processes (e.g., storage, manipulation) are domain general, by exploring the effect of modality and type of WM process (n-back, delayed discrimination, serial recall). Such knowledge is needed to inform current neurobiological models of WM and eventually contribute to guiding clinical interventions for WM deficits in adults and elderly populations.

Acknowledgements We thank all the participants. This study was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) (Grant No. 435774-2013) and from the Canadian foundation for innovation (Grant No. 31408) to P.T., who also holds a Career Awards from the ‘‘Fonds de Recherche du Québec – Santé’’ (FRQS) (Grant No. 35016). Support for MRI data acquisition was provided by the “Centre intégré en neuroimagerie et neurostimulation de Québec” (CINQ) via a platform support grant from the Brain Canada Foundation (#3456). ID is currently a faculty at Georgian College in Ontario, Canada.

Authors' contributions ID: Conceptualization, Methodology, Investigation, Project administration, Formal analysis, Visualization, Writing—Original Draft. MC: Investigation, Writing—Reviewing and Editing. AD: Writing—Reviewing and Editing. PT: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Resources, Writing—Reviewing and Editing, Visualization, Data Curation.

References

Aboitiz F (2018) A brain for speech. Evolutionary continuity in primate and human auditory-vocal processing. Front Neurosci. https://doi.org/10.3389/fnins.2018.00174

Atkinson RC, Shiffrin RM (1968) Human memory: a proposed system and its control processes. Psychol Learn Motiv 2:89–195

Awh E, Jonides EE, Schumacher EH, Koeppe RA, Katz S (1996) Dissociation of storage and rehearsal in verbal working memory: evidence from positron emission tomography. Psychol Sci 7:25–31

Baddeley A (1975) Word length and the structure of short-erm memory. J Verbal Learn Verbal Behav 14(6):575–589

Baddeley A (1986) Working Memory. Clarendon Press, Oxford

Baddeley A (1992) Working memory. Science 255(5044):556–559

Baddeley A (2003) Working memory and language: an overview. J Commun Disord 36(3):189–208

Baddeley A (2012) Working memory: theories, models, and controversies. Annu Rev Psychol 63:1–29

Baddeley A, Hitch GJ (1974) Working memory. In: Bower GA (ed) The psychology of learning and motivation: Advances in research and theory. Academic, New York

Barrouillet P, Bernardin S, Camos V (2004) Time constraints and resource sharing in adults' working memory spans. J Exp Psychol. https://doi.org/10.1037/0096-3445.133.1.83

Bergerbest D, Ghahremani DG, Gabrieli JD (2004) Neural correlates of auditory repetition priming: reduced fMRI activation in the auditory cortex. J Cogn Neurosci 16(6):966–977. https://doi.org/10.1162/0898929041502760

Boersma P, Weenink D (2011). Praat: doing phonetics by computer (5.1.42). 5.1.42

Buchsbaum BR, D'Esposito M (2008) The search for the phonological store: from loop to convolution. J Cogn Neurosci 20(5):762–778

Buchsbaum Bradley R, D'Esposito M (2019) A sensorimotor view of verbal working memory. Cortex. https://doi.org/10.1016/j.cortex.2018.11.010

Buchsbaum BR, Hickok G, Humphries C (2001) Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. Cogn Sci 25:663–678

Burton MW, Small SL, Blumstein SE (2000) The role of segmentation in phonological processing: an fMRI investigation. J Cogn Neurosci 12(4):679–690

Burton MW, Locasto PC, Krebs-Noble D, Gullapalli RP (2005) A systematic investigation of the functional neuroanatomy of auditory and visual phonological processing. Neuroimage 26(3):647–661. https://doi.org/10.1016/j.neuroimage.2005.02.024

Caspers S, Eickhoff SB, Geyer S, Scheperjans F, Mohlberg H, Zilles K, Amunts K (2008) The human inferior parietal lobule in stereotaxic space. Brain Struct Funct 212(6):481–495

Caspers S, Schleicher A, Bacha-Trams M, Palomero-Gallagher N, Amunts K, Zilles K (2012) Organization of the human inferior parietal lobule based on receptor architectonics. Cereb Cortex 12(4):679–690

Cattaneo L (2010) Tuning of ventral premotor cortex neurons to distinct observed grasp types: a TMS-priming study. Exp Brain Res 207(3–4):165–172. https://doi.org/10.1007/s00221-010-2454-5

Cattaneo Z, Rota F, Vecchi T, Silvanto J (2008) Using state-dependency of transcranial magnetic stimulation (TMS) to investigate letter selectivity in the left posterior parietal cortex: a comparison of TMS-priming and TMS-adaptation paradigms. Eur J Neurosci 28(9):1924–1929. https://doi.org/10.1111/j.1460-9568.2008.06466.x

Chao LL, Nielsen-Bohlman L, Knight RT (1995) Auditory event-related potentials dissociate early and late memory processes. Electroencephalogr Clin Neurophys. https://doi.org/10.1016/0168-5597(94)00256-E

Chen SH, Desmond JE (2005) Cerebrocerebellar networks during articulatory rehearsal and verbal working memory tasks. Neuroimage 24(2):332–338. https://doi.org/10.1016/j.neuroimage.2004.08.032

Cowan N (2001) The magical number 4 in short-term memory: a reconsideration of mental storage capacity. Behav Brain Sci. https://doi.org/10.1017/S0140525X01003922

Deschamps I, Baum SR, Gracco VL (2014) On the role of the supramarginal gyrus in phonological processing and verbal working memory: evidence from rTMS studies. Neuropsychologia 53:39–46. https://doi.org/10.1016/j.neuropsychologia.2013.10.015

Devlin JT, Matthews PM, Rushworth MF (2003) Semantic processing in the left inferior prefrontal cortex: a combined functional magnetic resonance imaging and transcranial magnetic stimulation study. J Cogn Neurosci 15(1):71–84

Fegen D, Buchsbaum BR, D'Esposito M (2015) The effect of rehearsal rate and memory load on verbal working memory. Neuroimage 105:120–131. https://doi.org/10.1016/j.neuroimage.2014.10.034

Fiez JA, Raife EA, Balota DA, Schwarz JP, Raichle ME, Petersen SE (1996) A positron emission tomography study of the short-term maintenance of verbal information. J Neurosci 16(2):808–822

Gaab N, Gaser C, Zaehle T, Jancke L, Schlaug G (2003) Functional anatomy of pitch memory–an fMRI study with sparse temporal sampling. Neuroimage 19(4):1417–1426

Gaskell MG, Marslen-Wilson WD (1997) Integrating form and meaning: a distributed model of speech perception. Lang Cogn Process 12(5–6):613–656

Gaskell MG, Marslen-Wilson WD (2002) Representation and competition in the perception of spoken words. Cogn Psychol 45(2):220–266

Gitelman DR, Nobre AC, Sonty S, Parrish TB, Mesulam MM (2005) Language network specializations: an analysis with parallel task designs and functional magnetic resonance imaging. Neuroimage 26(4):975–985. https://doi.org/10.1016/j.neuroimage.2005.03.014

Gough PM, Nobre AC, Devlin JT (2005) Dissociating linguistic processes in the left inferior frontal cortex with transcranial magnetic stimulation. J Neurosci 25(35):8010–8016

Gruber O (2000) Two different brain systems underlie phonological short-term memory in humans. Neuroimage 5(11):S407

Hartwigsen G, Baumgaertner A, Price CJ, Koehnke M, Ulmer S, Siebner HR (2010a) Phonological decisions require both the left and right supramarginal gyri. Proc Natl Acad Sci U S A 107(38):16494–16499

Hartwigsen G, Price CJ, Baumgaertner A, Geiss G, Koehnke M, Ulmer S, Siebner HR (2010b) The right posterior inferior frontal gyrus contributes to phonological word decisions in the healthy brain: evidence from dual-site TMS. Neuropsychologia 48(10):3155–3163

Hartwigsen G, Weigel A, Schuschan P, Siebner HR, Weise D, Classen J, Saur D (2016) Dissociating Parieto-Frontal Networks for Phonological and Semantic Word Decisions: a Condition-and-Perturb TMS Study. Cereb Cortex. https://doi.org/10.1093/cercor/bhv092

Hasson U, Egidi G, Marelli M, Willems RM (2018) Grounding the neurobiology of language in first principles: the necessity of non-language-centric explanations for language comprehension. Cognition 180:135–157. https://doi.org/10.1016/j.cognition.2018.06.018

Henson RN, Burgess N, Frith CD (2000a) Recoding, storage, rehearsal and grouping in verbal short-term memory: an fMRI study. Neuropsychologia 38(4):426–440

Henson RN, Shallice T, Dolan R (2000b) Neuroimaging evidence for dissociable forms of repetition priming. Science 287(5456):1269–1272

Herwig U, Abler B, Schönfeldt-Lecuona C, Wunderlich A, Grothe J, Spitzer M, Walter H (2003) Verbal storage in a premotor–parietal network: evidence from fMRI-guided magnetic stimulation. Neuroimage 20(2):1032–1041

Hickok G, Buchsbaum BR, Humphries C, Muftuler T (2003) Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. J Cogn Neurosci 15(5):673–682

Hitch GJ, Baddeley AD (1976) Verbal reasoning and working memory. The Quarterly Journal of Experimental Psychology 28(4):603–621

Hutchins S, Palmer C (2008) Repetition priming in music. J Exp Psychol Hum Percept Perform 34(3):693–707. https://doi.org/10.1037/0096-1523.34.3.693

Imm J-H, Kang E, Youn T, Park H, Kim J Il, Kang JI et al (2008) Different hemispheric specializations for pitch and audioverbal working memory. NeuroReport 19(1):99–103

Jacquemot C, Scott SK (2006) What is the relationship between phonological short-term memory and speech processing? Trends Cogn Sci 10(11):480–486

Jones DM, Macken WJ (1993) Irrelevant tones produce an irrelevant speech effect: implications for phonological coding in working memory. J Exp Psychol Learn Mem Cogn 19(2):369

Jones Dylan M, Hughes RW, Macken WJ (2007) The phonological store abandoned. Quarterly Journal of Experimental Psychology. https://doi.org/10.1080/17470210601147598

Kirschen MP, Davis-Ratner MS, Jerde TE, Schraedley-Desmond P, Desmond JE (2006) Enhancement of phonological memory following transcranial magnetic stimulation (TMS). Behav Neurol 17(3–4):187–194

Kirschen MP, Chen SHA, Desmond JE (2010) Modality specific cerebro-cerebellar activations in verbal working memory: an fMRI study. Behav Neurol. https://doi.org/10.3233/BEN-2010-0266

Koelsch S, Schulze K, Sammler D, Fritz T, Muller K, Gruber O (2009) Functional architecture of verbal and tonal working memory: an FMRI study. Hum Brain Mapp 30(3):859–873. https://doi.org/10.1002/hbm.20550

Kouider S, Dupoux E (2005) Subliminal speech priming. Psychol Sci 16(8):617–625. https://doi.org/10.1111/j.1467-9280.2005.01584.x

Kumar S, Joseph S, Gander PE, Barascud N, Halpern AR, Griffiths TD (2016) A brain system for auditory working memory. J Neurosci 36(16):4492–4505. https://doi.org/10.1523/JNEUROSCI.4341-14.2016

Liao DA, Kronemer SI, Yau JM, Desmond JE, Marvel CL (2014) Motor system contributions to verbal and non-verbal working memory. Front Hum Neurosci 753:1–8

Light LL, La Voie D, Kennison R (1995) Repetition priming of nonwords in young and older adults. J Exp Psychol Learn Mem Cogn 21(2):327–346

Macken WJ, Jones DM (2003) Reification of phonological storage. Q J Exp Psychol Sect A. https://doi.org/10.1080/02724980245000052

Macmillan NA, Creelman CD (1990) Response bias: characteristics of detection theory, threshold theory, and" nonparametric" indexes. Psychol Bull 107(3):401

Macmillan NA, Creelman CD (2004) *Detection theory: A user's guide*. Psychology press, London

Martin RC (2005) Components of short-term memory and their relation to language processing: evidence from neuropsychology and neuroimaging. Curr Direct Psychol Sci 14(4):204–208

Martin RC, Shelton JR, Yaffee LS (1994) Language processing and working memory: neuropsychological evidence for separate phonological and semantic capacities. J Mem Lang 33(1):83–111

Martin RC, Wu D, Freedman M, Jackson EF, Lesch M (2003) An event-related fMRI investigation of phonological versus semantic short-term memory. J Neuroling 16(4):341–360

Martinkauppi S, Rama P, Aronen HJ, Korvenoja A, Carlson S (2000) Working memory of auditory localization. Cereb Cortex 10(9):889–898

Marvel CL, Desmond JE (2012) From storage to manipulation: how the neural correlates of verbal working memory reflect varying demands on inner speech. Brain Lang 120(1):42–51. https://doi.org/10.1016/j.bandl.2011.08.005

McClelland JL, Elman JL (1986) The TRACE model of speech perception. Cogn Psychol 18(1):1–86

McDermott KB, Petersen SE, Watson JM, Ojemann JG (2003) A procedure for identifying regions preferentially activated by attention to semantic and phonological relations using functional magnetic resonance imaging. Neuropsychologia 41(3):293–303

Michon M, López V, Aboitiz F (2019) Origin and evolution of human language. Prog Brain Res. https://doi.org/10.1016/bs.pbr.2019.01.005

Miniussi C, Ruzzoli M, Walsh V (2010) The mechanism of transcranial magnetic stimulation in cognition. Cortex. https://doi.org/10.1016/j.cortex.2009.03.004

Mottaghy FM, Gangitano M, Krause BJ, Pascual-Leone A (2003) Chronometry of parietal and prefrontal activations in verbal working memory revealed by transcranial magnetic stimulation. Neuroimage 18(3):565–575

Nees MA (2016) Have we forgotten auditory sensory memory? Retention intervals in studies of nonverbal auditory working memory. Front Psychol 7:1892

Nixon P, Lazarova J, Hodinott-Hill I, Gough P, Passingham R (2004) The inferior frontal gyrus and phonological processing: an investigation using rTMS. J Cogn Neurosci 16(2):289–300

Norris D (1994) Shortlist: a connectionist model of continuous speech recognition. Cognition 52(3):189–234

Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9(1):97–113

Othman E, Yusoff AN, Mohamad M, Abdul Manan H, Giampietro V, Abd Hamid AI et al (2019) Low intensity white noise improves performance in auditory working memory task: an fMRI study. Heliyon. https://doi.org/10.1016/j.heliyon.2019.e02444

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

Pechmann T, Mohr G (1992) Interference in memory for tonal pitch: implications for a working-memory model. Mem Cognit 20(3):314–320

Rodriguez-Jimenez R, Avila C, Garcia-Navarro C, Bagney A, Aragon AM, Ventura-Campos N et al (2009) Differential dorsolateral prefrontal cortex activation during a verbal n-back task according to sensory modality. Behav Brain Res 205(1):299–302. https://doi.org/10.1016/j.bbr.2009.08.022

Romei V, Thut G, Silvanto J (2016) Information-based approaches of noninvasive transcranial brain stimulation. Trends Neurosci 39(11):782–795. https://doi.org/10.1016/j.tins.2016.09.001

Romero L, Walsh V, Papagno C (2006) The neural correlates of phonological short-term memory: a repetitive transcranial magnetic stimulation study. J Cogn Neurosci 18(7):1147–1155

Rossi S, Hallett M, Rossini PM, Pascual-Leone A (2009) Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. Clin Neurophysiol 120(12):2008–2039

Rossini PM, Barker AT, Berardelli A, Caramia MD, Caruso G, Cracco RQ et al (1994) Non-invasive electrical and magnetic stimulation of the brain, spinal cord and roots: basic principles and procedures for routine clinical application Report of an IFCN committee. Electroencephalogr Clin Neurophysiol 91(2):79–92

Ruchkin DS, Berndt RS, Johnson R Jr, Ritter W, Grafman J, Canoune HL (1997) Modality-specific processing streams in verbal working memory: evidence from spatio-temporal patterns of brain activity. Brain Res Cogn Brain Res 6(2):95–113

Rueckl JG (1990) Similarity effects in word and pseudoword repetition priming. J Exp Psychol Learn Mem Cogn 16(3):374–391

Salamé P, Baddeley A (1982) Disruption of short-term memory by unattended speech: implications for the structure of working memory. J Verbal Learn Verbal Behav 21(2):150–164

Salamé P, Baddeley A (1986) Phonological factors in STM: similarity and the unattended speech effect. Bull Psychon Soc 24(4):263–265

Salamé P, Baddeley A (1989) Effects of background music on phonological short-term memory. Q J Exp Psychol 41(1):107–122

Schaal NK, Williamson VJ, Banissy MJ (2013) Anodal transcranial direct current stimulation over the supramarginal gyrus facilitates pitch memory. Eur J Neurosci 38(10):3513–3518. https://doi.org/10.1111/ejn.12344

Schaal NK, Krause V, Lange K, Banissy MJ, Williamson VJ, Pollok B (2015a) Pitch memory in nonmusicians and musicians: revealing functional differences using transcranial direct current stimulation. Cereb Cortex 25(9):2774–2782. https://doi.org/10.1093/cercor/bhu075

Schaal NK, Williamson VJ, Kelly M, Muggleton NG, Pollok B, Krause V, Banissy MJ (2015b) A causal involvement of the left supramarginal gyrus during the retention of musical pitches. Cortex 64:310–317. https://doi.org/10.1016/j.cortex.2014.11.011

Schiller NO, Bles M, Jansma BM (2003) Tracking the time course of phonological encoding in speech production: an event-related brain potential study. Brain Res Cogn Brain Res 17(3):819–831

Schulze K, Zysset S, Mueller K, Friederici AD, Koelsch S (2011) Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. Hum Brain Mapp 32(5):771–783

Shen J, Zhang G, Yao L, Zhao X (2015) Real-time fMRI training-induced changes in regional connectivity mediating verbal working memory behavioral performance. Neuroscience 289:144–152. https://doi.org/10.1016/j.neuroscience.2014.12.071

Silvanto J, Pascual-Leone A (2008) State-dependency of transcranial magnetic stimulation. Brain Topogr 21(1):1–10. https://doi.org/10.1007/s10548-008-0067-0

Silvanto J, Muggleton NG, Cowey A, Walsh V (2007) Neural adaptation reveals state-dependent effects of transcranial magnetic stimulation. Eur J Neurosci 25(6):1874–1881. https://doi.org/10.1111/j.1460-9568.2007.05440.x

Sliwinska MW, Vitello S, Devlin JT (2014) Transcranial magnetic stimulation for investigating causal brain-behavioral relationships and their time course. J Visual Exp. https://doi.org/10.3791/51735

Stevens KN (1972) Segments, features, and analysis by synthesis

Tremblay P, Baroni M, Hasson U (2012) Processing of speech and non-speech sounds in the supratemporal plane: auditory input preference does not predict sensitivity to statistical structure. Neuroimage 66C:318–332. https://doi.org/10.1016/j.neuroimage.2012.10.055

Tremblay P, Perron M, Deschamps I, Kennedy-Higgins D, Houde J-C, Dick AS, Descoteaux M (2018) The role of the arcuate and middle longitudinal fasciculi in speech perception in noise in adulthood. Hum Brain Mapp 40(1):226–241

Vines BW, Schnider NM, Schlaug G (2006) Testing for causality with transcranial direct current stimulation: pitch memory and the left supramarginal gyrus. NeuroReport 17(10):1047–1050. https://doi.org/10.1097/01.wnr.0000223396.05070.a2

Wasserman EM (1998) 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 5–7, 1996. Electroencephalogr Clin Neurophysiol 108:1–16

Weissgerber TL, Milic NM, Winham SJ, Garovic VD (2015) Beyond Bar and Line Graphs: time for a New Data Presentation Paradigm. PLoS Biol. https://doi.org/10.1371/journal.pbio.1002128

Wig GS, Grafton ST, Demos KE, Kelley WM (2005) Reductions in neural activity underlie behavioral components of repetition priming. Nat Neurosci 8(9):1228–1233. https://doi.org/10.1038/nn1515

Zatorre RJ, Evans AC, Meyer E (1994) Neural mechanisms underlying melodic perception and memory for pitch. J Neurosci 14(4):1908–1919