

Common cortical architectures for phonological working memory identified in individual brains



Terri L. Scott^a, Tyler K. Perrachione^{b,*}

^a Graduate Program for Neuroscience, Boston University, USA

^b Department of Speech, Language, and Hearing Sciences, Boston University, USA

ARTICLE INFO

Keywords:

fMRI
Individual-subject analyses
Nonword discrimination
Nonword repetition
Phonological working memory
Pseudowords

ABSTRACT

Phonological working memory is the capacity to briefly maintain and recall representations of sounds important for speech and language and is believed to be critical for language and reading acquisition. Whether phonological working memory is supported by fronto-parietal brain regions associated with short-term memory storage or perisylvian brain structures implicated in speech perception and production is unclear, perhaps due to variability in stimuli, task demands, and individuals. We used fMRI to assess neurophysiological responses while individuals performed two tasks with closely matched stimuli but divergent task demands—nonword repetition and nonword discrimination—at two levels of phonological working memory load. Using analyses designed to address inter-subject variability, we found significant neural responses to the critical contrast of high vs. low phonological working memory load in both tasks in a set of regions closely resembling those involved in speech perception and production. Moreover, within those regions, the voxel-wise patterns of load-related activation were highly correlated between the two tasks. These results suggest that brain regions in the temporal and frontal lobes encapsulate the core neurocomputational components of phonological working memory; an architecture that becomes increasingly evident as neural responses are examined in successively finer-grained detail in individual participants.

1. Introduction

A major endeavor of cognitive neuroscience has been to identify the brain regions responsible for various mental functions; however, many important behaviors and cognitive abilities are actually complex, higher-order operations that integrate multiple systems and thus seem to defy the sort of strict functional localization found in, for instance, visual motion processing or face perception (e.g. Kanwisher et al., 1997; Tootell et al., 1995). Nonetheless, understanding the broad, integrated brain bases of such multiple-component processes is a critical step towards building neural models of human behavior and cognition. One such operation is phonological working memory – the ability to briefly maintain and manipulate the sounds important for speech and language – which likely involves broad integration of a variety of perception, language, cognition, and motor regions (Fiez, 2015). While reductionist approaches can investigate each of the potential constituent operations of phonological working memory in isolation, the clinical importance of this integrated faculty to language development (Adams and Gathercole, 1995, 1996; Baddeley et al., 1998; Dufva et al., 2001; Gathercole and

Baddeley, 1993; Gathercole et al., 2006; Martin, 2005; van der Schuit et al., 2011) – and its impairment in numerous developmental and communication disorders (Bowers et al., 2018; Byrd et al., 2015; Graf Estes et al., 2007; Kjelgaard and Tager-Flusberg, 2001; Lanfranchi et al., 2009; Peter et al., 2011) – makes understanding its holistic and nuanced neural basis of particular importance.

The most widely used theoretical framework for conceptualizing phonological working memory is Baddeley's theory of working memory (Baddeley and Hitch, 1974; Baddeley, 1986, 2003), which describes a “phonological loop” that supports the ongoing maintenance of phonological information in the brain through temporary storage and recurring subvocal articulation. This model has been successful in explaining many aspects of behavior, such as the diminished working memory capacity for longer words (i.e., word length effect; Baddeley et al., 1975); however, attempts to map components of the model onto distinct brain regions have yielded conflicting results. Early studies using sequences of letters or numbers as stimuli suggested that the phonological storage component is supported by inferior parietal areas and that articulatory rehearsal occurs in Broca's area/adjacent ventral premotor cortex (Awh et al.,

* Corresponding author. 635 Commonwealth Ave., Boston, MA, 02215, USA.
E-mail address: tkp@bu.edu (T.K. Perrachione).

1996; Baddeley, 2003; Paulesu et al., 1993; Smith and Jonides, 1998).

However, recent studies utilizing tasks requiring the manipulation and maintenance of specifically *speech sounds*, and therefore plausibly more relevant to language development (see Gathercole et al., 1994, 2006), indicate that bilateral superior temporal gyri (STG) play a considerable role in phonological working memory (Acheson et al., 2011; Buchsbaum et al., 2005; Koenigs et al., 2011; McGettigan et al., 2011; Perrachione et al., 2017; Strand et al., 2008), despite these areas being most often implicated in lower-level speech and phonological processing. Strand et al., (2008) argued that the presence of inferior parietal activation in past studies of phonological working memory is due to the visual presentation of verbal information and showed in an event-related design that during working memory maintenance of auditorily presented nonwords,¹ peak responses are located in left inferior frontal gyrus (LIFG), superior frontal gyrus, superior temporal sulcus, and left putamen. McGettigan et al., (2011) reported positive neural correlates of increased phonological working memory load in left and right planum temporale (PT), as well as left precentral gyrus (LPreCG) during nonword repetition. During nonword discrimination, a task designed to isolate phonological working memory processes from the motoric processes required for repetition, Perrachione et al., (2017) identified bilateral superior temporal gyri (STG), supplementary motor area (SMA) and LIFG as regions sensitive to increasing phonological working memory load. These studies all used tasks closely approximating clinical assessments of phonological working memory and have been unable to identify any parietal regions containing a classical “phonological store.” However, even though each study reported positive neural correlates of increased working memory load of nonwords, different brain areas were reported from one study to the next, and this lack of consistency across studies undercuts inferences about the functions supported by these areas.

An obvious source of variability that contributes to the breadth of results from previous studies is the choice of behavioral tasks. What we know about the developmental and clinical significance of phonological working memory comes from highly standardized assessments in which the examinee is asked to quickly repeat successively longer nonwords. However, the various neuroimaging studies of phonological working memory have been less consistent in how they operationalize this construct, limiting our ability to generalize the core (vs. task-specific) computational architecture for phonological working memory, likely because tasks have differed in many ways, such as construction of nonword stimuli, temporal structure, and response demands (e.g., repetition vs. discrimination). Presumably, common neurocomputational processes underlie phonological working memory across specific operationalizations, but without a method for exploring concurrence (as opposed to disjunction) in functional activation across tasks, we are left to the fallacy of reverse inference to suppose which regions are activated, or not, and why.

In order to discover these common components, we directly compared activation during neuroimaging from two tasks that placed similar demands on phonological working memory, but required different behavioral responses. In the scanner, participants completed a nonword repetition task that mirrored clinical assessments of phonological working memory, such as the *Children’s Test of Nonword Repetition* (CNRep; Gathercole et al., 1994; Gathercole and Baddeley, 1996), the *Comprehensive Test of Phonological Processing* (CTOPP; Wagner et al., 1999), and the *Nonword Repetition Test* (NRT; Dollaghan and Campbell, 1998). Scores on these types of assessments have been shown to be more highly correlated with vocabulary and reading comprehension abilities than classical working memory tasks, such as digit span (Gathercole

et al., 1994). Participants also completed a second task, nonword discrimination, that required the same temporary maintenance of phonological content, but eliminated the need for an overt verbal response. Both tasks were performed at two levels of phonological working memory load (operationalized by short vs. long nonwords), a manipulation believed to have the greatest diagnostic specificity for individuals with developmental language disorders (Graf Estes et al., 2007). We examined the conjunction of the contrasts between high and low memory loads in each task in order to establish which brain regions are similarly engaged when greater levels of phonological information must be maintained and recalled.

Traditional group-level conjunction analyses are commonly used to make inferences about the range of behaviors a brain region might support, but these do not reflect conjunction in activation across tasks in individual subjects, so it is unclear that the same tissue is actually supporting two different tasks in the same brain. Furthermore, even if the same tissue is recruited, conjunction based on activation magnitude does not reveal whether a structure is making the same neurocomputational contribution to each task. Despite the known heterogeneity in functional neuroanatomical organization between individuals, especially in association cortices and during high-level tasks (Fischl et al., 2008; Frost and Goebel, 2012; Tahmasebi et al., 2012), functional responses are remarkably consistent within individuals (Kanwisher et al., 1997; Mahowald and Fedorenko, 2016). This variability in the relationships between anatomical landmarks and functional activations reflects not only random noise to be averaged over, but also natural variance to be accounted for. By employing individual-subjects analysis techniques, brain regions with similar functional properties and similar spatial locations can be analyzed together without requiring strict voxel-wise correspondence across individuals (e.g., Basilakos et al., 2018; Epstein and Kanwisher, 1998; Fedorenko et al., 2010; Fedorenko et al., 2013; Kanwisher et al., 1997; Saxe and Kanwisher, 2003). This approach has been crucial for revealing highly heterogeneous functional responses within anatomically delineated regions that, using traditional group-averaging or meta-analyses, would appear to overlap and thus erroneously suggest shared neural origins (e.g. Deen et al., 2015; Fedorenko et al., 2012).

Here, instead of aiming to establish functional separation, we investigated *functional concurrence* in order to establish which regions are similarly engaged during our two phonological working memory tasks, and thus which regions support computations common to both. We identified regions with overlapping functional activation while accounting for the individual variability that would be obscured in a traditional group-averaged conjunction analysis. We employed group-constrained subject-specific (GCSS) analyses (Fedorenko et al., 2010; Julian et al., 2012) in order to (1) identify brain regions commonly modulated by phonological working memory load during nonword repetition and discrimination, (2) functionally define regions of interest (ROIs) in individual subjects, and (3) compare the evoked activity during both tasks in those regions. Additionally, we employed multi-voxel pattern analyses (MVPA; Haxby et al., 2001) to assess the similarity between task activation at the level of individual voxels. By demonstrating the similarity (or dissimilarity) of individuals’ responses to these two tasks in successively finer-grained detail, we aimed to identify brain regions supporting the core neurocomputational components of phonological working memory.

2. Materials and methods

2.1. Participants

Twenty adult participants (12 female, 8 male; age 19–32 years, $M = 24.1$ years) successfully completed this study. Participants were, by self-report, fluent speakers of American English who had no history of speech, hearing, reading, or language difficulties and no history of cognitive or motor developmental difficulties. All but two reported being

¹ Stimuli used in these tasks may be preferentially referred to as “pseudo-words” because they are pronounceable and mirror the structural and statistical properties of English without any linguistic meaning; however, in keeping with the nomenclature used in neuropsychological assessments and clinical literature, we will refer to these types of stimuli as “nonwords” throughout this paper.

native speakers of English; both bilingual speakers reported learning American English at or before the age of 5 years and primarily using English to communicate since childhood. Two participants reported a tendency towards left-handedness. Participants were recruited from the greater Boston University community. This study was approved by the Institutional Review Board at Boston University – Charles River Campus and the Massachusetts Institute of Technology Committee on the Use of Human Subjects as Experimental Subjects; all participants provided informed, written consent and received monetary compensation for their time.

In our previous study of phonological working memory, we measured effect sizes on the order of Cohen's $d = 0.9$ – 1.6 for differential fMRI activation in STG to discrimination of longer (4 syllable) vs. shorter (2 syllable) nonwords (Perrachione et al., 2017). Correspondingly, with $N = 20$ we have 96.5%–99.8% power to detect effect sizes in the published range and 80% power to detect effect sizes of $d \geq 0.66$.

2.2. Experimental design, materials, and procedures

Each participant performed the nonword repetition and nonword discrimination tasks during a single scanning session. During their visit, most participants completed additional tasks during functional imaging for separate but related experiments not reported here. Typical scanning sessions lasted approximately 2 h and included 20 min of structural data collection (including diffusion tensor imaging, data not included in this study) and 90–100 min of task-based fMRI, including 30 min of nonword task-based fMRI described in the present study. For two participants, the nonword repetition, nonword discrimination, and structural data reported here were acquired alone during a shorter, 1-h scanning session.

Audio recordings for nonword repetition target stimuli were produced by a female native speaker of standard American English. Stimuli for nonword discrimination were recorded by the same female speaker and by a male native speaker of American English. Both speakers were extensively familiarized with the nonwords in order to ensure natural, correct pronunciation. All stimuli were digitally recorded at 44.1 kHz, isolated, and RMS amplitude normalized to 70 dB using Praat (Boersma, 2001; <http://www.fon.hum.uva.nl/praat/>).

All experiments were written and presented using PsychoPy2 Experiment Builder v1.84.1 (Peirce et al., 2019; <http://www.psychopy.org/>). Participants were situated in the scanner in the head-first supine position. Auditory stimuli were presented over Sennheiser MRI-compatible Model S14 headphones, while participants viewed visual cues on a projector screen. Functional volumes were acquired using a simultaneous multi-slice sparse acquisition scheme. The use of sparse-acquisition (Hall et al., 1999) fMRI allowed the subject to hear the stimuli and respond during nonword repetition in silence, as is the case when this task is administered clinically. This acquisition scheme also reduced motion artifacts due to speaking during times when the scanner was acquiring data. Simultaneous multi-slice imaging with a rapid acquisition time (750 ms) minimized the delay between stimuli, approximating the timing of clinical nonword repetition tasks, which call for immediate repetition of the stimulus.

2.2.1. Nonword repetition task

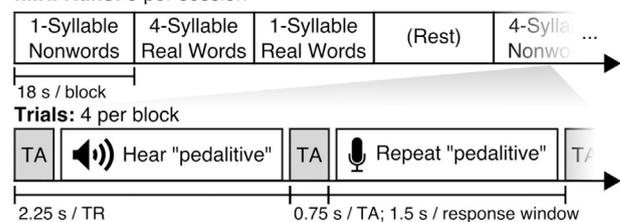
Participants heard nonwords and real English words and were asked to repeat them aloud in a block-design, sparse-sampled paradigm. Nonword stimuli were generated to closely parallel the structural and statistical properties of real English words. Stimuli belonged to either a low load (1-syllable) category or high load (4-syllable) category. The high load of four syllables was chosen to optimize the increase in task difficulty without reaching the level at which many errors might be made (typically six syllables; Gathercole et al., 1994). Forty-eight nonwords at each load were generated. An equal number of real English words were included as a control at both levels of difficulty. The nonwords were closely matched to the real words and did not significantly differ on average number of phonemes (independent-sample t -test of 1-syl.:

$p = 0.38$, 4-syl.: $p = 0.46$). In an analysis of variance (ANOVA), the segmental phonotactic probability (Vitevitch and Luce, 2004) of stimuli did not vary as a function of *word type* (real or nonword; $F_{1,188} = 0.008$, $p = 0.93$) or *syllable load* (1-, 4-syl.; $F_{1,188} = 0.020$, $p = 0.89$) with no significant interaction ($F_{1,188} = 1.02$, $p = 0.31$). In a second ANOVA, the biphone probability of stimuli was not found to vary with *word type* ($F_{1,188} = 1.12$, $p = 0.29$) or *syllable load* ($F_{1,188} = 0.63$, $p = 0.43$) with no significant interaction ($F_{1,188} = 0.84$, $p = 0.36$). Additionally, we tested for differences in lexical neighborhood density (Vitevitch and Luce, 2016) in an ANOVA and found significant main effects of *word type* ($F_{1,188} = 4.49$, $p = 0.04$) and *syllable load* ($F_{1,188} = 127.26$, $p \ll 0.0001$), and a significant interaction between the two ($F_{1,188} = 3.99$, $p = 0.05$). For 1-syllable stimuli, mean neighborhood density for nonwords was 7.90 and for real words 11.40 (post-hoc independent-sample t -test; $p = 0.04$). In the case of 4-syllable stimuli, all nonwords had a neighborhood density of 0, while real words had a mean neighborhood density of 0.10 (post-hoc independent-sample t -test; $p = 0.02$).

Each of the four experimental conditions (2 stimulus types \times 2 syllable loads) were presented in blocks of four trials, with four blocks of each condition per run. Fig. 1A illustrates example block and trial structures. Each trial began with a 750-ms scan, after which participants heard either a nonword or real word spoken by a female speaker during a 1500-ms silence. All stimuli were presented 150 ms after the volume acquisition to prevent masking by scanner noise. 1-syllable stimuli had an average duration of 703 ms and 4-syllable stimuli had an average duration of 982 ms. We measured the speech rate for our stimuli and found 1-syllable stimuli had an average rate of 1.42 syllables/s (5.56 phonemes/s) while 4-syllable stimuli had an average rate of 4.13 syllables/s (9.28 phonemes/s). (This rate difference parallels that seen in the recordings provided in standardized clinical assessments such as the nonword repetition subtest of the CTOPP.) After each stimulus was presented, a second 750 ms scan was acquired followed by a 1500 ms silence in which participants were prompted to repeat the previously presented nonword. Immediate repetition was used to reflect the design of neuropsychological tests of phonological working memory (Dollaghan and Campbell, 1998; Gathercole et al., 1994; Gathercole and Baddeley, 1996; Wagner et al., 1999). Participants saw a red fixation cross during scanning and stimulus presentation, which turned white to signal when

A. Nonword Repetition Design

fMRI Runs: 3 per session



B. Nonword Discrimination Design

fMRI Runs: 2 per session

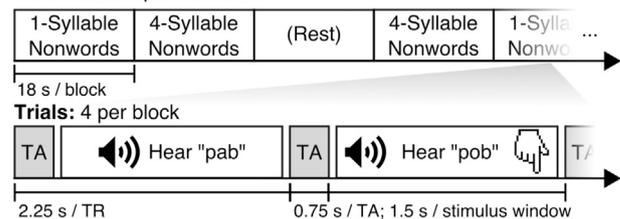


Fig. 1. Task design. (A) shows a portion of the block design structure used in nonword repetition and the structure and timing of an example trial. (B) depicts timing and structure of the nonword discrimination paradigm. Instead of repeating a word as in nonword repetition, subjects were instructed to press a button to indicate whether the second word that was presented in each trial matched the first.

to repeat. The screen background remained 50% gray throughout the experiment. Outside the scanner, audio recordings of the participant's response were collected using the in-scanner intercom system to assess nonword repetition accuracy. A single trial lasted 4500 ms.

Runs also included four blocks of "rest" during which no stimuli were presented and the fixation cross dimmed slightly to 25% gray. These silent trials were included in order to have a baseline against which to estimate responses to each condition. The block order was pseudo-randomized so that no two adjacent blocks belonged to the same condition and the transition probability between conditions, including rest, was equal across all runs. Participants completed three runs, each of which lasted 360s.

2.2.2. Nonword discrimination task

During nonword discrimination, participants were asked to decide whether a pair of nonwords matched, or if the second of the pair had been altered relative to the first. For this experiment, a second set of 48 1-syllable nonwords and 48 4-syllable nonwords were generated to meet the same criteria as those in the nonword repetition experiment. This set of stimuli did not significantly differ on average number of phonemes (independent-sample *t*-test of 1-syl.: $p = 0.65$, 4-syl.: $p = 0.75$), average segmental phonotactic probability (independent-sample *t*-test of 1-syl.: $p = 0.87$, 4-syl.: $p = 0.83$) or average biphone probability (independent-sample *t*-test of 1-syl.: $p = 0.64$, 4-syl.: $p = 0.72$) compared to the set used in nonword repetition.

Each nonword in this second set underwent a single alteration to produce 48 1-syllable and 48 4-syllable non-matching pairs. In every case the alteration was produced by replacing a single phoneme with another comparably probable phoneme. These changes were made to consonants and vowels, but never involved the first phoneme of the nonword (e.g., *pask/posk*, *motiliate/moniliate*). The altered nonwords did not significantly differ on average phonotactic probability (paired-sample *t*-test of 1-syl.: $p = 0.34$, 4-syl.: $p = 0.61$) or average biphone probability (paired-sample *t*-test of 1-syl.: $p = 0.39$, 4-syl.: $p = 0.73$) compared to the original set.

In a structure (illustrated in Fig. 1B) similar to the nonword repetition task, trials were presented in a block-design, sparse-sampled paradigm. Each trial began with a 750 ms volume acquisition followed by 1500 ms without scanner noise, during which the first nonword was heard spoken by a female speaker. Then, a second 750 ms scan was acquired after which a second nonword, either matching the first or slightly altered, was spoken by a male speaker. The use of separate speakers for the two nonwords ensured that subjects could not complete the task based on perception and retention of low-level acoustic features alone (Lim et al., 2015), and required processing at the level of phonological abstraction, as also required by nonword repetition. Stimuli spoken by the female speaker had an average duration of 686 ms (1-syllable) and 994 ms (4-syllable); those spoken by the male speaker were 662 ms (1-syllable) and 1071 ms (4-syllable) in duration. Speech rate for stimuli in this experiment was measured to be 1.48 syllables/s (5.66 phonemes/s) for 1-syllable stimuli, while 4-syllable stimuli had an average rate of 4.08 syllables/s (9.34 phonemes/s). Participants were asked to indicate with a button press of their right forefinger if the two words matched, or a button press of their right middle finger if the second word had been altered relative to the first. 50% of trials contained altered pairs presented in a pseudo-random order so that no more than three (non-) matching trials were presented in a row. Responses were collected via a USB in-scanner button box and recorded in PsychoPy. A fixation cross was displayed in red during the first nonword of the trial and during scans, then in white during presentation of the second nonword. Each trial lasted 4500 ms.

The two nonword discrimination conditions (working memory loads) were presented in blocks of four trials, with six blocks of each condition per run. Six blocks of "rest" were also included in each run during which the fixation cross dimmed to 25% gray. The block order was pseudo-randomized so that no two adjacent blocks belonged to the same

condition and the transition probability between conditions, including rest, was equal across all runs. Participants completed two runs of nonword discrimination, each lasting 324s.

2.3. fMRI acquisition and analyses

Data were acquired on a Siemens Trio 3T scanner with a 32-channel phased array head coil at the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research at MIT. Whole-head, high resolution structural images, including a T1-weighted, magnetization-prepared rapid gradient-echo (MPRAGE) anatomical volume (TR = 2530 ms, TE = [1.64, 3.50, 5.36, 7.22 ms], TI = 1400 ms, flip angle = 7.0°, voxel resolution = 1.0 mm isotropic, FOV = 256 × 256, 176 sagittal slices) and a T2-weighted anatomical volume (TR = 3200 ms, TE = 454 ms, voxel resolution = 1.0 mm isotropic, FOV = 256 × 256, 176 sagittal slices) were collected prior to functional imaging.

Functional data were acquired using a sparse-sampled, simultaneous multislice, T2*-weighted gradient-echo planar imaging (EPI) scans (TR = 2250 ms, TE = 30 ms, acquisition time (TA) = 750 ms, flip angle = 90°, voxel resolution = 3.0 mm isotropic, 10% slice gap, FOV = 72 × 72, 45 slices, 5 simultaneous slices). 162 volumes were acquired during each of the three runs of nonword repetition, and 146 volumes were acquired during each of the two runs of nonword discrimination.

Cortical reconstruction of the T1-weighted anatomical images was performed using the default processing stream in FreeSurfer v5.3.0 (Dale et al., 1999; <http://surfer.nmr.mgh.harvard.edu>). T2-weighted structural images were used to improve the estimation of the pial surface to aid reconstruction. Functional data were analyzed in FMRIB Software Library (FSL; Jenkinson et al., 2012; <https://fsl.fmrib.ox.ac.uk>) using workflows in Nipype v0.13 (Gorgolewski et al., 2011; <http://nipype.org/nipype>) implemented in Lyman v1.0.0 (<https://github.com/mwas/kom/lyman>). Image preprocessing consisted of motion correction within each run (i.e., rigid-body realignment to the mean EPI image) and spatial smoothing (6 mm FWHM kernel) using the SUSAN algorithm implemented in FSL (Smith and Brady, 1997). Motion and intensity outliers (functional volumes exceeding 1 mm in differential motion or differing from the mean image intensity by > 3 SD) were identified and included as nuisance regressors during modeling (Siegel et al., 2014). Model design was implemented using the *modelgen* algorithm in Nipype, and included four task regressors for nonword repetition (1-syl. and 4-syl. nonwords and real words) and two task regressors for nonword discrimination (1-syl. and 4-syl. nonwords), six motion parameters, and individual regressors for any outlier volumes. Vectors for task regressors were calculated by convolving a high temporal-resolution vector of event onsets with their durations, and convolving the resulting stimulation time series with a canonical hemodynamic response function to generate the hypothesized blood oxygenation level dependent response. To account for the discontinuous nature of sparse-sampling MR-signal acquisition, the hypothesized response vector was resampled over only those time points where MRI data were actually acquired – an approach that offers increased sensitivity to event-related activation in sparse-sampling fMRI by accounting for the neural response sampled across consecutive functional volumes (Perrachione and Ghosh, 2013). Contrasts of interest included each level of each condition vs. baseline and high vs. low working memory load each for nonword repetition, real-word repetition, and nonword discrimination. Within-subject estimation of the general linear model and contrasts was conducted within each run in participants' native EPI space.

Two transformations were calculated to bring the native EPI space volumes into a common space (the MNI152 template from FSL v5.0.7) for analysis. The coregistration transformation between each participant's mean functional EPI volume and their T1-weighted structural image was calculated using FreeSurfer's *BBRegister* program with FLIRT initialization (Greve and Fischl, 2009). The nonlinear warps between these high-resolution structural images and MNI space were calculated using nonlinear symmetric diffeomorphic mapping implemented in ANTS v.1.9

(Avants et al., 2008; <http://picsl.upenn.edu/software/ants>). Both the linear transformation matrix and nonlinear deformation field were applied concurrently to the contrast images from each participant's first-level analysis to ensure accurate coregistration between functional data and high-resolution anatomy. Transformed contrast images were then combined across runs in fixed-effects analyses, resulting in individual subjects' statistical maps for each contrast of interest.

2.4. Group-level univariate analyses

Group maps of working memory load contrasts (4-syl. > 1-syl.) for each condition (nonwords and real words in nonword repetition, nonwords in nonword discrimination) were computed using one-sample group mean tests on the fixed-effects individual subjects' contrasts. FMRIB Local Analysis of Mixed Effects (FLAME), included with FSL, was used to estimate group level statistics. Group statistical maps were then thresholded voxelwise at $p < 0.01$ and corrected for multiple comparisons by controlling the cluster-level family-wise error rate at $q = 0.05$, parameters shown to effectively control for Type I errors in block-design task fMRI (Eklund et al., 2016).

2.5. Group-constrained subject-specific (GCSS) analyses

In order to discover which broad brain regions captured similarly located phonological working memory load-related activations across our participants, a set of ROI "parcels" were generated using an algorithm similar to that in Fedorenko et al., (2010) and Julian et al., (2012). The procedure for identifying parcels is depicted in Fig. 2A. First, each individual subject's uncorrected statistical map showing the contrast of 4-syl. > 1-syl. nonwords was thresholded at $p < 0.01$ and binarized. Second, the binarized maps from all subjects were overlaid to create a probability map for activation in our subject group. This probability map was smoothed with a gaussian kernel of 6 mm FWHM and thresholded voxelwise at two subjects. Third, a watershed algorithm (Meyer, 1991) implemented in the SPM-SS toolbox (Nieto-Castañón and Fedorenko, 2012; https://www.nitrc.org/projects/spm_ss) was employed to find local maxima in the probability map and "fill in" the volumes around those local maxima, creating a parcellation of the probability map. To create this parcellation, voxels containing local maxima are assigned unique labels which are iteratively propagated to neighboring voxels until a local minima or zero-valued voxel is reached. The result is a map of labelled voxels in which each label denotes a 3-dimensional volume

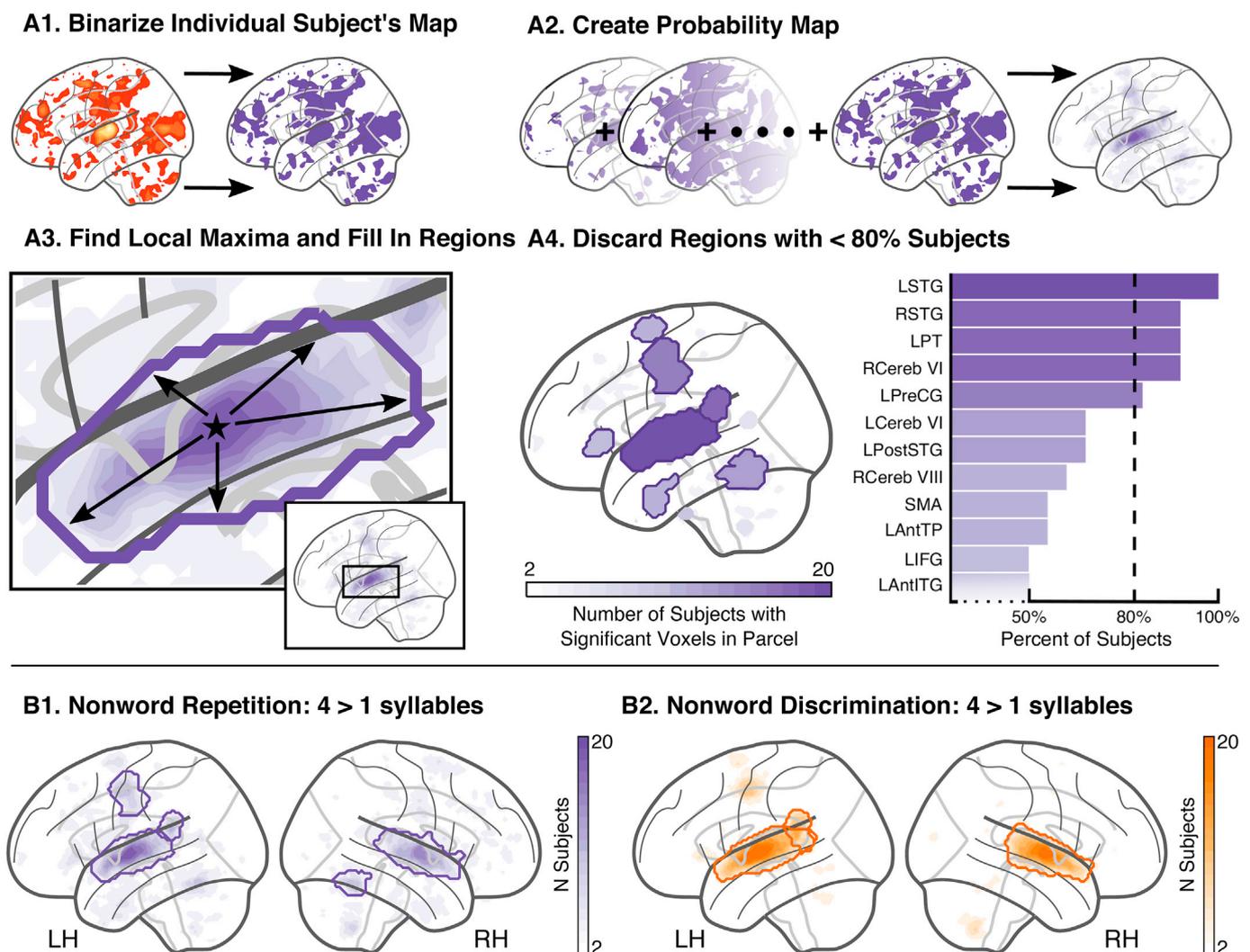


Fig. 2. Parcellation procedure and resulting parcels. The steps to algorithmically define parcels are: **A1.** Binarize each individual subject's uncorrected statistical map showing the contrast of 4-syl. > 1-syl. nonwords, thresholded at $p < 0.01$. **A2.** Overlay all the binarized maps to create a probability map. This map was smoothed and thresholded at two subjects (not shown). **A3.** Employ a watershed algorithm to locate local maxima and fill in regions around those maxima. **A4.** Calculate the number of subjects contributing voxels to each parcel and keep only the parcels that overlapped with significant voxels from greater than or equal to 80% of subjects (16 subjects). Results of these steps are shown for both the **(B1)** nonword repetition and **(B2)** nonword discrimination contrasts. Resulting parcel outlines are overlaid on the corresponding probability maps.

(“parcel”) encompassing a region where multiple participants exhibited suprathreshold activation, without requiring that this suprathreshold activation occur in the exact same voxels across participants. Fourth, the number of subjects contributing voxels to each parcel was calculated, and parcels that overlapped with significant voxels from greater than or equal to 80% of subjects (16 subjects) were chosen as representative of the common areas of activation. (All analyses were repeated in parcels in which the number of contributing subjects exceeded that which would be expected by chance with 95% confidence, and these data are included in the Supplementary Material.)

The GCSS analysis is advantageous over the one-sample group mean as it allows for nearby voxels from different subjects to be analyzed together, without requiring that those voxels overlap exactly in the common space. This approach also guards against the possibility that a few subjects with highly significant voxels drive mean activation in a region that isn't representative of the population.

2.5.1. Task response measurements

From these parcels we were able to define functional regions of interest (fROIs) and estimate responses for each individual subject. All fROIs were defined by intersecting individual-subject statistical maps of the 4-syl. > 1-syl. nonwords contrast with the resulting parcels (illustrated in Fig. S1A). The top 10% of voxels from the subject's statistical T-map within each parcel were defined as the fROI. In order to ensure that responses were not measured in voxels *selected because of their significance* (i.e., “double-dipping,” Vul and Kanwisher, 2010), we used independent data from separate runs of each task for defining and measuring responses within the fROIs. When estimating responses during nonword repetition, we employed a leave-one-out approach (Kreigeskorte et al., 2009) in which data from two runs of nonword repetition were used to define the fROI in each parcel and the response in that fROI was measured in the left out run. This procedure was repeated for all three splits of the data and then the measured responses were averaged together to give a single measure for the response in each parcel for each subject. Because functional data from the two tasks are statistically independent, statistical t-maps combining data from all three runs of nonword repetition were used to define the fROIs in which nonword discrimination responses were measured. We adopted a significance criterion $\alpha = 0.05$ for single tests and adjusted this level (via Bonferroni correction) to correct for multiple comparisons when performing the same test on multiple regions of interest.

2.5.2. Multi-voxel pattern analyses

It is possible that even though load-dependent differences in nonword discrimination may be significant when measured in fROIs defined using nonword repetition, there may be other regions within the broad volume of each parcel in which the difference in response magnitude between nonword discrimination conditions is greater. In this case, even though brain regions most sensitive to phonological working memory load in nonword repetition are modulated during nonword discrimination, these regions would not capture the most load-dependent voxels during nonword discrimination. To determine whether the pattern of activity in each parcel reflects similar task engagement during nonword repetition and nonword discrimination, we computed Pearson correlation coefficients between the 4-syl. > 1-syl. contrast images from both tasks across all voxels in each parcel, within individual subjects. Three possible scenarios involving different degrees of pattern overlap between tasks are illustrated in Figs. S1B–D. Correlations were computed on the spatially smoothed data used in all previous analyses, as well as on unsmoothed data that had undergone otherwise identical preprocessing and first-level analyses (to insure that spatial smoothing did not induce spuriously high correlations). We assessed the significance of these correlations across our participants first under a null hypothesis in which unrelated patterns of activity had a correlation of zero, and second under a null hypothesis that incidental correlations were similar in task relevant and irrelevant brain areas. For this second statistical test, we computed a

baseline distribution by randomly sampling the correlations within 3-voxel radius spheres 1000 times (i.e., “bootstrapping”), excluding voxels that fell within any task-positive parcels regardless of how many subjects had significant voxels within those parcels. We then computed a z-score with respect to this baseline distribution for each of our regions of interest and tested whether these z-scores differed from zero.

Finally, we tested how alike the patterns of activity in response to critical contrasts of 4-syl. > 1-syl. nonwords from both tasks were across an individual's entire brain. We performed whole-brain searchlight analyses in which the value of each voxel was determined by the correlation between task contrast maps within a 3-voxel radius sphere centered on that voxel. We then performed both univariate group (§2.4) and GCSS (§2.5) analyses across participants' correlation maps.

3. Results

3.1. In-scanner behavioral results

All subjects scored highly on both tasks, making few errors. A few trials during repetition were dropped because the subject was inaudible over the intercom system. This occurred in eight of the twenty subjects, never affected more than three trials for any one subject in any one condition, and summed to a total of fifteen omitted trials over all subjects and conditions. No significant effect of *word type* (real or nonword; $F_{1,19} = 0.68$, $p = 0.42$) or *syllable load* ($F_{1,19} = 0.073$, $p = 0.79$) on trial omission was identified. Percent correct was calculated for each subject based on the number of trials remaining. Descriptive statistics are reported in Table 1. We analyzed whether participants' repetition accuracy was affected by *word type* or *syllable load* using a generalized linear mixed-effects model for binomial data. The model's fixed-effect terms included the categorical factors *word type* (real vs. nonwords), *syllable load* (1- vs. 4-syllables), and the *word type* \times *syllable load* interaction. The model's random-effects terms included by-participant and by-item intercepts, as well as by-participant slopes for fixed factors *word type* and *syllable load*. The significance of fixed-effects terms was tested using deviation-coded contrasts on the model, with $\alpha = 0.05$. We did not find any main effect of *word type* ($\beta = -4.77$, $z = -0.12$, $p = 0.90$), *syllable load* ($\beta = -4.26$, $z = -0.11$, $p = 0.91$), or any interaction between the two ($\beta = -4.26$, $z = -0.11$, $p = 0.91$).

We tested the effect of *syllable load* on accuracy during nonword discrimination using a similar generalized linear mixed-effects model with a single fixed-effect term for the categorical factor *syllable load* and random-effects terms including by-participant and by-item intercepts, as well as *syllable load* by-participant slopes. Here, we again determined significance of fixed-effects using deviation-coded contrasts. We did not observe any main effect of *syllable load* ($\beta = -0.16$, $z = -0.71$, $p = 0.48$). Reaction time was measured from the onset of the second word of the trial. We modeled the effect of *syllable load* on reaction time using a linear mixed-effects model with a fixed-effects term for *syllable load* and random effects terms for by-participant and by-item intercepts, as well as *syllable load* by-participant slopes. The degrees of freedom were based on the Satterthwaite approximation. We found a highly significant main effect of *syllable load* ($\beta = -0.10$, $t = -8.16$, $p \ll 0.0001$); however, the duration of short and long nonwords also differed, and when we included the duration of each stimulus as a continuous covariate in this model, the

Table 1
Summary of behavioral results. Mean and standard deviations are given for task accuracy and reaction time measures across subjects.

Behavioral Measure	1-Syllable	4-Syllables
Accuracy (% correct)		
Nonword Repetition	98.12 \pm 0.022	97.80 \pm 0.027
Real Word Repetition	99.37 \pm 0.010	100. \pm 0.0
Nonword Discrimination	94.79 \pm 0.049	95.73 \pm 0.041
Reaction Time (ms)		
Nonword Discrimination	941 \pm 106	1153 \pm 103

effect of syllable load was no longer significant ($\beta = -0.006$, $t = -0.33$, $p = 0.74$). Stimulus duration, on the other hand, had a significant effect on reaction time ($\beta = 0.45$, $t = 6.19$, $p \ll 0.0001$) suggesting that much of the difference observed in reaction time between conditions was driven by stimulus duration.

3.2. Group-level univariate results

Our whole-brain group analyses tested the critical contrast of high vs. low phonological working memory load for each condition in each task (Table S1). During repetition of nonwords, we observed significant clusters of load-related activation in bilateral STG, LPT, and PreCG (Fig. 3A), as well as cerebellum (Fig. S2). No significant clusters were observed along the medial or ventral surfaces, nor within any subcortical structures. STG clusters extended into the white matter and reached the posterior portions of the lateral ventricles. Similar activation patterns were observed during repetition of real words, with additional small clusters located in the frontal lobe, posterior cingulate cortex, and basal ganglia (Fig. S3). As these results were not qualitatively different from repetition of nonwords, and understanding the effects of lexicality on phonological working memory was not the primary focus of this study, we do not report further analyses on the real-word repetition data here. During nonword discrimination, the critical contrast of 4-syl. > 1-syl. revealed significant clusters in bilateral STG and LPT (Fig. 3B), with a few small clusters in the cerebellum (Fig. S2).

In the conjunction of group maps from nonword repetition and

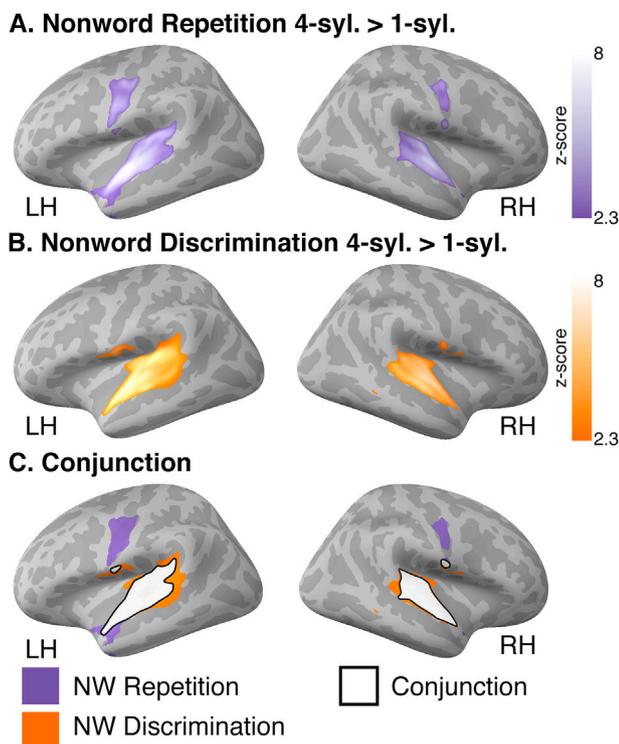


Fig. 3. Univariate group-averaged results. All maps were thresholded at $p < 0.01$ and cluster-corrected at a family-wise error rate of 0.05. Results were smoothed for display on inflated surfaces. There were no significant clusters on the ventral or medial surfaces. (A) The univariate group maps for the 4-syl. > 1-syl. contrasts for nonword repetition. Significant clusters included bilateral STG and bilateral PreCG. (B) The 4-syl. > 1-syl. contrast for nonwords during discrimination. The bilateral STG were the most prominent clusters. Cerebellar clusters were also observed in both experiments (see Fig. S2). (C) Both contrast maps from the two tasks were binarized at $p < 0.01$ and overlaid. The 4-syl. > 1-syl. contrast from nonword repetition is shown in purple, and nonword discrimination in orange. Places where both maps were significant are shown in white, and confined to the bilateral STG.

nonword discrimination (Fig. 3C, Table S1), we found significant overlap between the two tasks in bilateral superior temporal and surrounding cortices. Clusters in the cerebellum only overlapped in a small number of voxels in left cerebellar lobule VI.

We examined phonological working memory load-related activation to nonword repetition, nonword discrimination, and their conjunction in the brains of individual subjects (Fig. 4). Every subject showed activation in superior temporal cortex during both tasks. This activation varied across subjects in anterior-posterior extent (e.g., S3 vs. S13) and in the number of distinct clusters of activity observed (e.g., S1 vs. S7). Many subjects showed overlapping activation between the tasks in LPT, sometimes in a distinct cluster from activation elsewhere in STG (e.g., S1, S3, and S5). Over half of the subjects also showed overlapping activation in LPreCG (S1 through S11), where significant mean activation during nonword discrimination was not observed in the univariate group-level analysis.

3.3. GCSS results

3.3.1. Identification of parcels

In order to capture the intersubject variability and similarity observed in the individual-subject maps, GCSS analysis was used to identify and test individual-subject fROIs in our sample. From our parcellation of nonword repetition probability maps generated using individual-subject data (Fig. 2B1), five parcels emerged in which $\geq 80\%$ of subjects showed statistically significant differences in response to the syllable load manipulation: bilateral STG, LPT, LPreCG, and right cerebellum lobule VI (shown individually in Fig. 5A). (All nonword repetition parcels, including those with $< 80\%$ of subjects, are shown in Fig. S4).

The same procedure identified three parcels in nonword discrimination: bilateral STG and LPT, all of which had similar sizes and locations as those found in nonword repetition (Fig. 2B2). The probability maps for nonword discrimination showed some indication of overlap in LPreCG, as well as in the right cerebellum, but less than 80% of subjects had significant voxels in these areas. (All nonword discrimination parcels are shown in Fig. S5).

Within each broad parcel, we defined fROIs for each individual subject and then measured responses in left-out, independent data. Because our nonword repetition paradigm yielded a set of parcels which included all of the parcels identified in nonword discrimination, we chose to examine responses in the larger set. (Corresponding analyses for sub-threshold parcels obtained from the nonword repetition task, as well as all nonword discrimination parcels, are reported in Table S2, S3, and Fig. S6.)

Our investigation focused on several questions: (1) can we confirm that there is significant modulation by phonological working memory load during nonword repetition in each region identified, (2) are any of the regions identified in nonword repetition also significantly sensitive to phonological working memory load in nonword discrimination, and (3) are the patterns of activation across each parcel similar between nonword repetition and nonword discrimination. When performing statistical tests over each of the 5 regions, we adopted a Bonferroni-corrected significance criterion of $\alpha = 0.01$.

3.3.2. Task-induced response differences

In nonword repetition the greatest differences in activation between the 4-syllable and 1-syllable conditions were found in bilateral STG and LPT (Fig. 5B). In LPreCG and right cerebellum, differences were smaller though still significant (paired-sample t-tests, Bonferroni-corrected $\alpha = 0.01$).

During nonword discrimination, significant differences in neural responses to high vs. low phonological working memory load were found in each of the regions defined by their load-dependent sensitivity during nonword repetition, with the greatest modulation being found in bilateral STG and LPT (Fig. 5B). Though the overall magnitude of load-dependent response modulation was lower in LPreCG and RCereb,

Individual Subjects' Conjunction Maps - All Subjects

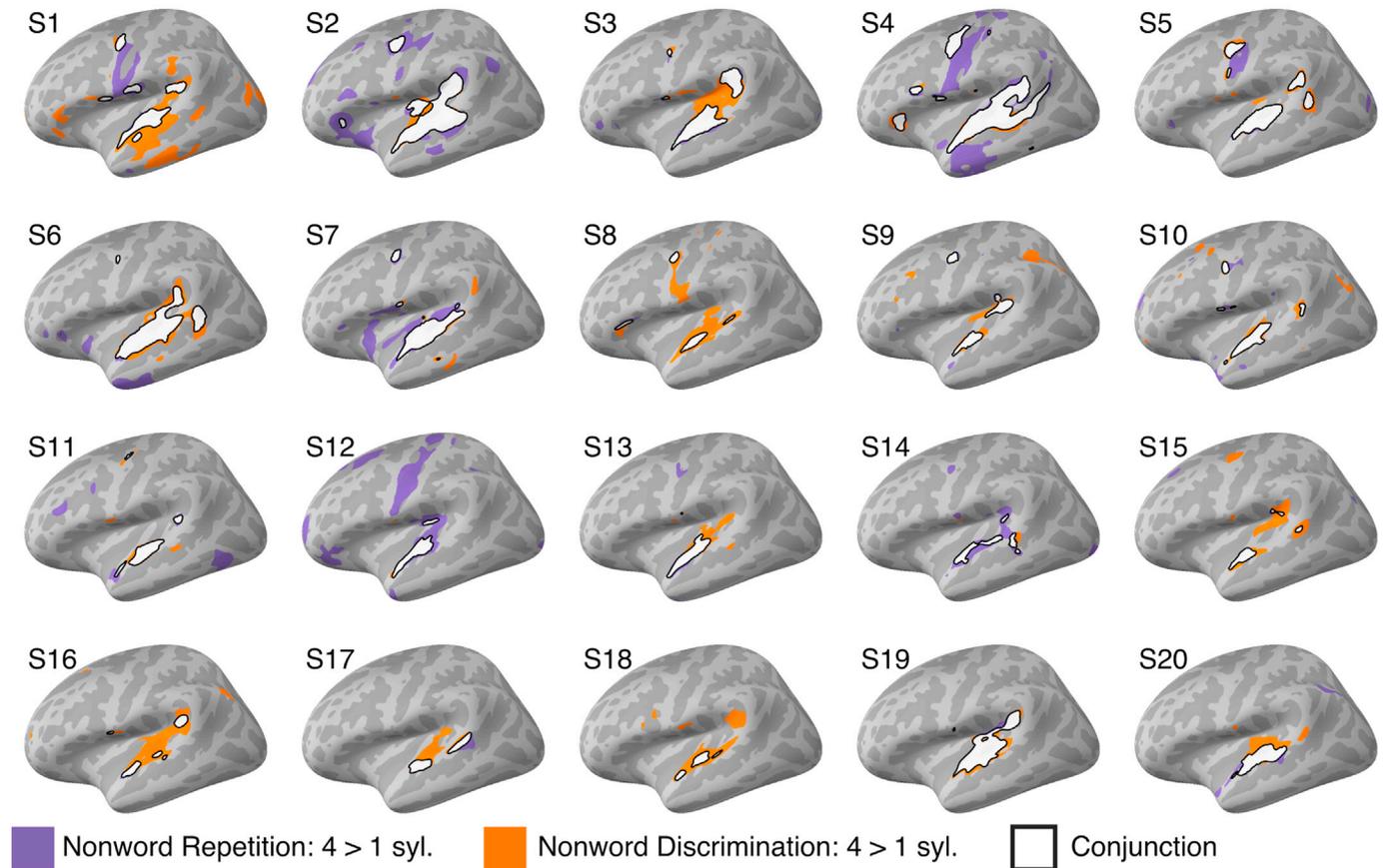


Fig. 4. Individual-subject conjunction maps. Each subject's 4-syl. > 1-syl. statistical maps from nonword repetition and nonword discrimination were thresholded at $p < 0.01$ (uncorrected) and binarized. Results were smoothed for display on left hemisphere inflated surfaces. Nonword repetition is shown in purple, and nonword discrimination in orange. Areas exceeding the threshold in both maps are shown in white. All 20 subjects are represented.

significant differences between nonword discrimination conditions were measured in fROIs in both of these areas, despite the fact that these regions did not emerge in the univariate group analysis. Right superior cerebellum showed the lowest level of activation and the smallest degree of modulation during nonword discrimination, but this difference was still significant (see Table 2 for all results).

3.3.3. Multi-voxel pattern analysis results

In order to determine whether the patterns of activation elicited by each task were similar across tasks in individual subjects, we performed a multi-voxel pattern analysis (Haxby, 2001). We calculated correlations between the voxels in each subject's contrast maps from nonword repetition and nonword discrimination within each parcel (Fig. 5C). The distributions of Fisher-transformed correlation coefficients across subjects were significantly greater than zero in all parcels (Bonferroni-corrected $\alpha = 0.01$) in both smoothed and unsmoothed contrast maps. We also tested whether correlations in each parcel exceeded subject-specific baseline ("bootstrapped") correlations calculated over brain regions that did not show task-relevant responses. We found that correlations in bilateral STG, LPT, and LPreCG were all significant with respect to this baseline; however, right cerebellum correlations did not reach significance across participants.

We performed several exploratory tests to discover whether any neural or behavioral variables were related to the range of pattern correlation values observed in LPreCG and cerebellum. The within-parcel/ between-task correlation coefficients across subjects were significantly correlated with participants' 4-syl. > 1-syl. contrast effect size during nonword discrimination (LPreCG: $r = 0.86$, $p \ll 0.0001$; right cerebellum:

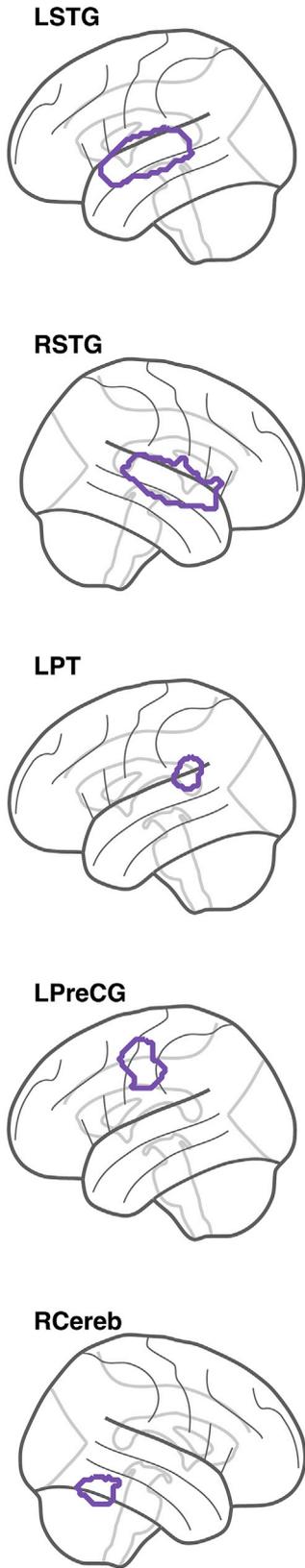
$r = 0.64$, $p < 0.01$), showing that the lowest correlations between tasks were measured in subjects in whom there was the least amount of task-related activity in these parcels. We tested for correlations relating between-task pattern similarity and (1) accuracy during nonword repetition, (2) accuracy during nonword discrimination, and (3) average reaction time during nonword discrimination, but we did not uncover any significant relationships (all $p > 0.05$).

Finally, we investigated local phonological working memory load-related pattern correlations across the whole brain in each individual. Using a 3-voxel radius spherical searchlight, we calculated the local correlation coefficient between the two tasks centered on each voxel. We summarized our findings first as an uncorrected group-average map of correlations (Fig. 6A), noting that areas of significant correlation ($p < 0.01$) were observed in bilateral STG and LPT. We then examined the probability map constructed from each individual's correlations thresholded at $p < 0.01$ (Fig. 6B). In a subsequent GCSS analysis based on this probability map, we identified four regions in which 80% or more of subjects showed significant correlations in voxel-wise activation between the two tasks: bilateral STG, LPT, and LPreCG (Fig. 6C).

4. Discussion

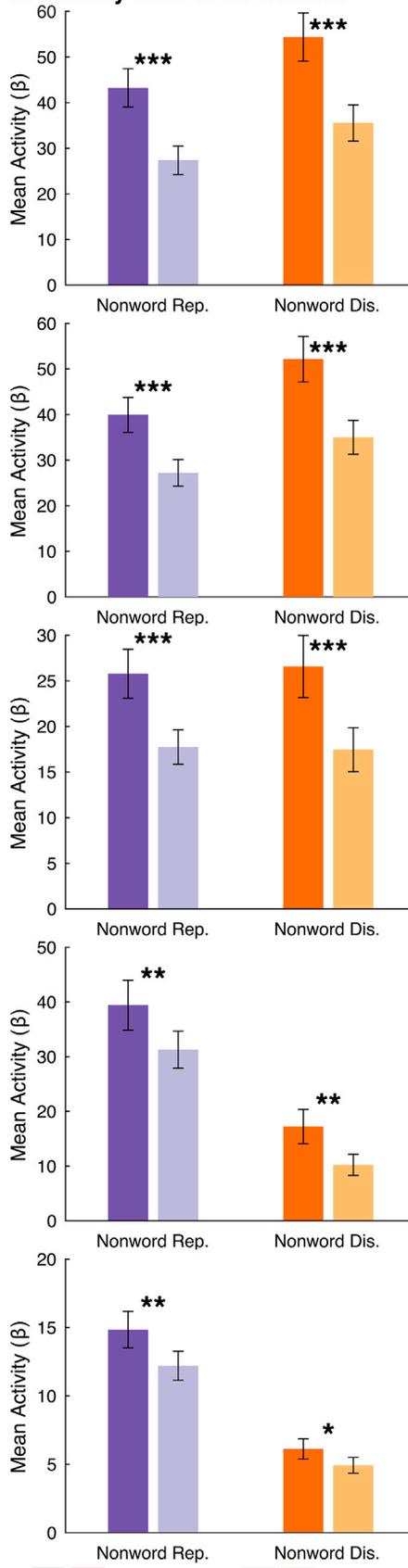
We sought to identify regions of the brain that are commonly recruited during phonological working memory. Our results point to four brain regions (bilateral STG, LPT, and LPreCG) that are consistently sensitive to phonological working memory load in the vast majority of our participants across both tasks. A fifth region, right cerebellar lobule VI, exhibited load-sensitive responses in a majority of participants during

A. Regions

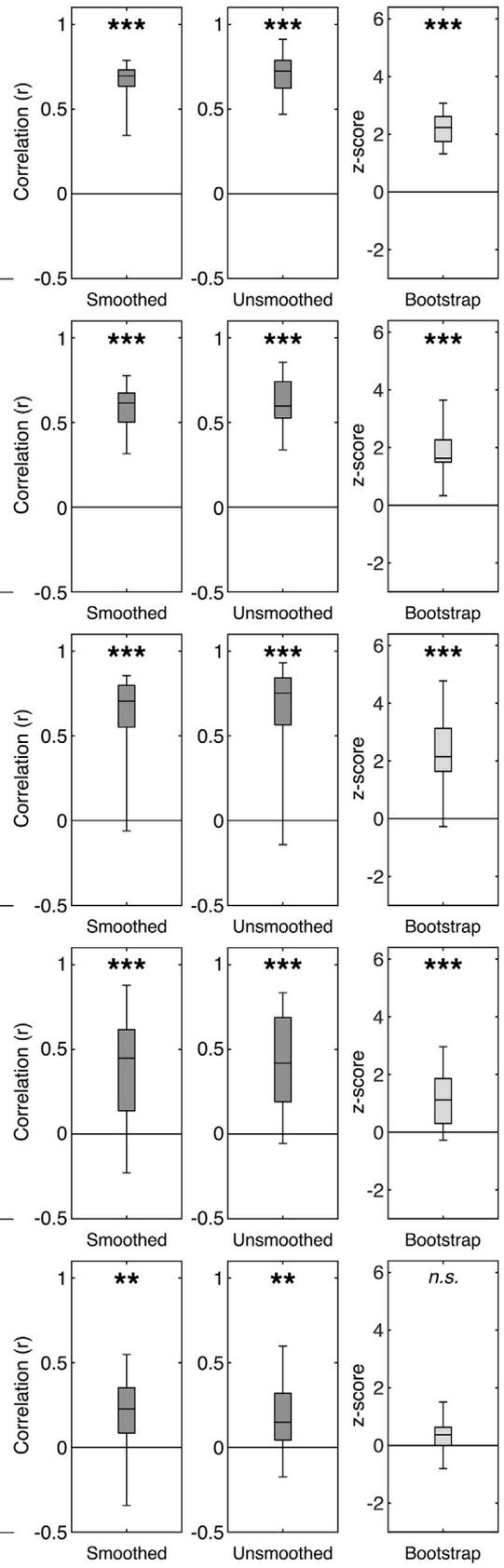


fROI Search Area

B. Activity Measured in fROIs



C. Correlation Analyses



4 syl. Nonwords
 1 syl. Nonwords
 *** $p < 0.0001$ ** $p < 0.001$ * $p < 0.01$

(caption on next page)

Fig. 5. Individual-subject fROI results. Column (A) shows the location and extent of each of the nonword repetition parcels. (B) shows the response magnitudes measured in fROIs defined as the top 10% of voxels in nonword repetition. Nonword repetition response magnitudes were measured using a leave-one-out procedure to ensure statistical independence from data used to define the fROIs. Nonword discrimination response magnitudes were measured in fROIs defined using data aggregated over all of the runs of nonword repetition. Column (C) shows the distribution of between-task correlations across subjects in the smoothed data (left side) presented thus far and unsmoothed data (middle) that had been otherwise identically analyzed. In both the left and middle boxplots, one-sample t-tests were performed on the Fisher-transformed values to determine if correlations significantly differed from zero across participants. In the right-most boxplots, z-scores were computed by comparing each subject's correlation to a bootstrapped subject-specific baseline. The resulting z-scores were then subjected to a one-sample t-test to determine significant difference from zero across participants. A significance level of $\alpha = 0.01$ was adopted to correct for multiple tests over the 5 regions.

Table 2

Response magnitude differences and voxel-wise correlations in regions of interest. For each identified parcel (Column 1) the total number of voxels in that parcel (Column 2) as well as the number of subjects with significant voxels within that parcel (Column 3) are shown. For each subject, the top 10% of voxels in each parcel are designated as their fROI. The magnitudes of the difference between 4-syllable and 1-syllable nonword conditions during nonword repetition (Column 4) and nonword discrimination (Column 5) are also given ($M \pm SE$). Pairwise, two-tailed t-tests were performed for each of these differences and effect sizes were computed (Cohen's d). Columns 6 and 7 give the median correlation values for each region, computed using smoothed and unsmoothed data respectively, as well as the result of one-sample, two-tailed t-tests on the Fischer transformed correlation values to determine if correlations across subjects differed significantly from zero. All correlations were performed within subjects comparing nonword repetition and nonword discrimination responses to 4-syl. > 1-syl. contrasts across all voxels within each parcel. A significance level of $\alpha = 0.01$ was adopted to correct for multiple tests over the 5 regions.

Region	# Voxels	# Subjects	4-syl > 1-syl. contrast		Correlations	
			Nonword Rep.	Nonword Dis.	Smoothed	Unsmoothed
LSTG	3358	20	15.88 \pm 1.50 $t(19) = 10.61$ $p = 2.00 \times 10^{-9}$ $d = 0.96$	18.79 \pm 1.58 $t(19) = 12.00$ $p = 2.59 \times 10^{-10}$ $d = 0.90$	$r = 0.70$ $p = 7.33 \times 10^{-15}$	$r = 0.72$ $p = 1.44 \times 10^{-12}$
RSTG	3248	18	12.70 \pm 1.83 $t(19) = 6.95$ $p = 1.28 \times 10^{-6}$ $d = 0.84$	17.06 \pm 1.66 $t(19) = 10.26$ $p = 3.48 \times 10^{-9}$ $d = 0.87$	$r = 0.61$ $p = 1.08 \times 10^{-12}$	$r = 0.60$ $p = 4.86 \times 10^{-10}$
LPT	585	18	8.03 \pm 1.22 $t(19) = 6.58$ $p = 2.66 \times 10^{-6}$ $d = 0.77$	9.10 \pm 1.20 $t(19) = 7.65$ $p = 3.24 \times 10^{-7}$ $d = 0.69$	$r = 0.70$ $p = 3.44 \times 10^{-8}$	$r = 0.75$ $p = 1.58 \times 10^{-7}$
LPreCG	925	16	8.13 \pm 1.78 $t(19) = 4.56$ $p = 2.12 \times 10^{-4}$ $d = 0.45$	6.83 \pm 1.69 $t(19) = 4.19$ $p = 4.97 \times 10^{-4}$ $d = 0.60$	$r = 0.45$ $p = 8.30 \times 10^{-5}$	$r = 0.42$ $p = 1.82 \times 10^{-5}$
RCereb	673	18	2.65 \pm 0.58 $t(19) = 4.54$ $p = 2.22 \times 10^{-4}$ $d = 0.49$	1.20 \pm 0.34 $t(19) = 3.56$ $p = 2.10 \times 10^{-3}$ $d = 0.40$	$r = 0.23$ $p = 1.84 \times 10^{-3}$	$r = 0.15$ $p = 1.33 \times 10^{-3}$

nonword repetition. While this region also showed significant sensitivity to phonological working memory load during nonword discrimination, it was the only region for which pattern analyses did not reliably reveal significant correlations between tasks across participants. In our approach, adapted from Fedorenko et al. (2010) and Julian et al. (2012), we created probability maps to summarize data beyond traditional univariate analyses at the group level, which allowed us to algorithmically identify constrained search areas for task-relevant activity and account for local variation of functional neuroanatomical organization across participants. We interrogated functional sensitivity to phonological working memory load in these fROIs by comparing responses to two separate phonological working memory tasks that were designed to similarly engage phonological working memory, but with divergent behavioral demands.

We found that all of our identified regions from the nonword repetition task were also sensitive to phonological working memory load in the nonword discrimination task, even though several of these regions were not observed in the univariate group conjunction. This suggests that these latter regions support phonological working memory in a way that is not unique to nonword repetition or the overt production of nonwords. We further investigated the similarity of the patterns of evoked activity during both tasks in individual subjects within the broader search areas and found significant correlations between tasks in all identified regions except the right cerebellar region. While these results may have been expected for regions such as STG, where phonological processing ought to have been similarly engaged due to the similarity of stimuli across tasks, finding highly similar patterns of spatial activity in the LPreCG during both repetition and discrimination of nonwords is surprising because activity in this region was not found to be significant during nonword discrimination in our traditional univariate group analysis.

4.1. Phonological working memory in regions associated with speech perception

Consistent with other recent studies of phonological working memory, the most robustly activated regions by either of our tasks were the bilateral STG. Activity in these regions has been previously, but separately, shown to be sensitive to phonological working memory load during both nonword repetition (McGettigan et al., 2011) and nonword discrimination (Perrachione et al., 2017). In predominant models of auditory working memory, bilateral STG are often associated with phonological encoding, not phonological working memory, and are traditionally considered to exist outside of the phonological loop (Baddeley, 2003). The strict separation of phonological working memory and speech perception has been recently challenged, however, by not only the aforementioned neuroimaging studies, but also by converging evidence from large- N lesion studies that link performance in both domains to structural damage in posterior temporal lobe (Koenigs et al., 2011; Leff et al., 2009). Indeed, clinical tests of phonological working memory that use increasing syllable load as a manipulation include a confound of not only placing greater demands on *maintenance* of additional information, but also *encoding* of additional information as well. Some groups have hypothesized that phonological working memory is inextricably linked with encoding (Barry et al., 2011; Jacquemot and Scott, 2006; Majerus, 2013), and therefore the phonological loop provides only a partial explanation for how phonological working memory is deployed during language and reading acquisition. One way in which we may begin to resolve the functional roles that these regions play during phonological working memory is by applying the methods developed here to examine the response of phonological working memory regions in individual subjects' brains to a wider variety of language and working memory tasks. In this way we may discover whether the computations performed

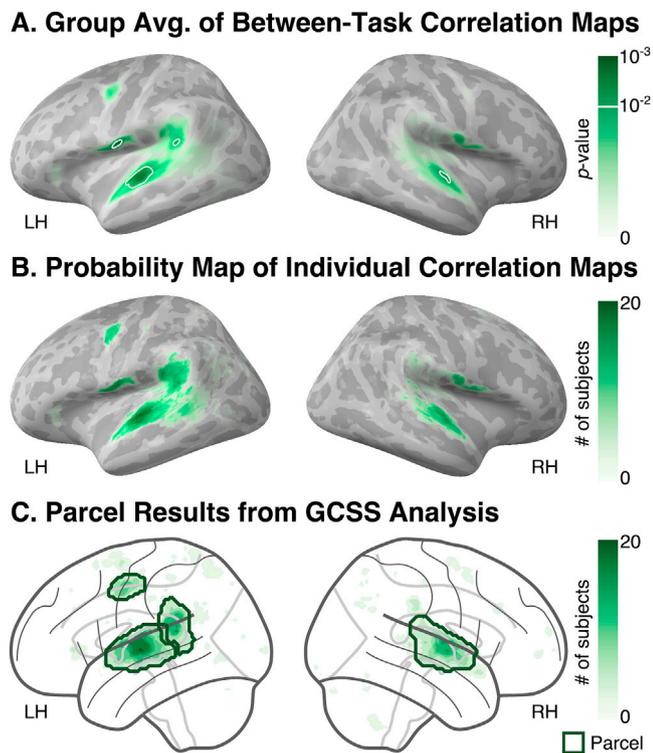


Fig. 6. Between-task correlation searchlight analysis. (A) shows the group average across all subjects' correlation maps. This map is uncorrected and unthresholded, with white contours indicating regions where the statistical value exceeded $p = 0.01$. (B) shows the probability maps generated from combining each subject's correlation maps. Darker areas indicate increased overlap across subjects. (C) shows the GCSS parcellation of (B) with contours drawn around parcels within which 80% or more of subjects had significantly correlated voxels ($p < 0.01$ uncorrected). The parcels identified are bilateral STG, LPT, and LPreCG.

in a particular region during nonword repetition are similar to its performance in potentially related tasks, such as digit span, verbal n-back, or natural speech perception. This approach would not only further elaborate the functional properties of these regions, it would represent a first step toward developing new cognitive neuroscience models that are based on the relationship between the computations that circumscribed areas of neural tissue are responsible for across diverse, complex tasks, rather than ones based on classic ideas about strict location-based selectivity (e.g., double dissociations, etc.).

Besides their demand on listeners to encode larger amounts of information, the longer (4-syllable) nonwords in our study also demand listeners encode speech information at a faster rate (syllables or phonemes/second) compared to shorter (1-syllable) nonwords, possibly introducing an additional perceptual load factor. There is some evidence that outside of primary auditory cortex, rapid adaptation to intelligible time-compressed speech allows processing in higher-level speech regions in STG and superior temporal sulcus to proceed without a commensurate increase in recruitment of neural resources (Adank and Devlin, 2010; Binder et al., 1994; Peele et al., 2004; Vagharchakian et al., 2012). However, Xie and Myers (2018) showed that STG is more sensitive to clear (slower, more articulated) than conversational (faster, less articulated) speech. Clinical assessments of phonological working memory, such as nonword repetition in the CTOPP, also include this confound, with the speech rate of their standardized stimuli also increasing for longer nonwords. Future work is necessary to directly ascertain what role encoding rate plays in phonological working memory demands, and research in this vein will be vitally important to better interpreting the relationship between clinically measured phonological working memory

deficits and language ability in children with developmental language disorders.

It is notable that the superior temporal areas were bilaterally activated, whereas other areas showed apparent lateralization during phonological working memory (e.g., cortical areas PreCG and PT were left lateralized, cerebellum was right lateralized). This is consistent with prior fMRI studies that also showed bilateral STG activation during phonological working memory tasks (McGettigan et al., 2011; Perrachione et al., 2017). Peele (2012) describes a hierarchical framework for hemispheric roles in speech processing in which unconnected speech is processed at the phoneme and word level largely bilaterally, whereas higher-level linguistic processing involving phrases, sentences, and narratives becomes primarily lateralized to the left hemisphere. New evidence for this idea is provided by a recent study (Pisoni et al., 2019) that compared performance on a variety of language and working memory tasks in patients before and after surgical resection of either left or right temporal areas. Their data showed that patients with left hemisphere damage performed significantly worse than those with right hemisphere damage on higher-level linguistic tasks such as sentence repetition or word comprehension, whereas performance on word and nonword repetition was not significantly different between the groups. However, a voxel-based lesion symptom mapping analysis done with these same patients revealed that only left-hemisphere temporal lobe damage was correlated with performance. This, along with other studies (e.g., Gainotti et al., 1982; Hickok et al., 2008), suggest that the right hemisphere is sufficient for low-level phonological processing and possibly phonological working memory, though whether or not it is necessary is still a matter for further study.

In addition to bilateral STG, LPT showed significant modulation with phonological working memory load during nonword repetition and discrimination in both our univariate-group conjunction and GCSS analyses. We note that while our group-averaged maps show activation in this area jointly with STG, most individual-participant maps show LPT as a separate cluster from the relatively more anterior STG activation, and this observation is only captured and summarized by the GCSS parcel delineation. Similar regions have been previously shown to be activated in nonword repetition and nonword discrimination tasks separately (Barry et al., 2011; McGettigan et al., 2010; Perrachione et al., 2017; Strand et al., 2008), as well as other paradigms designed to engage phonological working memory (Buchsbaum et al., 2001; Fegen et al., 2015). LPT, also referred to as area Spt (Sylvian fissure, parietal-temporal boundary), has been shown to play a causal role in speech repetition and working memory with voxel-based lesion symptom mapping (Buchsbaum et al., 2011; Koenigs et al., 2011; Rogalsky et al., 2015), transcranial magnetic stimulation (Acheson et al., 2011), and transcranial direct current stimulation (Savill et al., 2015), and has been theorized to serve as a neural hub for auditory-motor integration (Hickok et al., 2009). We have referred to this region as LPT in keeping with these findings, therefore leaning towards an interpretation that this region performs a critical function in speech perception that is shared with phonological working memory. However, several classic and recent studies (e.g. Paulesu et al., 1993; Salmon et al., 1996; Yue et al., 2019) identified a region in the inferior parietal lobe that actually closely overlaps our LPT parcel and is recruited during verbal working memory tasks. This region has been shown to be active during delays over which verbal information must be maintained and represents some information about stimuli, providing evidence that this region does in fact support a parietal phonological store, separate from brain areas responsible for speech comprehension. Future studies could directly compare, and parametrically manipulate, speech comprehension and working memory task responses in LPT/inferior parietal lobe in individual subjects in order to discover the scope of tasks that this region supports and whether or not it remains active during tasks with purely speech or working memory components.

4.2. Phonological working memory outside of the temporal lobe

LPreCG was identified as a significant cluster in our univariate group map for nonword repetition, but did not surpass the significance threshold for nonword discrimination. This finding may have been expected by comparing results across studies (found during nonword repetition by [McGettigan et al., 2011](#); not found during nonword discrimination by [Perrachione et al., 2017](#)) though it has not been shown before in the same subjects. Upon inspection of individual-subject conjunction maps, however, we found that just over half of participants showed overlap between the two tasks in this region at our chosen threshold, suggesting that evidence for common activation was lost during the group-averaging and cluster-thresholding procedures. In our GCSS analysis, LPreCG was once again identified in nonword repetition and not nonword discrimination, but when we defined individual-subject fROIs by high sensitivity to phonological working memory load in nonword repetition, we also found significant sensitivity to phonological working memory load in nonword discrimination in these same voxels. Furthermore, correlations between the two tasks across the entire LPreCG parcel in individual participants were high, suggesting that this area is similarly computationally engaged during both tasks, though there is greater overall activity during nonword repetition (e.g., [Haxby, 2001](#)). Finally, when we performed a whole-brain task correlation searchlight analysis and summarized the results using GCSS parcellation, this same area in LPreCG was shown to contain high correlations between our tasks in 80% or more of subjects, even though fewer than 80% showed significant activation there during nonword discrimination.

The insight we have gained from this discovery is twofold. First, it was surprising to find that this region, which in the past seemed to be recruited only in tasks with overt repetition, was indeed active in many subjects during a discrimination task. It may be that activation in this region reflects the articulatory rehearsal component of Baddeley's phonological loop and is necessary for phonological working memory tasks regardless of overt production. However, if PreCG subserves articulatory rehearsal, this would imply specificity in this region for working memory of speech sounds (although others have speculated that LIFG subserves this role, [Awh et al., 1996](#); [Paulesu et al., 1993](#); [Rogalsky et al., 2008](#)). Potentially analogous regions have been identified in the posterior middle frontal gyrus or precentral sulcus during a number of working memory and language tasks that do not involve overt responses, phonological working memory, or even speech/language stimuli ([Fedorenko et al. 2010, 2013](#); [Michalka et al., 2015](#); [Noyce et al., 2017](#)). Besides further probing the functional response profile of this region with neuroimaging, it will also be important to determine if this region has any causal influences on language and working memory tasks through noninvasive neurostimulation, such as transcranial magnetic or direct current stimulation, in order to further constrain our hypotheses about the function of this portion of LPreCG.

The discovery of LPreCG's common involvement in phonological working memory tasks further emphasizes the need for analysis methods that respect the spatial variability observed in individual subjects' functional activations. Had we not looked beyond the univariate group analysis, our study would have resulted in a strict replication of the [McGettigan et al. \(2011\)](#) and [Perrachione et al. \(2017\)](#) studies in which LPreCG modulation was observed during nonword repetition and not during nonword discrimination. Even in our whole-brain task correlation analysis, group averaging and thresholding obscured significant correlations in LPreCG that GCSS analysis readily identified. We conclude, then, that the types of individual subject analyses employed here offer an important and useful procedure for finding better agreement between fMRI studies that may appear to yield contradictory results, and will be of particular importance in studying phonological working memory – and its impairment – in clinical populations, where local functional neuroanatomy may be even more variable across individuals (e.g., [Kovelman et al., 2012](#)).

Our results in the right cerebellar ROI were less conclusive but also

demonstrate the utility of multiple complementary analyses. Overall mean activation was lower in the cerebellum than in the cortex, perhaps accounting for the null effects we encountered. While the region in lobule VI did demonstrate significant activity during nonword repetition in over 80% of participants and significant sensitivity to phonological working memory load in nonword discrimination across participants, we did not find that the correlations in this area exceeded subject-specific baselines calculated over other, non-task-positive, brain regions. The magnitude of correlations in right cerebellum was also found, in a post-hoc analysis, to be significantly related to the magnitude of activation in nonword discrimination. These mixed results suggest that this right cerebellar region supports functions that are somewhat task specific, driven more robustly during nonword repetition and weakly, or not at all in some participants, during nonword discrimination. There is evidence for right cerebellum lobule VI's involvement in speech and language, in particular overt and covert syllable repetition ([Ackermann et al., 2007](#)), and syllable sequence processing during verbal working memory ([Peterburs et al., 2019](#)), leading us to the hypothesis that the increased fidelity of information necessary to complete nonword repetition may have resulted in higher activation. We designed our tasks so that there would be few errors in order to examine responses when phonological working memory was employed successfully, but that does not exclude the possibility that one task was more difficult than the other for some participants. Follow-up studies will be necessary in order to better understand the effects of task difficulty as the current work contains only a very small range in participants' behavioral measures.

4.3. Advantages of accounting for individual subject variability

Though we focused here on brain areas activated in 80% or more of our participants, GCSS analyses allowed us to retain meaningful information about regions that showed load-dependent responses in a large fraction of participants, though not a majority as we have defined it. Activation in brain regions such as LIFG (present in 50% of our subjects) and SMA (55%) have been identified previously as showing parametric sensitivity to syllable load in phonological working memory (e.g., [Perrachione et al., 2017](#)). If only group-averaged results were reported here, our study might be interpreted as disagreeing with previous work stating the importance of the IFG and SMA in this faculty. By including this information, we are better able to evaluate agreement and disagreement across studies, as well as motivate further questions about how the presence or absence of activity in certain regions may be related to individual differences in behavior. As we have shown, in studies such as ours in which the per-subject signal-to-noise ratio is relatively high, it is possible for two separate but related tasks to yield differing results under group-averaged univariate analyses, but at the individual subject level the difference in activation is one of degree, not of kind. By reporting only results meeting or exceeding a hard threshold, we bias studies towards non-replicability and limit our ability to build predictions for future study.

Taken together, these results serve as a model to guide future analysis of how differences in brain activity may correlate with clinically significant impairment in nonword repetition performance and thus deficits in phonological working memory. By taking into account inter-subject variability in control subjects, we learn about the diverse range of typical brain activation during these tasks, and so are better prepared to assess differences in not just clinical populations, but impacted individuals. We can also broaden our space of possible hypotheses by not discarding subthreshold data and ask, for example, whether differences observed between groups are simply due to different levels of overall activation or more variable patterns of activation in commonly activated regions. In combining well-established neuropsychological assessments and neuroimaging analyses sensitive to individual differences, we hope to be able to narrow the gap between the clinical and neuroimaging literatures and better understand the neurogenic bases of developmental language disorders. Most studies comparing clinical populations and

typical participants use a single statistic to determine whether or not fMRI reveals significantly higher or lower activation in a region of interest. By describing the full set of activated regions observed in typical subjects in probabilistic terms, we can compute the likelihood of observed regional activation in an individual with a developmental language disorder, and possibly relate the typicality of that profile to their unique cognitive profile provided through standardized clinical testing.

Much of the neuroimaging literature has, until recently, focused on localizing circumscribed mental operations postulated by psychological models, which in turn, have been inferred by patient lesion studies. Although this approach has been successfully used to functionally define brain areas, the characterization of processing performed in many of these areas remains challenging. A way to begin to resolve many of the longstanding questions regarding the functioning of any one brain area, argued for by Genon et al. (2018), could be to build functional profiles of activity by aggregating data from many tasks. There are already several large-scale projects working towards this goal (BrainMap, Laird et al., 2005; NeuroSynth, Yarkoni et al., 2011). One significant complication, however, arises from the inability to account for intersubject variability in overlapping activations across studies. These databases are based on the final group-level contrast maps, meaning that all of the meaningful individual variability is lost. For example, if < 50% of subjects routinely activate SMA, this information, though of scientific importance, will never be reported in a meta-analysis based on group results, even though that activation is present in a plurality of participants. Here, we show that the same idea of building functional profiles of regions can be implemented on a much smaller scale with careful choice of behavioral tasks, and without sacrificing the specificity gained by making measurements on the individual-subject level.

5. Conclusion

In this study, we examined neural responses during two separate phonological working memory tasks, in order to better understand how phonological working memory is supported in the brain. By comparing activation from two tasks that theoretically require both similar and dissimilar kinds of operations (nonword repetition and nonword discrimination), we reasoned that brain regions demonstrating similar load-related modulation at the individual-subject level must reflect the computational architecture shared between both tasks, namely the encoding and maintenance of phonological information. Pattern analysis revealed that not only are STG, LPT, and LPreCG sensitive to phonological working memory load regardless of task demands, both tasks evoke highly similar patterns of activation in individual subjects in these regions, suggesting that the functions performed there during both tasks are also similar. We found that a traditional univariate conjunction analysis failed to identify empirically and theoretically important effects in LPreCG, while such effects were readily identified using analyses that take into account local variability in functional neuroanatomy. We also discovered intersubject variability in the collection of specific brain areas that were recruited in response to increased phonological working memory load, possibly explaining disparate results in the previous literature as well as motivating new questions about the links between functional activity and individual differences in behavior. Taken together, these results provide a fuller, more integrated picture of how phonological working memory is supported by the brain.

Funding

Research reported in this article was supported by the National Institute on Deafness and Other Communications Disorders at the National Institutes of Health under award number R03DC014045 and a NARSAD Young Investigator Grant to TP. TS was supported in part by the National Institute on Drug Abuse Computational Neuroscience Training Grant T90DA032484.

Acknowledgements

We thank Sara Dougherty, Ja Young Choi, Kristine Strand, Dana Gordon, and Deirdre McLaughlin. For technical support during scanning, the authors thank Atsushi Takahashi, Steve Shannon, and Sheeba Arnold. We also acknowledge the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research, MIT.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.neuroimage.2019.116096>.

References

- Acheson, D.J., Hamidi, M., Binder, J.R., Postle, B.R., 2011. A common neural substrate for language production and verbal working memory. *J. Cogn. Neurosci.* 23 (6), 1358–1367.
- Ackermann, H., Mathiak, K., Riecker, A., 2007. The contribution of the cerebellum to speech production and speech perception: clinical and functional imaging data. *Cerebellum* 6 (3), 202–213.
- Adams, A.M., Gathercole, S.E., 1995. Phonological working-memory and speech production in preschool-children. *J. Speech Hear. Res.* 38 (2), 403–414.
- Adams, A.M., Gathercole, S.E., 1996. Phonological working memory and spoken language development in young children. *Q. J. Exp. Psychol.* 49 (1), 216–233.
- Adank, P., Devlin, J.T., 2010. On-line plasticity in spoken sentence comprehension: adapting to time-compressed speech. *Neuroimage* 49 (1), 1124–1132.
- Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* 12, 26–41.
- Awh, E., Jonides, J., Smith, E.E., Schumacher, E.H., Koeppel, R.A., Katz, S., 1996. Dissociation of storage and rehearsal in verbal working memory: evidence from positron emission tomography. *Psychol. Sci.* 7 (1), 25–31.
- Baddeley, A.D., 1986. *Working Memory*. Oxford (UK): Clarendon Press. Oxford University Press, New York (NY).
- Baddeley, A.D., 2003. Working memory: looking back and looking forward. *Nat. Rev. Neurosci.* 4 (10), 829–839.
- Baddeley, A.D., Gathercole, S., Papagno, C., 1998. The phonological loop as a language learning device. *Psychol. Rev.* 105 (1), 158–173.
- Baddeley, A.D., Hitch, G., 1974. Working memory. *Psychol. Learn. Motiv.* 8, 47–89.
- Baddeley, A.D., Thomson, N., Buchanan, M., 1975. Word length and structure of short-term-memory. *J. Verb. Learn. Verb. Behav.* 14 (6), 575–589.
- Barry, J.G., Sabisch, B., Friederici, A.D., Brauer, J., 2011. Encoding: the keystone to efficient functioning of verbal short-term memory. *Neuropsychologia* 49 (13), 3636–3647.
- Basilakos, A., Smith, K.G., Fillmore, P., Fridriksson, J., Fedorenko, E., 2018. Functional characterization of the human speech articulation network. *Cerebr. Cortex* 28 (5), 1816–1830.
- Binder, J.R., Roa, S.M., Hammeke, T.A., Frost, J.A., Bandettini, P.A., Hyde, J.S., 1994. Effects of stimulus rate on signal response during functional magnetic resonance imaging of auditory cortex. *Cogn. Brain Res.* 2 (1), 31–38.
- Boersma, P., 2001. Praat, a system for doing phonetics by computer. *Glott Int.* 5, 341–345.
- Bowers, A., Bowers, L.M., Hudock, D., Ramsdell-Hudock, H.L., 2018. Phonological working memory in developmental stuttering: potential insights from the neurobiology of language and cognition. *J. Fluency Disord.* 58, 94–117.
- Buchsbaum, B.R., Baldo, J., Okada, K., Berman, K.F., Dronkers, N., D'Esposito, M., Hickok, G., 2011. Conduction aphasia, sensory-motor integration, and phonological short-term memory – an aggregate analysis of lesion and fMRI data. *Brain Lang.* 119 (3), 119–128.
- Buchsbaum, B.R., Hickok, G., Humphries, C., 2001. Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cogn. Sci.* 25 (5), 663–678.
- Buchsbaum, B.R., Olsen, R.K., Koch, P., Berman, K.F., 2005. Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron* 48 (4), 687–697.
- Byrd, C.T., McGill, M., Usler, E., 2015. Nonword repetition and phoneme elision in adults who do and do not stutter: vocal versus nonvocal performance differences. *J. Fluency Disord.* 44, 17–31.
- Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9 (2), 179–194.
- Deen, B., Koldewyn, K., Kanwisher, N., Saxe, R., 2015. Functional organization of social perception and cognition in the superior temporal sulcus. *Cerebr. Cortex* 25 (11), 4596–4609.
- Dollaghan, C., Campbell, T.F., 1998. Nonword repetition and child language impairment. *J. Speech Lang. Hear. Res.* 41 (5), 1136–1146.
- Dufva, M., Niemi, P., Voeten, M.J.M., 2001. The role of phonological memory, word recognition, and comprehension skills in reading development: from preschool to grade 2. *Read. Writ.* 14 (1–2), 91–117.
- Eklund, A., Nichols, T.E., Knutsson, H., 2016. Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. *Proc. Natl. Acad. Sci. U. S. A.* 113 (28), 7900–7905.

- Epstein, R., Kanwisher, N., 1998. A cortical representation of the local visual environment. *Nature* 392 (6676), 598–601.
- Estes, K.G., Evans, J.L., Else-Quest, N.M., 2007. Differences in the nonword repetition performance of children with and without specific language impairment: a meta-analysis. *J. Speech Lang. Hear. Res.* 50 (1), 177–195.
- Fedorenko, E., Duncan, J., Kanwisher, N., 2012. Language-selective and domain-general regions lie side by side within Broca's area. *Curr. Biol.* 22 (21), 2059–2062.
- Fedorenko, E., Duncan, J., Kanwisher, N., 2013. Broad domain generality in focal regions of frontal and parietal cortex. *Proc. Natl. Acad. Sci. U. S. A.* 110 (41), 16616–16621.
- Fedorenko, E., Hsieh, P.J., Nieto-Castañón, A., Whitfield-Gabrieli, S., Kanwisher, N., 2010. New method for fMRI investigations of language: defining ROIs functionally in individual subjects. *J. Neurophysiol.* 104 (2), 1177–1194.
- Fegen, D., Buchsbaum, B.R., D'Esposito, M., 2015. The effect of rehearsal rate and memory load on verbal working memory. *Neuroimage* 105, 120–131.
- Fiez, J.A., 2015. Neural basis of phonological short-term memory. In: Hickok, G., Small, S.L. (Eds.), *Neurobiology of Language*. Academic Press, London (UK), pp. 855–862.
- Fischl, B., Rajendran, N., Busa, E., Augustinack, J., Hinds, O., Yeo, B.T., Mohlberg, H., Amunts, K., Zilles, K., 2008. Cortical folding patterns and predicting cytoarchitecture. *Cerebr. Cortex* 18 (8), 1973–1980.
- Frost, M.A., Goebel, R., 2012. Measuring structural-functional correspondence: spatial variability of specialized brain regions after macro-anatomical alignment. *Neuroimage* 59 (2), 1369–1381.
- Gainotti, G., Miceli, G., Silveri, M.C., Villa, G., 1982. Some anatomo-clinical aspects of phonemic and semantic comprehension disorders in aphasia. *Acta Neurol. Scand.* 66, 652–665.
- Gathercole, S.E., Alloway, T.P., Willis, C., Adams, A.M., 2006. Working memory in children with reading disabilities. *J. Exp. Child Psychol.* 93 (3), 265–281.
- Gathercole, S.E., Baddeley, A.D., 1993. Phonological working-memory – a critical building-block for reading development and vocabulary acquisition. *Eur. J. Psychol. Educ.* 8 (3), 259–272.
- Gathercole, S.E., Baddeley, A.D., 1996. *The Children's Test of Nonword Repetition (CNRep)*. The Psychological Corporation, San Antonio (TX).
- Gathercole, S.E., Willis, C.S., Baddeley, A.D., Emslie, H., 1994. *The Children's Test of Nonword Repetition: a test of phonological working memory*. *Memory* 2 (2), 103–127.
- Genon, S., Reid, A., Langner, R., Amunts, K., Eickhoff, S.B., 2018. How to characterize the function of a brain region. *Trends Cogn. Sci.* 22 (4), 350–364.
- Gorgolewski, K., Burns, C.D., Madison, C., Clark, D., Halchenko, Y.O., Waskom, M.L., Ghosh, S.S., 2011. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Front. Neuroinf.* 5, 13.
- Greve, D.N., Fischl, B., 2009. Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* 48 (1), 63–72.
- Hall, D.A., Haggard, M.P., Akeroyd, M.A., Palmer, A.R., Summerfield, A.Q., Elliott, M.R., Gurney, E.M., Bowtell, R.W., 1999. Sparse temporal sampling in auditory fMRI. *Hum. Brain Mapp.* 7 (3), 213–223.
- Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293 (5539), 2425–2430.
- Hickok, G., Okada, K., Barr, W., Pa, J., Rogalsky, C., Donnelly, K., Barde, L., Grant, A., 2008. Bilateral capacity for speech sound processing in auditory comprehension: evidence from Wada procedures. *Brain Lang.* 107, 179–184.
- Hickok, G., Okada, K., Serences, J.T., 2009. Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *J. Neurophysiol.* 101 (5), 2725–2732.
- Jacquemot, C., Scott, S.K., 2006. What is the relationship between phonological short-term memory and speech processing? *Trends Cogn. Sci.* 10 (11), 480–486.
- Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M., 2012. FSL. *Neuroimage* 62 (2), 782–790.
- Julian, J.B., Fedorenko, E., Webster, J., Kanwisher, N., 2012. An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *Neuroimage* 60 (4), 2357–2364.
- Kanwisher, N., McDermott, J., Chun, M.M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17 (11), 4302–4311.
- Kjelgaard, M.M., Tager-Flusberg, H., 2001. An investigation of language impairment in autism: implications for genetic subgroups. *Lang. Cogn. Process.* 16 (2–3), 287–308.
- Koenigs, M., Acheson, D.J., Barbey, A.K., Solomon, J., Postle, B.R., Grafman, J., 2011. Areas of left perisylvian cortex mediate auditory-verbal short-term memory. *Neuropsychologia* 49 (13), 3612–3619.
- Kovelman, I., Norton, E.S., Christodoulou, J.A., Gaab, N., Lieberman, D.A., Triantafyllou, C., Wolf, M., Whitfield-Gabrieli, S., Gabrieli, J.D.E., 2012. Brain basis of phonological awareness for spoken language in children and its disruption in dyslexia. *Cerebr. Cortex* 22, 754–764.
- Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S.F., Baker, C.I., 2009. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat. Neurosci.* 12 (5), 535–540.
- Laird, A.R., Lancaster, J.L., Fox, P.T., 2005. BrainMap: the social evolution of a human brain mapping database. *Neuroinformatics* 3 (1), 65–78.
- Lanfranchi, S., Jerman, O., Vianello, R., 2009. Working memory and cognitive skills in individuals with down syndrome. *Child Neuropsychol.* 15 (4), 397–416.
- Leff, A.P., Schofield, T.M., Crinion, J.T., Seghier, M.L., Grogan, A., Green, D.W., Price, C.J., 2009. The left superior temporal gyrus is a shared substrate for auditory short-term memory and speech comprehension: evidence from 210 patients with stroke. *Brain* 132, 3401–3410.
- Lim, S.J., Wöstmann, M., Obleser, J., 2015. Selective attention to auditory memory neurally enhances perceptual precision. *J. Neurosci.* 35 (49), 16094–16104.
- Mahowald, K., Fedorenko, E., 2016. Reliable individual-level neural markers of high-level language processing: a necessary precursor for relating neural variability to behavioral and genetic variability. *Neuroimage* 139, 74–93.
- Majerus, S., 2013. Language repetition and short-term memory: an integrative framework. *Front. Hum. Neurosci.* 7, 357.
- Martin, R.C., 2005. Components of short-term memory and their relation to language processing – evidence from neuropsychology and neuroimaging. *Curr. Dir. Psychol. Sci.* 14 (4), 204–208.
- McGettigan, C., Warren, J.E., Eisner, F., Marshall, C.R., Shanmugalingam, P., Scott, S.K., 2011. Neural correlates of sublexical processing in phonological working memory. *J. Cogn. Neurosci.* 23 (4), 961–977.
- Meyer, F., 1991. *Un Algorithme Optimal Pour La Ligne De Partage Des Eaux*. Sme Congrès de Reconnaissance des Formes et Intelligence Artificielle, Lyon, France.
- Michalka, S.W., Kong, L., Rosen, M.L., Shinn-Cunningham, B.G., Somers, D.C., 2015. Short-term memory for space and time flexibly recruit complementary sensory-biased frontal lobe attention networks. *Neuron* 87, 882–892.
- Nieto-Castañón, A., Fedorenko, E., 2012. Subject-specific functional localizers increase sensitivity and functional resolution of multi-subject analyses. *Neuroimage* 63 (3), 1646–1669.
- Noyce, A.L., Cestero, N., Michalka, S.W., Shinn-Cunningham, B.G., Somers, D.C., 2017. Sensory-Biased and multiple-demand processing in human lateral frontal cortex. *J. Neurosci.* 37 (36), 8755–8766.
- Paulsen, E., Frith, C.D., Frackowiak, R.S.J., 1993. The neural correlates of the verbal component of working memory. *Nature* 362 (6418), 342–345.
- Peirce, J.W., Gray, J.R., Simpson, S., MacAskill, M.R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J., 2019. PsychoPy2: experiments in behavior made easy. *Behav. Res. Methods* 162 (1–2), 8–13.
- Peelle, J.E., 2012. The hemispheric lateralization of speech processing depends on what “speech” is: a hierarchical perspective. *Front. Hum. Neurosci.* 6, 309.
- Peelle, J.E., McMillan, C., Moore, P., Grossman, M., Wingfield, A., 2004. Dissociable patterns of brain activity during comprehension of rapid and syntactically complex speech: evidence from fMRI. *Brain Lang.* 91, 315–325.
- Perrachione, T.K., Ghosh, S.S., 2013. Optimized design and analysis of sparse-sampling fMRI experiments. *Front. Neurosci.* 7, 55.
- Perrachione, T.K., Ghosh, S.S., Ostrovskaya, I., Gabrieli, J.D.E., Kovelman, I., 2017. Phonological working memory for words and nonwords in cerebral cortex. *J. Speech Lang. Hear. Res.* 60 (7), 1959–1979.
- Peter, B., Raskind, W.H., Matsushita, M., Lisowski, M., Vu, T., Berninger, V.W., Wijsman, E.M., Brkanac, Z., 2011. Replication of CNTNAP2 association with nonword repetition and support for FOXP2 association with timed reading and motor activities in a dyslexia family sample. *J. Neurodev. Disord.* 3 (1), 39–49.
- Peterburg, J., Blevins, L.C., Sheu, Y.S., Desmond, J.E., 2019. Cerebellar contributions to sequence prediction in verbal working memory. *Brain Struct. Funct.* 224 (1), 485–499.
- Pisoni, A., Mattavelli, G., Casarotti, A., Comi, A., Riva, M., Bello, L., Papagno, C., 2019. The neural correlates of auditory-verbal short-term memory: a voxel-based lesion-symptom mapping study on 103 patients after glioma removal. *Brain Struct. Funct.* 224, 2199–2211.
- Rogalsky, C., Matchin, W., Hickok, G., 2008. Broca's area, sentence comprehension, and working memory: an fMRI study. *Front. Hum. Neurosci.* 2, 14.
- Rogalsky, C., Poppa, T., Chen, K.H., Anderson, S.W., Damasio, H., Love, T., Hickok, G., 2015. Speech repetition as a window on the neurobiology of auditory-motor integration for speech: a voxel-based lesion symptom mapping study. *Neuropsychologia* 71, 18–27.
- Salmon, E., VanderLinden, M., Collette, F., Delfiore, G., Maquet, P., Degueldre, C., Luxen, A., Franck, G., 1996. Regional brain activity during working memory tasks. *Brain* 119, 1617–1625.
- Savill, N., Ashton, J., Gugliuzza, J., Poole, C., Sim, Z., Ellis, A.W., Jefferies, E., 2015. tDCS to temporoparietal cortex during familiarization enhances the subsequent phonological coherence of nonwords in immediate serial recall. *Cortex* 63, 132–144.
- Saxe, R., Kanwisher, N., 2003. People thinking about thinking people. The role of the temporo-parietal junction in “theory of mind. *Neuroimage* 19 (4), 1835–1842.
- Siegel, J.S., Power, J.D., Dubis, J.W., Vogel, A.C., Church, J.A., Schlaggar, B.L., Petersen, S.E., 2014. Statistical improvements in functional magnetic resonance imaging analyses produced by censoring high-motion data points. *Hum. Brain Mapp.* 35, 1981–1996.
- Smith, E.E., Jonides, J., 1998. Neuroimaging analyses of human working memory. *Proc. Natl. Acad. Sci. U. S. A.* 95 (20), 12061–12068.
- Smith, S.M., Brady, J.M., 1997. SUSAN – a new approach to low level image processing. *Int. J. Comput. Vis.* 23 (1), 45–78.
- Strand, F., Forssberg, H., Klingberg, T., Norrelgen, F., 2008. Phonological working memory with auditory presentation of pseudo-words – an event related fMRI study. *Brain Res.* 1212, 48–54.
- Tahmasebi, A.M., Davis, M.H., Wild, C.J., Rodd, J.M., Hakyemez, H., Abolmaesumi, P., Johnsrude, I.S., 2012. Is the link between anatomical structure and function equally strong at all cognitive levels of processing? *Cerebr. Cortex* 22 (7), 1593–1603.
- Tootell, R.B.H., Reppas, J.B., Kwong, K.K., Malach, R., Born, R.T., Brady, T.J., Rosen, B.R., Belliveau, J.W., 1995. Functional-analysis of human MT and related visual cortical areas using magnetic-resonance-imaging. *J. Neurosci.* 15 (4), 3215–3230.
- Vagharchakian, L., Dehaene-Lambertz, G., Pallier, C., Dehaene, S., 2012. A temporal bottleneck in the language comprehension network. *J. Neurosci.* 32 (26), 9089–9102.
- van der Schuit, M., Segers, E., van Balkom, H., Verhoeven, L., 2011. How cognitive factors affect language development in children with intellectual disabilities. *Res. Dev. Disabil.* 32 (5), 1884–1894.

- Vitevitch, M.S., Luce, P.A., 2004. A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behav. Res. Methods Instrum. Comput.* 36, 481–487.
- Vitevitch, M.S., Luce, P.A., 2016. Phonological neighborhood effects in spoken word perception and production. *Annu. Rev. Ling.* 2 (1), 75–94.
- Vul, E., Kanwisher, N., 2010. Begging the question: the non-independence error in fMRI data analysis. In: Hanson, S.J., Bunzi, M. (Eds.), *Foundational Issues in Human Brain Mapping*. MIT Press, Cambridge (MA).
- Wagner, R.K., Torgesen, J.K., Rashotte, C.A., 1999. *Comprehensive Test of Phonological Processing (CTOPP)*. Pro-Ed, Austin (TX).
- Xie, X., Myers, E., 2018. Left inferior frontal gyrus sensitivity to phonetic competition in receptive language processing: a comparison of clear and conversational speech. *J. Cogn. Neurosci.* 30 (3), 267–280.
- Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., Wager, T.D., 2011. Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods* 8 (8), 665–670.
- Yue, Q., Martin, R.C., Hamilton, A.C., Rose, N.S., 2019. Non-perceptual regions in the left inferior parietal lobe support phonological short-term memory: evidence for a buffer account? *Cerebr. Cortex* 29, 1398–1413.