Distributed networks for auditory memory contribute differentially to recall precision

Sung-Joo Lim[1,2,3,4]*, Christiane Thiel[5], Bernhard Sehm[2], Lorenz Deserno[2], Jöran Lepsien[2], and Jonas Obleser[1,2]*

1. Department of Psychology, University of Lübeck, 23562, Lübeck, Germany
2. Max Planck Institute for Human Cognitive and Brain Sciences, 04103, Leipzig, Germany
3. Department of Psychology, Binghamton University, State University of New York, Vestal, NY, USA
4. Department of Speech, Language, and Hearing Sciences, Boston University, 02215, Boston, MA, USA
5. Department of Psychology, Carl von Ossietzky University of Oldenburg, 26129, Oldenburg, Germany

Running title
Distributed auditory memory representations

***Correspondence:**
Sung-Joo Lim (sungjoo@binghamton.edu)
Department of Psychology, Binghamton University, State University of New York
4400 Vestal Parkway East, Science 4, Binghamton NY 13902, USA

Jonas Obleser (jonas.obleser@uni-luebeck.de)
Department of Psychology, University of Lübeck
Maria-Goeppert-Str 9a, Lübeck, 23562, Germany

Conflict of Interest: The authors declare no competing financial interests.

Acknowledgements: Research was supported by the Max Planck Society (Max Planck research grant to JO) and the European Research Council (ERC Consolidator grant AUDADAPT, no. 646696, to JO). The authors also thank Steven Kalinke, Sören Nikolaus, and Dunja Kunke for their help conducting the study.

Abstract

The representations held in working memory are inherently noisy, but attention directed to relevant objects can effectively enhance their fidelity. While recent working memory models suggest that memory representations are distributed across sensory and cognitive-control brain regions, it remains unknown how multiple brain networks generate this attentional gain in fidelity. Here, we investigated the contributions of the distinct brain networks in maintaining and enhancing memory representations using psychophysical modeling and fMRI. Human listeners performed an auditory syllable pitch-discrimination task, in which they received valid (vs. neutral) retro-active cues to selectively attend to one of the two syllable categories maintained in memory. Valid (vs. neutral) retro-cues facilitated task performance, eliciting faster recall and enhanced recall precision of syllables in memory. Valid retro-cues also led to increased neural activation in fronto-parietal and cingulo-opercular networks, but not in sensory-specific superior temporal cortex. Multivariate pattern analysis as a proxy for representational fidelity in memory revealed that attended syllable objects were maintained in distributed areas across superior temporal, frontal, parietal, and sensorimotor brain areas. However, neural fidelity in left superior temporal sulcus and its enhancement through attention-to-memory best predicted the ensuing individual gain in recall precision of auditory objects from memory. These results demonstrate that maintaining versus attentionally enhancing auditory memory representations are functionally separable mechanisms across distributed brain regions.

Significance Statement

Working memory is distributed across sensory and cognitive-control brain regions. But how do these brain networks enhance working memory precision when attention is re-directed to memory? We here investigate the contributions of distinct brain networks in maintaining and enhancing auditory memory representations through attention-to-memory using fMRI. We demonstrate that re-directing attention to the relevant auditory memory objects mainly recruits higher-order cognitive-control networks. Among the multiple brain regions retaining memory representations, however, attentional enhancement of the neural fidelity in superior temporal sulcus best predicts the individual gain in recall precision of auditory objects from memory. This study provides evidence of the interplay among the discrete, functionally specialized brain regions in maintaining and attentionally enhancing working memory representations.

Introduction

Working memory is a short-term mental storage that maintains perceptual information even in the absence of sensory input. A hallmark of working memory is in its limited capacity—that is, perceptual information held in memory are inherently imprecise and noisy (e.g., Cowan, 2001; Bays and Husain, 2008; Luck and Vogel, 2013; Ma et al., 2014). Selective attention thus plays a crucial role in overcoming the bottleneck of the limited capacity. Directing attention to the relevant information in memory can reduce inherent noise and enhance representations of the attended memory items—a process referred as retrospective attention (Griffin and Nobre, 2003; Postle, 2006; Gazzaley and Nobre, 2012; Serences and Kastner, 2014; Myers et al., 2017). However, the neural processes that pose a neural substrate of the behavioral benefit from retrospective attention remains unclear.

Research on the neural bases of working memory and selective attention implicates that multiple functional brain networks support active maintenance of the attended memory items. Neuroimaging research across various modalities suggests that both domain-general networks related to attention (i.e., fronto-parietal network) and cognitive control (i.e., cingulo-opercular network), and the sensory-specific neural regions are involved in maintaining items held in working memory (e.g., Goldman-Rakic, 1995; Harrison and Tong, 2009; Higo et al., 2011; Riggall and Postle, 2012; Kiyonaga and Egner, 2013; D'Esposito and Postle, 2015; Ester et al., 2015). For example, auditory and verbal working memory research demonstrated contributions of the domain-general networks and language-related brain regions (e.g., the inferior/middle frontal and premotor regions; (Baddeley, 2003; Hickok and Poeppel, 2007; Hickok, 2009; Koelsch et al., 2009; Fedorenko et al., 2011), and the bilateral auditory cortices retaining stimulus-specific features of the auditory information held in memory (Linke et al., 2011; Linke and Cusack, 2015; Kumar et al., 2016).

Collectively, the converging evidence suggests that neural traces of memory representations under focused attention are maintained across distributed brain networks. However, we do not know whether and how the discrete networks support the neural enhancement from attention directed to memory. In particular, how do these large-scale brain networks support retrospective attention leading to enhancement? How do the memory representations maintained in these brain networks directly affect the attentional gain in improving the mnemonic fidelity? These questions are especially important in understanding auditory memory representations as selective attention plays a crucial role in maintaining precise auditory objects in memory while suppressing irrelevant noise often intermixed with task-relevant acoustic signals (Fritz et al., 2007; Shinn-Cunningham, 2008; Wilsch and Obleser, 2016).

Prior auditory working memory work has demonstrated that retrospective attention enhances representational precision of the attended versus unattended objects held in memory (Lim et al., 2015; 2018). Importantly, such precision benefit has been reflected in EEG neural signatures (i.e., enhanced sustained negativity and neural alpha (~10 Hz) oscillatory power), indicating that greater cognitive resource is allocated to the attended memory object, which ultimately enhances its representational precision (Lim et al., 2015). However, the limited spatial resolution of EEG cannot delineate how distinct brain networks—cognitive control vs. auditory processing regions—contribute to precision benefit from retrospective attention.

To investigate the neural underpinning of retrospective attentional gain in mnemonic fidelity, we conducted an fMRI experiment, in which participants performed a previously established auditory working memory task—a delayed recall of syllable pitch discrimination task (Lim et al., 2015; 2018; Alavash et al., 2018). In this task, listeners encoded two speech syllables into memory, and received one of the two types of retroactive cues that either directed attention to one of the syllable objects or indicated to maintain both syllables in memory.

So how should neural activities in domain-general and modality-specific brain regions be modulated when selective attention is re-directed to memory? Here, we used a multivariate classifier approach and fMRI data to quantify whether the memory representations were retained in aforementioned brain regions, and which of these brain regions contribute to the attentional gain in mnemonic fidelity. Given the existing controversy regarding whether different brain regions retain abstracted information versus stimulus-specific details in working memory (Christophel et al., 2012; Lee et al., 2013; Ester et al., 2015; Christophel et al., 2017; Long and Kuhl, 2018), we expected that the domain-general cognitive control and sensory-specific brain regions would be differentially related to the benefits when attention is directed to objects maintained in memory.

Our results demonstrate that selective attention to auditory memory leads to increased neural activity in domain-general cognitive control networks, but not in the auditory-sensitive superior temporal cortex. Using a multivariate analysis approach, we further show that while the neural representation of the attended memory object is distributed across the brain, these regions differentially contribute to the attentional gain in improving representational fidelity. The neural fidelity and its enhancement through attention-to-memory in left superior temporal sulcus and frontal cortex, but not in parietal cortex, are shown to predict the ensuing individual gain in recall precision of auditory objects from memory.

Materials and Methods

Participants

Twenty-two participants (mean age: 27.9 years; SD: 2.75 years; age range: 25–35; 12 females) took part in the experiment that has been a part of a previously published study (Alavash et al., 2018). Participants were recruited from the participant database of the Max Planck Institute for Human Cognitive and Brain Sciences (Leipzig, Germany). Two additional participants underwent the study, but were excluded from the all data analyses due to excessive head movements (>5 mm). All subjects reported normal hearing, and no history of neurological or psychiatric disorders. None of the participants were under medication. All participants except one were right handed.

All participants gave their informed written consent, and received financial compensation for their participation. The study procedure was in accordance with the Declaration of Helsinki. The local ethics committee (University of Leipzig) approved the study.

Stimuli

The stimulus material was identical to the ones used in previously published studies (Lim et al., 2015; 2018; Alavash et al., 2018). In brief, we used two German syllables, /da/ and /ge/. Each syllable category consisted of six naturally spoken tokens, truncated from German words recorded by a trained female speaker. All sound tokens were digitized at 44.1 kHz, and 200-ms in duration. For each token, 3-ms and 30-ms linear ramps were applied at sound onset and offset, respectively.

For each syllable category, two tokens served as to-be-probed syllables in the syllable-pitch discrimination task. The pitch dimension (i.e., the fundamental frequency; F0) of each token was parametrically manipulated in four steps: ±0.125 and ±0.75 semitones from its F0. The original token was presented during the syllable encoding phase, and the F0-manipulated tokens were presented as probes. The average F0 of the to-be-probed /da/ tokens was 162.8 Hz [min = 154.0, max = 171.7]. For syllable category /ge/, the average F0 was 175.5 Hz [min = 169.4, max = 183.5]. In order to increase acoustic variability during the syllable encoding phase, we created additional 12 tokens per syllable category based on the six utterances from the same speaker. These sounds served as unprobed tokens: they were presented during encoding, but were not used for discriminating pitch change at probe. On average, unprobed tokens of /da/ and /ge/ had F0's of 163.7 Hz [min= 152.8, max = 181.7] and 175.9 Hz [min = 163.7, max = 186.9], respectively. These unprobed sounds were used to ensure that participants were exposed to acoustically-variable sounds beyond the fixed set of to-be-

probed tokens, but within the similar F0-range such that to-be-probed and unprobed sounds were indistinguishable during the encoding phase.

All stimuli were normalized to equivalent amplitude (root-mean-squared dB full scale; RMS dBFS). Praat version 5.3 was used to manipulate the F0 dimension.

Experimental Task

In the MRI scanner, participants performed a previously established auditory working memory task—a syllable-pitch discrimination task with retroactive cues (Lim et al., 2015; 2018; Alavash et al., 2018). Figure 1A illustrates the task structure. On each trial, participants first heard two syllable tokens (i.e., /da/ and /ge/) in a random order sequence, separated by a 1- or 2-s inter-stimulus interval (ISI). After a delay period of 3–5 s, participants saw a visual retro-cue for 1 s displayed on the screen. With an additional delay of 5–7 s following the visual cue, participants heard a probe syllable. The probe was one of the two syllables heard during encoding with a slight manipulation in the pitch (F0). Within a 4-s time window, participants responded whether the pitch of the probe syllable was higher or lower than the corresponding syllable (i.e., the same category) heard during encoding in the beginning of the trial. Response feedback was provided on the screen for 0.5 s. Trials were separated by a 3–7 s silence period with a fixation cross on the screen.

During the retention period on each trial, participants saw one of the two retro-cues: valid or neutral. A valid retro-cue indicated the identity of the probe syllable category; a written syllable (e.g., “da”) was presented on the screen for 1 s in order to direct participants’ attention to one of the syllables held in memory prior to hearing the probe. On the neutral retro-cue trials, participants were presented with “xx” on the screen. This cue did not provide any information about the upcoming probe syllable. Thus, listeners must maintain both syllable sounds in memory until hearing a probe.

Participants went through a total of eight blocks of 16 trials. The valid and neutral retro-cue trials appeared on an equal probability on each block; thus, participants were unaware of trial type until seeing a visual retro-cue. For every two blocks, unique combinations of 2 syllable categories × 2 syllable positions during encoding being probed × 2 retro-cue types × 4 probed pitch step sizes were presented. It is of note that pitch (F0’s) of the syllables presented during the encoding phase varied continuously across trials; thus, listeners must remember specific sound instance of the syllable in order to perform the task.

Procedure

Prior to fMRI scanning, all participants completed a short practice session of the syllable pitch discrimination task on a laptop in a separate behavioral testing room. The practice session involved 18 trials (9 trials for each cue type), which only probed easily detectable syllable pitch changes (i.e., ±1.25 semitones). This practice session was given to ensure that participants understood the task.

In the scanner, participants used left and right index fingers to provide “high” and “low” responses. The mapping between the responses and hands were counterbalanced across participants. Experimental trials were controlled with Presentation software (Neurobehavioral Systems). Auditory stimuli were delivered through MR-compatible headphones (MrConfon, Magdeburg, Germany), and Music safe pro earplugs (Alpine Hearing Protection) provided an additional attenuation (on average 20-dB across 125–8000 Hz range). Prior to the main experimental task, a short hearing test was conducted with the auditory syllable stimuli to ensure that participants could hear the syllables with an on-going scanner noise.

After the eight scanning runs of the syllable-pitch discrimination task, participants went through two additional scanning runs of an auditory habituation (~16 min), involving listening to a sequence of four syllable tokens while watching a silent movie. However, excessive head movements were observed in at least 7/22 participants (> 4.5 mm) during this task; therefore, data from this task was not further reported.

Note that the data and results reported here have been acquired as part of a larger, pharmacological fMRI study such that all participants went through two identical fMRI sessions, separated by at least one week (EudraCT number 2015-002761-33). The study design was a double-blind, counterbalanced repeated-measures of the intake of levodopa (L-dopa) and placebo. Because our current research focus is on the neural basis of retrospective attention to auditory working memory (rather than the effect of L-dopa), we report the results only from the placebo session. Of significant note, in all our analyses, we controlled a potential confound in the placebo session order differences across participants. The effect of L-dopa on auditory working memory, corresponding functional neural organization, and the detailed procedure of the pharmacological fMRI study has been described elsewhere and can be found in Alavash et al. (2018).

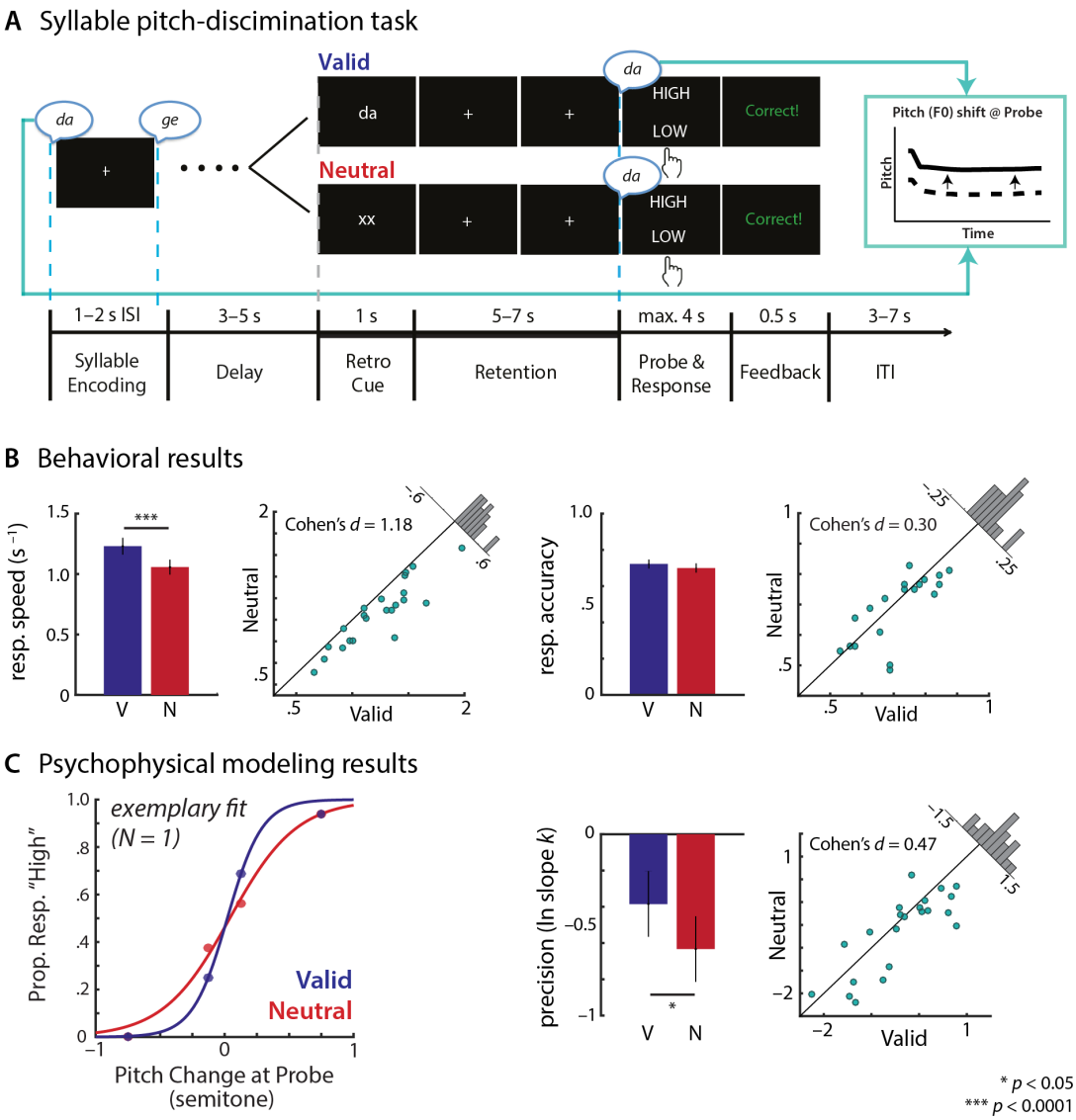

Figure 1. Illustration of the syllable pitch-discrimination task and the summary of task performance and psychophysical modeling results for validly cued (V) and neutrally cued (N) trials. (A) Trial structure of the task. On each trial, participants encoded two syllables (each syllable 0.2 s in duration) presented in a random order separated by a 1- or 2-s inter-stimulus interval (ISI). After the first delay period (3–5 s), a visual retro-cue appeared on the screen for 1 s, which was followed by a stimulus-free retention period (with duration jittered from 5–7 s). After the retention period, participants heard one of the encoded syllables again as a probe, the pitch of which was slightly manipulated. Participants judged whether the pitch of the probe syllable was higher or lower than the corresponding syllable heard during encoding (i.e., beginning of the trial). Participants were given a maximum of 4 s to provide their response. On every trial, participants received either a valid or neutral retro-cue, which respectively informed or did not inform the identity of the to-be-probed syllable category. (B) The behavioral performance results as a function of retro-cues. Left, response speed (1/response time in seconds) of the correctly responded trials. Right, response accuracy. Bar plot shows the average performance measure in the valid and neutral retro-cue conditions. Error bars indicate ± 1 standard error of the mean (SEM). Scatter dots represent the corresponding behavioral measure with respect to the cue conditions of individual participants. The 45-degree lines represent the performances in the two retro-cue conditions are identical. The histograms represent the distributions of the cue-related behavioral difference across N=22 participants. (C) The parameter estimates of psychophysical modeling results. Left, an exemplary

psychophysical modeling fit of a single participant's response patterns in the valid and neutral retro-cue conditions. Each dot indicates the proportion responses "high" as a function of syllable pitch change at probe, with respect to the matching reference syllable presented during the encoding phase. The lines denote the logistic model fits. Right, individual's perceptual precision estimates (ln slope *k*) with respect to the cue conditions. The illustration scheme is identical to panel B.

Data Analysis

Behavioral measures and psychophysical modeling analysis

Response speed of the correctly responded-to trials (i.e., inverse of correct-trials response time) and accurate syllable pitch judgment (i.e., a binary response accuracy measure) served as overall behavioral performance measures. Beyond these two behavioral performance measures, we also obtained a more fine-grained perceptual precision measure using a psychophysical modeling approach (e.g., Bays and Husain, 2008; Zhang and Luck, 2008; Murray et al., 2013). We estimated individual participant's perceptual precision in judging syllable pitch change by fitting each participant's response patterns along the varying levels of pitch changes occurred at probe. Following the previous work (Lim et al., 2015; 2018), we used a nonlinear least square curve fitting procedure (MATLAB *lsqcurvefit*) to fit individual's response patterns with a logistic (sigmoid) function,

$$y = \frac{1}{1+e^{-k(x-m)}}$$

where x indicates the amount of pitch changes occurred at probe relative to the syllable (4 step sizes), y indicates the proportion of "high" responses, k indicates the slope of the curve, and m indicates the mid-point/inflection point of the function along the x-axis. The parameter m estimates bias in each participant's response, and the slope k provides an estimate of the perceptual precision in pitch judgment in the task: higher perceptual precision is indicated by steeper slope (see Figure 1C, left for a single subject exemplary model fit).

For statistical analyses, the two single-trial behavioral measures (i.e., correct response speed and response accuracy) and the psychophysical modeling estimates of perceptual precision slope k and bias m were separately analyzed using a (generalized) linear mixed-effects model framework (*lme4*, ver. 3.3) in R environment. The model included a fixed factor of retro-cue (Valid vs. Neutral) and the random-effects term included by-participant intercepts. In order to remove any confounding effects due to the differences in the session order, the session order was centered and entered as a regressor of no interest. The significance of the effects was determined based on Wald χ^2 tests (at $p < 0.05$).

FMRI data acquisition and analysis

Imaging data were acquired on a 3-Tesla Siemens MAGNETOM Prisma^{fit} scanner (TimTrio Upgrade) at the Max Planck Institute for Human Cognitive and Brain Sciences. Whole-brain functional images, sensitive to the blood-oxygen-level-dependent (BOLD) contrast, were acquired with a 20-channel head/neck coil using a T2*-weighted echo-planar imaging (EPI) sequence (repetition time (TR) = 2000 ms; echo time (TE) = 26 ms; flip angle (FA) = 90°). Each volume had 40 oblique axial slices (ascending) parallel to the anterior commissure–posterior commissure (AC–PC) line (acquisition matrix = 64 × 64; field of view (FOV) = 192 mm × 192 mm; voxel size = 3 × 3 × 3 mm; interslice gap = 0.3 mm). For each block of the syllable-pitch discrimination task runs, a total of 181 volumes were acquired.

Structural images of (n = 15) participants were available from the Institute's brain database acquired from previous MR studies. A magnetization prepared rapid gradient echo (MP-RAGE) T1-weighted sequence was used to acquire the structural images (TR = 2300 ms; TE = 2.01–2.98 ms; FA = 9°; 176 saggital slices with 1 × 1 × 1 mm resolution). For the participants (n = 7) whose structural images did not exist in the database, a high-resolution T1-weighted structural image was acquired using a MP-RAGE sequence (TR = 2300 ms; TE = 2.98 ms; FA = 9°; 176 saggital slices with 1 × 1 × 1 mm resolution) at the end of functional scans.

Imaging data were preprocessed using the Analysis of Functional NeuroImages (AFNI; ver. 01-2017) software (Cox, 1996). The first two volumes of each run were discarded to allow the scanner to approach net magnetization equilibrium. Preprocessing steps included 1) slice-time correction, 2) motion correction of all volumes based on the first volume of the syllable-pitch discrimination task run using a rigid-body alignment, 3) spatial alignment of functional data to each participant's structural anatomy, and 4) the signal for each voxel was scaled to a global mean of 100. For univariate analysis, images were spatially smoothed with an 8-mm full-width at half maximum (FWHM) Gaussian kernel. For multivariate classification analysis, data were not smoothed. Spatial normalization of functional images to Montreal Neurological Images (MNI–152) was performed prior to the group-level statistical analyses.

Univariate analysis. For each participant, a voxel-wise general linear model (GLM) was constructed to estimate hemodynamic responses during different phases of the syllable-pitch discrimination task across all scanning runs. The BOLD response to the trial onset (i.e., the syllable encoding phase) was modeled using a canonical gamma variate function (GAM) with a pairwise amplitude-modulated regressor that accounted for the ISI between presentation of the two syllables. The second regressor captured the entire hemodynamic response of the visual retro-cue and the following stimulus-free retention phase up to auditory

probe phase using a finite impulse response (FIR) function (TENT by AFNI) as a change in BOLD activation across 0–20 s from the retro-cue onset. To account for the varying durations of the retention phase (Figure 1A), a parametric amplitude-modulated regressor was entered as a trial-wise covariate. To model the BOLD response related to the participants' response and visual feedback, we used a canonical GAM. All task-related regressors were separately modelled for the valid and neutral retro-cue trials in the GLM. The baseline activity was modelled by trends from linear up to fourth-degree polynomials in order to remove the slow scanner signal drift (i.e., high-pass filter with an approximate cut-off of 240 s). The six additional continuous head motion parameters were entered as regressors of no interest.

The primary purpose of the univariate analysis was to identify brain regions that were modulated by selective vs. non-selective attention directed to auditory working memory objects informed by retro-cues. Thus, we conducted a group-level analysis to contrast the neural activities during the cue and the following retention phase in the valid vs. neutral retro-cue conditions. For each subject, the cue-retention phase activity was quantified by averaging the β-parameter estimate of the FIR during 4–8 s from the cue onset, separately for the valid and neutral retro-cue conditions (Figure 2B).

The group-level analysis was conducted using a voxel-wise linear mixed-effects modeling approach (3dLME by AFNI). The model involved one within-subjects fixed factor of interest (retro-cue: Valid vs. Neutral) with subjects as a random factor, and regressor of no interest (i.e., placebo session order) was centered, and entered as covariates in the model. Planned contrast was established to differentiate activities during the cue–retention phase in the valid vs. neutral retro-cue conditions. The group-level maps were initially threshold at voxel-level $p = 0.001$ ($Z = 3.277$ for 2-tailed) and corrected for multiple comparisons at the cluster-level $p < 0.05$ using 10,000 Monte Carlo simulations implemented in the AFNI software (3dClustSim) accounting for the spatial autocorrelation function (ACF) estimating the smoothness of the residual images (Eklund et al., 2016).

ROI-based analysis. We conducted an additional ROI-based analysis to (*i*) complement and to confirm the whole-brain univariate analysis with regards to whether the valid vs. neutral retro-cues modulate the brain activities in the domain-general top-down networks established by previous literature; and, more importantly, to (*ii*) investigate whether the neural activity modulations in these distinct networks are related to behavior (see below). To this end, we extracted the cue–retention phase neural activities from the brain regions organized by intrinsic connectivity networks (e.g., Dosenbach et al., 2007; 2008; Power et al., 2011; Sadaghiani and D'Esposito, 2014). A 6-mm spherical radius seed was placed at the center of network coordinates defined in previous literature (see Figure 2A).

For the fronto-parietal network, the seeds were placed over the bilateral dorsolateral prefrontal cortex (MNI coordinates: [–44, 27, 33] and [46, 28, 31]) and the bilateral inferior parietal lobules ([-53 -50 39], and [54 -44 43]) as shown in Dosenbach et al. (2007) and Power et al. (2011). Seeds representing the cingulo-opercular network were placed based on Dosenbach et al. (2007) in the bilateral insula/frontal operculum ([–36, 18, 2] and [38, 21, –1]), and the supplementary motor area (SMA)/dorsal anterior cingulate cortex at [0, 15, 45]. Default mode network seeds were placed at the posterior cingulate cortex [1, –51, 29] and medial prefrontal cortex [–1, 61, 22] based on findings by Power et al. (2011) and Raichle et al. (2001). In order to observe neural activity modulated by the retro-cues in the sensory processing (i.e., auditory) regions, we also placed spherical seeds, centered over the peak activations during encoding of the two syllables at the onset of trial in the current experiment; these were placed in the bilateral auditory regions (superior temporal gyrus/sulcus) at [–58, –24, 12] and [62, 22, 2].

To test whether and how the retro-cues modulate brain activity in each of the four network ROIs, we first extracted the average β estimate during the cue and the following memory retention phase from the voxels within the seeds representing each network. The average β estimate data of each network were submitted to a linear mixed-effects model with a fixed factor of retro-cue (Valid vs. Neutral), a nuisance regressor of the session order, and subject-wise random intercept.

Multivariate analysis. The main purpose of the multi-voxel pattern analysis (MVPA) was to examine (*i*) whether the attended syllable category information in working memory (i.e., the retention phase) can be decoded from the activity patterns across voxels, and (*ii*) whether the decoding performance for syllable information differs between the valid vs. neutral retro-cue trials.

We trained pattern classifiers to decode syllable category information (i.e., /da/ vs. /ge/) from the brain activity patterns when participants “heard” one of the two syllables at the end of the trial (i.e., the auditory probe phase). We then tested whether the trained classifier can successfully decode syllable category information from the brain activity patterns during the memory maintenance phase (i.e., the retro-cue and the following stimulus-free retention period). Of important note, in order to avoid potential syllable category-specific bias when encountering the auditory probe following the valid retro-cues, classifier training was performed only on the auditory probe phase of the neutral retro-cue trials, in which participants did not have any expectation about the upcoming probe syllable. Thus, training and testing phases for the classifier were separate and temporally independent from each other (especially, for the valid retro-cue trials).

In order to find brain regions, in which patterns of neural activity were sensitive to syllable category decoding, we performed a whole-brain searchlight analysis with a searchlight sphere of 10 mm radius (~123 voxels; Kriegeskorte et al., 2006; Haynes et al., 2007). For each retro-cue condition, a separate binary classification of category syllable (/da/ vs. /ge/) was performed using linear support vector machine (Chang and Lin, 2011). This approach fits a hyperplane $w^T x + b = 0$: given a feature vector x (i.e., activations in voxels in a searchlight sphere) of the training data points, a linear SVM finds the weight vector w and an offset b to separate data into two classes (i.e., /da/ vs. /ge/). To this end, we constructed a single subject-level GLM for each functional run (i.e., eight runs per participant). As in the univariate analysis, a 0–20 s FIR was used to model the hemodynamic response ranging from the retro-cue onset, the following retention, to the auditory probe phases. Four, 20-s FIRs were included in the GLM to separately model the hemodynamic response of 2 cue types (i.e., valid and neutral) × 2 to-be-probed syllable categories (i.e., /da/ and /ge/). From each run-wise GLM, condition-specific activity patterns were estimated with *t*-estimates for the MVPA (Misaki et al., 2010).

For cross-validation of the generalization of classifier performance, eight functional runs were divided into training and test datasets. The classifier was trained on the average *t*-estimates of the auditory probe phase (i.e., 10–12s in the FIR) of the neutral-cue trials of five randomly selected functional runs, and the classifier was tested on the average *t*-estimates of the cue to retention phase (i.e., 4–8s in the FIR) of each cue condition from the remaining three functional runs. The cross-validation iteration was performed 28 times (i.e., a half of all possible combinations of five training and three test datasets), and the decoding accuracy was calculated based on the averaged across all iterations, separately for each cue condition.

Searchlight MVPA was performed on unsmoothed functional data in each participant's native space. Assessment of classification performances of individuals was tested against chance (i.e., 50%). The resulting decoding accuracy maps of participants were normalized to standard MNI space, and subjected to the group-level analysis. As in the univariate analysis, we used the voxel-wise linear mixed-effects modeling approach (3dLME) for the group-level analysis of the retro-cue conditions as a within-subjects factor and subjects as a random factor while accounting for regressor of no interest (i.e., the fMRI session order). Using this model, we evaluated the accuracy of the syllable-specific decoding of brain regions from each cue condition separately (1-tailed test, accuracy above chance at 50%). From these regions, we extracted syllable-specific decoding performances for the valid and neutral retro-cue conditions. Statistical thresholds for the multivariate group-level analysis were set at voxel-level $p < 0.001$ ($Z = 3.091$ for 1-tailed contrast) and cluster-size threshold was estimated by using 10,000 Monte Carlo simulations to correct for multiple comparisons problem at $p < 0.05$.

Brain–Behavior relationships. In order to understand how the network of task-relevant brain regions contribute to the representational precision benefits from valid retro-cues (i.e., Valid – Neutral: Δ ln slope k), we performed a multiple linear regression analysis to examine the relationship between the degrees of neural modulations and the representational precision modulations from the retro-cues (i.e., Valid vs. Neutral) across individuals. The two types of neural modulation measures, one from the univariate and the other from the multivariate results, were analyzed separately to explain the extent of the precision enhancement with valid (vs. neutral) retro-cues.

Starting from a model that includes all factors representing neural modulations of multiple brain regions, we used a step-wise model comparison to find the best model that explains the precision modulations from valid vs. neutral cues. The model comparison was based on the AIC of the model (*StepAIC* from the package *MASS* ver. 7.3-49 in R) as well as χ^2 testing to deduce into a simplest model that explains the cue-related modulations of the perceptual precision.

In analyzing the univariate BOLD data, we performed the model comparison, starting from the full model—that is, the neural activity modulations derived from the four ROIs, including the bilateral auditory region (Figure 2), to explain the degree of the perceptual precision modulations (Δ ln slope k):

Δ ln slope k = Δ Auditory + Δ Fronto-Parietal + Δ Cingulo-Opercular + Δ Default-mode

where Δ represents the degree of cue-related modulation (i.e., Valid – Neutral).

Using a similar approach, we analyzed the multivariate data to understand whether the neural decoding accuracy modulations in the task-relevant brain regions were related to the extent of precision modulation from valid vs. neutral cues. To this end, we started the model comparison from seven task-relevant cortical regions that exhibited significantly accurate decoding of to-be-probed syllable categories during the memory retention period (Figure 3; Table 2). These regions were selected based on whether they fell within the language processing regions (Baddeley, 2003; Hickok and Poeppel, 2007; Hickok, 2009; Koelsch et al., 2009), auditory regions in the temporal cortex, domain-general attentional network regions (i.e., fronto-parietal network). We used a step-wise model comparison to deduce from the full-factor model to find the simplest/best model that explains the extent of precision modulation (Δ ln slope k: ln slope k_{Valid} – ln slope k_{Neutral}). To this end, we started from the full-factor model that included the neural decoding accuracy modulations from the seven regions (Table 2).

Note that all of the cue-related modulation data (i.e., the perceptual precision modulation and neural modulations across the brain regions) were z-scored prior to the model testing. In order to ensure the validity of the final model resulted from the model comparison

step, we performed a leave-one-subject-out cross-validation (i.e., 22 CV's). On each CV step, we used n=21 subjects' data to fit the model parameters, and tested on the remaining one subject's data. We compared the cross-validated fitted values vs. the actual data outcome using a correlation test; and we report the average adjusted R^2 from 22 cross-validation steps.

Results

Valid retro-cues facilitate syllable-pitch recall performance

We first examined the effect of retro-cues (valid vs. neutral) on behavioral performances. A linear mixed effects model on participants' response speed (i.e., an inverse of response time) of the correct trials revealed a significant main effect of retro-cues ($M_{Valid-Neutral} = 0.17$, $\chi^2(1) = 101.45$, $p << 0.0001$). As shown in Figure 1B (left), participants were significantly faster in correctly judging the pitch of syllables when they received valid retro-cues compared to the neutral cues. However, we did not observe similar robust benefit from valid retro-cues on listeners' response accuracy; while we observed 13/22 participants were more accurate in recalling syllable pitch with valid than neutral retro-cues, a logistic mixed effects model did not reveal a significant effect of retro-cues on response accuracy ($\chi^2(1) = 1.73$, $p = 0.19$; Figure 1B, right).

We further examined the effect of retro-cues on listeners' perceptual precision and response bias, estimated by the slope (k) and inflection/mid-point (m) parameters of logistic function fit from psychophysical modeling (see Materials and Methods; Figure 1C, left). A linear mixed-effects model on the log-transformed perceptual precision estimate (ln slope k) revealed a significant effect of retro-cue condition ($M_{Valid-Neutral} = 0.25$, $\chi^2(1) = 4.56$, $p = 0.033$). Consistent with the prior findings (Lim et al., 2015; 2018), valid retro-cues led to more precise recall of syllable pitch compared to the neutral retro-cue trials (Figure 1C, right). For the response bias estimate (m), a linear mixed-effects model revealed no significant effect of retro-cue condition ($\chi^2(1) = 1.22$, $p = 0.27$); moreover, neither of the cue conditions did exhibit significant bias (one sample t-test against 0, ts ≤ 1.11, ps ≥ 0.28).

It is of note that for any of these measures, we did not find any significant main or interaction effects related to the factor of no interest—that is, the order of the two separate scanning sessions taking place for each participant (all χ^2s < 0.36, all ps > 0.55).

Valid and neutral retro-cues differentially modulate top-down networks during memory maintenance

To identify brain regions that were modulated by retrospective auditory attention, we compared the brain activity of valid vs. neutral retro-cue trials during the memory retention period following visual retro-cues. As shown in Figure 2A, we found significantly greater recruitment of broad regions across the whole brain when participants received valid than neutral retro-cues (Table 1). While the effect was more prevalent in the left hemisphere, these regions include the bilateral inferior frontal gyri extending into the dorsolateral prefrontal cortex, and the bilateral inferior parietal regions, the left superior parietal lobule, and bilateral supramarginal gyri. In addition, we found increased activity for the valid than neutral retro-cues trials in the supplementary motor area (SMA) extending into the bilateral superior and medial frontal gyri, and cingulate regions, the bilateral insula as well as the putamen. Similar patterns of the effect were also observed in the bilateral precuneus/superior parietal lobule (BA 7).

Furthermore, the greater activations in the valid vs. neutral retro-cues was accompanied by stronger deactivations in the brain regions, typically identified as a default-mode network. As expected, the whole-brain analysis revealed significantly stronger deactivation for the valid than neutral retro-cues in the medial prefrontal gyrus (BA 10), the anterior cingulate (BA 32) and posterior cingulate cortex (BA 31), as well as bilateral angular gyrus regions (Figure 2A).

In addition to the whole-brain analysis, we also conducted a ROI-based analysis in order to examine the effect of valid vs. neutral retro-cues on the domain-general top-down attentional networks established in previous literature (e.g., Dosenbach et al., 2007; Sadaghiani and D'Esposito, 2014). The results from the ROI-based analysis were consistent with the result patterns revealed in the whole-brain analysis. While both retro-cue conditions recruited the task-relevant attentional networks (i.e., fronto-parietal and cingulo-opercular ROIs; Figure 2C), the corresponding linear mixed-effects model revealed that both of these networks exhibited a significantly higher activation in the valid than neutral retro-cues during the follow-up memory retention period (fronto-parietal ROI: $\chi^2(1) = 23.62$, $p << 0.0001$; cingulo-opercular ROI: $\chi^2(1) = 31.42$, $p << 0.0001$). As expected, the default mode network was significantly deactivated during the memory retention phase of the trial for both cue conditions, but valid retro-cues led to a significantly greater deactivation of the default mode network compared to the neutral retro-cue condition ($\chi^2(1) = 23.43$, $p << 0.0001$).

While the broad regions along the auditory cortex (i.e., the bilateral temporal cortex) were significantly recruited when listeners encode the auditory speech syllables at the trial

onset (Figure S1), these regions were not significantly activated during the memory retention period. The bilateral auditory cortical ROIs, defined by the regions maximally active when listeners encode the auditory syllables, were deactivated when listeners were maintaining syllable information during the memory retention period (Figure 2B–C; one-sample *t*-test against 0; Valid: M = –0.082, $t(21) = 1.98$, $p = 0.060$, Neutral: M = –0.090, $t(21) = 2.77$, $p = 0.012$). Furthermore, these auditory cortical regions did not exhibit significantly different BOLD activation in the valid vs. neutral retro-cue trials during memory retention (Figure 2B–C; $\chi^2(1) = 0.26$, $p = 0.61$). Likewise, the whole-brain analysis did not reveal any significant activation differences in the auditory cortex based on the cue conditions, except in the right middle temporal gyrus (BA 21) showing a relatively greater activation in the valid than neutral retro-cue trials; nevertheless, this region was not strongly active during the memory retention period (one-sample *t*-test against 0; Valid: M = 0.058, $t(21) = 2.38$ $p = 0.027$, Neutral: M = 0.016, $t(21) = 0.62$, $p = 0.55$).

We further examined whether the retrocue-related modulations in the BOLD activation in the brain networks, including the fronto-parietal, cingulo-opercular, default-mode, and the auditory networks defined by ROIs, were related to the retrocue-related enhancement of the perceptual precision measure (i.e., ln slope k_{Valid} – ln slope k_{Neutral}). The step-wise model comparison approach starting from the maximal model did not reveal any significant effect of BOLD modulations on the cue-related perceptual precision modulations; the simplest and the best model, which was revealed by this analysis, included a marginal effect of the cue-related BOLD modulation of the default-mode network ROI on the recall precision enhancement ($\beta = 0.38$, $t = 1.82$, $p = 0.085$) explaining 9.85% of the variance of the model (i.e., adjusted R^2). The leave-one-subject-out cross-validation of this final model exhibited that the average model predicted outcomes did not exhibit significant relationship to the observed outcome of cue-related precision modulation (Pearson's correlation $r = 0.15$, $p = 0.051$; average adjusted R^2s from 22 cross-validations = 0.098).

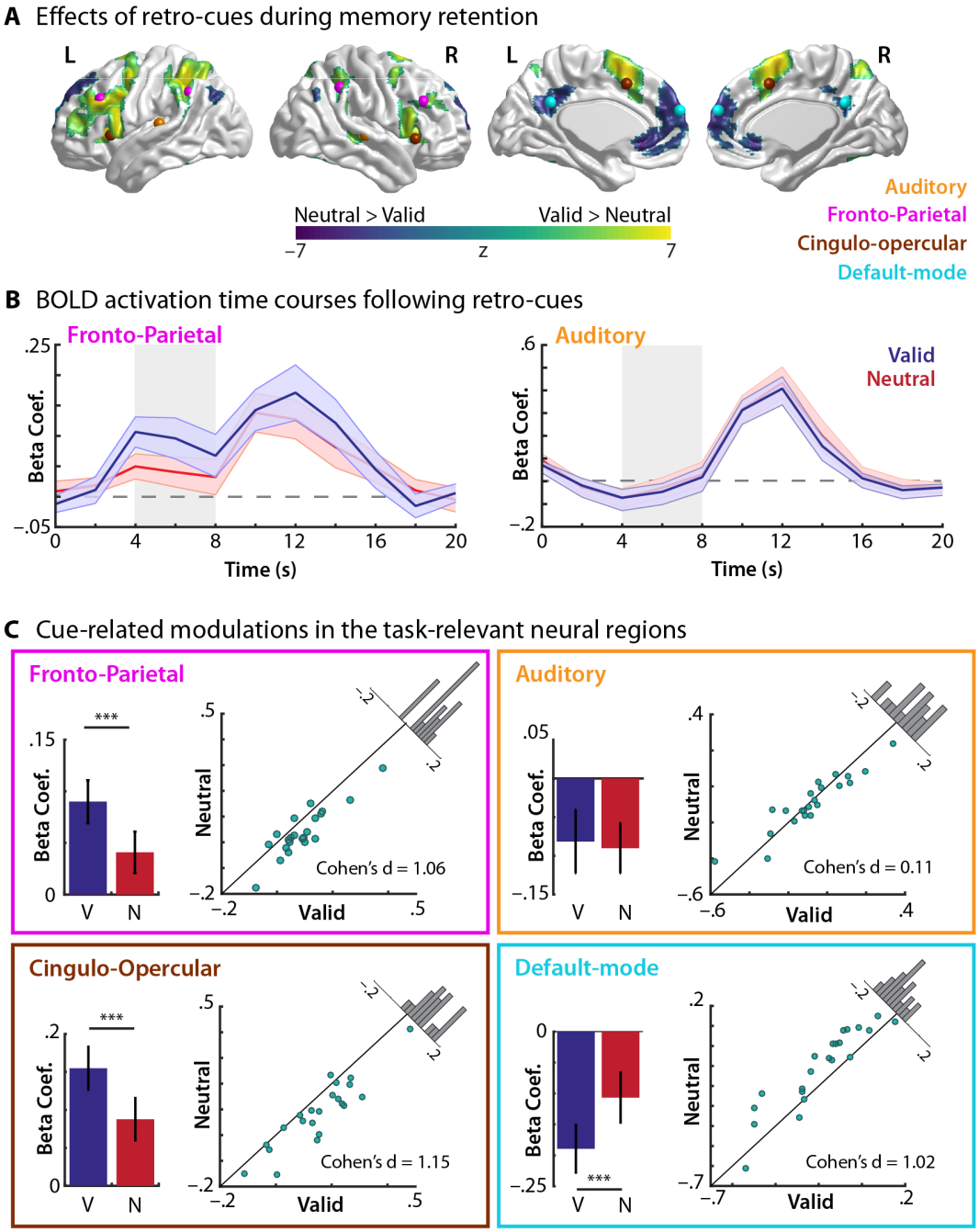

Figure 2. Result of the univariate analysis of Valid vs. Neutral retro-cues on modulating the brain activation during the post-cue memory retention phase. (A) Brain regions exhibiting significant effects of retro-cues (Valid, V – Neutral, N), thresholded at p <0.001 and corrected for multiple comparisons at p < .05. ROI seeds of the four networks are illustrated with the spheres of the corresponding colors. Visualization of the cortical surface is rendered with BrainNetViewer (Xia et al., 2013). L, left; R, right. (B) BOLD activation time courses of the Valid vs. Neutral retro-cue conditions from the cue onset. The BOLD time courses were extracted from the exemplary regions (fronto-parietal and auditory ROIs). The grey boxes denote the time windows of the BOLD activity during the cue and the following memory retention phase (4–8s; see Methods and Materials). (C) Cue-related modulation of the BOLD activation in each of the task-relevant ROIs. Bar graphs depict the average BOLD activations (i.e., beta coefficient) during the time points (4–8 s) of Valid and Neutral retro-cue conditions. Scatter plots represent the average BOLD activations of the corresponding ROIs with respect to the retro-cue conditions of individual participants. The 45-degree lines indicate identical magnitude of activation of the two cue

conditions. The histograms depict the distributions of the cue-related BOLD activity differences across N=22 participants. *** indicates $p < 0.0001$. Error bars in panels B and C indicate ± 1 SEM.

Table 1. Brain regions exhibiting a significant contrast of Valid vs. Neutral retro-cue trials during the cue and following memory retention period

Brain region	MNI peak coordinate			*Z*-score	Voxels
	x	*y*	*z*		
Valid > Neutral					
L Pre-SMA	–6	18	54	7.93	3692
L Inferior frontal gyrus (BA 44)	–48	8	26		
L Middle frontal gyrus (BA 6)	–42	2	50		
R Middle FG (BA 6)	44	2	54		
R Middle frontal gyrus	38	42	32		
L Insula/Inferior frontal gyrus (BA 24)	–34	24	6		
R Insula /Inferior frontal gyrus	36	24	2		
L Putamen	–20	6	8		
L Inferior parietal lobule (BA 40)	–36	52	50	7.40	897
L Superior parietal lobule	–32	–60	54		
R Inferior parietal lobule	36	–46	48	5.16	319
R Supramarginal gyrus	50	–42	30		
R Precuneus / Superior parietal lobule	8	–66	56	5.93	131
R Putamen	20	8	6	5.16	65
L Cerebellum	–34	–66	–24	5.02	127
R Cerebellum	38	–60	–28	5.89	241
R Cerebellum (semi lunar lobule)	24	–70	–48	4.36	47
R Middle temporal gyrus	54	–34	–4	4.62	59
Left postcentral gyrus (BA 2)	–60	–22	30	3.86	7
Neutral > Valid					
R Medial frontal gyrus (BA 10)	12	66	8	–6.27	1338
L Anterior cingulate (BA 32)	–2	30	–8		
L Cingulate gyrus (BA 31) / Precuneus	–6	–46	32	–5.54	317
L Intraparietal lobule / Precuneus	–42	–72	38	–4.51	64
R Angular gyrus	54	–66	36	–4.17	22

Note: L, left; R, right; BA, Brodmann area

Valid retro-cues enhance neural representations of syllable category in memory

Using the multivariate classification analysis, we investigated whether the speech syllable object retained in memory can be decoded from the patterns of neural activation during memory retention period upon the presentation of retro-cues. To this end, we trained the classifier to discriminate auditorily presented syllables (i.e., /da/ vs. /ge/) when participants heard them as a probe (i.e., one syllable presentation) only in the neutral cue trials in order to prevent confounds from listeners' expectation about the upcoming probe syllable informed by valid retro-cues. We then tested generalization of the trained classifier whether it could decode syllable object from patterns of neural activation during the cue–retention phase (i.e., in the absence of auditory presentation) of the task.

As illustrated in Figure 3, the multivariate analysis revealed the brain regions, in which attended syllable category objects can be decoded from neural activities during the cue–retention phase when listeners received valid retro-cues (Figure 3; Table 2). Specifically, we found that neural activity patterns in the left inferior frontal gyrus/dorsolateral prefrontal cortex, an anterior portion of the left superior temporal sulcus (STS), and right precentral gyrus adjoining the lateral Heschl's gyrus exhibited decoding accuracy significantly above chance (i.e., >50%) when a valid retro-cue was presented. In addition, activity patterns in the right precentral region and the left parietal cortical region can accurately distinguish syllable category representations during the cue–memory retention phase. As retro-cues were presented visually on the screen prior to the memory retention period, we also found that the visual processing regions, such as the left inferior temporal/fusiform gyrus region and left occipital/cuneus area exhibited above-chance decoding accuracy for syllable category information with valid retro-cues. However, as expected, the same analysis did not reveal any significant brain clusters exhibiting above-chance decoding accuracy for syllable objects under neutral retro-cue trials.

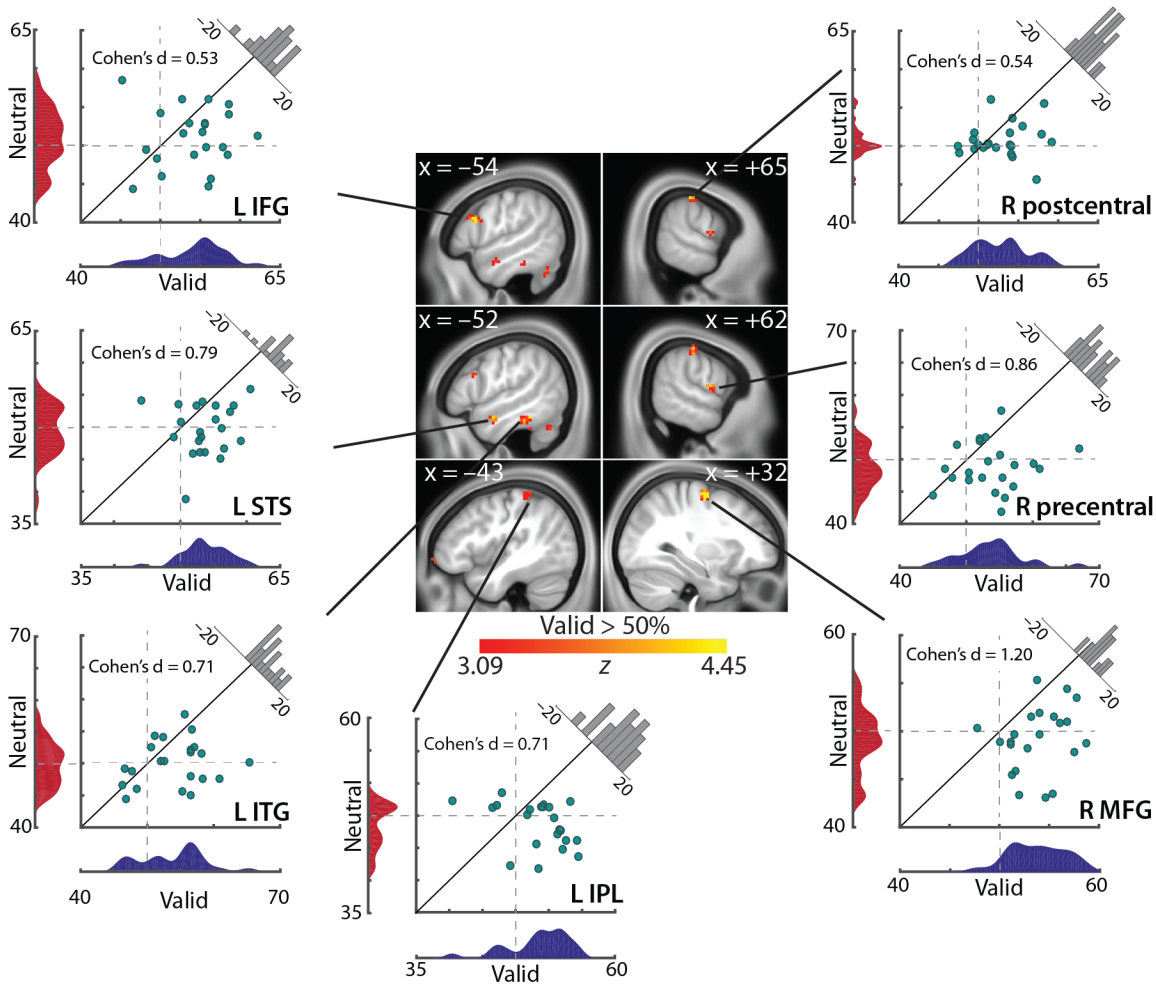

Figure 3. Multivariate classification results on decoding syllable category held in memory. The classifier was trained to decode syllable categories based on the auditorily presented probe sounds in the neutral retro-cue trials; and the classifier was tested to decode the neural activity patterns during the memory retention phase following retro-cues. The brain regions exhibit above-chance classifier accuracies on decoding the cued syllable category (i.e., indicated by valid retro-cues; see Methods and Materials for further details). Each scatter plot depicts individuals' decoding accuracies of the valid vs. neutral retro-cue conditions extracted from the corresponding brain regions that retain neural representations of syllable category. The distribution of decoding accuracy of each cue condition is shown next to the corresponding axis of the scatter plot. The histogram represents the distributions of the cue-related decoding accuracy modulations (Valid – Neural) of all participants. The 45-degree line would denote identical decoding accuracies in the two cue conditions. Dashed lines indicate chance level (50%) neural decoding performances.

Table 2. MNI coordinates of significant brain clusters exhibiting neural decoding accuracy for syllable categories

Brain region	MNI peak coordinate			*Z*-score	Voxels
	x	*y*	*z*		
L Inferior frontal gyrus (BA 9) [a]	–54	18	26	4.60	26
R Precental gyrus / Heschl's gyrus [a]	62	–4	12	4.11	16
R Postcentral gyrus [a]	66	–24	48	4.31	15
R Middle frontal gyrus [a]	32	–10	60	5.80	34
L Superior temporal sulcus (BA 21) [a]	–52	–4	–18	4.25	13
L Inferior parietal lobule [a]	–46	–42	62	3.70	13
L Inferior temporal gyrus (BA 20) [a]	–52	–42	–18	4.34	25
L Middle occipital gyrus (BA 18)	–18	–88	12	3.94	25
L Medial frontal gyrus (BA 25)	–4	26	–18	4.48	20
R Medial frontal gyrus	14	66	–16	3.57	20
L Middle/Superior frontal gyrus (BA 11)	–24	42	–12	4.14	30
L Middle frontal gyrus (BA 11)	–39	57	–9	4.01	16
R Cingulate gyrus	18	21	36	5.59	20
R Cerebellum	6	–78	–30	3.97	23

Note: L, left; R, right; BA, Brodmann area; [a]indicates the brain regions selected to examine brain–behavior relationships (see Materials and Methods).

Based on these set of task-relevant neural regions that exhibited significantly accurate decoding of syllable categories during the memory retention phase (Figure 3; Table 2), we further examined the relationship to the behavior. Specifically, we examined whether cue-related modulation of decoding accuracy in these brain regions was related to the representational precision benefit from valid retro-cues (i.e., ln slope *k* Valid vs. ln slope *k* Neutral). As done for the ROI-based univariate analysis (BOLD), we used a stepwise multiple linear regression analysis approach to start and deduce from the seven task-relevant brain regions (Table 2; see Methods for details). Among the seven brain regions, this analysis revealed significant contributions of four neural regions in predicting the degree of perceptual precision benefit from valid retro-cues (adjusted $R^2 = 0.60$; $F(4,17) = 8.82$; $p = 0.00048$; Figure 4, left).

Neural decoding accuracy modulations in the left STS exhibited a significant relationship to the precision gain from valid retro-cues ($\beta = 0.72$, $p = 0.00047$; Table 3). We also found significant relationships to the precision benefit with the neural decoding modulations in the left inferior frontal gyrus/dorsolateral prefrontal cortex, left inferior temporal gyrus, and the right middle frontal gyrus (Table 3). However, we did not find any significant

contributions of the right motor regions in the precentral and postcentral cortex, nor the left parietal lobule.

Confirming the robustness of the model, we performed 22 iterations of a leave-one-participant-out cross-validation of the final model to predict each subject's precision modulation (ln *k* Valid vs. ln *k* Neutral) based on the remaining 21 participants' neural modulations. We found that this model-predicted precision modulation had a significant relationship to the observed outcome of the cue-related precision modulation (Figure 4, right; Pearson's correlation $r = 0.63$, $p = .0018$; average adjusted R^2 (M ± SD) = 0.60 ± 0.036).

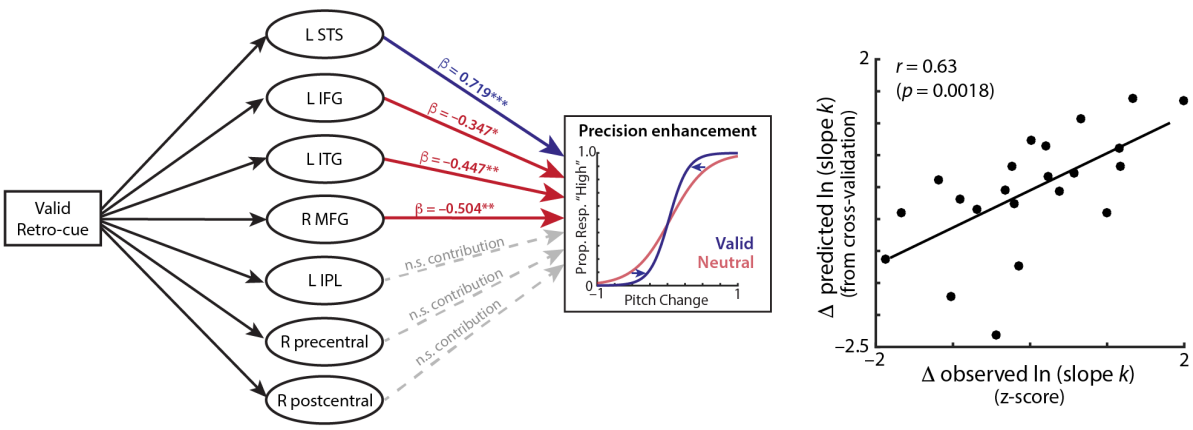

Figure 4. The relationship between the neural decoding for syllable category representations in memory and behavioral modulations. Left, estimated contributions of the brain regions' neural decoding modulations are resulted from the step-wise multiple linear regression model, which significantly predicted the extent of perceptual precision modulations (quantified as log-transformed slope *k* from psychophysical modeling) from Valid vs. Neutral cue conditions; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Right, correlation between observed and the cross-validated model-predicted cue-related modulations in perceptual precision (ln slope *k*) of the two cue conditions (Δ: Valid – Neutral). Individual's model-predicted perceptual precision modulation obtained from model fitted with data from remaining *n*=21 participants. All data points are z-scored prior to the model fitting.

Table 3. Multiple linear regression modeling result on predicting cue-related modulation of perceptual precision from the neural decoding accuracy modulations from retro-cues

Predictors	β	t	p
L Superior temporal sulcus (BA 21)	0.72	4.46	0.00035
R Middle frontal gyrus	–0.50	–3.10	0.0065
L Inferior temporal gyrus (BA 20)	–0.45	–3.23	0.0049
L Inferior frontal gyrus (BA 9)	–0.35	–2.22	0.040

Note: L, left; R, right; BA, Brodmann area

Discussion

How are multiple networks engaged in directing attention to auditory memory, and how do the memory representations held in these networks contribute to enhancing representational precision of the attended objects? The present study investigated the role of various distinct, sensory-specific as well as modality-general neocortical brain regions, in supporting the recall benefit from attention directed to working memory objects. We here have utilized psychophysical modeling and a combination of univariate and multivariate analyses on the fMRI data collected during an auditory retro-cueing working memory task with speech syllables.

Attention directed to auditory memory object enhances its representational precision

We found that retrospective attention facilitated auditory working memory performance. Consistent with previous work on how retrospective attention benefits memory recall (e.g., Griffin and Nobre, 2003; Lepsien et al., 2005; Lepsien and Nobre, 2007; Kumar et al., 2013; Pertzov et al., 2013; Backer et al., 2015), valid retro-cues led to faster responses than uninformative neutral cues. Furthermore, replicating the prior findings (Lim et al., 2015; 2018), our psychophysical modeling results demonstrate that directing attention to a specific syllable in memory enhances the representational precision of the attended object even under adverse listening conditions (i.e., auditory syllables presented under background noise). Overall, our findings are in line with the notion that retrospective attention enhances memory representations (Lepsien et al., 2011; Rerko and Oberauer, 2013; Peters et al., 2015).

Notably, we observed the facilitatory effects of retrospective attention even though valid retro-cues provided information orthogonal to behavioral response planning. That is, valid cues provided information only about the upcoming probe syllable category, whereas the task required decisions on the acoustic feature (i.e., pitch) of the syllable. Consistent with the object-based account of attention implicated in the auditory and visual domains (Duncan, 1984; Desimone and Duncan, 1995; Alain and Arnott, 2000; Shinn-Cunningham, 2008), our results suggest that selective attention directed to a specific memory object enhances the representations across varying levels of information, including lower-level features associated with the object.

Retrospective attention engages top-down cognitive control networks

Our univariate results revealed that retrospective attention to auditory objects held in memory recruited domain-general task control brain regions. Compared to neutral cue trials, valid retro-cues led to increased activations in broad cortical areas, including bilateral frontal and parietal

cortices, as well as SMA/ACC and bilateral insula. These brain regions largely overlap with two top-down networks: (1) the frontoparietal network—comprised of prefrontal and parietal cortices—has been implicated in executive control, including selective attention and working memory maintenance across various modalities (Rowe et al., 2000; Wager and Smith, 2003; Owen et al., 2005; Bledowski et al., 2009; Rottschy et al., 2012); (2) the cingulo-opercular network, consisted of SMA/ACC and bilateral anterior insula, is known to play a role in sustained maintenance of task control and goals (Dosenbach et al., 2007; Seeley et al., 2007).

Prior neuroimaging work has observed increased activations in these top-down networks with greater task demands. For instance, increased activation in the cingulo-opercular network is observed when detecting unpredictable vs. predictable events (Sadaghiani and D'Esposito, 2014), or during effortful comprehension of highly degraded speech (Wild et al., 2012; Erb et al., 2013; Vaden et al., 2013). Similarly, the fronto-parietal network activity also scales with task difficulty, such as with increasing working memory load and attentional demands (Todd and Marios, 2004; Xu and Chun, 2006; Emrich et al., 2013; Kiyonaga et al., 2014; 2017). However, based on these evidences, our results seem counterintuitive—that is, while the valid retro-cues consistently facilitated task performance, we observed increased activations in these networks in the valid vs. neutral cue conditions. Our results may suggest a potential neural mechanism by which retrospective attention facilitates memory recall: increased activations in the top-down networks might indicate that re-directing attention to a specific memory object does not reduce memory load by removing unattended objects; rather, maintaining the attended object in memory actively engages cognitive resources.

Our results concur with prior EEG findings that demonstrated that valid retro-cues engaged neural resources to actively maintain the cued item, and the greater resources allocated to the cued item enhances its representational precision (Lim et al., 2015). This interpretation is consistent with the view that retrospective attention engages additional neural processes (Griffin and Nobre, 2003; Lepsien et al., 2005; Lepsien and Nobre, 2007), manifested in the involvement of fronto-parietal and cingulo-opercular networks for selecting and prioritizing the cued item in memory to be readily available for upcoming tasks (Wallis et al., 2015; Myers et al., 2017). Although we did not find a direct relationship between the brain activations and the enhancement of the representational precision of the cued syllable, the active engagement of the neural resources allocated to the cued item in memory might be indirectly related to enhancing its representational precision.

Distributed neural representations of auditory memory object differentially contribute to representational precision benefit from attention

The multivariate analysis revealed that attended syllable held in memory was represented in distributed areas of the brain. When listeners directed their attention to a particular speech syllable guided by a valid retro-cue, we found that auditory representations of the cued syllable category could be decoded from the activity patterns in modality-general and sensory-specific brain areas. These areas include lateral prefrontal and parietal areas, part of the domain-general cognitive control regions known to support both visual and auditory working memory (Duncan and Owen, 2000; Postle et al., 2000; Todd and Marios, 2004; Lewis-Peacock et al., 2012; Noyce et al., 2017; Uluç et al., 2018) and verbal working memory (Awh et al., 1996; Smith et al., 1996; 1998; Baddeley, 2003). We also found that speech- and language-sensitive cortical areas along the left STS and sensorimotor cortices maintained the cued syllable information during memory retention. These results are consistent with recent neural models of working memory: multiple brain regions, in the prefrontal, parietal, and sensory-specific areas represent information maintained in memory (Ester et al., 2015; Christophel et al., 2017; Long and Kuhl, 2018).

Interestingly, our results further suggest that widely distributed memory representations do not uniformly contribute to retrospective attentional benefit in enhancing representational precision of the cued item. The regression analysis in predicting individual differences in the memory precision gain from valid retro-cues (Δ slope k) revealed that the extent of individuals' neural fidelity enhancements in the left STS, the prefrontal cortex, and right MFG exhibited significant relationships to enhancing representational precision of the cued syllable in memory, and the left STS exhibited the best predictive relationship. However, no such relationships were found with the parietal and sensorimotor regions (Figure 4). This pattern indicates that although the distributed memory representations across broad cortical regions work together in enhancing representational precision of the attended object held in working memory, they do so by exerting differential contributions.

One potential explanation of our findings is that while the multiple brain regions redundantly represent the objects held in working memory, these regions may not necessarily maintain the same level of representations of the objects. Consistent with a view that there is a division of labor between sensory-specific and domain-general cortices in efficient maintenance of working memory (Spitzer et al., 2014; Sreenivasan et al., 2014; Christophel et al., 2017), it is possible that prefrontal cortex holds object-level representations of speech sounds (i.e., syllable category), whereas the speech-processing area in the left STS retains

both the object-level syllable category and stimulus-specific featural information bound to the object.

This interpretation is in line with functional specializations of these cortical regions. Many of prior work demonstrate that frontal areas are sensitive to between-category, but not within-category differences (e.g., Freedman et al., 2001; 2003; Myers et al., 2009). In addition, the frontal cortex might maintain memory representations in an abstract format that can be generalized across sensory modalities (Spitzer et al., 2014; Uluç et al., 2018) and be flexibly transformed based on the task-relevant goals (Lee et al., 2013; Stokes et al., 2013; Long and Kuhl, 2018).

On the contrary, as we will discuss in more detail below, the left STS has been prominently implicated in speech sound processing and categorization (Formisano et al., 2008; Leech et al., 2009; Obleser and Eisner, 2009; Mesgarani et al., 2014; Lim et al., 2019). The left STS region may thus be suitable for retaining acoustic details of speech sounds held in memory. Overall, our results may suggest that while both domain-general fronto-parietal networks and domain-specific regions maintain the working memory contents (Ester et al., 2015; Bettencourt and Xu, 2016; Christophel et al., 2017; Bracci et al., 2017), the representations held in the distinct regions might be differentially sensitive to top-down task-related goals vs. bottom-up stimulus-specific features (Long and Kuhl, 2018), thereby exerting differential influences in attentional enhancement of representational precision.

The role of left STS in maintaining auditory memory representations

During memory maintenance of auditory syllables, we neither observed significant engagement of the auditory/speech-sensitive brain regions along the superior temporal cortex, nor found differential recruitment of this region during memory maintenance in the valid vs. neutral cue condition. However, MVPA revealed that the left STS retained the cued syllable item held in memory, which seems to be directly related to the extent of precision enhancement gained from retrospective attention. Although there is conflicting evidence about the engagement of the auditory cortex during auditory working memory (e.g., Gaab et al., 2003; Linke et al., 2011; Linke and Cusack, 2015; Kumar et al., 2016), our results are broadly in line with the view that sensory-specific brain regions retain working memory contents, irrespective of its activations (Harrison and Tong, 2009; Serences et al., 2009; Riggall and Postle, 2012; Emrich et al., 2013).

One difference to note is that while prior work found the primary auditory cortex to maintain pure tones in memory (Linke et al., 2011; Kumar et al., 2016), we found the left STS maintaining auditory speech syllables. This discrepancy might lie in the nature of the stimulus

objects held in memory. Although listeners had to judge pitch of speech syllables, the current task cannot be performed simply by extracting pitch information from syllables without maintaining syllable category representation in memory. And this task demand might have induced the auditory memory representations to reside in the left STS.

Prior neuroimaging studies in speech perception demonstrate that the left superior temporal regions functionally subserve speech category representation and processing (Formisano et al., 2008; Obleser and Eisner, 2009; Lim et al., 2019). Especially, the anterior STS/STG is shown to be sensitive to tonal speech categories (Feng et al., 2018) and speech vowels (Obleser et al., 2006) that contain talkers' voice/pitch (Bachorowski and Owren, 1999). Thus, accessing the auditory syllable mnemonic representations in the left STS/STG might allow readout of the featural-level information (e.g., pitch) of the object, which can be enhanced by attention directed to the syllable object in memory.

Furthermore, the contribution of speech-sensitive brain areas seems consistent with an emerging evidence that speech working memory is not independent from sensory/auditory processing (cf. Paulesu et al., 1993; Smith et al., 1998; Baddeley, 2003); rather, both auditory cortical regions, responsible for speech perception, and sensorimotor cortical areas support speech working memory (Jacquemot and Scott, 2006; Perrachione et al., 2017; Scott and Perrachione, 2019).

Potential limitations

At first glance, one might be concerned with a potential circularity in finding brain regions that retained auditory syllable memory representations. We defined these regions based on the neural activity patterns in brain areas exhibiting above-chance decoding accuracy of the syllable category information provided by the visual valid cues. These regions might be biased to exhibit higher decoding accuracies for memory content in the valid than the neutral cue conditions. Nevertheless, the extent of individuals' neural decoding accuracy modulations (i.e., valid vs. neutral) in predicting the extent of behavioral (precision slope *k*) modulations cannot be solely driven by the decoding accuracy in the valid cue condition. Furthermore, our main focus was on examining the relative contributions of distributed brain regions' enhanced representational precision, which should be relatively unaffected by the overall bias that favored the valid condition.

The second potential issue concerns the chosen retro-cues themselves. Because valid retro-cues were written syllables, all activation and representation differences observed between valid and neutral cues might at least in parts be by-products of reading written text. To minimize this concern, our classification approach specifically aimed to find the brain

regions maintaining auditory traces of speech syllables held in memory. The classifier was trained to learn syllable sounds when they were presented as auditory probes, and tested to decode the neural activity patterns during memory maintenance (i.e., no auditory input). Furthermore, the classifier was trained only on the auditory probes of the neutral cue trials to further avoid potential bias or expectations of a particular syllable informed by valid retro-cues. Of course, we cannot completely rule out the possibility that reading written syllables on valid cues itself yields reliable neural decoding. For instance, we found that the left inferior temporal gyrus, known to be sensitive to written texts (e.g., Cohen et al., 2002; Cohen and Dehaene, 2004), exhibited above-chance accuracy in decoding syllable category maintained in memory. Nevertheless, the current task required listeners to maintain the pitch information of syllables held in memory; thus, reliance on the visually presented syllable category alone neither can facilitate the task performance, nor can explain the enhancement of representational precision from valid retro-cues.

Conclusions

As demonstrated here, working memory representations across the brain, ranging from sensory and higher-order cognitive-control regions differentially contribute to the retrospective attentional gain in the fidelity of memory representations. Here, we show that while re-directing attention to the relevant auditory memory objects mainly recruits higher-order cognitive-control networks, the attentional enhancement in the neural fidelity in superior temporal sulcus and frontal cortical regions was related to the individual gain in recall precision of auditory objects from memory, with the superior temporal sulcus best predicting the attentional gain. Together with previous work on working memory, our findings provide evidence of the interplay among the discrete, functionally specialized brain regions in maintaining and attentionally enhancing working memory representations.

References

Alain C, Arnott SR (2000) Selectively attending to auditory objects. Front Mol Biosci 5:d202–d212.

Alavash M, Lim S-J, Thiel C, Sehm B, Deserno L, Obleser J (2018) Dopaminergic modulation of hemodynamic signal variability and the functional connectome during cognitive performance. Neuroimage 172:341–356.

Awh E, Jonides J, Smith EE, Schumacher EH, Koeppe RA, Katz S (1996) Dissociation of Storage and Rehearsal in Verbal Working Memory: Evidence From Positron Emission Tomography. Psychol Sci 7:25–31.

Bachorowski J-A, Owren MJ (1999) Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. J Acoust Soc Am 106:1054–1063.

Backer KC, Binns MA, Alain C (2015) Neural dynamics underlying attentional orienting to auditory representations in short-term memory. J Neurosci 35:1307–1318.

Baddeley A (2003) Working memory: looking back and looking forward. Nat Rev Neurosci 4:829–839.

Bays PM, Husain M (2008) Dynamic shifts of limited working memory resources in human vision. Science 321:851–854.

Bettencourt KC, Xu Y (2016) Decoding the content of visual short-term memory under distraction in occipital and parietal areas. Nat Rev Neurosci 19:150–157.

Bledowski C, Rahm B, Rowe JB (2009) What “Works” in Working Memory? Separate Systems for Selection and Updating of Critical Information. J Neurosci 29:13735–13741.

Bracci S, Daniels N, Op de Beeck H (2017) Task Context Overrules Object- and Category-Related Representational Content in the Human Parietal Cortex. Cereb Cortex 84:81.

Chang C-C, Lin C-J (2011) LIBSVM: A Library for Support Vector Machines. ACM Transactions on Intelligent Systems and Technology 2:27:1–27:27.

Christophel TB, Hebart MN, Haynes J-D (2012) Decoding the Contents of Visual Short-Term Memory from Human Visual and Parietal Cortex. J Neurosci 32:12983–12989.

Christophel TB, Klink PC, Spitzer B, Roelfsema PR, Haynes J-D (2017) The Distributed Nature of Working Memory. Trends Cogn Sci 21:111–124.

Cohen L, Dehaene S (2004) Specialization within the ventral stream: the case for the visual word form area. Neuroimage 22:466–476.

Cohen L, Lehericy S, Chochon F, Lemer C, Rivaud S, Dehaene S (2002) Language-Specific Tuning of Visual Cortex? Functional Properties of the Visual Word Form Area. Brain 125:1054–1069.

Cowan N (2001) Metatheory of storage capacity limits. Behav Brain Sci 24:154–176.

Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 29:162–173.

D'Esposito M, Postle BR (2015) The cognitive neuroscience of working memory. Annual review of psychology 66:115–142.

Desimone R, Duncan J (1995) Neural Mechanisms of Selective Visual Attention. Annu Rev Neurosci 18:193–222.

Dosenbach NUF, Fair DA, Cohen AL, Schlaggar BL, Petersen SE (2008) A dual-networks architecture of top-down control. Trends Cogn Sci 12:99–105 Available at: http://linkinghub.elsevier.com/retrieve/pii/S1364661308000272.

Dosenbach NUF, Fair DA, Miezin FM, Cohen AL, Wenger KK, Dosenbach RAT, Fox MD, Snyder AZ, Vincent JL, Raichle ME, Schlaggar BL, Petersen SE (2007) Distinct brain networks for adaptive and stable task control in humans. Proc Natl Acad Sci U S A 104:11073–11078.

Duncan J (1984) Selective Attention and the Organization of Visual Information. J Exp Psychol Gen 113:51–517.

Duncan J, Owen AM (2000) Common regions of the human frontal lobe recruited by diverse cognitive demands. Trends in neurosciences 23:475–483.

Eklund A, Nichols TE, Knutsson H (2016) Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. Proc Natl Acad Sci U S A 113:7900–7905.

Emrich SM, Riggall AC, LaRocque JJ, Postle BR (2013) Distributed Patterns of Activity in Sensory Cortex Reflect the Precision of Multiple Items Maintained in Visual Short-Term Memory. J Neurosci 33:6516–6523.

Erb J, Henry MJ, Eisner F, Obleser J (2013) The brain dynamics of rapid perceptual adaptation to adverse listening conditions. J Neurosci 33:10688–10697.

Ester EF, Sprague TC, Serences JT (2015) Parietal and Frontal Cortex Encode Stimulus-Specific Mnemonic Representations during Visual Working Memory. Neuron 87:893–905.

Fedorenko E, Behr MK, Kanwisher N (2011) Functional specificity for high-level linguistic processing in the human brain. Proc Natl Acad Sci U S A 108:16428–16433.

Feng G, Yi H-G, Chandrasekaran B (2018) The Role of the Human Auditory Corticostriatal Network in Speech Learning. Cereb Cortex 29:4077–4089.

Formisano E, De Martino F, Bonte M, Goebel R (2008) "Who" is saying “what?” Brain-based decoding of human voice and speech. Science 322:970–973 Available at: http://www.ncbi.nlm.nih.gov/pubmed/18988858.

Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2001) Categorical representation of visual stimuli in the primate prefrontal cortex. Science 291:312–316.

Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2003) A Comparison of Primate Prefrontal and Inferior Temporal Cortices during Visual Categorization. J Neurosci 23:5235–5246 Available at: http://www.jneurosci.org/content/23/12/5235.full.

Fritz JB, Elhilali M, David SV, Shamma SA (2007) Auditory attention — focusing the searchlight on sound. Curr Opin Neurobiol 17:437–455.

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

Gazzaley A, Nobre AC (2012) Top-down modulation: bridging selective attention and working memory. Trends Cogn Sci 16:129–135.

Goldman-Rakic PS (1995) Cellular Basis of Working Memory Review. Neuron 14:477–485.

Griffin IC, Nobre AC (2003) Orienting Attention to Locations in Internal Representations. J Cogn Neurosci 15:1176–1194.

Harrison SA, Tong F (2009) Decoding reveals the contents of visual working memory in early visual areas. Nature 458:632–635.

Haynes J-D, Sakai K, Rees G, Gilbert S, Frith C, Passingham RE (2007) Reading hidden intentions in the human brain. Current Biology 17:323–328.

Hickok G (2009) The functional neuroanatomy of language. Physics of Life Reviews 6:121–143.

Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev Neurosci 8:393–402.

Higo T, Mars RB, Boorman ED, Buch ER, Rushworth MFS (2011) Distributed and causal influence of frontal operculum in task control. Proc Natl Acad Sci U S A 108:4230–4235.

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

Kiyonaga A, Dowd EW, Egner T (2017) Neural Representation of Working Memory Content Is Modulated by Visual Attentional Demand. J Cogn Neurosci 29:2011–2024.

Kiyonaga A, Egner T (2012) Working memory as internal attention: Toward an integrative account of internal and external selection processes. Psychon Bull Rev 20:228–242.

Kiyonaga A, Korb FM, Lucas J, Soto D, Egner T (2014) Dissociable causal roles for left and right parietal cortex in controlling attentional biases from the contents of working memory. Neuroimage 100:200–205.

Koelsch S, Schulze K, Sammler D, brain TFH, 2009 (2009) Functional architecture of verbal and tonal working memory: an FMRI study. Hum Brain Mapp 30: 859-873.

Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. Proc Natl Acad Sci U S A 103:3863–3868.

Kumar S, Joseph S, Gander PE, Barascud N, Halpern AR, Griffiths TD (2016) A Brain System for Auditory Working Memory. J Neurosci 36:4492–4505.

Kumar S, Joseph S, Pearson B, Teki S, Fox ZV, Griffiths TD, Husain M (2013) Resource allocation and prioritization in auditory working memory. Cogn Neurosci 4:12–20.

Lee S-H, Kravitz DJ, Baker CI (2013) Goal-dependent dissociation of visual and prefrontal cortices during working memory. Nat Neurosci 16:997–999.

Leech R, Holt LL, Devlin JT, Dick F (2009) Expertise with artificial nonspeech sounds recruits speech-sensitive cortical regions. J Neurosci 29:5234–5239.

Lepsien J, Griffin IC, Devlin JT, Nobre AC (2005) Directing spatial attention in mental representations: Interactions between attentional orienting and working-memory load. Neuroimage 26:733–743.

Lepsien J, Nobre AC (2007) Attentional Modulation of Object Representations in Working Memory. Cereb Cortex 17:2072–2083.

Lepsien J, Thornton I, Nobre AC (2011) Modulation of working-memory maintenance by directed attention. Neuropsychologia 49:1569–1577.

Lewis-Peacock JA, Drysdale AT, Oberauer K, Postle BR (2012) Neural Evidence for a Distinction between Short-term Memory and the Focus of Attention. J Cogn Neurosci 24:61–79.

Lim S-J, Fiez JA, Holt LL (2019) Role of the striatum in incidental learning of sound categories. Proc Natl Acad Sci U S A:1–10.

Lim S-J, Wöstmann M, Geweke F, Obleser J (2018) The Benefit of Attention-to-Memory Depends on the Interplay of Memory Capacity and Memory Load. Front Psychol 9:146.

Lim S-J, Wöstmann M, Obleser J (2015) Selective Attention to Auditory Memory Neurally Enhances Perceptual Precision. J Neurosci 35:16094–16104.

Linke AC, Cusack R (2015) Flexible Information Coding in Human Auditory Cortex during Perception, Imagery, and STM of Complex Sounds. J Cogn Neurosci 27:1322–1333.

Linke AC, Vicente-Grabovetsky A, Cusack R (2011) Stimulus-specific suppression preserves information in auditory short-term memory. Proc Natl Acad Sci U S A 108:12961–12966.

Long NM, Kuhl BA (2018) Bottom-up and top-down factors differentially influence stimulus representations across large-scale attentional networks. J Neurosci 38:2724–17–2504.

Luck SJ, Vogel EK (2013) Visual working memory capacity: from psychophysics and neurobiology to individual differences. Trends Cogn Sci 17:391–400.

Ma WJ, Husain M, Bays PM (2014) Changing concepts of working memory. Nat Rev Neurosci 17:347–356.

Mesgarani N, Cheung C, Johnson K, Chang EF (2014) Phonetic feature encoding in human superior temporal gyrus. Science 343:1006–1010.

Misaki M, Kim Y, Bandettini PA, Kriegeskorte N (2010) Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. Neuroimage 53:103–118.

Murray AM, Nobre AC, Clark IA, Cravo AM, Stokes MG (2013) Attention Restores Discrete Items to Visual Short-Term Memory. Psychol Sci 24:550–556.

Myers EB, Blumstein SE, Walsh E, Eliassen J (2009) Inferior frontal regions underlie the perception of phonetic category invariance. Psychol Sci 20:895–903.

Myers NE, Stokes MG, Nobre AC (2017) Prioritizing Information during Working Memory: Beyond Sustained Internal Attention. Trends Cogn Sci 21:449–461.

Noyce AL, Cestero N, Michalka SW, Shinn-Cunningham BG, Somers DC (2017) Sensory-Biased and Multiple-Demand Processing in Human Lateral Frontal Cortex. J Neurosci 37:8755–8766.

Obleser J, Boecker H, Drzezga A, Haslinger B, Hennenlotter A, Roettinger M, Eulitz C, Rauschecker JP (2006) Vowel sound extraction in anterior superior temporal cortex. Hum Brain Mapp 27:562–571.

Obleser J, Eisner F (2009) Pre-lexical abstraction of speech in the auditory cortex. Trends Cogn Sci 13:14–19.

Owen AM, McMillan KM, Laird AR, Bullmore E (2005) N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. Hum Brain Mapp 25:46–59.

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

Perrachione TK, Ghosh SS, Ostrovskaya I, Gabrieli JDE, Kovelman I (2017) Phonological Working Memory for Words and Nonwords in Cerebral Cortex. Journal of Speech, Language, and Hearing Research 60:1959–1979.

Pertzov Y, Bays PM, Joseph S, Husain M (2013) Rapid forgetting prevented by retrospective attention cues. J Exp Psychol Hum Percept Perform 39:1224–1231.

Peters B, Kaiser J, Rahm B, Bledowski C (2015) Activity in Human Visual and Parietal Cortex Reveals Object-Based Attention in Working Memory. J Neurosci 35:3360–3369.

Postle BR (2006) Working memory as an emergent property of the mind and brain. Neuroscience 139:23–38.

Postle BR, Stern CE, Rosen BR, Corkin S (2000) An fMRI Investigation of Cortical Contributions to Spatial and Nonspatial Visual Working Memory. Neuroimage 11:409–423.

Power JD, Cohen AL, Nelson SM, Wig GS, Barnes KA, Church JA, Vogel AC, Laumann TO, Miezin FM, Schlaggar BL, Petersen SE (2011) Functional Network Organization of the Human Brain. Neuron 72:665–678.

Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL (2001) A default mode of brain function. Proc Natl Acad Sci U S A 98:676–682.

Rerko L, Oberauer K (2013) Focused, unfocused, and defocused information in working memory. J Exp Psychol Learn Mem Cogn 39:1075–1096.

Riggall AC, Postle BR (2012) The Relationship between Working Memory Storage and Elevated Activity as Measured with Functional Magnetic Resonance Imaging. J Neurosci 32:12990–12998.

Rottschy C, Langner R, Dogan I, Reetz K, Laird AR, Schulz JB, Fox PT, Eickhoff SB (2012) Modelling neural correlates of working memory: A coordinate-based meta-analysis. Neuroimage 60:830–846.

Rowe JB, Toni I, Josephs O, Frackowiak RS, Passingham RE (2000) The prefrontal cortex: response selection or maintenance within working memory? Science 288:1656–1660.

Sadaghiani S, D'Esposito M (2014) Functional Characterization of the Cingulo-Opercular Network in the Maintenance of Tonic Alertness. Cereb Cortex:1–11.

Scott TL, Perrachione TK (2019) Common cortical architectures for phonological working memory identified in individual brains. Neuroimage 202:116096.

Seeley WW, Menon V, Schatzberg AF, Keller J, Glover GH, Kenna H, Reiss AL, Greicius MD (2007) Dissociable Intrinsic Connectivity Networks for Salience Processing and Executive Control. J Neurosci 27:2349–2356.

Serences JT, Ester EF, Vogel EK, Awh E (2009) Stimulus-specific delay activity in human primary visual cortex. Psychol Sci 20:207–214.

Serences JT, Kastner S (2014) A multi-level account of selective attention. In: The Oxford Handbook of Attention (Nobre AC, Kastner S, eds), pp 76–104. Oxford University Press.

Shinn-Cunningham BG (2008) Object-based auditory and visual attention. Trends Cogn Sci 12:182–186.

Smith EE, Jonides J, Koeppe RA (1996) Dissociating Verbal and Spatial Working Memory Using PET. Cereb Cortex 6:11–20.

Smith EE, Jonides J, Marshuetz C, Koeppe RA (1998) Components of verbal working memory: Evidence from neuroimaging. Proc Natl Acad Sci U S A 95:876–882.

Spitzer B, Fleck S, Blankenburg F (2014) Parametric Alpha- and Beta-Band Signatures of Supramodal Numerosity Information in Human Working Memory. J Neurosci 34:4293–4302.

Sreenivasan KK, Gratton C, Vytlacil J, D'Esposito M (2014) Evidence for working memory storage operations in perceptual cortex. Cogn Affect Behav Neurosci 14:117–128.

Stokes MG, Kusunoki M, Sigala N, Nili H, Gaffan D, Duncan J (2013) Dynamic Coding for Cognitive Control in Prefrontal Cortex. Neuron 78:364–375.

Todd JJ, Marios R (2004) Capacity limit of visual short-term memory in human posterior parietal cortex. Nat Neurosci 428:751–754.

Uluç I, Schmidt TT, Wu Y-H, Blankenburg F (2018) Content-specific codes of parametric auditory working memory in humans. Neuroimage 183:254–262.

Vaden KI, Kuchinsky SE, Cute SL, Ahlstrom JB, Dubno JR, Eckert MA (2013) The Cingulo-Opercular Network Provides Word-Recognition Benefit. J Neurosci 33:18979–18986.

Wager TD, Smith EE (2003) Neuroimaging studies of working memory: A meta-analysis. Cogn Affect Behav Neurosci 3:255–274.

Wallis G, Stokes M, Cousijn H, Woolrich M, Nobre AC (2015) Frontoparietal and Cingulo-opercular Networks Play Dissociable Roles in Control of Working Memory. J Cogn Neurosci 15:1–16.

Wild CJ, Yusuf A, Wilson DE, Peelle JE, Davis MH, Johnsrude IS (2012) Effortful listening: the processing of degraded speech depends critically on attention. J Neurosci 32:14010–14021.

Wilsch A, Obleser J (2016) What works in auditory working memory? A neural oscillations perspective. Brain Res 1640:193–207.

Xia M, Wang J, He Y (2013) BrainNet Viewer: A Network Visualization Tool for Human Brain Connectomics. PLoS One 8:e68910.

Xu Y, Chun MM (2006) Dissociable neural mechanisms supporting visual short-term memory for objects. Nature 440:91–95.

Zhang W, Luck SJ (2008) Discrete fixed-resolution representations in visual working memory. Nature 453:233–235.