HHS Public Access

Cingulo-opercular activity affects incidental memory encoding for speech in noise

Kenneth I. Vaden Jr, Ph.D.[1], **Susan Teubner-Rhodes, Ph.D.**[1], **Jayne B. Ahlstrom, M.S.**[1], **Judy R. Dubno, Ph.D.**[1], and **Mark A. Eckert, Ph.D.**[1]

[1]Hearing Research Program, Department of Otolaryngology-Head and Neck Surgery, Medical University of South Carolina

Abstract

Correctly understood speech in difficult listening conditions is often difficult to remember. A long-standing hypothesis for this observation is that the engagement of cognitive resources to aid speech understanding can limit resources available for memory encoding. This hypothesis is consistent with evidence that speech presented in difficult conditions typically elicits greater activity throughout cingulo-opercular regions of frontal cortex that are proposed to optimize task performance through adaptive control of behavior and tonic attention. However, successful memory encoding of items for delayed recognition memory tasks is consistently associated with increased cingulo-opercular activity when perceptual difficulty is minimized. The current study used a delayed recognition memory task to test competing predictions that memory encoding for words is enhanced or limited by the engagement of cingulo-opercular activity during challenging listening conditions. An fMRI experiment was conducted with twenty healthy adult participants who performed a word identification in noise task that was immediately followed by a delayed recognition memory task. Consistent with previous findings, word identification trials in the poorer signal-to-noise ratio condition were associated with increased cingulo-opercular activity and poorer recognition memory scores on average. However, cingulo-opercular activity decreased for correctly identified words in noise that were not recognized in the delayed memory test. These results suggest that memory encoding in difficult listening conditions is poorer when elevated cingulo-opercular activity is not sustained. Although increased attention to speech when presented in difficult conditions may detract from more active forms of memory maintenance (e.g., sub-vocal rehearsal), we conclude that task performance monitoring and/or elevated tonic attention supports incidental memory encoding in challenging listening conditions.

Keywords

Address Correspondence To: Mark A. Eckert (Eckert@musc.edu) or, Kenneth Vaden (Vaden@musc.edu), Department of Otolaryngology – Head and Neck Surgery, Medical University of South Carolina, 135 Rutledge Avenue, MSC 550, Charleston, S.C. 29425-5500, Telephone: 00-1-843-792-2774.

Introduction

Poor recall of words occurs in difficult listening conditions, such as in background noise, even when those words were understood (Murphy et al., 2000; Pichora-Fuller et al., 1995; Rabbitt, 1968; Ward et al., 2016). One explanation for poor recall of speech in noise is that attention is diverted away from memory encoding to help extract the speech signal from noise, at least for serial recall tasks (Heinrich et al., 2008; Rabbitt, 1991, 1968; Tun et al., 2009; Wild et al., 2012). For example, increased cingulo-opercular cortex activity is frequently observed during speech identification tasks (Eckert et al., 2009; Erb and Obleser, 2013; Harris et al., 2009; Vaden et al., 2013). The limited neural resource hypothesis would predict that this attention-related activity results in poorer memory encoding in challenging listening conditions. However, elevated cingulo-opercular activity is consistently associated with successful recall when memory encoding occurs in less perceptually demanding conditions (Kim, 2011; Spaniol et al., 2009). We examined the extent to which the engagement of cingulo-opercular cortex during a word identification in noise task was associated with relatively better or worse recognition memory[1].

Cingulo-opercular activity is pronounced during word identification in noise tasks (Eckert et al., 2009; Harris et al., 2009; Erb and Obleser, 2013; Wild et al., 2012; see meta-analyses and reviews: Adank, 2012; Eckert et al., 2016). These anatomically distinct cingulo-opercular regions are differentially sensitive to errors and response uncertainty (mid-cingulate; Ullsperger and von Cramon, 2004), response selection demands (left inferior frontal gyrus; Thompson-Schill et al., 1997; Goghari and Macdonald, 2005; Moss et al., 2005), autonomic responses (insula; Cechetto, 2014), and inhibition (right inferior frontal gyrus; Aron et al., 2003; Aron, 2007; Hughes et al., 2013). Together, these diverse functions could be integrated to provide adaptive control, which consists of performance monitoring and flexible modifications of attention and behavior.

An adaptive control framework for cingulo-opercular function is supported by evidence of consistently elevated cingulo-opercular activity across varied perceptual and response demands of different tasks (Dosenbach et al., 2006). Activity increases following response errors and uncertainty, when the benefit from adaptive control (Shenhav et al., 2013) is not restricted by conditions that result in particularly poor or good performance (Eckert et al., 2016; Poldrack et al., 2001; Zekveld et al., 2006). Consistent with this framework, elevated cingulo-opercular activity has been associated with subsequent improvements in task performance (Botvinick et al., 2004; Carter et al., 2000, 1998; Eichele et al., 2008; Kerns et al., 2004; Sheth et al., 2012), including word identification in noise (Vaden et al., 2015, 2013). Complementing the premise that cingulo-opercular functions optimize task performance, attentional lapses that occur with lower cingulo-opercular activity have been associated with a subsequent increase in the likelihood of response errors (Eichele et al., 2008; Weissman et al., 2006).

---

[1]To limit terminology confusion, *word identification* refers to understanding aurally presented words and *recognition memory* refers to familiarity-memory for an item after a delay of minutes or longer between presentation and memory test.

Cingulo-opercular engagement during memory encoding has also been consistently linked to successful recognition memory in neuroimaging studies (see two large-scale meta-analyses: Kim, 2011; Spaniol et al., 2009). Specifically, cingulo-opercular activity has been shown to increase during the encoding phase for items that were correctly remembered in delayed recognition memory tests. Together with observations that lateral prefrontal cortex lesions impair associative memory and limit the use of common memory strategies, these findings support the proposal that prefrontal cortices engage attention control to enhance processing of task-relevant information and memory formation (Blumenfeld and Ranganath, 2007).

The existing evidence on memory encoding and memory for speech in noise sets up competing predictions for the role of attention, as reflected in cingulo-opercular activity. On the one hand, cingulo-opercular activity during encoding has been linked to both 1) correct word identification in noise on a trial-by-trial basis (Vaden et al., 2013) and 2) successful memory encoding of items in recognition memory tasks, albeit in the absence of perceptual difficulty manipulations (Kim, 2011; Spaniol et al., 2009). Taken together, this evidence predicts that increased cingulo-opercular activity benefits memory encoding for correctly identified words in noise through changes in tonic attention and behavior that support speech understanding. On the other hand, observations that listening to speech in noise results in both 1) elevated cingulo-opercular activity (Eckert et al., 2009; Harris et al., 2009; Erb and Obleser, 2013; Wild et al., 2012) and 2) poorer serial recall (Heinrich et al., 2008; Rabbitt, 1991, 1968; Tun et al., 2009) could also indicate that increased attention required to aid speech understanding limits the resources available for memory encoding. Under this view, increased cingulo-opercular activity for words identified in noise would be predicted to result in poorer memory encoding.

To date, there is no direct evidence that increased attention-related cingulo-opercular activity accounts for poorer memory for speech in noise. Moreover, there is extensive neuroimaging evidence from delayed recognition memory studies supporting the opposite conclusion. We predicted that elevated cingulo-opercular activity during correct word identification in noise improves memory encoding, resulting in better delayed recognition memory task performance. Because cortical attention systems are often engaged variably across trials despite consistent task demand, this prediction also means that failures to engage or maintain attention (i.e., lapses in cingulo-opercular activity) could result in poorer memory encoding.

Materials and methods

During two consecutive fMRI runs, participants performed a word identification in noise task (Task 1: 25 m 48 s) that was immediately followed by a delayed recognition memory task (Task 2: 21 m 30 s). Neuroimaging data from Task 1 were analyzed to examine changes in activity during the word identification in noise task (i.e., during memory encoding) that were associated with delayed recognition memory. Although neuroimaging data from Task 2 was not a focus of the current study, the Task 2 memory hits or misses were used in the functional imaging analyses of Task 1 words that were correctly identified.

Results from the signal-to-noise ratio (SNR) conditions during word identification (Task 1) were previously reported in Vaden et al. (2013). The neural memory encoding effects during Task 1 and delayed recognition memory results (Task 2) are the focus of the current study, and have not been reported previously. Additional details about the Task 2 method and results are presented in the Supplementary Materials.

Participants

Twenty healthy, young adults (10 females, average age = 29.8 ± 5.9 years) with normal hearing participated in the current study. They were recruited as part of a larger study on age-related changes in hearing and communication. The final sample included 20 participants, after excluding participants older than 41 years and one participant with noted movement in the scanner and related artifacts (e.g. ghosting). Pure-tone thresholds were measured with a Madsen OB922 audiometer and TDH-39 headphones (American National Standards Institute, 2004, 2010). Each participant had mean pure-tone thresholds < 12 dB HL from 500 to 2000 Hz (better ear), with less than 7 dB difference between right and left ears. All participants demonstrated normal immittance measures. The participants were all native English speakers, with an average of 16.4 ± 2.2 years of education (M ± SD). Handedness preference scores of 70.3 ± 48.0 indicated that the sample was largely right-handed (possible range = -100, strongly left-handed, to 100, strongly right-handed; Oldfield, 1971). None of the participants reported a history of neurological or psychiatric events. Informed consent was obtained in compliance with the Institutional Review Board at the Medical University of South Carolina (MUSC), and experiments were conducted in accordance with the Declaration of Helsinki.

Experimental design

Task 1: Word identification in noise—For Task 1, each participant was instructed to listen to a single monosyllabic word presented in multitalker babble and repeat the word out loud, or say "nope" if they could not recognize the word. The word recordings from a male talker were originally prepared by Dirks et al. (2001), and the multitalker babble recordings were from Kalikow et al. (1977). Each trial had an 8.6 s inter-trial interval (ITI), which was the length of time between consecutive scans in the sparse fMRI acquisition sequence. A word was presented 3.1 s into each trial during the relatively quiet period following the scanner offset. This design allowed for greater control over the SNR for the calibrated speech and babble stimuli than if the stimuli were presented in scanner noise. Words were presented through piezoelectric headphones (Sensimetrics) at 85 or 92 dB SPL with continuous babble presented at 82 dB SPL, which resulted in a +3 dB or +10 dB SNR. Words were presented in the same SNR for 4-6 consecutive trials, with a total of 60 words in each SNR across Task 1. Participant responses were recorded using an MRI-compatible microphone (Resonance Technology, Inc.), during an interval (4.1-6.1 s) cued by a crosshair that changed colors and was viewed through a headcoil-mounted periscope. The timing of the response interval was designed to reduce head motion during image acquisitions. The experimental block design included 2 epochs of word identification task trials (60 trials in each SNR), 3 rest epochs during which no sounds were presented (10 trials each), and 2 epochs of multitalker babble with no task demands (15 trials each) to allow participants to habituate to the babble (shown in Figure 1, Vaden et al., 2013).

Task 2: Delayed recognition memory—Instructions for Task 2 were given in the scanner immediately after the word identification task, which prevented rehearsal or other maintenance strategies that affect memory. Participants were instructed to listen to a word on each trial and respond with a button press to indicate whether they 1) remembered hearing that word, 2) did not remember hearing that word in the earlier task, or 3) could not understand the word, to disambiguate memory from speech intelligibility. On average, there was a 29 ± 2 min delay interval between presentations of each word for Task 1 and Task 2.

A total of 120 band-pass filtered words[2] were presented at 87 dB SPL in quiet for the recognition memory task, so that words were the salient memory-related stimuli rather than multitalker babble from Task 1. The digital band-pass filter (BPF) applied to each word recording for Task 2 had a 200-Hz lower cutoff frequency and four upper cutoff frequencies (400, 1000, 1600, 3150 Hz). Thirty words were presented in each of the BPF conditions, with 28 words that were presented in Task 1 and two foil words that were not presented in Task 1. Words from each SNR condition were selected for the BPF conditions to maximize the number of correctly understood words across both tasks. From the 60 words previously presented in the more advantageous +10 dB SNR, Task 2 presented 28 words low-pass filtered at 1600 Hz and 28 words at 3150 Hz. From the 60 words presented in the less advantageous +3 dB SNR, Task 2 presented 28 low-pass filtered words at 400 Hz and 28 words at 1000 Hz. Although the words were presented in conditions that made identification easier or more difficult in both Tasks 1 and 2, restricting the fMRI memory analyses to correctly understood words provided a means to test predictions about cingulo-opercular activity and delayed recognition memory, as described later. Words that were not correctly identified in Task 1 were treated as additional foils in our analyses because misidentification during Task 1 meant that participants had, in essence, never been exposed to these items. The addition of these foils (proportion of foils = $21 \pm 6\%$) also more closely approximated a 50-50 split of targets versus foils that is typically used for recognition memory tasks.

Recognition memory analysis—Signal detection theory measures were calculated to characterize the effects of the speech intelligibility manipulations on memory sensitivity and response bias. Memory responses from Task 2 (button presses to indicate: "remember" or "don't remember") were analyzed to characterize the extent to which word identification during Task 1 affected the likelihood of delayed recognition memory for each word. For each participant, any word that was not responded to during either Task 1 or Task 2 was excluded from memory analyses. Word identification scores[3] were used in conjunction with recognition memory responses to determine whether each response was a hit, miss, false alarm, or correct rejection. Memory *hits*, indicative of successful recognition memory, were defined by "remember" button responses for words that were correctly repeated in Task 1. Memory *misses*, indicative of words that were not recognized after the delay, were defined by "don't remember" responses for words that were correctly identified in Task 1. *False*

---

[2]The purpose of the speech intelligibility manipulation was to characterize activity changes across a broad range of word identification scores, based on previous bandpass filter manipulations (Eckert et al., 2008; Harris et al., 2009). The current study focuses on memory encoding effects, but the intelligibility analyses and results for both tasks are included in the Supplementary Materials.
[3]Word identification responses from Task 1 were scored as correct only when words were repeated exactly as presented, and unintelligible or missing responses were excluded from the analyses. The two raters who scored participant responses were in 96.1% agreement, and each disagreement was resolved by listening to participant recordings from the experiment.

alarms were defined by "remember" button responses and *correct rejections* were defined by "don't remember" responses for words that were either not presented or not correctly identified in Task 1 (i.e. foils).

Nonparametric estimates of sensitivity (A') and bias (B''_D) were used because they are more independent (Donaldson, 1992) and robust to differing or non-normal signal and noise distributions (Stanislaw and Todorov, 1999), compared to traditional parametric estimates (i.e., d' and c). *Sensitivity* was estimated using A', which can range from 0.5 for chance detection to 1 for perfect detection. *Response bias* was estimated using B''_D, with 0 representing no bias, $B''_D = -1$ for the maximum false-positive bias, and $B''_D = 1$ for the maximum false-negative bias (Donaldson, 1992).

Nonparametric statistical tests were performed to characterize SNR-related differences in memory sensitivity and bias, since parametric assumptions are unlikely to be met for these measures (Stanislaw and Todorov, 1999). Bootstrapping tests were used to resample sensitivity (A') and bias (B''_D) scores from the subjects for 10,000 iterations to compute bootstrapped 95% confidence intervals (BCI) for SNR-related differences (R version 3.3.1; R-package: boot, version 1.3.18).

Image acquisition—The imaging data were collected with a Siemens 3T Trio scanner with a 32-channel headcoil at the MUSC Center for Biomedical Imaging. The T1-weighted structural images were collected using an MPRAGE sequence with 160 slices, 256×256 matrix, TR = 8.13 ms, TE = 3.7 ms, 8 degree flip angle, 1.0 mm slice thickness, and zero slice gap. Functional image data were collected (Task 1: 180 volumes, Task 2: 150 volumes) with a T2*-weighted sequence with single-shot echo-planar imaging (EPI), 36 slices, 64×64 matrix, TR = 8.6 s, TE = 35 ms, 90 degree flip angle, acquisition time = 1647 ms, slice thickness = 3.0 mm, slice gap = 0, sequential order, GRAPPA-parallel imaging with acceleration factor = 2. The EPI sequence was designed to exclude four volumes (i.e. dummy scans) prior to data collection. The functional images consisted of 3 mm isomorphic voxels.

Image preprocessing—The Advanced Normalization Tools software (ANTS version 2.1) was used to create a study-specific template that reflects the average space defined by the brains of the study participants (Avants and Gee, 2004). Each participant's anatomical T1-weighted image and coregistered functional images were spatially transformed to match the study template space using ANTS. Voxel coordinates for statistic peaks in Montreal Neurological Institute (MNI) space were also determined by using ANTS to spatially transform the study template to match the MNI template, then applying those transformations to statistical maps in the study template space.

Functional BOLD images collected during word identification in noise (Task 1) were preprocessed using SPM8 procedures (www.fil.ion.ucl.ac.uk/spm): realignment and unwarping, coregistration of functional images to the native T1 structural image, spatial normalization (ANTS), and spatial smoothing (Gaussian kernel with FWHM = 8 mm). Because the image data were spatially transformed prior to smoothing, all subject and group level image analyses were performed in the space of the study-specific template. Global

BOLD signal fluctuations were residualized from each voxel level BOLD time series (Macey et al., 2004). Four nuisance regressors were entered into the General Linear Model (GLM) to summarize three-dimensional head position and movements based on the application of the Pythagorean Theorem to SPM motion correction output (Kuchinsky et al., 2012; Wilke, 2012; http://www.nitrc.org/projects/pythagoras).

Image analyses—Whole brain subject-level regression analyses were performed using a general linear model (SPM8 software) to identify changes in BOLD contrast during the Task 1 memory encoding for words identified in noise. Separate GLMs were performed so that we could appropriately estimate variance attributed to memory-misses or memory-hits (Mumford et al., 2015). Each GLM was used to predict BOLD contrast based on events that were convolved with the hemodynamic response function (HRF) for: 1-2) word presentations in +3 and +10 dB SNR; 3) babble trials with no task demands; and 4) transitions between blocks of task trials and trials with no task demands (i.e. salient transitions). For each SNR condition, a parametric modulator was included for correct or incorrect word identification in Task 1. Similarly, separate parametric modulators were included for memory-hits (1 hit, 0 non-hit) and memory-misses[4] (1 miss, 0 non-miss) during Task 2. For example, the memory-misses variable was entered last in the GLM when examining the degree to which misses were related to brain activity. Thus, these analyses were designed to characterize activity changes during memory encoding for words that were later memory-hits or memory-misses with the other trials in each SNR condition as baseline, which were adjusted for effects related to incorrect responses and head motion (detailed above).

Group-level statistics were performed on the contrast maps produced by the subject-level GLMs to identify consistent BOLD contrast effects across participants. The contrasts generated from the word identification trial parameters included the following for each subject: 1) Hits > Misses, 2) Hits > Non-Hits, 3) Misses < Non-Misses. All of the group-level statistic maps were submitted to an uncorrected voxel statistic threshold ($Z = 3.09$, $p_{UNC} = 0.001$), then permutation tests were performed using BROCCOLI (Eklund et al., 2014; https://github.com/wanderine/BROCCOLI) to determine the family-wise error corrected $p_{FWE} < 0.05$ cluster threshold. The permutation method was chosen to identify significant clusters based on superior false positive error rate control compared to standard parametric tests (Eklund et al., 2016). Image processing, analysis methods, and fMRI results for Task 2 are presented in Supplementary Materials.

Results

The results from manipulating speech level and BPF in each task demonstrated that, as expected, lower SNR and lower BPF cutoff frequency was associated with poorer word identification and changes in cingulo-opercular activity (Supplementary Figures 1 and 2).

---

[4]Despite the complementary nature of hits and misses, 54% of the word identification task trials were neither hits nor misses on average (e.g. words that were not understood in Task 2), which limited the collinearity of the memory-hit and miss parameters.

Noise effects on memory encoding

The memory sensitivity A' results indicated that words encoded in the +10 dB SNR were correctly recognized (i.e. remembered) more often than words in the +3 dB SNR (+10 dB SNR: $A' = 0.86$; +3 dB SNR: $A' = 0.64$; SNR-difference bootstrap 95% BCI = [0.16, 0.28]; see Figure 1). The results from the bias analysis demonstrated a conservative response bias regardless of SNR condition (+10 dB SNR: $B''_D = 0.83$; +3 dB SNR: $B''_D = 0.84$; SNR-difference bootstrap 95% BCI = [-0.09, 0.05]). This false negative bias indicates that subjects were more likely to respond that they did not remember a word, regardless of whether the word was previously correctly identified.

Cingulo-opercular activity effects on memory encoding

We tested the hypothesis that trial-level changes in cingulo-opercular activity during correct word identification (Task 1; i.e. encoding) were associated with recognition memory in Task 2, following the delay. As shown in Table 1 and Figure 2, the results of the group-level test on the [Misses < Non-Misses] contrasts indicated that BOLD contrast in cingulo-opercular regions was significantly lower than baseline when a word was correctly identified, but not remembered after the delay. Post-hoc tests demonstrated that those regions exhibited lower BOLD contrast during miss trials [Misses < Non-Misses] for the +3 dB SNR and +10 dB SNR conditions, and that there was no significant difference in BOLD contrast between these conditions (+3 dB SNR: $t(19) = -2.35$, 95% BCI = [-4.03, -0.42]; +10 dB SNR: $t(19) = -5.84$, 95% BCI = [-7.66, -2.33]; SNR-difference: $t(19) = 1.74$, 95% BCI = [-1.51, 3.68]). Because lower cingulo-opercular activity was observed for correctly identified words that were not remembered following the delay, regardless of SNR, memory encoding was assumed to be poorer during lapses in cingulo-opercular activity. There were no significant clusters observed for memory hits [Hits > Non-Hits] or for memory hits compared to misses across SNR conditions [Hits > Misses]. No areas outside of the cingulo-opercular cortex exhibited significant changes in activation that related to subsequent recognition memory.

Discussion

The current study tested competing predictions that the engagement of attentional resources, reflected in increased cingulo-opercular activity, would either limit or support memory for speech understood in difficult listening conditions. The results demonstrated that lapses in cingulo-opercular activity during correct word identification in noise were associated with memory misses in the delayed recognition memory task. While cingulo-opercular activity is typically high during speech in noise tasks (Eckert et al., 2009; Erb and Obleser, 2013; Harris et al., 2009; Vaden et al., 2013), it is the variation in this elevated activity that was critical for testing the attentional resources hypothesis. The results are consistent with a large memory literature that has shown better memory encoding with increased cingulo-opercular activity. We conclude that speech is more susceptible to forgetting when support from cingulo-opercular cortex wanes in difficult listening conditions.

Large scale meta-analyses of delayed recognition memory results reveal that cingulo-opercular activity is increased during encoding for memory hits versus misses (26 studies, Spaniol et al., 2009; 74 studies, Kim, 2011). The results of both meta-analyses indicate that

delayed recognition memory is more likely for items accompanied by elevated cingulo-opercular activity during encoding. These memory encoding effects are more typically observed in the left inferior frontal cortex (Spaniol et al., 2009), particularly when encoding of verbal stimuli is compared to pictorial stimuli (Kim, 2011). The spatial pattern of these effects is consistent with the results from the current study.

While the current pattern of results is broadly consistent with previous findings that higher cingulo-opercular activity during the encoding phase is associated with a higher likelihood of successful recognition memory (Kim, 2011; Spaniol et al., 2009), our results were not an exact replication. Our results indicate that reduced cingulo-opercular activity during encoding was associated with poorer recognition memory for words that were previously identified in difficult multitalker babble listening conditions. We did not observe the predicted activity increases for words that were later memory hits nor the predicted activity differences for hits compared to misses. This subtle difference in the current findings from the extant recognition memory literature could reflect the sustained, elevated cingulo-opercular activity that typically accompanies performance of a difficult word identification in noise task, including the current experiment. In other words, the speech SNR manipulation could have limited the upper range of cingulo-opercular activity that was predicted to relate to memory-hits during encoding. Memory studies often present words visually to limit perceptual demands during the encoding phase (e.g., Buckner et al., 2001; Otten et al., 2001; Clark and Wagner, 2003; Wimber et al., 2010).

Decreased cingulo-opercular activity in the current study could also reflect attention lapses (Eichele et al., 2008; Weissman et al., 2006) or drifting tonic attention (Coste and Kleinschmidt, 2016; Sadaghiani et al., 2009; Sadaghiani and D'Esposito, 2015) during the relatively long (26 min) word identification task, wherein a single word was presented every 8.6 s to accommodate the sparse fMRI acquisition timing. A distraction interpretation is also consistent with evidence of poorer memory encoding of speech in noise when participants were cued to attend to a concurrent visual task (Wild et al., 2012).

The engagement of cingulo-opercular regions appears to support speech understanding in difficult listening conditions, perhaps due to greater attention control (Eckert et al., 2016). More extensive cingulo-opercular activation has also been related to more detailed or more deeply encoded memories, based on recognition memory studies that did not manipulate perceptual difficulty (Otten et al., 2001; Ritchey et al., 2011). Because participants were unaware of the recognition memory task goals during Task 1, adjustments in attention would optimize word identification in noise and benefit memory encoding indirectly. This means that the current neuroimaging results and previous similar findings are related to *incidental memory encoding*, passive and unintentional transfer of information to long-term memory through cognitive operations performed on that information (Craik and Tulving, 1975). Within that framework, our current finding that words were more poorly encoded when activity decreased could reflect less extensive cognitive processing despite correct identification in noise.

Memory for speech in noise is consistently poorer when attention is divided between difficult speech listening conditions and memory maintenance or rehearsal (Heinrich et al.,

2008; Pichora-Fuller et al., 1995; Rabbitt, 1968). This negative effect of noise on memory has been hypothesized to result, in part, from the limits of attention-related resources that appear to facilitate perceptual repair or extraction of the speech signal from noise at the expense of memory encoding or maintenance (Heinrich et al., 2008; Pichora-Fuller et al., 1995). This hypothesis is largely supported by evidence from serial recall tasks and continuous speech tasks with high working memory requirements (e.g., Piquado et al., 2010). Our current results suggest that attention to the task and attendant cingulo-opercular activity facilitates rather than limits memory for speech, at least for incidental encoding that transfers information to memory without active rehearsal or maintenance[5].

We note that the current experiment did not manipulate working memory or the engagement of a fronto-parietal system that supports working memory. We also did not explicitly manipulate attention and cingulo-opercular activity within SNR conditions. However, the activity of cingulo-opercular regions reflects performance monitoring (Ullsperger and von Cramon, 2004), response selection (Moss et al., 2005), inhibition (Aron, 2007), and evaluation of the expected value from performance (Shenhav et al., 2013; for review see Eckert et al., 2016). Thus, our results support the premise that activity of a system of brain regions that is important for monitoring and optimizing task performance supports rather than limits incidental memory encoding. It is unclear the extent to which recognition memory for speech encoded in noise would be negatively affected by increased or decreased activity of a fronto-parietal working memory system.

The association between cingulo-opercular activity and delayed recognition memory was observed for both SNRs during Task 1, which indicates that this activity was important for incidental memory encoding regardless of the difficulty of the listening task. Nevertheless, delayed recognition memory tasks consistently demonstrate that encoding is poorer for speech listening task conditions that are more difficult (e.g., Wild et al., 2012; Zekveld et al., 2013; Gilbert et al., 2014; Van Engen and Peelle, 2014). The present study replicates these behavioral results: recognition memory was worse for words that had been presented in a lower compared to a higher SNR in Task 1. However, we could not differentiate these well-established intelligibility effects during encoding from potential retrieval differences in the current behavioral results, because of the collinear SNR (Task 1) and BPF (Task 2) conditions. Despite that limitation, our neuroimaging results clearly indicate that trial-by-trial fluctuations in cingulo-opercular activity during the difficult listening task were predictive of delayed recognition memory task performance.

The results of the current study also demonstrate the importance of trial-level analyses for evaluating an attentional resources hypothesis. Within-subject, trial-level activity associations with delayed memory performance were revealed that would otherwise be obscured by relating mean activity changes and memory differences between SNR conditions. Because cingulo-opercular activity during encoding was higher on average (see Supplementary Figure 2) and recognition memory was poorer for words in the more difficult

---

[5]Our results do not rule out the possibility that attention control limits encoding when participants are instructed to remember speech presented in noise, which requires attention to understand speech and actively maintain representations (i.e., rehearsal) at the same time.

SNR, cingulo-opercular activity changes in response to SNR could lead to an erroneous interpretation that this activity is limiting memory encoding rather than benefitting it. The results of our trial-level analyses support the opposite conclusion: cingulo-opercular activity was lower for words that were not remembered after the delay period. The trial-level effect of cingulo-opercular activity across SNR conditions is more consistent with previous evidence for incidental memory encoding related to cingulo-opercular activity, as well as a proposed adaptive control function during speech listening in noise tasks (Eckert et al., 2016; Kim, 2011; Spaniol et al., 2009; Vaden et al., 2013).

Conclusions

The current results suggest that cingulo-opercular regions support incidental memory encoding for words correctly repeated in difficult listening conditions. Listeners may have difficulty sustaining elevated cingulo-opercular performance monitoring activity during a long and challenging task, which would limit the identification of degraded speech stimuli and incidental memory encoding for correctly understood speech. Characterizing how and when to optimally engage attention control systems could enhance memory encoding for speech in noisy environments.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Adank P. The neural bases of difficult speech comprehension and speech production: two Activation Likelihood Estimation (ALE) meta-analyses. Brain Lang. 2012; 122:42–54. DOI: 10.1016/j.bandl. 2012.04.014 [PubMed: 22633697]

American National Standards Institute. Specification for Audiometers ANSI S3.6-2004. American National Standards Institute; New York: 2004.

American National Standards Institute. Specification for Audiometers ANSI S3.6-2010. American National Standards Institute; New York: 2010.

Aron AR. The neural basis of inhibition in cognitive control. Neuroscientist. 2007; 13:214–28. [PubMed: 17519365]

Aron AR, Fletcher PC, Bullmore ET, Sahakian BJ, Robbins TW. Stop-signal inhibition disrupted by damage to right inferior frontal gyrus in humans. Nat Neurosci. 2003; 6:115–6. [PubMed: 12536210]

Avants BB, Gee JC. Geodesic estimation for large deformation anatomical shape averaging and interpolation. Neuroimage. 2004; 23(l):S139–S150. [PubMed: 15501083]

Blumenfeld RS, Ranganath C. Prefrontal cortex and long-term memory encoding: an integrative review of findings from neuropsychology and neuroimaging. Neuroscience. 2007; 13:280–291.

Botvinick MM, Cohen JD, Carter CS. Conflict monitoring and anterior cingulate cortex: an update. Trends Cogn Sci. 2004; 8:539–546. [PubMed: 15556023]

Buckner RL, Wheeler ME, Sheridan MA. Encoding processes during retrieval tasks. J Cogn Neurosci. 2001; 13:406–415. [PubMed: 11371316]

Carter CS, Braver TS, Barch DM, Botnivick MM, Noll D, Cohen JD. Anterior cingulate cortex, error detection, and the online monitoring of performance. Science. 1998; 280:747–749. [PubMed: 9563953]

Carter CS, Macdonald AM, Botvinick M, Ross LL, Stenger VA, Noll D, Cohen JD. Parsing executive processes: strategic versus evaluative functions of the anterior cingulate cortex. Proc Natl Acad Sci USA. 2000; 97:1944–1948. [PubMed: 10677559]

Cechetto DF. Cortical control of the autonomic nervous system. Exp Physiol. 2014; 99:326–331. DOI: 10.1113/expphysiol.2013.075192 [PubMed: 24121283]

Clark D, Wagner AD. Assembling and encoding word representations: fMRI subsequent memory effects implicate a role for phonological control. Neuropsychologia. 2003; 41:304–317. DOI: 10.1016/S0028-3932(02)00163-X [PubMed: 12457756]

Coste CP, Kleinschmidt A. Cingulo-opercular network activity maintains alertness. Neuroimage. 2016; 128:264–272. DOI: 10.1016/j.neuroimage.2016.01.026 [PubMed: 26801604]

Craik FIM, Tulving E. Depth of processing and the retention of words in episodic memory. J Exp Psychol Gen. 1975; 104:268–294.

Dirks DD, Takayanagi S, Moshfegh A, Noffsinger PD, Fausti SA. Examination of the neighborhood activation theory in normal and hearing-impaired listeners. Ear Hear. 2001; 22:1–13. [PubMed: 11271971]

Donaldson W. Measuring recognition memory. J Exp Psychol Gen. 1992; 121:275–277. [PubMed: 1402701]

Dosenbach NUF, Visscher KM, Palmer ED, Miezin FM, Wenger KK, Kang HC, Burgund ED, Grimes AL, Schlaggar BL, Petersen SE. A core system for the implementation of task sets. Neuron. 2006; 50:799–812. [PubMed: 16731517]

Eckert MA, Menon V, Walczak A, Ahlstrom JB, Denslow S, Horwitz A, Dubno JR. At the heart of the ventral attention system: the right anterior insula. Hum Brain Mapp. 2009; 30:2530–2541. [PubMed: 19072895]

Eckert MA, Teubner-Rhodes S, Vaden KI. Is listening in noise worth it? The neurobiology of speech recognition in challenging listening conditions. Ear Hear. 2016; 37:101S–110S. [PubMed: 27355759]

Eckert MA, Walczak A, Ahlstrom J, Denslow S, Horwitz A, Dubno JR. Age-related effects on word recognition: reliance on cognitive control systems with structural declines in speech-responsive cortex. J Assoc Res Otolaryngol. 2008; 9:525–259.

Eichele T, Debener S, Calhoun VD, Specht K, Engel AK, Hugdahl K, von Cramon DY, Ullsperger M. Prediction of human errors by maladaptive changes in event-related brain networks. Proc Natl Acad Sci USA. 2008; 105:6173–8. [PubMed: 18427123]

Eklund A, Dufort P, Villani M, Laconte S, Cheng X, Ben-Shalom R. BROCCOLI: software for fast fMRI analysis on many-core CPUs and GPUs. Front Neuroinf. 2014; 8doi: 10.3389/fninf. 2014.00024

Eklund A, Nichols TE, Knutsson H. Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. Proc Natl Acad Sci. 2016; doi: 10.1073/pnas.1602413113

Erb J, Obleser J. Upregulation of cognitive control networks in older adults' speech comprehension. Front Syst Neurosci. 2013; 7(116):1–13. [PubMed: 23420631]

Gilbert RC, Chandrasekaran B, Smiljanic R. Recognition memory in noise for speech of varying intelligibility. J Acoust Soc Am. 2014; 135:389–399. DOI: 10.1121/1.4838975 [PubMed: 24437779]

Goghari VM, Macdonald AW. The neural basis of cognitive control: response selection and inhibition. Cereb Cortex. 2005; 71:72–83. DOI: 10.1016/j.bandc.2009.04.004

Harris KC, Dubno JR, Keren NI, Ahlstrom JB, Eckert MA. Speech recognition in younger and older adults: a dependency on low-level auditory cortex. J Neurosci. 2009; 29:6078–6087. [PubMed: 19439585]

Heinrich A, Schneider BA, Craik FIM. Investigating the influence of continuous babble on auditory short-term memory performance. Q J Exp Psych. 2008; 61:735–751. DOI: 10.1080/17470210701402372

Hughes ME, Johnston PJ, Fulham WR, Budd TW, Michie PT. Stop-signal task difficulty and the right inferior frontal gyrus. Behav Brain Res. 2013; 256:205–213. DOI: 10.1016/j.bbr.2013.08.026 [PubMed: 23973765]

Kalikow DN, Stevens KN, Elliott LL. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. J Acoust Soc Am. 1977; 61:1337–1351. [PubMed: 881487]

Kerns JG, Cohen JD, MacDonald AW, Cho RY, Stenger VA, Carter CS. Anterior cingulate conflict monitoring and adjustments in control. Science. 2004; 303:1023–6. [PubMed: 14963333]

Kim H. Neural activity that predicts subsequent memory and forgetting: a meta-analysis of 74 fMRI studies. Neuroimage. 2011; 54:2446–2461. DOI: 10.1016/j.neuroimage.2010.09.045 [PubMed: 20869446]

Kuchinsky SE, Vaden KI, Keren NI, Harris KC, Ahlstrom JB, Dubno JR, Eckert MA. Word intelligibility and age predict visual cortex activity during word listening. Cereb Cortex. 2012; 22:1360–1371. DOI: 10.1093/cercor/bhr211 [PubMed: 21862447]

Macey PM, Macey KE, Kumar R, Harper RM. A method for removal of global effects from fMRI time series. Neuroimage. 2004; 22:360–6. [PubMed: 15110027]

Moss HE, Abdallah S, Fletcher P, Bright P, Pilgrim L, Acres K, Tyler LK. Selecting among competing alternatives: selection and retrieval in the left inferior frontal gyrus. Cereb Cortex. 2005; 15:1723–1735. DOI: 10.1093/cercor/bhi049 [PubMed: 15728742]

Mumford JA, Poline JB, Poldrack RA. Orthogonalization of regressors in fMRI models. PLoS One. 2015; doi: 10.1371/journal.pone.0126255

Murphy DR, Craik FIM, Li KZH, Schneider BA. Comparing the effects of aging and background noise on short-term memory performance. Psychol Aging. 2000; 15:323–334. DOI: 10.1037//0882-7974.15.2.323 [PubMed: 10879586]

Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia. 1971; 9:97–113. [PubMed: 5146491]

Otten LJ, Henson RNA, Rugg MD. Depth of processing effects on neural correlates of memory encoding: relationship between findings from across-and within-task comparisons. Brain. 2001; 124:399–412. [PubMed: 11157567]

Pichora-Fuller MK, Schneider BA, Daneman M. How young and old adults listen to and remember speech in noise. Acoust Soc Am. 1995; 97:593–608.

Piquado T, Cousins KAQ, Wingfield A, Miller P. Effects of degraded sensory input on memory for speech: behavioral data and a test of biologically constrained computational models. Brain Res. 2010; 1365:48–65. DOI: 10.1016/j.brainres.2010.09.070 [PubMed: 20875801]

Poldrack RA, Temple E, Protopapas A, Nagarajan S, Tallal P, Merzenich M, Gabrieli JD. Relations between the neural bases of dynamic auditory processing and phonological processing: evidence from fMRI. J Cogn Neurosci. 2001; 13:687–97. DOI: 10.1162/089892901750363235 [PubMed: 11506664]

Rabbitt P. Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ. Acta Otolaryngol. 1991; 476:167–175.

Rabbitt PMA. Channel capacity, intelligibility, and immediate memory. Q J Exp Psychol. 1968; 20:241–248. [PubMed: 5683763]

Ritchey M, Labar KS, Cabeza R. Level of processing modulates the neural correlates of emotional memory formation. J Cogn Neurosci. 2011; 23:757–771. [PubMed: 20350176]

Sadaghiani S, D'Esposito M. Functional characterization of the cingulo- opercular network in the maintenance of tonic alertness. Cereb Cortex. 2015; 25:2763–73. [PubMed: 24770711]

Sadaghiani S, Hesselmann G, Kleinschmidt A. Distributed and antagonistic contributions of ongoing activity fluctuations to auditory stimulus detection. J Neurosci. 2009; 29:13410–7. [PubMed: 19846728]

Shenhav A, Botvinick MM, Cohen JD. The expected value of control: an integrative theory of anterior cingulate cortex function. Neuron. 2013; 79:217–240. [PubMed: 23889930]

Sheth SA, Mian MK, Patel SR, Asaad WF, Williams ZM, Dougherty DD, Bush G, Eskandar EN. Human dorsal anterior cingulate cortex neurons mediate ongoing behavioural adaptation. Nature. 2012; 488:218–221. [PubMed: 22722841]

Spaniol J, Davidson PSR, Kim ASN, Han H, Moscovitch M, Grady CL. Event-related fMRI studies of episodic encoding and retrieval: meta-analyses using activation likelihood estimation. Neuropsychologia. 2009; 47:1765–1779. DOI: 10.1016/j.neuropsychologia.2009.02.028 [PubMed: 19428409]

Stanislaw H, Todorov N. Calculation of signal detection theory measures. Behav Res Methods, Instruments, Comput. 1999; 31:137–149.

Thompson-Schill SL, D'Esposito M, Aguirre GK, Farah MJ. Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. Proc Natl Acad Sci USA. 1997; 94:14792–14797. [PubMed: 9405692]

Tun PA, McCoy S, Wingfield A. Aging, hearing acuity, and the attentional costs of effortful listening. Psychol Aging. 2009; 24:761–766. DOI: 10.1037/a0014802.Aging [PubMed: 19739934]

Ullsperger M, von Cramon DY. Neuroimaging of performance monitoring: error detection and beyond. Cortex. 2004; 40:593–604. [PubMed: 15505969]

Vaden KI, Kuchinsky SE, Ahlstrom JB, Dubno JR, Eckert MA. Cortical activity predicts which older adults recognize speech in noise and when. J Neurosci. 2015; 35:3929–3937. DOI: 10.1523/JNEUROSCI.2908-14.2015 [PubMed: 25740521]

Vaden KI, Kuchinsky SE, Cute SL, Ahlstrom JB, Dubno JR, Eckert MA. The cingulo-opercular network provides word-recognition benefit. J Neurosci. 2013; 33:18979–86. DOI: 10.1523/JNEUROSCI.1417-13.2013 [PubMed: 24285902]

Van Engen KJ, Peelle JE. Listening effort and accented speech. Front Hum Neurosci. 2014; 8:577.doi: 10.3389/fnhum.2014.00577 [PubMed: 25140140]

Ward CM, Rogers CS, Van Engen KJ, Peelle JE. Effects of age, acoustic challenge, and verbal working memory on recall of narrative speech. Exp Aging Res. 2016; 42:97–111. [PubMed: 26683044]

Weissman DH, Roberts KC, Visscher KM, Woldorff MG. The neural bases of momentary lapses in attention. Nat Neurosci. 2006; 9:971–978. [PubMed: 16767087]

Wild CJ, Yusuf A, Wilson DE, Peelle JE, Davis MH, Johnsrude IS. Effortful listening: the processing of degraded speech depends critically on attention. J Neurosci. 2012; 32:14010–14021. [PubMed: 23035108]

Wilke M. An alternative approach towards assessing and accounting for individual motion in fMRI timeseries. Neuroimage. 2012; 59:2062–2072. [PubMed: 22036679]

Wimber M, Heinze HJ, Richardson-Klavehn A. Distinct frontoparietal networks set the stage for later perceptual identification priming and episodic recognition memory. J Neurosci. 2010; 30:13272–13280. DOI: 10.1523/JNEUROSCI.0588-10.2010 [PubMed: 20926653]

Zekveld AA, Heslenfeld DJ, Festen JM, Schoonhoven R. Top-down and bottom-up processes in speech comprehension. Neuroimage. 2006; 32:1826–36. [PubMed: 16781167]

Zekveld AA, Rudner M, Johnsrude IS, Rönnberg J. The effects of working memory capacity and semantic cues on the intelligibility of speech in noise. J Acoust Soc Am. 2013; 134:2225–2234. DOI: 10.1121/1.4817926 [PubMed: 23967952]
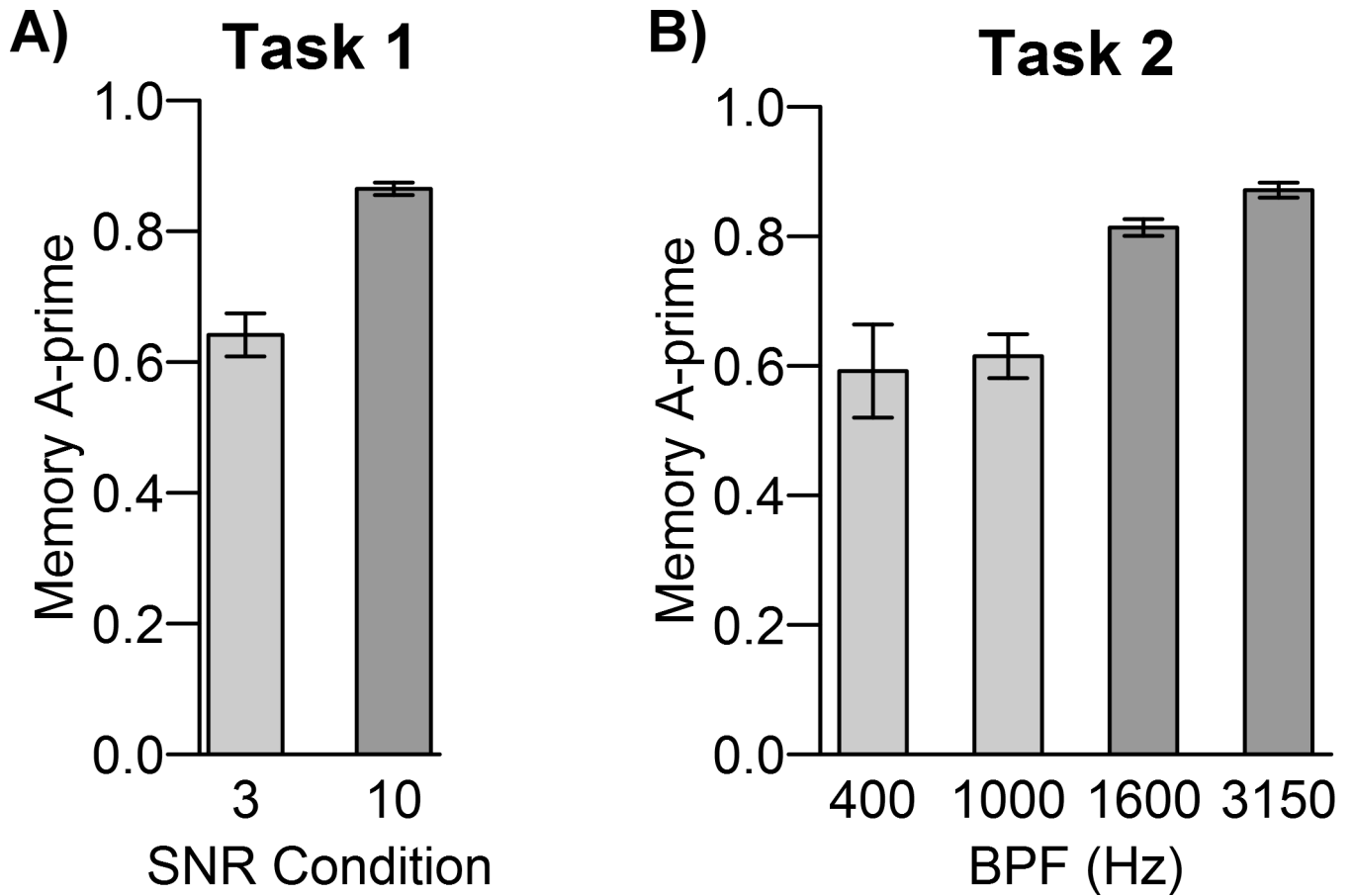
Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.
Left: Memory sensitivity (A') scores during the delayed recognition memory test were significantly higher for speech presented in the higher SNR condition during the encoding phase. Average values are shown with bootstrap 95% confidence intervals. The asterisk signifies that the bootstrap 95% BCI for sensitivity differences between SNR conditions did not include zero, indicating $p < 0.05$. Sensitivity within each SNR was also significantly above chance (A' 95% BCI did not include 0.5). Right: The false-negative response bias was not significantly different across SNR conditions. There was a significantly conservative bias within each SNR (B''_D 95% BCI did not include 0).

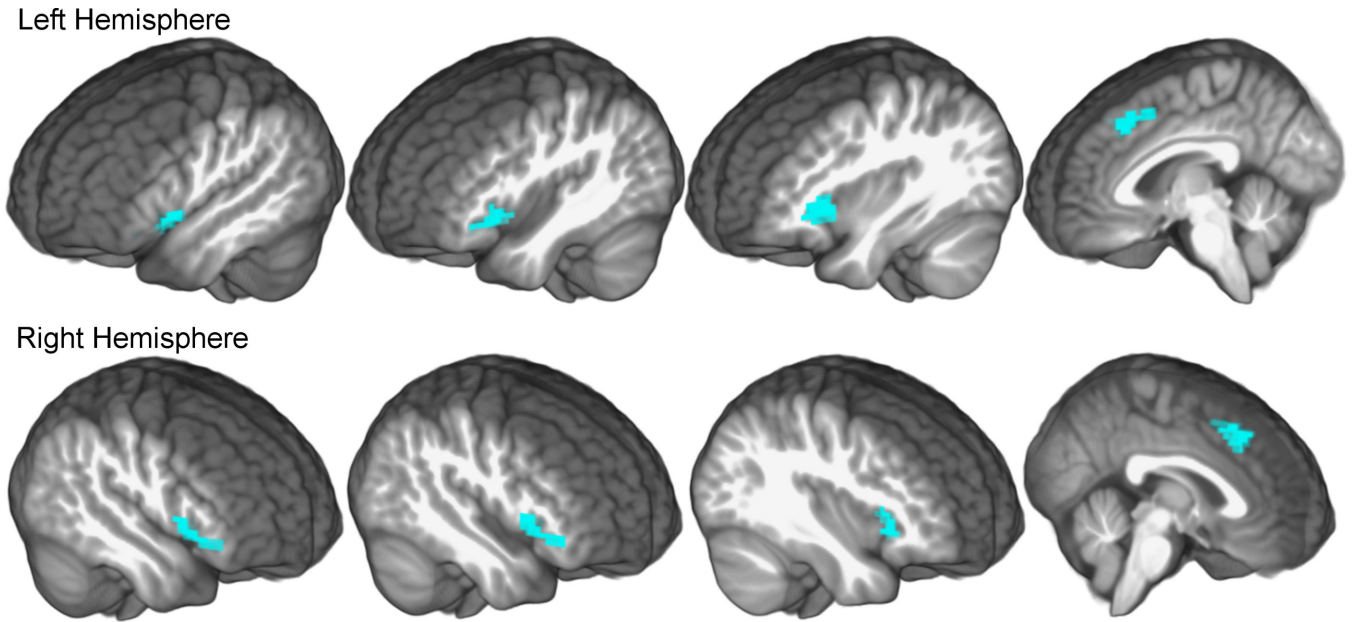## Left Hemisphere

Right Hemisphere

Figure 2.
Significantly decreased BOLD contrast was observed during the encoding phase for correctly understood words that were not remembered after the delay period [contrast: Misses < Non-Misses]. Cingulo-opercular BOLD activity was significantly lower for words that were later memory-misses in the delayed recognition memory task (cyan), compared to the other words in each SNR condition. Combined voxel statistic: $Z = 3.09$, $p_{UNC} = 0.001$ and permutation-corrected cluster extent: $p_{FWE} < 0.05$ thresholds were applied to the statistic map, which is shown superimposed on study-specific average brain template.

Table 1
Memory-Related BOLD Changes during Word Identification in Noise (Task 1)

Description of contrast, cluster extent	Peak Z	# Voxels	Peak MNI
Misses < Non-Misses			
Dorsal cingulate / paracingulate	4.54	89	1, 27, 47
L. inf. frontal gyrus, L. ant. insula	4.31	133	-42, 24, -1
R. ant. Insula, R. inf. frontal gyrus	4.05	130	55, 18, 5

Note: MNI: Montreal Neurological Institute coordinates; L: left, R: right, otherwise bilateral; Ant: anterior, Inf: inferior. The Misses < Non-Misses contrast tested whether BOLD contrast was lower during the identification of words that were subsequent memory misses, compared to the other words in each SNR condition.