

Human Dorsal and Ventral Auditory Streams Subserve Rehearsal-Based and Echoic Processes during Verbal Working Memory

Bradley R. Buchsbaum,* Rosanna K. Olsen,
Paul Koch, and Karen Faith Berman
Section on Integrative Neuroimaging
Clinical Brain Disorders Branch
National Institute of Mental Health
National Institutes of Health
Intramural Research Program
Department of Health and Human Services
Bethesda, Maryland 20892

Summary

To hear a sequence of words and repeat them requires sensory-motor processing and something more—temporary storage. We investigated neural mechanisms of verbal memory by using fMRI and a task designed to tease apart perceptually based (“echoic”) memory from phonological-articulatory memory. Sets of two- or three-word pairs were presented bimodally, followed by a cue indicating from which modality (auditory or visual) items were to be retrieved and rehearsed over a delay. Although delay-period activation in the planum temporale (PT) was insensitive to the source modality and showed sustained delay-period activity, the superior temporal gyrus (STG) activated more vigorously when the retrieved items had arrived to the auditory modality and showed transient delay-period activity. Functional connectivity analysis revealed two topographically distinct fronto-temporal circuits, with STG coactivating more strongly with ventrolateral prefrontal cortex and PT coactivating more strongly with dorsolateral prefrontal cortex. These argue for separate contributions of ventral and dorsal auditory streams in verbal working memory.

Introduction

In his 1874 monograph on the aphasias, Carl Wernicke argued that the posterior part of the STG constitutes the storehouse for “auditory word images.” He maintained that the neural representation of sound could last longer than the external stimulus impinging on the senses and, moreover, that sensory impressions could be revived independently of the stimulus that gave rise to them. To Wernicke, auditory-verbal memory was simply the persistence of perception whereby the same cortical structures that enable the perception of a sound also provide for its retention via an internal recapitulation of the original sensory event (Wernicke, 1874).

One hundred years later, Baddeley and Hitch, in their 1974 Working Memory (WM) model, offered a view of memory based on an emerging consensus (Shiffrin and Atkinson, 1969) in cognitive psychology and neuropsychology (Warrington and Shallice, 1969) that memories for the immediate (short-term) and distant (long-term) past depend on functionally distinct and largely

independent cognitive structures. Moreover, WM distinguished between the perceptual processes required for stimulus identification and the short-term storage of the mental representations derived from the perception of those same stimuli. Whereas Wernicke argued that the same cortical structures that underlie direct perception also underlie memory, more modern cognitive memory constructs often postulated separate processing units for stimulus recognition (perception), short-term storage, and long-term storage.

Evidence for a functional separation between short- and long-term memory emerged from the discovery of H.M., a patient who could no longer form new long-term memories after surgery to remove the hippocampus, though his capacity to repeat strings of letters or digits was left largely intact (Corkin, 2002). A double dissociation between short- and long-term memory was established with the addition of the case of J.B., a patient whose auditory-verbal span was reliable for only one or two items but whose verbal long-term memory, auditory perception, and motor-speech abilities were relatively unaffected (Shallice and Warrington, 1977). In addition, because J.B. showed no impairment in either speech perception or speech production, the case provided *prima facie* evidence for a separation between auditory-verbal short-term memory and the perceptual identification, comprehension, and production of speech sounds.

On the basis of these key results from neuropsychology as well as certain verbal short-term memory phenomena known to occur irrespective of stimulus-input modality (e.g., the phonological similarity effect, the word-length effect, the irrelevant sound effect, and the effect of articulatory suppression [Baddeley, 1986]), Baddeley and colleagues argued that retention of verbal items in memory depends upon a collaborative interaction between a short-term “phonological store” and an articulatory control system that, through the process of inner rehearsal, refreshes and revives the decaying traces in the phonological store (Baddeley, 1986). This model, known as the phonological loop, served as the guiding theoretical framework for most neuroimaging studies of verbal STM over the last 20 years (Awh et al., 1996; Fiez et al., 1996; Jonides et al., 1998; Paulesu et al., 1993). Several recent studies in cognitive neuroscience, however, may contradict the idea of a separation between perception and memory, supporting the early ideas of Wernicke that short- and long-term memory representations depend largely on the same cortical structures as perception itself (e.g., Postle et al., 2003; Ruchkin et al., 2003).

Although some functional neuroimaging studies of verbal WM have identified activation associated with phonological storage in posterior parietal cortex (Awh et al., 1996; Jonides et al., 1998) or Brodmann Area 40 (BA 40)—an area far afield of those superior temporal regions known to be important for speech perception (Hickok and Poeppel, 2000)—several recent studies with event-related fMRI have identified sites in the superior temporal cortex (STC) that activate during short-term

*Correspondence: buchsbab@intr.nimh.nih.gov

retention of verbal information (Buchsbaum et al., 2001; Hickok et al., 2003; Postle et al., 1999; Sakai and Passingham, 2003; Stevens, 2004). Moreover, it has been shown (Buchsbaum et al., 2001; Hickok et al., 2003) that there are at least two anatomically separable regions in posterior STC, one in the superior temporal sulcus (STS) bilaterally and the other in the left posterior PT (area Spt, i.e., Sylvian-parietal-temporal) that activate during both perception and silent rehearsal of auditory-verbal stimuli. The location of these regions in the vicinity of superior temporal structures implicated in neuroimaging studies of speech perception (Belin et al., 2000; Scott et al., 2000; Vouloumanos et al., 2001) suggests that some aspects of perception and short-term memory (STM) depend on the same underlying neural structures.

Further evidence for a perceptually based verbal memory code comes from cognitive psychology (Watkins and Watkins, 1973). It has often been demonstrated that when performance on verbal STM tasks is scored as a function of the serial position of the item in a word stimulus set, similar levels of performance are observed for visual and auditory presentation across all serial positions except for the last one or two items, in which a significant advantage is observed for the auditory modality. This *auditory recency* effect has been thought to be attributable to the added benefit auditory items receive from “echoic memory,” a form of perceptual memory specific to the auditory modality and unavailable to items presented in the visual domain (Cowan, 1984; Penney, 1989; Watkins and Watkins, 1980). Indeed, there is evidence (Takayama et al., 2004) from a neurological patient with a focal lesion to the left STG (sparing the PT and supramarginal gyrus) that “echoic memory” can be selectively impaired. The patient showed a negative recency effect (poorer recall for the last two items) for auditory presentation of digits and a significantly larger span for items presented to the visual (mean span = 5.8) than the auditory (mean span = 4.3) modality (Takayama et al., 2004). The eradication in this patient of the auditory recency effect, coupled with a relatively preserved digit span in the visual modality, suggests that echoic memory may be partly attributable to the functioning of the cortex of the superior temporal region (STG/STS) and that modality-neutral memory mechanisms, likely based on a phonological-articulatory code, also play a role in the retention of verbal items presented to the auditory and visual modalities.

The difference between perceptual (echoic) and phonological-articulatory (motor) memory may be conceptualized as emerging from the separate contribution of dorsal and ventral anatomical streams. Based on non-human primate work, it has been suggested that as in the visual system (Ungerleider and Mishkin, 1982), information flow from auditory cortex diverges into segregated anatomical streams—a ventral “what” stream and a dorsal “where” stream (Rauschecker and Tian, 2000; Romanski et al., 1999). Other authors, drawing on human lesion evidence and functional neuroimaging data (Hickok and Poeppel, 2000, 2004; Scott and Wise, 2004; Wise et al., 2001), have proposed that the auditory dorsal stream in humans predominantly involves action planning—principally for the production of speech—with the posterior portion of the PT (area Spt) serving as an auditory-motor interface (Hickok et al., 2003).

We might argue that modality-neutral phonological-articulatory memory depends on the action circuits (“how”) of the auditory dorsal stream and that echoic memory involves the perceptual (“what”) modules of the auditory ventral stream. However, little direct system-level neurobiological evidence in humans exists to support these contentions.

In the present study, we sought to assess the extent to which temporal lobe regions previously identified as playing a role in verbal STM—the STG/STS and Spt—are sensitive to the input modality of stimuli that are to be maintained in memory over a brief delay. We hypothesize modality-specific echoic memory and modality-neutral phonological-articulatory memory, with the former located ventrally in the STG/STS and the latter located dorsally in Spt. We further postulate that if this functional-anatomical separation is confirmed, interregional functional connectivity analyses will reveal differential patterns of connectivity between these temporal lobe sites and areas in the frontal lobe, with the specific prediction that as has been shown in the monkey (Romanski et al., 1999), ventral temporal sites project preferentially to ventrolateral prefrontal cortex (VLPFC) and dorsal sites project preferentially to dorsal prefrontal regions.

To test our hypotheses, we adapted a paradigm (Broadbent, 1956) that uses bisensory presentation of verbal stimuli and requires subjects to selectively recall either items that were heard (MEMORY trial; EAR condition), those that were (silently) read (MEMORY trial; EYE condition), or no items at all (perception/encoding without retrieval/rehearsal; NO-MEMORY control condition). Each visually presented word was paired with a different word that was presented auditorally. After a sequence of word pairs was presented, subjects saw a cue that indicated which set of words (i.e., those they heard, those they read, or neither) they should recall. The format of this task dictates that the set of items to be retrieved—either auditory, visual, or neither—not be known until after the presentation of the stimuli. Thus, the stimulus-encoding portion of each trial was identical across all conditions, and the experimental manipulation was whether words heard or words read were to be retrieved and maintained in memory after stimulus presentation (see Figure 1 for event-related structure). If items in memory were solely stored in an amodal phonological-articulatory code, then there would be no observed differences in activation as a function of the modality in which the items to be retrieved were originally presented. On the other hand, if the neural representation of the items were modality specific, then activation would be observed in locales that vary as a function of the modality of the items to be retrieved. If modality-neutral and modality-specific codes both contributed to STM performance, however, recall of items from the auditory modality should selectively enhance delay-period activation in the STC that we would hypothesize as echoic, but not phonological-articulatory, memory. Moreover, because echoic memory rapidly decays a few seconds (Cowan, 1984) after stimulus offset, whereas (subspan) rehearsal-based verbal memory can be maintained indefinitely (Baddeley and Hitch, 1974), we expect that over the course of a long (12 s) retention interval, neural activation supporting echoic memory would have

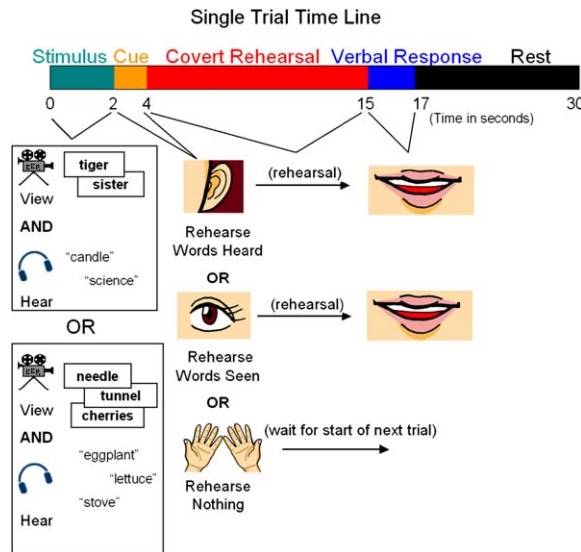


Figure 1. Graphical Depiction of Memory Task with Bisensory Stimulus Presentation

In the first 2–3 s of each trial, the subject is presented with two or three pairs of words, pairs being presented simultaneously with one member of each pair presented visually and the other presented auditorially. Thus, the total number of words presented is either four (two-pair condition) or six (three-pair condition). These bisensory word pairs are presented sequentially, one pair per second. Immediately after the stimulus presentation phase, the subject sees one of the three retrieval cues displayed for 2 s. If the retrieval cue is an EAR, the subject retains in memory the words that he or she has heard (e.g., “candle,” “science”). If the cue is an EYE, the subject retains the words that he or she has read (e.g., “tiger,” “sister”). If the cue is HANDS, the subject simply rests until the next trial. In the EAR and EYE conditions, the subject must covertly rehearse the correct words across a 12 s delay period, after which a MOUTH appears, signaling to the subject to say aloud the words currently retained in memory. In the NO-MEMORY control condition (signaled by HANDS cue), subjects are instructed simply to watch the fixation cross and wait for the next trial.

returned to baseline early in the delay period, whereas activation associated with sustained subvocal rehearsal processes would be observed across the entire retention interval.

Results

Behavioral Performance in fMRI Experiment

Performance was assessed from recordings of subjects’ verbal responses during the scans. Subjects performed the task at near ceiling, with percent correct averaged for all subjects above 90% (LOAD = 2 words, EAR = 95.3%, EYE = 96.48%; LOAD = 3 words, EAR = 88.28%, EYE = 91.4%). Wilcoxon signed rank tests on each subject’s percent correct scores showed no significant modality effect for a given load (LOAD 2, EAR > EYE, $p = 0.67$; LOAD 3, EAR > EYE, $p = 0.40$). A significant overall effect of LOAD was observed ($p < 0.009$), with higher percent-correct scores observed for two item trials than for three item trials. No LOAD by MEMORY interaction was observed, however (computed as a double subtraction, EAR [LOAD 3 – LOAD 2] – EYE [LOAD 3 – LOAD 2]; Wilcoxon signed rank test, $p = 0.8328$).

Imaging Results

Activation Across Trial Phases

Each trial was modeled with four regressors, each of which sampled a distinct “phase” of the 30 s trials (Figure 2). Figure 3 presents an overview of the activation profile for each of these four trial phases (in chronological order: ENCODING, EARLY delay, LATE delay, and RESPONSE) for the left hemisphere, with the delay period split according to the modality of the items to be retrieved (top, EAR; bottom, EYE). During stimulus perception (ENCODING), substantial portions of the temporal, parietal, and frontal lobes showed robust activity. In the EARLY and LATE delay periods, while frontal and parietal sites remained active, temporal lobe activation was greatly reduced. Islands of delay-period activity were, however, detected in Spt (both EAR and EYE conditions), lateral STG/STS (EAR condition only), and middle/inferior gyri of temporal cortex (both EAR and EYE, though this area was not active during ENCODING). During the RESPONSE phase, considerable activation was observed in posterior frontal cortex (motor and sensorimotor regions) and the STC, the latter likely because of auditory input provided by the subject’s own voice.

Delay-Period Activation

Activation observed during the EARLY and LATE delay periods for each of the MEMORY (EAR and EYE) conditions is shown for the left hemisphere in Figures 4A and 4B, with the delay-period images masked to contain only regions that are also significantly active during ENCODING ($p < 0.001$, uncorrected). We used a lower threshold ($p < 0.005$ and cluster size $> 50 \text{ mm}^2$) for the circumscribed region of the STC as defined by the ENCODING contrast (see Figure 3) for the assessment of delay-period activity in auditory cortical regions including Heschl’s gyrus, PT, STG, STS, and MTG.

In the EAR condition, we saw EARLY activity in the STG/STS and LATE activity in the posterior PT. For the EYE condition, significant delay-period activity was seen, for both EARLY and LATE phases, but only in the posterior PT. To assess whether either of the two temporal lobe sites that showed delay-period activity when compared to a no-memory control condition also showed greater activity for auditory-verbal retrieval (EAR) than for visual-verbal retrieval (EYE), we performed the following conjunction: (EAR > NO-MEMORY \cap EAR > EYE) for both EARLY and LATE trial phases. This conjunction was designed specifically to test our hypothesis that the STG/STS, but not the PT, would be associated with modality-specific echoic memory. For all conjunctions, regions were only considered significant if p values for both inputs were less than 0.005. The only temporal lobe region showing enhanced activity during auditory retrieval is in the STG and STS, for the EARLY period only (Figure 5D). Additional regions in the inferior frontal gyrus (IFG) and the anterior insula extending into the frontal operculum also showed auditory-selective EARLY delay-period activity. The complementary conjunction (EYE > NO-MEMORY \cap EYE > EAR) did not reveal any significant areas in the STC; though two regions outside auditory cortex, one in the ventral temporal lobe (fusiform gyrus) and one in the parieto-occipital sulcus, did show a visually selective memory response during the EARLY period of the trial.

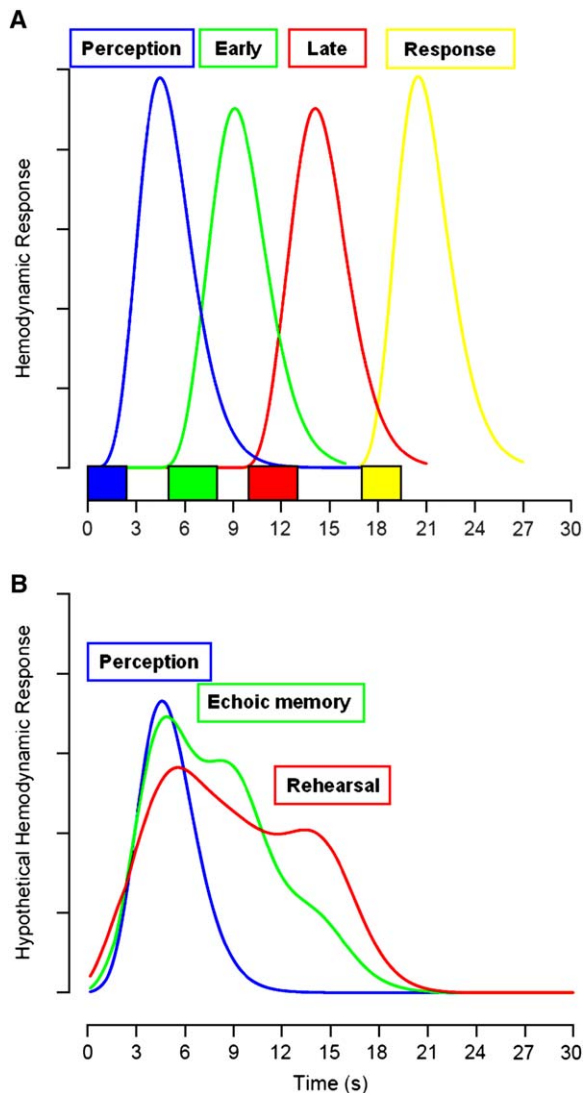


Figure 2. Modeling the Hemodynamic Response across the Three Phases within MEMORY Trials

Top, (A) basis functions separately modeling (1) ENCODING (blue), (2) EARLY memory response (green), (3) LATE memory response, and (4) overt RESPONSE for trials with two-word pairs. Box-car functions for convolution with impulse response function indicated on x axis. Bottom, (B) Hypothetical hemodynamic response functions for a cortical region subserving (1) perceptual processes/encoding only (blue), (2) transient echoic memory (green), and (3) sustained rehearsal processes (red). For simplicity, potential hemodynamic response during overt response not incorporated.

In the right hemisphere, significant delay-period activity was not observed in the STC, though EARLY trend-level activity ($p < 0.05$, uncorrected) was evident in the right STS extending into the middle temporal gyrus and subthreshold, and LATE trend-level activity ($p < 0.05$, uncorrected) was observed in posterior STG at the lateral-most edge of the PT. Thus, right hemisphere delay-period activity was characterized by a qualitatively similar, though statistically depressed, pattern of activity. A main effect of LOAD (collapsing across delay-period phase and retrieval modality) was generally seen in all regions in which delay-period activity was evident,

except for the STG/STS (Figure S1); nor did this region show load sensitivity when restricting the test to the EARLY delay-period phase. None of these load-sensitive regions, however, showed LOAD \times MEMORY interactions.

The two left temporal lobe sites that showed delay-period activity, the STG/STS and Spt, had response properties that differed as a function of time and modality. These contrasting profiles are seen in the group-averaged time courses and bar graph (Figure 5E) of the mean t statistics for both memory conditions across all phases of the trial. Whereas the temporal activation profile in the STS/STG appeared sensory weighted (with a preference for the auditory modality), Spt had a response profile that seemed rehearsal weighted. That is, the delay-period activation in STG/STS “decayed” rapidly with the offset of sensory stimulation, whereas activity in Spt was sustained throughout the delay period. We confirmed this observation formally with a repeated measures ANOVA with independent variables REGION (STG/STS and Spt) and PERIOD (EARLY and LATE) by using single-subject t statistics for the EAR condition as the dependent variable, showing a REGION \times PERIOD interaction ($F[1,15] = 30.266$, $p < 0.0001$). The same interaction was also found when the dependent variable was the t statistics from the EYE condition ($F[1,15] = 42.3$, $p < 0.0001$). Thus, activity in the STG/STS and Spt differed significantly such that the former activated in the early part of the delay period and latter showed a more sustained delay-period response profile (see Figure 5E). To ensure that delay-period activity in STG/STS and Spt could not be explained by activation differences already present during encoding, we performed, for both regions, a series of paired t tests between each of the separate ENCODING t statistics ($\text{ENCODING}_{\text{EAR}} > \text{ENCODING}_{\text{NO-MEMORY}}$, $\text{ENCODING}_{\text{EYE}} > \text{ENCODING}_{\text{EAR}}$, $\text{ENCODING}_{\text{EAR}} > \text{ENCODING}_{\text{EYE}}$). None of the six t tests (three for each region) were significant at $p < 0.05$ (minimum p value = 0.4235).

Functional Connectivity and the Dorsal and Ventral Auditory Streams

The two temporal lobe regions that showed delay-period activity correspond anatomically to areas in the monkey that are known to project to different regions of the prefrontal cortex. In the monkey, the rostral belt and parabelt areas in auditory cortex (corresponding approximately to our STG/STS cluster) project to sites in the ventral prefrontal cortex, whereas the caudal belt and parabelt (corresponding to our Spt cluster) project to more dorsally situated regions of the monkey prefrontal cortex (Romanski et al., 1999). We carried out a functional connectivity analysis to explore differences in interregional connectivity between the (ventral) STG/STS and the (dorsal) Spt clusters evincing delay-period activity in the temporal lobe. The STG/STS cluster was defined as the region active during the EARLY phase of the EAR condition. The Spt cluster was defined as the area active during the LATE phase of the EAR condition. We used the LATE phase to define the Spt cluster because the extent of activation in the LATE phase was more robust than during the EARLY phase. The EAR condition was used to define both clusters because the STS/STG response in the EYE condition, though

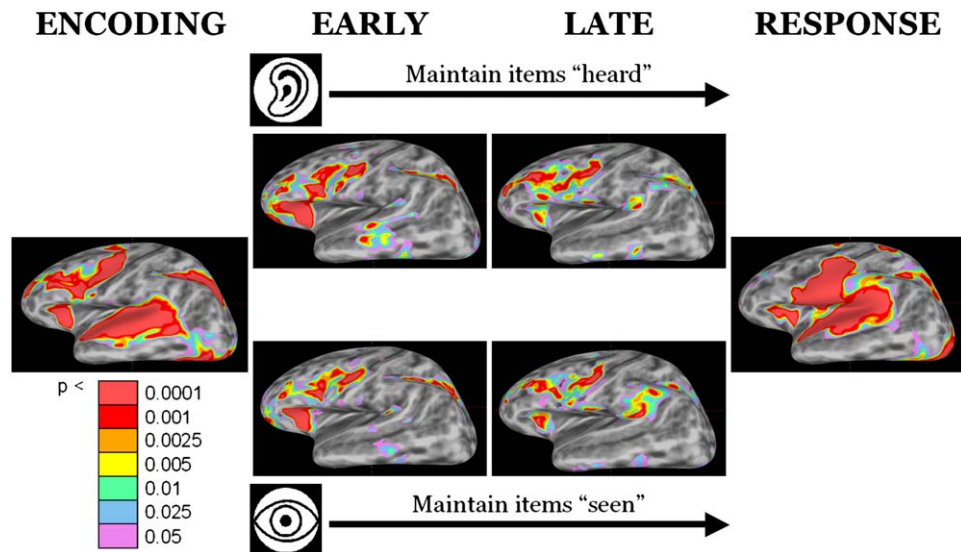


Figure 3. Left Hemisphere Group Surface-Based fMRI Activation across Four Trial Phases

Images show multisubject functional activation for the ENCODING, EARLY, LATE, and RESPONSE periods. Middle four panels (illustrating post-encoding task divergence as a function of retrieval modality) show EARLY and LATE delay period responses for EAR and EYE memory conditions, respectively.

elevated with respect to the control condition, did not reach statistical significance ($p < 0.005$) and a cluster size of 50 mm^2 (see Figure 3). As is evident in Figure 6, the ventral STG/STS site was generally more highly correlated with regions in the ventral prefrontal cortex than was Spt, which itself was more highly correlated with dorsally situated perirolandic regions, including premotor, primary motor, and sensorimotor cortices, as well as an area in the superior frontal gyrus.

Discussion

We examined the extent to which verbal memory representations in the STC are based on (1) an auditory-specific “echoic” code, (2) an abstract modality-neutral phonological code, or (3) a dual code in which both phonological and auditory memory jointly contribute to verbal memory maintenance over the short term. By using bisensory stimulus presentation while requiring items to be retrieved from only one modality at a time, we forced a degree of reliance on the original sensory context of the verbal items. The strategy was successful in demonstrating dissociations between modality-sensitive and insensitive verbal storage sites in the STC. Specifically, the left STG/STS showed enhanced delay-period activation when the to-be-retrieved items were delivered to the auditory modality, whereas Spt showed delay-period activity that was unaffected by stimulus modality. In addition, Spt exhibited elevated activity extending more evenly across the delay interval, suggesting a central role for this area in sustained rehearsal processes. On the other hand, the STG/STS showed a pattern of delay-period activation that was weighted toward the early portion of the delay interval—a profile consistent with the behaviorally estimated temporal durability (4–5 s) of echoic memory (Cowan, 1984). The findings suggest that the delay-period activation in STG/STS, with its transient pattern, is most im-

portant during initial memory retrieval, whereas the more posterior, dorsally situated Spt is involved throughout the temporally extended covert rehearsal process. Our interpretation of this pattern is that the auditory-perceptual echoic codes of the STG/STS are relatively short lived and difficult to reactivate in the absence of external stimuli, whereas phonological (or prearticulatory) codes of Spt are regenerable, as they can be reactivated in a top-down fashion by frontal motor systems. The finding of load effects in Spt and frontal systems may reflect an increased demand (e.g., faster rate of subvocalization) on rehearsal that, because it is sustained across the retention interval, leads to a robust increase in neural activity. On the other hand, access to rapidly decaying, fixed-capacity auditory sensory memory may reflect a one-time operation that does not necessarily scale (in terms of neural activity) with informational load. Thus, the finding of no load effects in the STG/STS may be attributed to a “fixed cost” for access to the contents of echoic memory. Alternatively, if we regard access to the echoic store as a transient event and rehearsal as a sustained process requiring extended neural activity, failure to find load effects in the STG/STS may simply reflect reduced statistical power (related to the difference in statistical power between “block” and “event-related” designs).

Although literal representations of auditory-verbal stimuli are represented in the STG/STS, the codes stored therein are subject to decay, and maintenance of verbal material over extended delay periods favors an internal, speaker-generated phonological-articulatory code that can be repeatedly refreshed by the action of fronto-temporal sensorimotor circuits (Hickok and Poeppel, 2004). We suggest that the modality-neutral activity of Spt is associated with a process that translates between the features of speech as captured by the sensory systems (auditory system for “heard” speech, visual system for “read” speech) and the basic

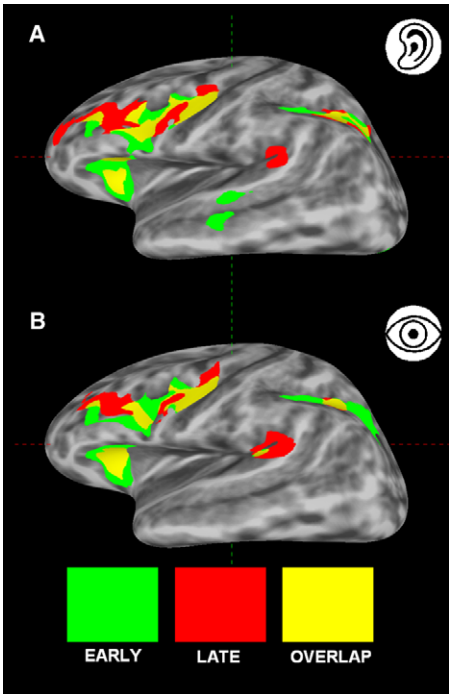
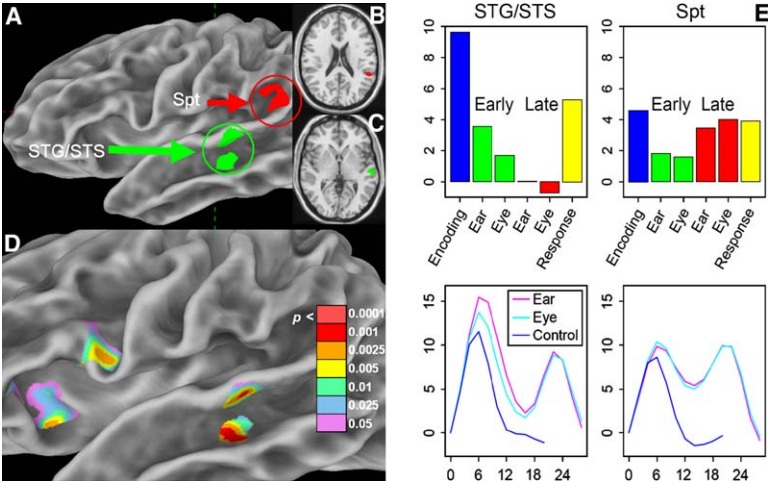


Figure 4. View of Left Hemisphere Showing Delay Period Group Activation during Memory Conditions
(A) Significant delay period activity for EAR condition. EARLY delay (green; $p < 0.005$), LATE delay (red; $p < 0.005$), and conjunction EARLY \cap LATE (yellow; $p < 0.005$ and $p < 0.005$).
(B) Significant delay period activity for EYE condition (color scheme same as in [A]).

articulatory codes required for reproduction of that same speech. Through the transfer of a stream of information from the time-limited memory store of the auditory system to the endlessly regenerable codes of the motor system, a more efficient means of maintaining verbal information in an accessible state is achieved.

According to Hickok and Poeppel (2004), Spt serves to interface structured representations of sound sequences including speech, music, and rhythmic patterns, with the particular effector system (e.g., articulatory, manual) required for the reproduction of those sequences. For example, an fMRI study by Mustovic et al. (2003) showed that unexpected silent gaps embedded in otherwise predictable sound sequences are associated with increased activity in the posterior PT bilaterally. If this response were viewed as a manifestation of an acoustic trace, it would seem to contradict our finding of “echoic memory” in the more anterior/ventral STG/STS instead of the PT. An alternative explanation of activation elicited by embedded gaps in stimulus sequences is that it reflects processing associated with a mismatch between a forward model of a structured sequence and the actual (anomalous) auditory event, rather than a mismatch between the anomalous event and a transient auditory memory. Schubotz and colleagues (Schubotz et al., 2003; Schubotz and von Cramon, 2004) have shown that the critical region for the prediction of events in a structured sequence is the ventral premotor cortex, a region that is tightly correlated with area Spt in our functional connectivity analysis.

Though we have used the term “echoic memory” to refer to the phenomenon of auditory replay and “phonological-articulatory memory” to refer to the kind of rehearsal mechanism described in the “phonological loop” of Baddeley and colleagues, this does not rule out the possibility that storage in the STG/STS, which we have associated with echoic memory, involves a “phonological code.” Indeed, psycholinguistic models of speech perception and speech production often postulate separate phonological “input” (for perception) and “output” (for production) systems (Levelt et al., 1999; Martin and Saffran, 2002). Thus, the dissociation we observed between the STG/STS and Spt might best be described as a difference between the respective roles these two areas play in speech perception (phonological input) and production (phonological



NO-MEMORY control condition. Bottom, left, group trial-averaged time course plots for EAR (magenta), EYE (cyan), and NO-MEMORY (blue) conditions in STG/STS cluster. Right, time course plots for Spt (same color scheme). Time course are based on raw intensity values with the value a $t = 0$ subtracted out for each subject.

Figure 5. Response Properties of Ventral (STG/STS) and Dorsal (Spt) Temporal Lobe Memory Sites
(A) Surface rendering of left temporal regions of interest, defined by EARLY (green) and LATE (red) components of EAR contrast.
(B and C) Axial slice cutting through Spt cluster (tal. x, y, z = -53, -35, 21) overlaid on ICBM single-subject MRI (C) Axial slice cutting through STG/STS cluster (tal. x, y, z = -60, -15, -1).
(D) Regions active for EAR_EARLY > EYE_EARLY (see color scale for threshold scale) after masking with EAR_EARLY > NO-MEMORY_EARLY contrast ($p < 0.005$).
(E) Top, left, mean group t statistic within STG/STS cluster across four trial phases and two memory conditions. Top, right, mean group t statistics in Spt cluster. Note that ENCODING and RESPONSE bars represent a contrast against baseline, whereas intervening bars are contrasts against a

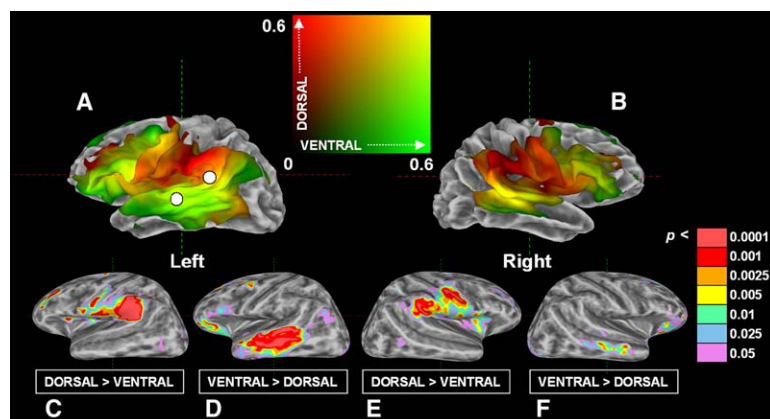


Figure 6. Functional Connectivity with Ventral and Dorsal Temporal Memory Sites

(A) Left hemisphere surface view showing average correlation coefficients between two seed regions, shown as white circles. Correlation with STG/STS shown as green colors (scaled 0–0.60); correlation with Spt shown as red colors (scaled 0–0.6); blending of red and green channels shows degree of mutual correlation between the two seed regions. Areas without at least one correlation coefficient >0.1 are transparent.

(B) Right hemisphere correlations (same color scheme as in [A]).

(C) Surface view of areas significantly more correlated with Spt than STG/STS (t statistic map; graded threshold scale at far right).

(D) Areas significantly more correlated with STG/STS than Spt.

(E) Same as (C) but for right hemisphere.

(F) Same as (D) but for right hemisphere.

output). From these data alone, however, it is not possible to determine at what stage of the acoustic-phonological processing stream echoic memories arise. Indeed, it has long been debated in the cognitive psychological literature whether echoic memory involves access to pre- or postcategorical storage mechanisms. Crowder and Morton (1969) maintained that auditory recency reflects the operation of a limited-capacity peripheral auditory store from which sensory information can later be retrieved in an unprocessed, “precategorical” state. Frankish (1989), on the other hand, has argued that, at least in the context of STM for verbal material, echoic memory reflects postcategorical speech perception processes, not a low-level sensory register such as that suggested by Crowder and Morton (1969). This interpretation is consistent with our observation of delay period activation in the STG/STS—a region critical for speech perception (Hickok and Poeppel, 2000)—and with our failure to find delay-period activation in primary auditory cortex, which would be the predicted locus of an auditory sensory register. Furthermore, the finding of enhanced responsivity during retrieval of spoken speech in this higher-order auditory region is consistent with our interpretation of echoic memory for speech as emerging from the ventral stream, rather than from early or peripheral auditory regions from which dorsal and ventral streams originate. Indeed, Kraus and Nicol (2005) argue that the auditory dorsal and ventral streams may emerge early in the processing hierarchy from specializations already present at the subcortical level. Taken together, then, the observation of enhanced delay-period activity during selective retrieval of words presented to the auditory modality in a region known to be important for perception, suggests that memory “storage” should be viewed as an endpoint or residue of the processing mechanisms that underlie perception itself. It should be made clear, however, that the concept of “echoic memory” no doubt encompasses more than speech-specific STM and, taken in its broadest sense, refers to any kind of auditory-specific memory phenomenon, including, for instance, the auditory mismatch negativity in electrophysiological research (Naatanen et al., 2005; Winkler and Cowan, 2005) and low-level memory for sound intensity (Lu et al., 1992a) or change detection (Mustovic et al.,

2003)—all phenomena generally associated with primary auditory cortex (but see Lu et al., 1992b). Thus, our results may be primarily relevant to an aspect of auditory-verbal memory that has, with many other possibly unrelated auditory memory phenomena, fallen under the semantic umbrella of “echoic memory.”

In light of recent work showing that the STS is an important region for multimodal integration (Beauchamp, 2005; Calvert et al., 2000) and that the simultaneous visual and auditory presentation used in our study may have placed unnatural demands on this system, one might argue that our finding of enhanced delay-period activation actually reflects multisensory processing. For instance, the simultaneous presentation of auditory and visual words might have induced a conflict or competition between the two stimulus streams. Because this was the case for all trials, irrespective of the modality from which items were retrieved, it is difficult to see how such a competition would lead to an asymmetrical response after stimulus presentation. It is conceivable, however, that the differences we observed reflect serial processing of the stimulus sets whereby the visual items are always processed first, whereas the processing of auditory items is postponed—and resuming thereafter only on an as-needed basis. But this hypothetical deferred processing of auditory stimuli is difficult to distinguish, at least at a conceptual level, from what we have referred to as “echoic memory.”

We also demonstrated that the two memory sites in the temporal lobe have different patterns of interregional connectivity with the rest of the brain. These patterns strongly resemble the intrinsic anatomical connections that have been shown in the monkey between rostral and caudal regions in the STC and ventral and dorsal counterparts in the frontal cortex. Our results, then, support the emerging consensus that there exist—as in the visual system—separate dorsal and ventral processing pathways originating in auditory cortex. There is controversy as to the functional significance of these two streams; some have argued for a “what/where” dichotomy as in the visual system (Arnott et al., 2004; Romanowski et al., 1999), whereas others have suggested a perceptual/motor dichotomy, in which the dorsal stream is concerned with imitable acoustic signals—that is, sound

stimuli that can both be perceived and produced—and the ventral stream initiates the analysis of sound for meaning (Hickok and Poeppel, 2000; Wilson, 2001). Although our study does not speak to the auditory-spatial dimension, the finding that Spt maintains a steady level of activation throughout the delay period and is highly correlated with premotor and motor cortices argues for its role as a “gateway to the motor system” and that it serves as an interface system that connects the sounds of speech with the motor commands necessary to reproduce those sounds.

It has been argued that in the visual system, “conscious access” to perceptual information is mediated by the ventral stream (Milner and Goodale, 1995). If the auditory system were similarly organized, conscious access to auditory-perceptual information would be mediated by the ventral auditory stream. Recent neuroimaging work has shown that VLPFC is important for episodic retrieval and “source memory,” for example, the recollection of contextual details associated with a past experience (Buckner et al., 1998; Dobbins et al., 2002). We suggest, then, that conscious access to just-perceived auditory events may rely on the same retrieval processes that have been associated with source memory and appear to be mediated by the VLPFC. The tight functional connectivity between VLPFC and STG/STS may reflect a domain-general role for the VLPFC in episodic retrieval and a domain-specific role for the STG/STS in auditory-perceptual STM.

Conclusion

By presenting items in a verbal WM task simultaneously to both auditory and visual modalities—and only requiring retrieval from one source at a time—we identified modality-specific and modality-neutral storage sites in auditory association cortex. A region in the lateral portion of the STG/STS showed greater early delay-period activity when the items to be retrieved and then maintained in memory arrived from the auditory modality. In contrast, area Spt was insensible to the modality of the retrieved items and showed a sustained delay-period activity profile. Finally, although the posterior and dorsally located Spt showed greater functional connectivity with frontal premotor, motor, and sensorimotor cortices, the STG/STS correlated more strongly with the VLPFC. Taken together, these data argue for a bidimensional view of auditory-verbal WM in which initial retrieval of an episodic or perceptually based code depends on a ventral “what” stream, and sustained rehearsal processes depend on a dorsal-going “action” pathway, with area Spt serving as the temporal lobe gateway to the motor system.

Experimental Procedures

Subjects

Sixteen subjects (seven women; 20–42 years old, mean age = 29.6), all native English speakers, participated in the study after giving informed written consent, as approved by the National Institute of Mental Health Institutional Review Board. No subject had past history of psychiatric or neurological diseases. All subjects, as assessed by the Edinburgh Handedness Inventory, were determined to be strongly right handed.

Task

The paradigm consisted of three basic experimental conditions (see Figure 1). Every trial began with the presentation of a sequence of paired stimuli in which for each pair, one word was presented to the auditory modality, and the other word was simultaneously presented visually. Each word pair was separated, onset to onset, by 1 s. The number of word pairs in a trial was two for odd-numbered runs and three for even-numbered runs. Set size was alternated by scanning run and always began with a set size of two pairs. After stimulus presentation (which lasted 2 s for two-pair sets and 3 s for three-pair sets), one of three retrieval cues was displayed on the screen for 2 s. If the cue was a picture of an ear (EAR), subjects were to retrieve and then maintain in memory, through silent verbal rehearsal, only those items they had heard; if the cue was a picture of an eye (EYE), subjects were to retrieve and then maintain in memory only those items that they had (silently) read; if the picture was of hands (NO-MEMORY), then subjects were to watch the fixation cross and wait for the next trial. Note that there was no special reason for the use of the “hands” picture as the NO-MEMORY cue except that it was a body part and a good match for the EAR and EYE cues. For the two memory (MEMORY) conditions, EAR and EYE, the 12 s retention period was followed by a second cue, a picture of a mouth (MOUTH; 2 s duration), which instructed subjects to say aloud the words that they had been rehearsing. The subjects' responses were recorded with a Phone-OR optical microphone placed inside the head coil approximately 1 in above the subject's mouth. On MEMORY trials, 13 s intervened between the onset of the MOUTH cue and the start of the next trial. On NO-MEMORY trials, 13 s intervened between the onset of the picture of hands and the start of the next trial. No response was required on NO-MEMORY trials. A required response in the NO-MEMORY trials might have induced subvocal preparatory activity during the delay period, affecting our ability to detect activity related to rehearsal processes in the MEMORY conditions. Trial type was pseudorandomized with the restriction that within each scanning run, there were four trials in each condition (for 12 total trials). Each run lasted 6 min, and most subjects (12 of 16) were scanned over eight runs (for a total scanning time of 48 min), netting 32 trials per condition (96 total trials). The other four subjects completed six runs, 24 trials per condition (72 total trials).

Stimuli

All words in the experiment were taken from a set of 92 common nouns (mean number of syllables = 1.94, range = 1–5; mean number of phonemes = 4.94, range = 2–11; Francis-Kucera written frequency: median = 37 per million, range = 1–312 per million; concreteness = 366–662, median = 602; all data were taken from the MRC psycholinguistic database, <http://www.psy.uwa.edu.au/mrcdatabase/mrc2.html>). A male speaker with Cool Edit Pro 2000 made audio recordings of each word. The mean duration of each word was 788 ms (range = 480–960 ms). In 96 trials (over eight experimental runs), 480 words were presented, and each member of the 90 word set was repeated an average of 5.33 times (SD = 1.19). There were no statistically significant differences between the distributions of words across conditions on any of the following dimensions: written frequency, concreteness, number of phonemes, and syllables.

MRI Data Acquisition

Functional and structural images were acquired with a 3.0 Tesla GE Signa scanner (Milwaukee, WI) with a GE birdcage head coil. Each subject performed either six (four subjects) or eight (12 subjects) scanning runs, each of which lasted 340 s. Functional images were collected with a gradient echo echoplanar imaging sequence (TR = 2 s; TE = 25 ms; FOV = 24 cm; flip angle = 90; 64 × 64 matrix). Image volumes were acquired in 24 axial slices (thickness = 5 mm; in-plane resolution = 3.75 × 3.75 mm). In addition, high-resolution MP-RAGE structural images were acquired in 124 axial slices (thickness = 1.2 mm, in-plane resolution = 0.975 × 0.975 mm). The experimental paradigm was programmed with Presentation software version 5.5 (Neurobehavioral Systems) and ran on a Dell laptop. Visual stimuli were rear projected onto a translucent screen outside the bore of the magnet and viewed via a mirror system attached to the head coil. Auditory stimuli were delivered via air conductance tubes

connected to magnet-safe headphones (Avotec model SS-3100) placed around the subject's ears.

Data Preprocessing

The images of every scanning run were concatenated to form, for each subject, a set of four-dimensional (ANALYZE 7.5) data files. Within-run motion was corrected with MCFLIRT (Jenkinson et al., 2002), in which the middle volume (the 90th image) served as the registration reference. After motion correction, the mean volume for each run was aligned with FLIRT (Jenkinson et al., 2002) to the mean volume of the first run, and the resulting transformation matrices were used to coregister all of the four-dimensional fMRI volumes. For each time series within a run, low-frequency noise was modeled by a locally (with a window of 50 time points) weighted polynomial regression and then subtracted from the data (Loader, 1999). Low-pass temporal smoothing was carried out with natural smoothing splines (with 68 degrees of freedom) that reduce sharp scan-to-scan deviations by imposing roughness penalties on the second derivative (Ramsay and Silverman, 1997).

Cortical-Surface Generation and Intersubject Registration

Freesurfer version 0.9 (Dale et al., 1999; Fischl et al., 1999) was used to segment gray and white matter and create smooth white matter and pial surface representations from the high-resolution anatomical images of each participant. These surface representations consisted of large numbers of points, or nodes, typically 150,000, connected in a triangular mesh. Each individual's surface mesh was then inflated to a sphere and registered to a spherical template representing the average sulcal and gyral curvature across a sample of normal brains. The program Maplesahedron from the AFNI/SUMA (<http://afni.nimh.nih.gov/>) analysis package resampled each subject's spherically registered mesh onto a regularly sampled icosahedron to achieve a one-to-one mapping between the nodes of each subject's spherically aligned surface. Volumetric data could then, for each subject, be mapped to this standard regularly sampled surface and statistics computed for every node on the mesh.

Activation Analysis

Because trials of every condition began identically with a stimulus-encoding phase, all relevant differences between the hemodynamic response as a function of experimental condition (EAR, EYE, NO-MEMORY) had to occur after the presentation of the retrieval cue. We predicted, however, that some memory-related areas (e.g., those related to "echoic memory") would show an enhanced response only during the early part of the rehearsal period, whereas other areas would maintain a steady level of activation throughout the rehearsal period. Thus, we modeled the hemodynamic response with four regressors (shown in Figure 2): one for the period of stimulus presentation (ENCODING), one for the early (EARLY) part of the rehearsal period, one for the late (LATE) part of the rehearsal period, and one for the overt response (RESPONSE). Although the ENCODING phase was identical for all conditions of the same load, we chose to model this phase with separate regressors (one for each of the three experimental conditions) so that spurious between-condition differences present before the appearance of the memory instruction could be ruled out as an explanation for differences observed after the memory instruction.

The regressors were generated by convolving a box-car representation of each of the trial phases (ENCODING, EARLY, LATE) with a γ probability density function (shape = 6, rate = 1.5). To reduce collinearity between each of the regressors, we introduced a 2 s "gap" between each of the box-car input functions (see Figure 2). Analyses were performed in the statistical language R (Ihaka and Gentleman, 1996), with the function *gls* for linear models with serially correlated errors (Pinheiro and Bates, 2000). Multiple regression was performed for each subject in the native space of the functional images, and contrasts yielding *t* statistics were computed to detect differences in activation levels for the three retrieval modes (EAR, EYE, NO-MEMORY) during each of the two delay period phases (EARLY, LATE) and item loads (LOAD-2, LOAD-3). Activation during the encoding phase was assessed by contrasting the ENCODING_{NO-MEMORY} regression coefficient against the implicit resting baseline. Differential activation during encoding was assessed by contrasting each of the ENCODING_{MEMORY} coefficients with the ENCODING_{NO-MEMORY} re-

gression coefficients; though this contrast was performed only as a check against spurious precue differences in encoding activation. Activity during overt recall was modeled with a single RESPONSE regressor, and contrasts were performed only with respect to an implicit baseline (as between-condition differences in the RESPONSE phase were not of theoretical interest).

Each *t* statistic image was projected onto the subject's cortical surface representation with the AFNI program 3dVol2Surf. Group *t* statistics with subject as a random effect could then be computed for each node of the regularly sampled average surface template. Previous work in our laboratory has shown that a large proportion of the temporal cortex, especially along the middle and anterior portions of supratemporal plane, deactivates during silent speech production (Buchsbbaum et al., 2005). With spatial smoothing and multi-subject averaging, these deactivating regions can "wash out" the pockets of positive activity typically observed in the posterior PT and STS/STG. Thus, in the region of the STC, statistical significance is defined as any cluster exceeding 50 mm² (a surface area that corresponds to approximately 10–15 contiguous 3.75 × 3.75 × 5 mm voxels) after thresholding the activation map at *p* < 0.005, uncorrected. For all other regions, a standard *p* < 0.001 (uncorrected) threshold is used.

Hypothesis Testing

In this task, all relevant differences that might be ascribed to memory-related processes are reflected in the estimated coefficients for the EARLY and LATE delay-period regressors. Thus, because the retrieval cue (EAR, EYE, NO-MEMORY) occurs after stimulus presentation, the estimated slopes for the ENCODING regressor should, in principle, be identical across all conditions of the same set size. Because there is some blurring of the hemodynamic signal, however, it is possible that memory-related activation could be captured in the coefficient of the ENCODING regressor (Postle et al., 2000). Even so, we would expect strong delay-period activation to be captured primarily by some linear combination of the EARLY and LATE regressors. Dividing the delay period in to two phases allows for the detection of both transient and sustained delay period activity within a single regression model. Thus, to test for memory-related effects, we looked for positive differences between the relevant estimated coefficients (EARLY, LATE) of the MEMORY conditions (EAR, EYE) and the corresponding coefficients for the NO-MEMORY condition. By adding a no-memory condition that requires encoding but not retrieval/rehearsal, we were able to test explicitly whether responses measured during the delay period of trials requiring memory differed significantly from the delay-period signal detected during trials that did not require memory. In the analyses reported herein, delay-period activation is expressed as the difference between the EARLY and LATE coefficients in the two MEMORY (EAR, EYE) conditions and the corresponding coefficients in the NO-MEMORY condition. Thus, unless otherwise noted, the contrast referred to by, for instance, EAR_{EARLY} is equivalent to the subtraction EAR_{EARLY} > NO-MEMORY_{EARLY}.

Functional Connectivity

Two regions of interest, one ventral (STG/STS) defined on the EAR_{EARLY} > NO-MEMORY_{EARLY} surface contrast and one dorsal (Spt) defined on the EAR_{LATE} > NO-MEMORY_{LATE} surface contrast, were used as seed regions for an analysis of functional connectivity. Each of the seed regions was defined by the cluster extent of connected supra-threshold nodes on the group-averaged surface template. As these surface clusters were defined on a spherically registered average surface template, they were first back-projected (with AFNI program 3dSurf2Vol) onto each subject's native fMRI volumetric space. For the correlational analysis, we used the "β-series" technique described in Rissman et al. (2004). In this method, the raw time series are substituted with regression coefficients that are computed trial by trial, so that one can assess correlated variations in amplitude that directly relate to the independent variables used in the standard activation analysis. In our case, we used β coefficients for the EARLY delay period because (1) this was the trial phase of primary interest and (2) β coefficients for different phases of the same trial are correlated, and, therefore, restricting the analysis to use only one β coefficient per trial allows for the assumption of independence between the successive (across-trial) coefficients.

In addition, only MEMORY trials were used for the β series correlation. So, for each seed cluster, the average β series was computed for each subject. Differences attributable to MEMORY condition (EAR, EYE) or LOAD (2 items, 3 items) were regressed out of the seed and nonseed β series (otherwise correlations between areas could be attributed merely to differential activation across task conditions). Correlation coefficients were then computed separately, voxel-by-voxel, for both seed clusters, yielding one correlation image map for the STG/STS and one for area Spt in each subject. These volumes were then projected back onto each subject's surface, and random effects group statistics were performed on the raw correlation coefficients. To assess differential dorsal versus ventral connectivity, we computed paired *t* tests on the nodes of the standard surface and thresholded for significance ($p < 0.001$, uncorrected) and display ($p < 0.05$ to $p < 0.0001$, uncorrected).

Supplemental Data

The Supplemental Data for this article can be found online at <http://www.neuron.org/cgi/content/full/48/4/687/DC1/>.

Acknowledgments

We thank Aaron Bonner-Jackson and Deepak Sarpal for research assistance; and Philip Kohn, Shane Kippenhan, Dylan Wint, and Andreas Meyer-Lindenberg for helpful discussions. This work was supported by the Department of Health and Human Services/National Institutes of Health/National Institute of Mental Health/Intramural Research Program.

Received: May 5, 2005

Revised: August 17, 2005

Accepted: September 22, 2005

Published: November 22, 2005

References

- Arnott, S.R., Binns, M.A., Grady, C.L., and Alain, C. (2004). Assessing the auditory dual-pathway model in humans. *Neuroimage* 22, 401–408.
- Awh, E., Jonides, J., Smith, E.E., Schumacher, E.H., Koeppel, R.A., and Katz, S. (1996). Dissociation of storage and rehearsal in working memory: PET evidence. *Psychol. Sci.* 7, 25–31.
- Baddeley, A.D. (1986). *Working Memory* (Oxford, Oxfordshire: Clarendon Press).
- Baddeley, A.D., and Hitch, G.J. (1974). Working memory. In *The Psychology of Learning and Motivation*, G. Bower, ed. (New York, NY: Academic Press), pp. 47–90.
- Beauchamp, M.S. (2005). See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. *Curr. Opin. Neurobiol.* 15, 145–153.
- Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., and Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature* 403, 309–312.
- Broadbent, D.E. (1956). Successive responses to simultaneous stimuli. *Q. J. Exp. Psychol.* 8, 145–152.
- Buchsbaum, B., Hickok, G., and Humphries, C. (2001). Role of left superior temporal gyrus in phonological processing for speech perception and production. *Cogn. Sci.* 25, 663–678.
- Buchsbaum, B.R., Olsen, R.K., Koch, P.F., Kohn, P., Kippenhan, J.S., and Berman, K.F. (2005). Reading, hearing, and the planum temporale. *Neuroimage* 24, 444–454.
- Buckner, R.L., Koutstaal, W., Schacter, D.L., Wagner, A.D., and Rosen, B.R. (1998). Functional-anatomic study of episodic retrieval using fMRI. I. Retrieval effort versus retrieval success. *Neuroimage* 7, 151–162.
- Calvert, G.A., Campbell, R., and Brammer, M.J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr. Biol.* 10, 649–657.
- Corkin, S. (2002). What's new with the amnesic patient H.M.? *Nat. Rev. Neurosci.* 3, 153–160.
- Cowan, N. (1984). On short and long auditory stores. *Psychol. Bull.* 96, 341–370.
- Crowder, R.G., and Morton, J. (1969). Precategorical Acoustic Storage. *Percept. Psychophys.* 5, 365–373.
- Dale, A.M., Fischl, B., and Sereno, M.I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9, 179–194.
- Dobbins, I.G., Foley, H., Schacter, D.L., and Wagner, A.D. (2002). Executive control during episodic retrieval: multiple prefrontal processes subservise source memory. *Neuron* 35, 989–996.
- Wernicke, C. (1874). Der aphasische symptomkomplex: eine psychologische studie auf anatomischer basis. In *Wernicke's Works on Aphasia: A Sourcebook and Review*, G.H. Eggbert, ed. (The Hague: Mouton), pp. 91–145.
- Fiez, J.A., Raife, E.A., Balota, D.A., Schwarz, J.P., Raichle, M.E., and Petersen, S.E. (1996). A positron emission tomography study of the short-term maintenance of verbal information. *J. Neurosci.* 16, 808–822.
- Fischl, B., Sereno, M.I., and Dale, A.M. (1999). Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage* 9, 195–207.
- Frankish, C. (1989). Perceptual Organization and Precategorical Acoustic Storage. *J. Exp. Psychol. Learn. Mem. Cogn.* 15, 469–479.
- Hickok, G., and Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *J. Cogn. Neurosci.* 4, 131–138.
- Hickok, G., and Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92, 67–99.
- Hickok, G., Buchsbaum, B., Humphries, C., and Muftuler, T. (2003). Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *J. Cogn. Neurosci.* 15, 673–682.
- Ihaka, R., and Gentleman, R. (1996). R: a language for data analysis and graphics. *J. Comput. Graph. Statist.* 5, 299–314.
- Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17, 825–841.
- Jonides, J., Schumacher, E.H., Smith, E.E., Koeppel, R.A., Awh, E., Reuter-Lorenz, P.A., Marshuetz, C., and Willis, C.R. (1998). The role of parietal cortex in verbal working memory. *J. Neurosci.* 18, 5026–5034.
- Kraus, N., and Nicol, T. (2005). Brainstem origins for cortical 'what' and 'where' pathways in the auditory system. *Trends Neurosci.* 28, 176–181.
- Levelt, W.J.M., Roelofs, A., and Meyer, A.S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–38.
- Loader, C. (1999). *Local Regression and Likelihood* (New York, NY: Springer).
- Lu, Z.L., Williamson, S.J., and Kaufman, L. (1992a). Behavioral lifetime of human auditory sensory memory predicted by physiological measures. *Science* 258, 1668–1670.
- Lu, Z.L., Williamson, S.J., and Kaufman, L. (1992b). Human auditory primary and association cortex have differing lifetimes for activation traces. *Brain Res.* 572, 236–241.
- Martin, N., and Saffran, E.M. (2002). The relationship of input and output phonological processing: an evaluation of models and evidence to support them. *Aphasiology* 16, 107–150.
- Milner, A.D., and Goodale, M.A. (1995). *The Visual Brain in Action* (Oxford: Oxford University Press).
- Mustovic, H., Scheffler, K., Di Salle, F., Esposito, F., Neuhaus, J.G., Hennig, J., and Seifritz, E. (2003). Temporal integration of sequential auditory events: silent period in sound pattern activates human planum temporale. *Neuroimage* 20, 429–434.
- Naatanen, R., Jacobsen, T., and Winkler, I. (2005). Memory-based or afferent processes in mismatch negativity (MMN): a review of the evidence. *Psychophysiology* 42, 25–32.
- Paulesu, E., Frith, C.D., and Frackowiak, R.S. (1993). The neural correlates of the verbal component of working memory. *Nature* 362, 342–345.

- Penney, C.G. (1989). Modality effects and the structure of short-term verbal memory. *Mem. Cognit.* 17, 398–422.
- Pinheiro, J.C., and Bates, D.M. (2000). *Mixed-Effects Models in S and S-Plus* (New York, NY: Springer).
- Postle, B.R., Berger, J.S., and D'Esposito, M. (1999). Functional neuroanatomical double dissociation of mnemonic and executive control processes contributing to working memory performance. *Proc. Natl. Acad. Sci. USA* 96, 12959–12964.
- Postle, B.R., Zarahn, E., and D'Esposito, M. (2000). Using event-related fMRI to assess delay-period activity during performance of spatial and nonspatial working memory tasks. *Brain Res. Brain Res. Protoc.* 5, 57–66.
- Postle, B.R., Druzgal, T.J., and D'Esposito, M. (2003). Seeking the neural substrates of visual working memory storage. *Cortex* 39, 927–946.
- Ramsay, J.O., and Silverman, B.O. (1997). *Functional Data Analysis* (New York, NY: Springer).
- Rauschecker, J.P., and Tian, B. (2000). Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proc. Natl. Acad. Sci. USA* 97, 11800–11806.
- Rissman, J., Gazzaley, A., and D'Esposito, M. (2004). Measuring functional connectivity during distinct stages of a cognitive task. *Neuroimage* 23, 752–763.
- Romanski, L.M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P.S., and Rauschecker, J.P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat. Neurosci.* 2, 1131–1136.
- Ruchkin, D.S., Grafman, J., Cameron, K., and Berndt, R.S. (2003). Working memory retention systems: a state of activated long-term memory. *Behav. Brain Sci.* 26, 709–728.
- Sakai, K., and Passingham, R.E. (2003). Prefrontal interactions reflect future task operations. *Nat. Neurosci.* 6, 75–81.
- Schubotz, R.I., and von Cramon, D.Y. (2004). Sequences of abstract nonbiological stimuli share ventral premotor cortex with action observation and imagery. *J. Neurosci.* 24, 5467–5474.
- Schubotz, R.I., von Cramon, D.Y., and Lohmann, G. (2003). Auditory what, where, and when: a sensory somatotopy in lateral premotor cortex. *Neuroimage* 20, 173–185.
- Scott, S.K., and Wise, R.J. (2004). The functional neuroanatomy of prelexical processing in speech perception. *Cognition* 92, 13–45.
- Scott, S.K., Blank, C.C., Rosen, S., and Wise, R.J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123, 2400–2406.
- Shallice, T., and Warrington, E.K. (1977). Auditory-verbal short-term-memory impairment and conduction aphasia. *Brain Lang.* 4, 479–491.
- Shiffrin, R.M., and Atkinson, R.C. (1969). Storage and retrieval processes in long-term memory. *Psychol. Rev.* 76, 179–193.
- Stevens, A.A. (2004). Dissociating the cortical basis of memory for voices, words and tones. *Brain Res. Cogn. Brain Res.* 18, 162–171.
- Takayama, Y., Kinomoto, K., and Nakamura, K. (2004). Selective impairment of the auditory-verbal short-term memory due to a lesion of the superior temporal gyrus. *Eur. Neurol.* 51, 115–117.
- Ungerleider, L., and Mishkin, M. (1982). Two cortical visual systems. In *Analysis of Visual Behavior*, D. Ingle, M.A. Goodale, and R.J.W. Mansfield, eds. (Cambridge, MA: MIT Press), pp. 549–586.
- Vouloumanos, A., Kiehl, K.A., Werker, J.F., and Liddle, P.F. (2001). Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. *J. Cogn. Neurosci.* 13, 994–1005.
- Warrington, E., and Shallice, T. (1969). Selective impairment of auditory verbal short-term memory. *Brain* 92, 885–896.
- Watkins, M.J., and Watkins, O.C. (1973). The postcategorical status of the modality effect. *J. Exp. Psychol.* 99, 226–230.
- Watkins, O.C., and Watkins, M.J. (1980). The modality effect and echoic persistence. *J. Exp. Psychol. Gen.* 109, 251–278.
- Wilson, M. (2001). Perceiving imitable stimuli: consequences of isomorphism between input and output. *Psychol. Bull.* 127, 543–553.
- Winkler, I., and Cowan, N. (2005). From sensory to long-term memory: evidence from auditory memory reactivation studies. *Exp. Psychol.* 52, 3–20.
- Wise, R.J., Scott, S.K., Blank, S.C., Mummery, C.J., Murphy, K., and Warburton, E.A. (2001). Separate neural subsystems within 'Wernicke's area'. *Brain* 124, 83–95.