Manipulating stored phonological input during verbal working memory

Gregory B. Cogan[1], **Asha Iyer**[2], **Lucia Melloni**[3,5], **Thomas Thesen**[3], **Daniel Friedman**[3], **Werner Doyle**[4], **Orrin Devinsky**[3,4], and **Bijan Pesaran**[1,*]

[1]Center for Neural Science, New York University, New York, NY USA

[2]Icahn School of Medicine at Mount Sinai, New York, NY USA

[3]Department of Neurology, NYU School of Medicine, New York, NY USA

[4]Department of Neurosurgery, NYU School of Medicine, New York, NY USA

[5]Department of Neurophysiology, Max-Planck Institute for Brain Research, Frankfurt, Germany

Abstract

Verbal working memory (vWM), involves storing and manipulating information in phonological sensory input. An influential theory of vWM proposes that manipulation is carried out by a central executive while storage is performed by two interacting systems: A phonological input buffer that captures sound-based information and an articulatory rehearsal system that controls speech motor output. Whether, when, and how neural activity in the brain encodes these components remains unknown. Here, we read-out the contents of vWM from neural activity in human subjects as they manipulate stored speech sounds. As predicted, we identify storage systems that contain both phonological sensory and articulatory motor representations. Surprisingly however, we find that manipulation does not involve a single central executive but rather involves two systems with distinct contributions to successful manipulation. We propose, therefore, that multiple subsystems comprise the central executive needed to manipulate stored phonological input for articulatory motor output in vWM.

Introduction

The ability to hold short term information in working memory underlies a range of core cognitive functions including reasoning, planning, and creative thinking[1,2]. Verbal working

memory (vWM) involves both holding and manipulating speech-specific content so that it can be accessed by higher-order cognitive functions[3]. These can include, but are not limited to, operations necessary to support language such as syntactic/semantic operations[4–7], and vocabulary acquisition during development[2,8]. Patients with vWM impairments can have difficulty with syntactic operations that involve maintaining verbal items in vWM until they can be resolved. For instance, in sentences with center embedded clauses such as 'The man that the women knew liked cake', patients would have difficulty determining who liked cake[4,9]. Patients with vWM deficits can also show difficulty learning a foreign language as holding the phonological representation in vWM cannot occur[10]. Intriguingly, when machines process natural language input they also display difficulty with syntactic operations[11], which could be due to the architecture of their working memory systems[12,13], and is an area in which rapid progress is being made[14].

The dominant view is that storage and manipulation are modular cognitive operations. Storing items in vWM involves both a sensory and a motor component. Baddeley's influential phonological loop model for instance, postulates that storage involves an auditory-based phonological short term store that decays over time, and a motor-based phonological output buffer that supports the maintenance of the phonological short term store via active motor-based rehearsal[2,15,16]. Other models similarly advocate that the storage/maintenance of vWM involves a tight interplay between separate phonological components for auditory input and production output[17–21]. This modality-specific store is proposed to interact with a second module, the central executive system[2,16]. Manipulation of the modality-specific store by the central executive is necessary to support more abstract cognitive processing[2,22–25]. Despite broad acceptance of this view, the neural architecture of vWM remains unknown. Evidence for vWM modules comes primarily from patients with brain lesions that reveal patterns of specific behavioral impairments[26]. In contrast, non-invasive functional activations reported by hemodynamic studies have not been able to unambiguously assign a specific modular architecture to vWM[27–29]. In particular, the neural computations performed by the working memory modules, the timing of those computations, and their role in successful performance, remains unresolved.

To more clearly resolve the modular architecture of vWM, we measured electrocorticography (ECoG) directly from the surface of the brain while subjects performed a match-mismatch task requiring a remembered sensory-motor manipulation. The match-mismatch task required subjects to manipulate speech items in vWM by varying the mapping from an auditory sensory input to a speech motor output according to an abstract rule. This allowed us to define and separately examine storage and manipulation processes. Our results support previous models of storage: We found distinct sensory input and output processes, akin to a phonological input buffer and motor rehearsal components, respectively [16]. Surprisingly, however, we found evidence for two manipulation processes. These manipulation processes operated concurrently but encoded information differently and predicted errors in task performance differently. These results divide vWM into multiple subsystems for storage and manipulation.

Results

We recorded subdural electrocorticographic (ECoG) neural signals from 8 patients (3 males; mean age 30 years, see Table S1) undergoing treatment for pharmacologically-resistant epilepsy as they completed a match-mismatch task. In this task, the incoming speech stimuli consisted of two non-words ('kig' and 'pob') that were chosen to differ maximally on their articulatory dimensions [30]. Subjects received a visual cue 1.5 seconds prior to an auditory presentation of a non-word that instructed them that after a variable delay (1.5 – 2 s), they were to follow one or the other rules. On Match trials, they were to say the non-word they had been presented with (Match - 'kig' to 'kig' and 'pob' to 'pob' – Fig 1a) after the go cue. On Mismatch trials, they were to say the non-word they hadn't been presented with (Mismatch – 'kig' to 'pob' and 'pob' to 'kig' – Fig 1b). By interleaving match and mismatch conditions trial-by-trial (see Fig 1a,b, **Experimental Procedures**), we could assess the maintenance and manipulation processes in working memory. The match-mismatch task is similar in spirit to anti-saccade tasks[31] and delayed pair association tasks[32,33]. Despite the difficulty of the task, subjects performed the task correctly on 83% of trials. Errors were predominantly errors of omission in which subjects did not speak after the go cue (11%) instead of saying the incorrect non-word (6%; Table S2).

Characterizing persistent, delay activity

A prominent feature of neural activity during WM is sustained, persistent activity during a remembered delay. To identify such persistent activity, we first performed a permutation test using the high gamma band neural response (70–160 Hz). Activity in this frequency range is associated with both the BOLD response measured non-invasively, and multi-unit responses measured invasively [34,35]. We compared activity during the baseline period (−2000 ms to −1500 ms pre Auditory Onset – see Fig 1a,b,) with activity during the delay period (1000 ms – 1500 ms post Auditory Onset - see Fig 1a,b). Significant delay activity (p< 0.05 FDR-corrected permutation test, p-threshold = 0.01) was present for 7/8 Subjects (36 total electrodes, range: 1 – 9 delay electrodes per subject, mean = 5, see Fig 1c,d).

To isolate the underlying modules that may compose vWM, we used a functional localizer repetition task to isolate speech perception from production according to three response profiles (see **Experimental Procedures**, Cogan et al. 2014[36]). Auditory responses (green electrodes; 29 total electrodes: 6 LH, 23 RH) were predominately located in the temporal cortex (21/29, 72%). Production responses (blue electrodes; 43 total electrodes:20 LH, 23 RH) were predominately located in the pre- and post-central gyri and the parietal lobe (31/43, 63%). Sensory-motor responses (red electrodes; 18 electrodes: 11 LH, 7 RH) were located in multiple sites.

Since robust delay activity was generally not present during the repetition task, we assessed delay activity on the match-mismatch task, which presumably places a larger demand on vWM. We observed persistent delay activity in the responses of all three functional classes as well as in an additional 10 delay responses localized to the prefrontal cortex that we term "Delay-Only" that were not associated with any of the auditory, production and sensory-motor functional classes described above (Fig 2a,b, Fig S3). The response profiles were located in the same regions of the brain as their non-delay counterparts (6 Delay+Auditory

electrodes, 9 Delay+Production electrodes, and 10 Delay+Sensory-Motor electrodes; see Fig 3a,b, Fig S1). The remaining Delay-Only electrodes did not exhibit explicit sensory or motor properties suggesting a role in more abstract functioning. These results, therefore, suggest that vWM may be composed of at least four sub-components: Auditory, Production, Sensory-Motor, and Abstract.

Delay Representation

While robust neural delay activity is present in each sub-system, it is unclear how each of the electrode classes represents information during the vWM task. We hypothesized that each electrode class would represent information during the sensory and motor epoch reflective of their functional localization[36]. Specifically, the Delay+Auditory electrodes should demonstrate a sensory response that differentiates between the two sensory inputs regardless of the motor outputs (i.e. grouping 'kig' to 'kig' with 'kig' to 'pop' and 'pob' to 'pob' with 'pob' to 'kig' – see Fig 4a - top row). The Delay+Production electrodes on the other hand, should demonstrate responses during the motor epoch that differentiate between the motor responses (i.e. grouping 'kig' to 'kig' with 'pob' to 'kig' and 'pob' to 'pob' with 'kig' to 'pob' – see Fig 4a - second row). The Delay+Sensory-Motor should demonstrate both a response in the sensory epoch and a response in the motor epoch, following sensory-motor transformations (see Fig 4a third row). Lastly, if the prefrontal cortex electrodes are representing the rule, then their neural response should reflect a differentiation between the match ('kig' to 'kig' and 'pob' to 'pob') and mismatch ('kig' to 'pob' and 'pob' to 'kig') conditions (see Fig 4a - bottom row).

The neural responses in the sensory and motor epochs qualitatively followed the predictions. The Delay+Auditory electrodes appeared to represent information in sensory space (Fig 4b - top row), the Delay+Production electrodes appeared to represent information in motor space during the utterance (Fig 4b - second row). The Delay+Sensory-Motor electrodes displayed both sensory and motor responses (Fig 4b - third row), and the Delay-Only electrodes appeared to represent the rule. What remains unclear however, is how this information is represented in vWM and how this information evolves over time.

To assess the information present in each sub-class of neural delay activity, we trained a linear classifier on the singular value decomposition (SVD) of the high gamma response (70–160 Hz). Using a leave-one-out validation method, we decoded the neural responses of each of the electrode categories as they evolved in time (see **Experimental Procedures**). The vWM information during the task can be encoded in at least four different kinds of classifier responses (see Fig 5a): A sensory classifier response would confuse trials in which the incoming sensory signal is the same regardless of the output ('kig' to 'kig' with 'kig' to 'pob' and 'pob' to 'pob' with 'pob' to 'kig' – leftmost model Fig 5a), whereas a motor classifier response would confuse trials with the same motor output ('kig' to 'kig' with 'pob' to 'kig' and 'pob' to 'pob' with 'kig' to 'pob' – second model Fig 5a). A transformation classifier response would track both the incoming sensory signal and the motor output, forming an intermediate response that reflects both sensory and motor properties (third model – Fig 5a). Lastly, a rule-based classifier response would track the abstract rule and confuse trials within either 'match' or 'mismatch' conditions regardless of the token

presented (see Fig 1a, Last model - Fig 5a). Analyzing these four response types, therefore, allowed us to decode the information present in different vWM components.

Each response model had support and each sub-system represented vWM differently (Fig 5b). To quantify how well each model described the responses, we performed a modified KL-analysis on the time-course of each vWM component (see **Experimental Procedures**). The Delay-Auditory electrodes display delay representations consistent with their auditory representations (see Fig 5c). During the delay, Delay-Production responses also encode the upcoming motor act (see Fig 5c). These responses differed from auditory and production electrodes that displayed no delay activity and no discernable representational information during the delay (see Fig S4). These responses are also insensitive to the manipulation demands of the task. Therefore, we propose these two response profiles demonstrate the storage components of vWM in that they store incoming or outgoing information during the delay similar to the phonological buffer and motor rehearsal stages [2,16].

We also observed responses that were sensitive to the manipulation demands of the task. The Delay+Sensory-Motor responses displayed a representation that evolved in time and changed from an auditory, followed by a sensory-motor, and finally to production representation, consistent with a transformation of information required by the task (Fig 5c). This dynamic transformation contrasts with the static representations of both the Delay +Auditory and Delay+Production responses (as well as the Sensory-Motor electrodes with no delay activity – see Fig S3). The results indicate that the Delay+Sensory-Motor system reflects a vWM component that links perceptual and motor delay processes. Delay-Only responses demonstrated a different pattern and represented the abstract rule presented on each trial (i.e. 'match' vs. 'mismatch' – See Fig 5c). Interestingly, this rule representation was not present during the early auditory period (see Fig 5c), or the Cue period (data not shown), indicating that this representation was not driven by the visual presentation of the cue itself. Instead, the timing of the rule representation coincided with the epoch just prior to and during the transformation in the Delay+Sensory-Motor responses during which time information in vWM was likely being manipulated (see Fig 5c). This suggests that there is a tight interplay between manipulation systems. This result supports theories that posit that the prefrontal cortex is involved in manipulation of items as opposed to their storage [37–39].

Error Analysis

We next sought to elucidate the storage and manipulation components required for vWM by analyzing neural responses on trials in which subjects failed to complete the task correctly. Most error trials were ones in which the subject did not register a response (see Table S2). We trained a 2-way linear classifier on the SVD of the high gamma neural activity in each of the delay electrode types (Delay+Auditory, Delay+Production, and Delay+Sensory-Motor, and Delay-Only) on trials in which the task was performed correctly and the error trials in which no response was made. It is important to note that during the task, subjects were being closely monitored for task engagement. Consequently, errors or omissions during task completion likely reflect internal rather than external factors.

Analyzing errors let us separate out the processes associated with perception from those due to vWM production because subjects heard the sounds but did not speak. As expected, both

the Delay+Sensory-Motor and the Delay+Production electrodes could successfully decode utterances vs no utterances as measured by FDR-corrected chi-square tests (Delay +Production – Early Motor epoch $\chi(1) = 6.54$, Delay+Sensory-Motor – Early Motor epoch $\chi(1) = 14.04$, p threshold = 0.02, see Fig 6). Furthermore, during the delay, all four electrode classes could distinguish between responses and no responses, (Delay-Auditory $\chi(1) = 7.8$, Delay+Production $\chi(1) = 5.98$, Delay+Sensory-Motor $\chi(1) = 5.42$, Delay-Only $\chi(1) = 8.13$, p threshold = 0.02, see Fig 6), This suggests that on trials when subjects made errors, the incoming sensory stimuli was not entering the vWM sub-components.

To more directly test whether putative modules for manipulation are distinct, we asked when during each trial we could predict a failure of vWM task performance. In the sensory epoch, activity in the Delay-SM was the only electrode class able to distinguish between successful and error trials ($\chi(1) = 7.72$, p threshold = 0.02, see Fig 6). Successful decoding was also present in the cue period, and was present in both manipulation systems - the Delay +Sensory-Motor and the Delay-Only ($\chi(1) = 11.71$, $\chi(1) = 8.69$, respectively, p threshold = 0.02, see Fig 6). Even during the baseline before the cue had been presented, neural activity for both the Delay+Sensory-Motor and Delay-Only electrodes could predict the success or failure of an upcoming utterance ($\chi(1) = 8.07$, $\chi(1) = 7.04$, respectively, p threshold = 0.02). Thus both manipulation systems, but not the storage systems, demonstrate an early role in the ability to successfully manipulate remembered phonological input, and each manipulation system represents a distinct modularity as shown by the different error patterns.

Discussion

Here we decoded the contents of vWM using neural recordings from the brain surface of patients performing a match-mismatch task. We co-localized four different response types involving persistent activity. These responses were associated with speech perception, production, sensory-motor transformation and rule-related processing. Storage appeared to be supported by persistent perceptual and production-related processes. By contrast, manipulation appeared to be performed by distinct sensory-motor transformation and abstract rule processes. One manipulation process in the prefrontal cortex encoded the abstract rule independently of the stimulus. Interestingly, this response was engaged only after the presentation of the stimulus, suggesting an active role in manipulation. The second manipulation process was reflected in a sensory-motor transformation system whose activity appeared to track the transformation of information from sensory to motor representations. Activity in both the storage (phonological input buffer and motor rehearsal) and manipulation (abstract and sensory-motor) systems predicted errors during the delay period. However, only the manipulation processes distinguished between utterances and non-utterances before the phonological input had been presented. These results support earlier work proposing that vWM is composed of multiple storage processes and extend this to demonstrate that at least two separable components also comprise manipulation.

Components of storage functions

We found two kinds of persistent activity whose properties were consistent with a storage function. During the delay, the input properties of the phonological stimulus were statically encoded by the Delay+Auditory responses and the upcoming utterance was statically encoded by the Delay+Production responses. Since these representations were static and did not change in time, the manipulation of phonological input to guide the response could not be directly attributed to these responses. Furthermore, the responses on error trials revealed storage functions: both Delay+Auditory and Delay+Production responses clearly predicted failures to respond during the delay, consistent with a maintenance role. Interestingly, Delay +Auditory responses did not distinguish between successful and unsuccessful trials during the sensory period, suggesting that the initial sensory component of this system passively tracked the incoming auditory input. Also, not surprisingly, the Delay+Production responses clearly revealed the failure to respond following the go cue consistent the lack of maintenance of a motor plan.

These results are consistent with previous models that posit that storage is composed of a phonological input buffer and a motor rehearsal system[2,16]. The auditory input buffer is reflected in the Delay+Auditory electrodes in which the information is encoded in the incoming phonological space. The Delay+Production electrodes reflect the rehearsal system that could underlie the cycling of the motor output plan. Taken together, therefore, we propose that these two components appear to maintain static input and output, respectively, and so are consistent with a storage system.

Components of verbal manipulation: rule processing and sensory-motor processing

Manipulation also appeared to be carried out by two separate systems: Delay-Only and Delay+Sensory-Motor, but these responses appeared to be distinct from the maintenance of incoming phonological input and motor output.

The Delay-Only responses, localized to the prefrontal cortex appeared to reflect the application of the abstract rule required to perform the necessary transformation during the delayed-sensory motor task. These responses appeared in an abstract system. They do not encode the identity of the phonological input nor the motor response, and cannot reflect phonological input or motor rehearsal. Neural activity present in this abstract response encodes the rule and so may, instead reflect storage of the rule. The presence of activity at these prefrontal sites that predicts failures to respond during the cue delay period is consistent with a failure on those trials associated with storage of the rule. We also propose that the prefrontal cortex responses are not involved in rule storage alone due to the temporal properties of its coding. When the initial visual stimulus which indicates what condition the subjects are in (see Fig 1) is presented, prefrontal activity demonstrates discernable rule selectivity neither during the early auditory period (see Fig 5c) nor during the initial Cue period (data not shown). These temporal properties mean that the Delay-Only responses do not maintain a static representation of the rule following the visual cue, and so cannot be interpreted as solely due to storage of the rule. Furthermore, the prediction of the failure to

respond is not present in the interval between the initial cue encoding and the delay activity. This also indicates that a failure to respond begins as a failure to initially encode the rule. We propose that the Delay-Only responses appear to be involved in processing necessary to support manipulation: they do not encode the rule until after the early auditory period. The rule encoding is present only when phonological input is being manipulated to support subsequent motor output.

The sensory-motor responses also appear to be best considered as a component of the manipulation process. Persistent activity in the Delay+Sensory-Motor responses displayed dynamic coding properties from sensory, to sensory-motor and finally to motor-based representations. The dynamic coding properties indicate that these responses are not simply involved in maintenance because they do not strictly store either incoming or outgoing information. The sensory-motor system has properties necessary to link the phonological input buffer to the motor rehearsal system through a transformation in representational coordinates from sensory to sensory-motor and finally to motor. Interestingly, the onset of rule selectivity in the Delay-Only responses coincides with the initial change in representation present in the sensory-motor system.

Furthermore, the successful classification of failure to respond in the sensory epoch of the Delay+Sensory-Motor electrodes indicates that the sensory-motor system is active rather than passive, and the processing of the incoming information is not automatic. A striking aspect of the Delay+Sensory-Motor responses during trials in which there is a failure to respond is that they can be distinguished throughout the trial, even before the phonological input. This is not true for any other response profile and indicates that the sensory-motor system is involved in setting up the manipulation of phonological input and does not merely reflect its transformation. The successful classification during the Cue period for both the Delay+Sensory-Motor and abstract rule encoding system argues for privileged role for these two components of the manipulation systems as they initially process the rule.

Storage and manipulation interactions

Our results suggest that both the storage and manipulation components of vWM are active and not a direct/mandatory consequence of sensory processing. This is in contrast to theories that propose that vWM is an automatic process associated with perception[6]. While these results demonstrate activity associated with sensory, motor, and sensory-motor processing, it is still unclear how these systems interact with each other. Baddeley[2,16], for instance, proposed that verbal working memory was composed of two components: an auditory-based phonological short term store that decays over time, and a motor-based phonological output buffer that supports the maintenance of the phonological short term store via rehearsal. The interaction between these two components is therefore the maintenance of the phonological store via a motor-based rehearsal. The sensory-motor system may reflect the transformation of information and is involved in the persistent maintenance of items in vWM. The state of the sensory-motor system, be it phonological input or articulatory output, may be updated according to the activation of the articulatory rehearsal and the phonological buffer. Studying neural activity during longer 30s + delays could assess this putative mechanism.

Input-output buffers

Evidence for the existence of a phonological short-term store comes from three main sources. Behavioral results demonstrate that words with similar phonological structure (e.g. car, cab, cat) are more difficult to maintain than words with dissimilar phonological structure (e.g. hit, bite, rat[40]). Clinically, patients with lesions in the posterior temporal and inferior parietal lobes have demonstrated deficits associated with verbal working memory than cannot be attributed to impaired analysis of incoming sensory information or outgoing articulation[41]. Lastly, functional imaging studies have implicated portions of the parietal[27,29] and posterior temporal cortices[19,42,43] as potential substrates for vWM. Hickok, Buchsbaum, and Humphries[19] have shown that regions in the posterior temporal region are active both during a sensory period and a maintenance period, suggesting a degree of overlap between the sensory – input systems and the working memory-delay activity. This discrete compartmentalization contrasts with models that posit a more distributed localization for the components of vWM[44].

Evidence for the usage of the articulatory system for maintenance comes from behavioral results that indicate a small drop in performance if the articulatory system is being used for something else during the short retention period (articulatory suppression)[15]. What is unclear in this model is how the motor based rehearsal interacts with the phonological store.

Sensory-motor working memory

Recent models have proposed that the sensory-motor system underlies verbal working memory[17,18,20,21,43]. Specifically, transforming a sensory-based coordinate system to a motor-based coordinate system creates the representation of verbal working memory. Jacquemot and Scott [17] propose that verbal working memory operates using three separate sub-systems: a phonological input buffer that converts sensory information into sensory/ phonology representations, a phonological output buffer that converts motor-based phonological information into motor output, and an operation/system to convert between the two phonological systems. The cycling of this conversion underlies vWM. Buchsbaum, Hickok, and Humphries[43] instead posit that that the input and output phonological systems partially overlap in the left inferior parietal lobe/posterior temporal lobe. This overlapping cortex mediates between sensory areas (temporal lobe) and motor areas (inferior frontal and premotor/motor cortex). vWM is thereby mediated by this shared phonological system.

The present work supports elements of both classes of model. We find that sensory systems that display delay activity encode speech elements in sensory space, production delay responses encode information about the upcoming motor act, and the sensory-motor system encodes a dynamic transformative representation from sensory to sensory-motor, and finally to production. Our work therefore supports the Jacquemot and Scott model through the presence of these three systems, as well as the Buchsbaum, Hickok, and Humphries model which proposes that the sensory-motor system underlies the intersection of the input and output phonological systems.

What remains unclear is the degree to which the Delay+Auditory activity and the Delay +Production activity reflect phonological input and phonological output information respectively. While the Delay+Auditory responses display representations that are consistent with incoming sensory information, the level of abstraction within this representation is unclear. The same is true of the Delay+Production responses. It is possible that each of these response categories reflects multiple levels of abstraction and is only bound by their relationship to either the input or the output.

A similar issue relates to the Delay+Sensory-Motor responses. While we demonstrate that there is a translation between sensory and motor based representations during verbal working memory, the exact nature of the coordinate space in the input, output, and transformation is not yet known. As in the auditory and production systems, there may be multiple levels of abstraction bound together by a transformation.

We have analyzed persistent delay period activity to demonstrate a parcellation of vWM. We show that there is an input sensory buffer that maintains information in sensory space as well as an output buffer that represents information in motor space, similar to the canonical models for vWM[2,16]. We also show however, that there are multiple subsystems that comprise the central executive: a sensory-motor based system that is responsible for transforming phonological sensory input into motor output, and a prefrontal-based rule abstract system that encodes higher order rules. Combined, these systems underlie the storage and manipulation of stored phonological input for articulatory motor output in vWM.

A recent study using the same behavioral tasks as the current study demonstrated a bilateral locus for sensory-motor transformations for speech[36]. Our work builds upon and departs from that work by focusing on persistent delay activity to assess the neural organization of vWM. We show that delay responses in the sensory-motor system appear to be crucial for vWM processing and reveal a manipulation process distinct from sensory and motor storage functions. However, since the spatial sampling of ECoG recordings remains sparse, we cannot make strong anatomical statements about the localization of the different components of vWM.

Other aspects of working memory

While we highlight the manipulation and storage systems, other systems may also be important to vWM. A great deal of work supports a role for attention and cognitive control in WM function[46]. How the different components of WM are related and interact remains unclear. For example, the manipulation system we demonstrate shares similarities to cognitive control given the localization to the prefrontal cortex and further work is needed to establish the nature of any link. The results of the error analysis also suggest a role for attention in vWM. Since the neural response could not significantly differentiate between success and error trials during the sensory period but could during the delay period in all electrode classes, vWM could be contingent on attention. This interpretation is further supported by the ability to differentiate successful trials in the cue period, but only for the two classes of electrodes that form the manipulation system. While this latter result suggests

that attention is necessary to encode the rule, the former result suggests that perhaps like the spatial working memory system, vWM and speech/auditory attention share a common neural substrate[47].

We should also emphasize that we explicitly defined vWM activity according to trial-averaged neural responses recorded at the cortical surface. Additional analysis investigating the underlying nature of persistent activity is warranted. In particular, while single neurons can show clear sustained responses on single trials[48], the picture of delay period activity obtained by analyzing single neuron activity on single trials can significantly differ from that obtained by the trial-average response[49,50]. In particular, responses that are transient on single-trials can appear sustained after trial-averaging. Furthermore, since the high-gamma neural response we analyze closely relates to multi-unit activity[35], different single neurons could code different temporal components of vWM. An important topic for future work, therefore, involves analyzing the single-trial properties of delay period responses measured at the cortical surface and, when possible, within the brain. This analysis will be critical to understanding the extent to which the ability of vWM to bridge events in time arises from the persistent dynamics of neural activity and not necessarily sustained responses.

Online Methods

Participants

We obtained electrocortiographic (ECoG) recordings from 8 human patients undergoing clinically motivated subdural recordings for treatment for pharmacologically resistant epilepsy (2 males, mean age 30, range 17–44, see Table S1). Recordings were done at the New York University School of Medicine Comprehensive Epilepsy Center. Informed consent was obtained from each patient in accordance with the Institutional Review Board at the New York University Langone Medical Center. Patients were selected based on neuropsychological testing results and had to be within the normal limits for cognitive and language abilities (see Table S1). Electrodes within the seizures onset zone were not analyzed.

Behavioral Task and Recordings

All behavioral recordings were done via on a computer on the service tray of a hospital bed using Presentation Software (NeuroBehavioral Systems). Audio recordings were obtained using a computer microphone and were synched to the onset of the Go Cue.

All participants completed a match-mismatch task in which the incoming speech stimuli consisted of two non-words ('kig' and 'pob') that were chosen to differ maximally on their articulatory dimensions [30]. Subjects received a visual cue ('Match Listen' or 'Mismatch Listen') 1.5 seconds prior to an auditory presentation of a non-word. After a variable delay (1.5 – 2 s), they were to instructed via another visual cue ('Speak') to either say the non-word they had been presented with (Match - 'kig' to 'kig' and 'pob' to 'pob') or say the non-word they hadn't been presented with (Mismatch – 'kig' to 'pob' and 'pob' to 'kig'). In this way, we could separately assess the sensory, motor, and sensory-motor contributions (see Fig 1a,b). The total number of trials presented ranged from 256 to 309 trials.

To further clarify the role of sensory, motor, and sensory-motor processing and its role in verbal working memory, we had each subject carry out a localizer task in which 7 CVC tokens ('hit','heat','hat','hoot','hot','het','hut') were presented auditorily, and had to be repeated following a short delay. Subjects were presented with a visual cue ('Listen') 1.5 seconds prior to the presentation of a CVC. After a short delay, another visual cue instructed them to either repeat ('Speak'-listen-speak), or mime ('Mime' - move their articulators to mouth the words, but without vocal cord vibrations – listen-mime). Using these two conditions, we were able to ensure that we were measuring the neural activity associated with motor production and not simply the neural activity associated with the auditory processing of the participant's own voice. A third condition contained a different initial visual cue (':=:') which instructed subjects to simply listen to the word (listen). This task was completed in separate blocks between blocks of the main task, and the range of trials presented to each subject was 252 to 315 times. Within each block (localizer and match-mismatch), trials were randomized across conditions.

Neural Recordings and Preprocessing

Electroencephalographic (ECoG) activity was recorded from intracranially implanted subdural electrodes (AdTech Medical Instrument Corp.) in patients undergoing monitoring as part of treatment for pharmacologically resistant epilepsy. Electrode placement was clinically selected to localized seizure activity and eloquent tissue during stimulation mapping. Recordings included grids (8×8 contacts), depth (1×8 contacts), and strip (1×4 to 1×12 contacts) electrode arrays. Each electrode had a diameter of 4 mm (2.3 mm exposure), and the space between electrodes was 6 mm (10 mm center to center). Neural signals were recorded on a 128-channel Nicolet One EEG system with a sampling rate of 512 Hz and bandpass filtered between 0.5 and 250 Hz. Line-noise filtering centered at 60, 120 and 180 Hz was also performed offline.

Electrode localization and Surface Generation

Electrode localization was done by first obtaining both pre- and post-surgical T1-weighted magnetic resonance imaging (MRI) scans from each patient. These images were then co-registered with each other[51] and then normalized to an MNI-152 template. Electrode locations were then extracted in MNI space using the co-registered image. A three-dimensional reconstruction of each participant's brain was used to generate Fig 3 and Fig S2[52].

Data Analysis

Spectral Analysis—Spectral decomposition was carried out using a multi-taper spectral analysis[53] using a 500 ms analysis window with +/– 5 Hz smoothing and 50 ms stepping between spectral estimates. Trials were removed from the analysis if they exceeded 8 standard deviations from the mean across the trial pool. Noisy channels were also removed via visual inspection.

Delay Analysis—Delay activity was analyzed by comparing the high gamma power (70–160 Hz) during the delay period of the match-mismatch task (1 – 1.5 seconds post auditory onset) with the power prior to the visual cue (–2 – –1.5 seconds post auditory onset).

Statistical significance was computed using a 10 000 iteration one-sided permutation test. Significance was corrected to an alpha level of 0.05 using a false discovery rate procedure (FDR − p threshold = 0.01).[54]

Localizer Task and Electrode Categorization—For the localizer test, neural data (70–160 Hz – high gamma band) was analyzed in either the auditory epoch (250 – 750 ms post auditory onset), or production epoch (500–1000 ms post Go Cue) for all three task conditions (listen-speak, listen-mime, and listen) and the same permutation test was used as in the match-mismatch task. Auditory electrodes were defined as electrodes that contained significant activations in the auditory epoch in all three task conditions. Production electrodes were defined as electrodes that contained neural activity in the production epoch in both the Listen-Speak and Listen-Mime conditions. Sensory-motor electrodes were defined as electrodes that contained significant neural activity in both the auditory and production epochs.

The localizer task was used to define the sub-sets of delay activity electrodes. Delay +Auditory electrodes were electrodes that were defined as an auditory electrode using the localizer task and had delay activity present in the match-mismatch task. Likewise, Delay +Production electrodes were defined as being a production electrode in the localizer task and having significant delay activity in the match-mismatch task, and Delay+Sensory-Motor electrodes were defined as being sensory-motor electrodes in the localizer task and having significant delay activity in the match-mismatch task. The remaining significant delay electrodes that were localized to the prefrontal cortex were categorized as Delay-Only.

Representation/Classifier Analysis—To track the temporal evolution of the representations within each class of electrode, we first divided each of the six electrode categories into six time epochs of interest: Early Auditory (0–750ms post Auditory Onset), Late Auditory (750–1000ms post Auditory Onset), Delay (1000–1500ms post Auditory Onset), Late Delay (500–0ms pre Go cue), Early production (0–500ms post Go Cue), and Late Production (500–1000ms r post Go Cue). Since the delay was randomized between 1.5 and 2 s, on some trials, there is some overlap in time between the Early and Late Delay period.

We used a linear discriminant analysis (LDA) classification algorithm performed on the single value decomposition (SVD – 15 components) of the high gamma neural response activity (70 – 160 Hz) for each electrode category, during each time epoch. Validation was performed using a leave-one-out validation method. To avoid over-fitting due to within trial noise correlations, we also shuffled the trial order within our training and testing set. An equal number of trials per condition and electrode was used for the training set. One hundred iterations of classification were performed for each epoch and electrode category. The percentage of trials in which the neural response from an actual condition (vertical axis) is decoded as a particular condition (horizontal axis) generates the confusion matrices, depicted in Fig 5a,b. The four trial conditions ('kig' to 'kig', 'pob' to 'pob', 'kig' to 'pob', 'pob' to 'kig') yielded a 4×4 confusion matrix.

The prediction for these confusion matrices varied according to the neural process being represented, and the templates for these predictions are illustrated in Fig 5a. An auditory representation, would represent the auditory input and therefore confuse ('kig' to 'kig' with 'kig' to 'pob' and 'pob' to 'pob' with 'pob' to 'kig'. A production representation would track the production response and confuse 'kig' to 'kig' with 'pob' to 'kig' and 'pob' to 'pob' with 'kig' to 'pob'. Lastly, a sensory-motor representation should track both the input and the output and therefore not confuse any of the conditions with each other (see Fig 5a).

In order to assess significance of these classification patterns, we first assessed significance by performing a permutation test by creating a surrogate distribution for each electrode category and each of the six time epochs, by shuffling the classifier training set. Significance was then empirically assessed by comparing the actual classification values to the surrogate distribution. The results of this analysis can be seen in Fig 5b.

To further assess the quality of each of the model predictions and relate the support for each model to the classifier results, we used the Kullback-Leibler Divergence, which quantifies the amount of information lost in bits when Q (the model) is used to approximate P (the data):

$$D_{KL}(P\|Q) = \sum_i P(i) log_2 \left(\frac{P(i)}{Q(i)} \right)$$

We constructed four models of the neural responses to the match-mismatch task: An Auditory model, a Production Model, a Sensory-Motor model, and a chance model (Fig 5a). For the model response values, the mean values within each positive (black) and negative (white) entry were used. Smaller values represent a smaller distance between the classification of the neural response and the idealized model of that neural response. To test for statistical significance, we compared the values within each confusion matrices for each electrode class to a distribution of values that were shuffled before being entered in the K-L analysis. Statistical values were then corrected for multiple comparisons using a one-sided FDR procedure with an alpha level of 0.05 (p threshold = 0.005). To create a normalized KL index for display purposes, we plotted the negative logarithm of each p-value (Fig 5 and Fig S4).

Since these results were obtained by pooling within class of electrodes across subjects, these effects could in theory be driven largely by individual subjects. To control for this, we repeated our classifier/model analyses using subsets of our subjects (N−1). Qualitatively similar results were obtained for each subject removal and for each electrode class (see Fig S5).

Response versus No Response/Classifier Analysis—The no-response trials were first identified via the audio recordings. Response versus No Response analysis was carried out using the same linear classifier as listed above on the SVD of the high gamma band neural response for the same epochs as in the representation classifier analysis. All successful responses were collapsed into a single Response condition and all non-responses were collapsed in a single no response condition. The maximum number of trials present

was used both for the training set and testing set using the same leave one out validation method. Since there were only 2 conditions, this resulted in a 2×2 confusion matrix. The mean entry over the 100 iterations was multiplied by the number of trials in the testing set and the entered as the observed values into a FDR-corrected one-sided chi-squared analysis (expected value = 50%) with an alpha level set at 0.05 (p threshold = 0.02) to assess significance. Significant Chi-Square values ranged from 5.42 to 19.99.

Statistical Methods

Whenever possible/appropriate, we used nonparametric methods for comparisons: permutation tests for significance calculation for high gamma power, as well as for assessing the significance of our model strength using the normalized Kullback-Leibler divergence (KLn). When required, we corrected for multiple comparisons using an FDR procedure with an alpha of level of 0.05. For our error analysis (Response versus No Response), we used a chi-square analysis. Data distribution was assumed to be normal for these tests, but this was not formally tested. No statistical methods were used to pre-determine sample sizes but our sample sizes are similar/greater than those reported in previous publications[39]. Both data collection and analysis were not performed blind to the conditions of the experiments.

Code Availability

Code can be made available by contacting the corresponding author.

Data Availability

Data can be made available by contacting the corresponding author.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Logie RH. Spatial and Visual Working Memory: A Mental Workspace. Psychol Learn Motiv - Adv Res Theory. 2003; 42:37–78.

2. Baddeley A. Working memory: looking back and looking forward. Nat Rev Neurosci. 2003; 4:829–39. [PubMed: 14523382]

3. Atkinson, RC., Shiffrin, RM. The Psychology of learning and motivation: Advances in research and theory. Spence, KW., Spence, JT., editors. New York: Academic Press; 1968. p. 89-195.

4. Vallar G, Baddeley AD. Phonological short-term store, phonological processing and sentence comprehension : A neuropsychological case study. Cogn Neuropsychol. 1984; 1:121–141.

5. Martin, RC., Lesch, M. Models of short-term memory. Gathercole, S., editor. Erlbaum; 1996. p. 149-178.

6. Saffran, EM., Martin, N. Neuropsychological Impairments of Short-term Memory. Vallar, G., Shallice, T., editors. Cambridge University Press; 1990. p. 428-447.

7. Jacquemot C, Dupoux E, Decouche O, Bachoud-Levi AC. Misperception in sentences but not in words: Speech perception and the phonological buffer. Cogn Neuropsychol. 2006; 23:949–71. [PubMed: 21049361]

8. Gupta P, MacWhinney B. Vocabulary acquisition and verbal short-term memory: computational and neural bases. Brain Lang. 1997; 59:267–333. [PubMed: 9299067]

9. King J, Just MA. Individual differences in syntactic processing: The role of working memory. J Mem Lang. 1991; 30:580–602.

10. Papagno C, Valentine T, Baddeley AD. Phonological Short-Term Memory and Foreign-Language Vocabulary Learning. J Mem Lang. 1991; 30:331–347.

11. Hirschberg J, Manning CD. Advances in natural language processing. Science (80–). 2015; 349:261–266.

12. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997; 9:1735–80. [PubMed: 9377276]

13. Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural networks. 2005; 18:602–10. [PubMed: 16112549]

14. Sutskever, I., Vinyals, O., Le, QVV. Sequence to Sequence Learning with Neural Networks; Advances in Neural Information Processing Systems. 2014. p. 3104-3112.at <http://papers.nips.cc/paper/5346-information-based-learning-by-agents-in-unbounded-state-spaces>

15. Baddeley AD, Thomson N, Buchanan M. Word length and the structure of short-term memory. J Verbal Learning Verbal Behav. 1975; 14:575–589.

16. Baddeley A. Working Memory. Science (80–). 1992; 255:556–559.

17. Jacquemot C, Scott SK. What is the relationship between phonological short-term memory and speech processing? Trends Cogn Sci. 2006; 10:480–6. [PubMed: 16997610]

18. Buchsbaum BR, D'Esposito M. The search for the phonological store: from loop to convolution. J Cogn Neurosci. 2008; 20:762–78. [PubMed: 18201133]

19. Hickok G, Buchsbaum B, Humphries C, Muftuler T. Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. J Cogn Neurosci. 2003; 15:673–82. [PubMed: 12965041]

20. Hickok G, Poeppel D. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition. 2004; 92:67–99. [PubMed: 15037127]

21. Hickok G, Poeppel D. Towards a functional neuroanatomy of speech perception. Trends Cogn Sci. 2000; 4:131–138. [PubMed: 10740277]

22. Atkinson RC, Shiffrin RM. Human Memory: A Proposed System and its Control Processes. Psychol Learn Motiv - Adv Res Theory. 1968; 2:89–195.

23. Shallice, T. Principles of the frontal lobe function. Stuss, DT., Knight, RT., editors. Oxford University Press; 2002. p. 261-284.

24. Goldman-Rakic PS. The prefrontal landscape: implications of functional architecture for understanding human mentation and the central executive. Philos Trans R Soc Lond B Biol Sci. 1996; 351:1445–53. [PubMed: 8941956]

25. Wallis JD, Anderson KC, Miller EK. Single neurons in prefrontal cortex encode abstract rules. Nature. 2001; 411:953–6. [PubMed: 11418860]

26. Vallar G, Baddeley AD. Fractionation of working memory: Neuropsychological evidence for a phonological short-term store. J Verbal Learning Verbal Behav. 1984; 23:151–161.

27. Paulesu E, Frith CD, Frackowiak RSJ. The neural correlates of the verbal component of working memory. Nature. 1993; 362:342–345. [PubMed: 8455719]

28. Veltman DJ, Rombouts SaR, Dolan RJ. Maintenance versus manipulation in verbal working memory revisited: an fMRI study. Neuroimage. 2003; 18:247–256. [PubMed: 12595179]

29. Becker JT, MacAndrew DK, Fiez Ja. A comment on the functional localization of the phonological storage subsystem of working memory. Brain Cogn. 1999; 41:27–38. [PubMed: 10536084]

30. Ladefoged, P. A course in Phonetics. Wadsworth Publishing; 2005.

31. Zhang M, Barash S. Neuronal switching of sensorimotor transformations for antisaccades. Nature. 2000; 408:971–5. [PubMed: 11140683]

32. Sakai K, Miyashita Y. Neural organization for the long-term memory of paired associations. Nature. 1991; 343:152–155.

33. Tomita H, Ohbayashi M, Nakahara K, Hasegawa I, Miyashita Y. Top-down signal from prefrontal cortex in executive control of memory retrieval. Nature. 1999; 401:699–703. [PubMed: 10537108]

34. Nir Y, et al. Coupling between Neuronal Firing Rate, Gamma LFP, and BOLD fMRI Is Related to Interneuronal Correlations. Curr Biol. 2007; 17:1275–1285. [PubMed: 17686438]

35. Ray S, Crone NE, Niebur E, Franaszczuk PJ, Hsiao SS. Neural correlates of high-gamma oscillations (60–200 Hz) in macaque local field potentials and their potential implications in electrocorticography. J Neurosci. 2008; 28:11526–11536. [PubMed: 18987189]

36. Cogan GB, et al. Sensory-motor transformations for speech occur bilaterally. Nature. 2014; 507:94–8. [PubMed: 24429520]

37. D'Esposito M, et al. The Neural Basis of the Central Executive System of Working Memory. Nature. 1995; 378:279–281. [PubMed: 7477346]

38. Rowe JB, Toni I, Josephs O, Frackowiak RS, Passingham RE. The prefrontal cortex: response selection or maintenance within working memory? Science (80–). 2000; 288:1656–60.

39. Voytek B, et al. Oscillatory dynamics coordinating human frontal networks in support of goal maintenance. Nat Neurosci. 2016; 18:1318–1324.

40. Baddeley AD. The influence of acoustic and semantic similarity on long-term memory for word sequences. Q J Exp Psychol. 1966; 18:302–9. [PubMed: 5956072]

41. Vallar, G., Papagno, C. The Handbook of Memory Disorders. Baddeley, AD.Kopelman, MD., Wilson, BA., editors. John Wiley & Sons, Ltd; 2002. p. 249-270.

42. Postle BR, Berger JS, D'Esposito M. Functional neuroanatomical double dissociation of mnemonic and executive control processes contributing to working memory performance. Proc Natl Acad Sci U S A. 1999; 96:12959–64. [PubMed: 10536030]

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

44. Jonides J, et al. Verbal Working Memory Load Affects Regional Brain Activation as Measured by PET. J Cogn Neurosci. 1997; 9:462–75. [PubMed: 23968211]

45. Baddeley A, Lewis V, Vallar G. Human Experimental Psychology Exploring the articulatory loop. Q J Exp Psychol. 1984; 36:233–252.

46. Awh E, Vogel EK, Oh SH. Interactions between attention and working memory. Neuroscience. 2006; 139:201–208. [PubMed: 16324792]

47. Ikkai A, Curtis CE. Common neural mechanisms supporting spatial working memory, attention and motor intention. Neuropsychologia. 2011; 49:1428–1434. [PubMed: 21182852]

48. Markowitz, Da, Curtis, CE., Pesaran, B. Multiple component networks support working memory in prefrontal cortex. Proc Natl Acad Sci USA. 2015; 112:11084–11089. [PubMed: 26283366]

49. Latimer KW, Yates JL, Meister MLR, Huk AC, Pillow JW. Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. Science (80–). 2015; 349:37–54.

50. Harvey CD, Coen P, Tank DW. Choice-specific sequences in parietal cortex during a virtual-navigation decision task. Nature. 2012; 484:62–68. [PubMed: 22419153]

51. Yang AI, et al. Localization of dense intracranial electrode arrays using magnetic resonance imaging. Neuroimage. 2012; 63:157–165. [PubMed: 22759995]

52. Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis. I: Segmentation and Surface Reconstruction. Neuroimage. 1999; 9:179–194. [PubMed: 9931268]

53. Mitra P, Pesaran B. Analysis of Dynamic Brain Imaging Data. Biophys J. 1999; 76:691–708. [PubMed: 9929474]

54. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc Ser B. 1995; 57:289–300.
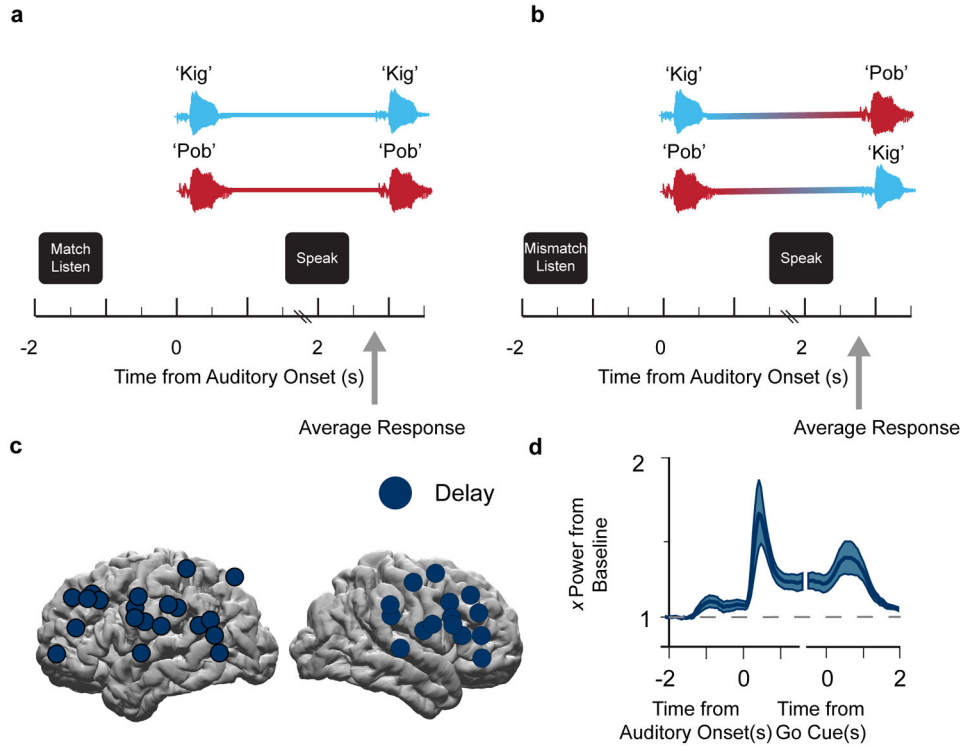
Figure 1. Match-mismatch task

Participants performed trials that were either in the **a**. match condition or **b**. mismatch condition. In the match condition (**a**), subjects first saw a visual cue ('Match Listen') and then 1.5 seconds later were presented with one of two words auditorily ('kig' or 'pob'). After a delay (1.5 to 2 s), the participants saw another cue ('Speak') which instructed them to say the non-words they had heard ('kig' to 'kig' and 'pob' to 'pob'). In the mismatch condition (**b**), participants were presented with a different visual cue ('Mismatch Listen') which instructed them that they were to speak the non-word they hadn't heard (e.g. 'kig' to 'pob' and 'pob' to 'kig'). This setup allowed us to isolate the sensory and motor processes and their activity during the delay period. Gray arrows indicate the average response times for the Match (832 ms) and Mismatch (773 ms). **c.** Electrodes with significant delay activity (1000 ms – 1500 m post Auditory Onset – see Experimental Procedures) are denoted in purple and selectively localized to the prefrontal cortex. **d.** The time course of delay activity electrodes aligned to the Auditory Onset. Task epochs are denoted in different colors (see legend). Note the elevated delay activity, as well as the activity in the sensory and motor epoch. Error values are SEM of power across electrodes (N = 36 electrodes), one sided p threshold = 0.01.
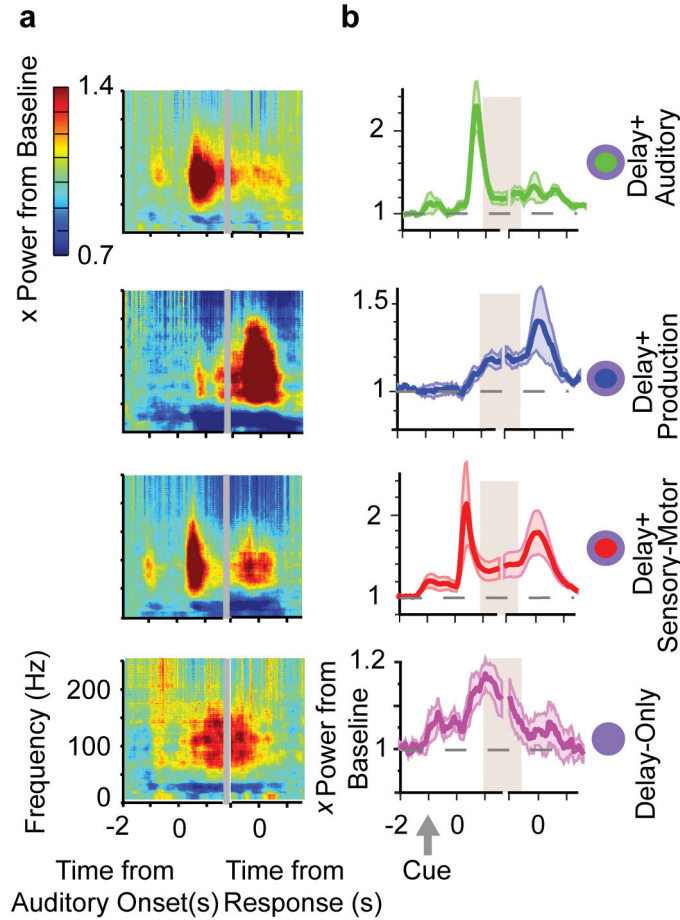
Figure 2. Example responses
a. Example spectrograms of the neural responses for Delay Electrodes. A Delay+Auditory electrode (top row) during the match-mismatch task: A high gamma (70 Hz +) neural response is seen in the auditory epoch as well as during the delay epoch. In an example Delay+Production electrode (second row) significant neural activity is present in the delay and the production epoch. In an example Delay+Sensory-Motor electrode (third row) significant neural activity is seen in the auditory, delay, and production epochs. Lastly, in an example Delay-Only electrode (bottom row) delay activity is present along with an absence of activity in the auditory and motor epoch. Gray bar reflects the variable delay period (1.5 – 2 seconds post auditory onset). Example electrodes without delay can be seen in Fig S2a and locations of the electrodes can be seen in Fig S2c. While representative, these kinds of neural responses can be seen across subjects (see Fig S3). **b**. Average high gamma power traces (70 – 160 Hz) are shown for each electrode class as shown in **a**: Delay+Auditory, Delay+Production, Delay+Sensory-Motor, and Delay-Only. The gray arrows indicate the onset of the cue and average response time. Error values are SEM of power across electrodes (Delay+Auditory: N = 6 Electrodes, Delay+Production: N = 9, Delay+Sensory-Motor: N = 11, Delay-Only: N = 10).
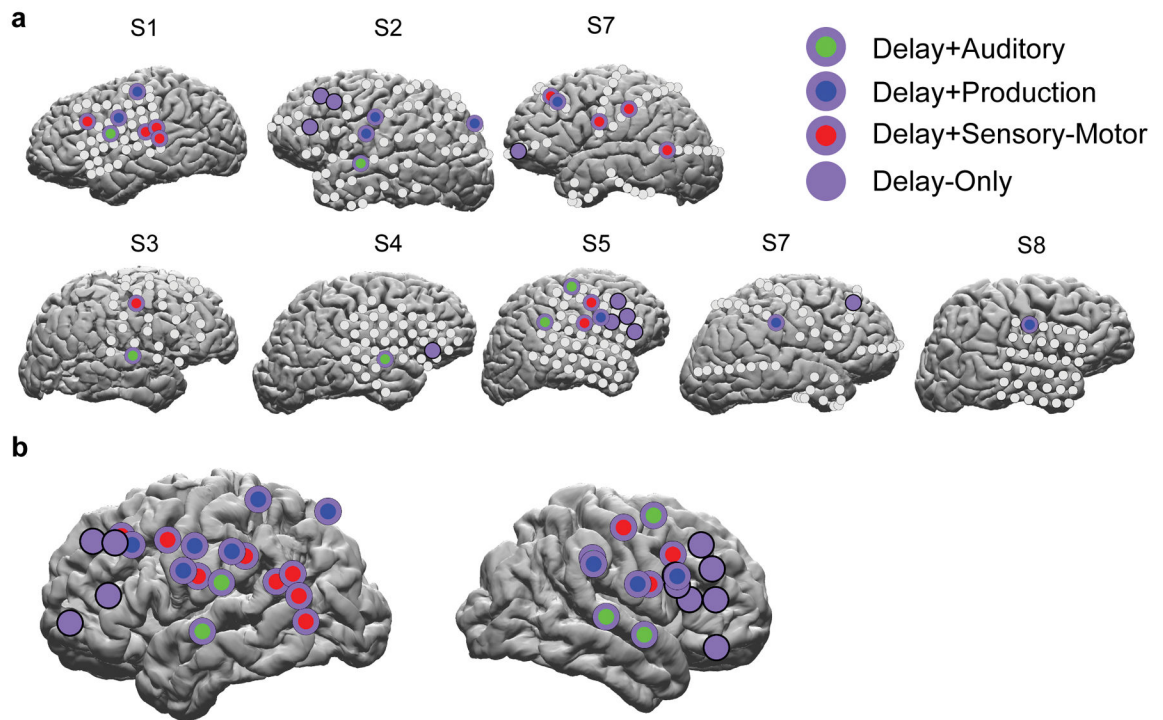
Figure 3. Subdivisions of Delay Activity

a. For each subject, we classified delay activity by a combination of an electrode's response to a localizer task (see **Experimental Procedures**) and the delay response during the Sensory-Motor Mismatch Task. Delay+Auditory electrodes (green with purple outline) are active during the delay and auditory presentation of the non-word. Delay+Production electrodes (blue with purple outline) during the delay and articulation of the utterance. Delay +Sensory-Motor electrodes (red with purple outline) are active both during the auditory presentation and the articulation. The Delay-Only electrodes (purple) were active during the delay but not the sensory or motor epoch. **b.** Localization of electrodes with delay activity across subjects. Color convention the same as in **a.** One sided p threshold = 0.01
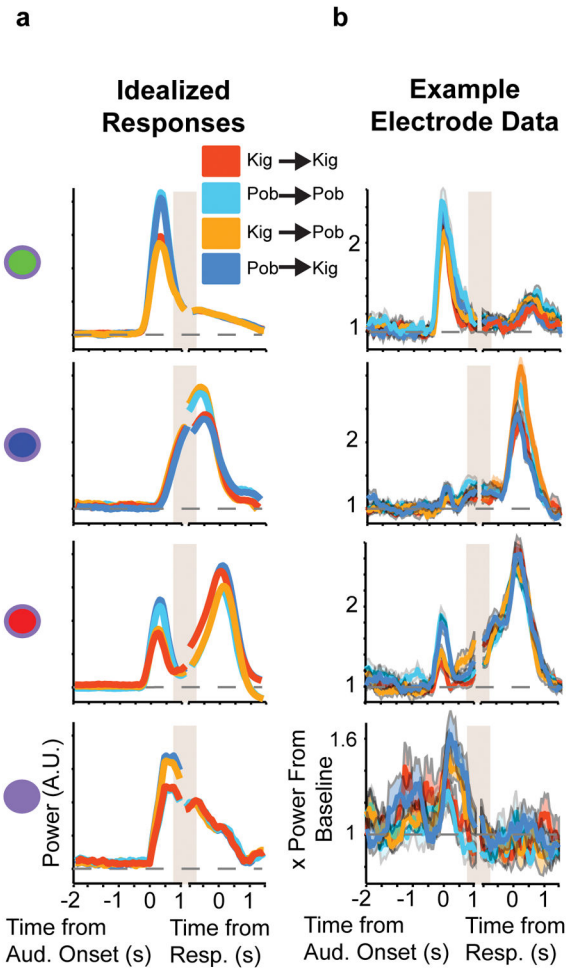
Figure 4. Idealized and representative responses

a. Idealized neural responses should demonstrate patterns in the sensory and motor epochs that represent the class that they belong to. Delay+Auditory (top row) should demonstrate a sensory neural response profile whereas Delay+Production electrodes (second row) should demonstrate a motor neural profile. Delay+Sensory-Motor (third row) should demonstrate a conjunction of a sensory representation in the sensory epoch and a motor representation in the motor epoch. If the Delay-Only responses track the abstract rule representation, they should reflect a differentiation between the 'match' and 'mismatch' conditions. **b.** Representative electrodes demonstrate the above hypothesized patterns during the sensory and motor epochs. Error values are SEM of power across trials (Delay+Auditory: N = 51 Electrodes, Delay+Production: N = 48, Delay+Sensory-Motor: N = 57, Delay-Only: N = 50).
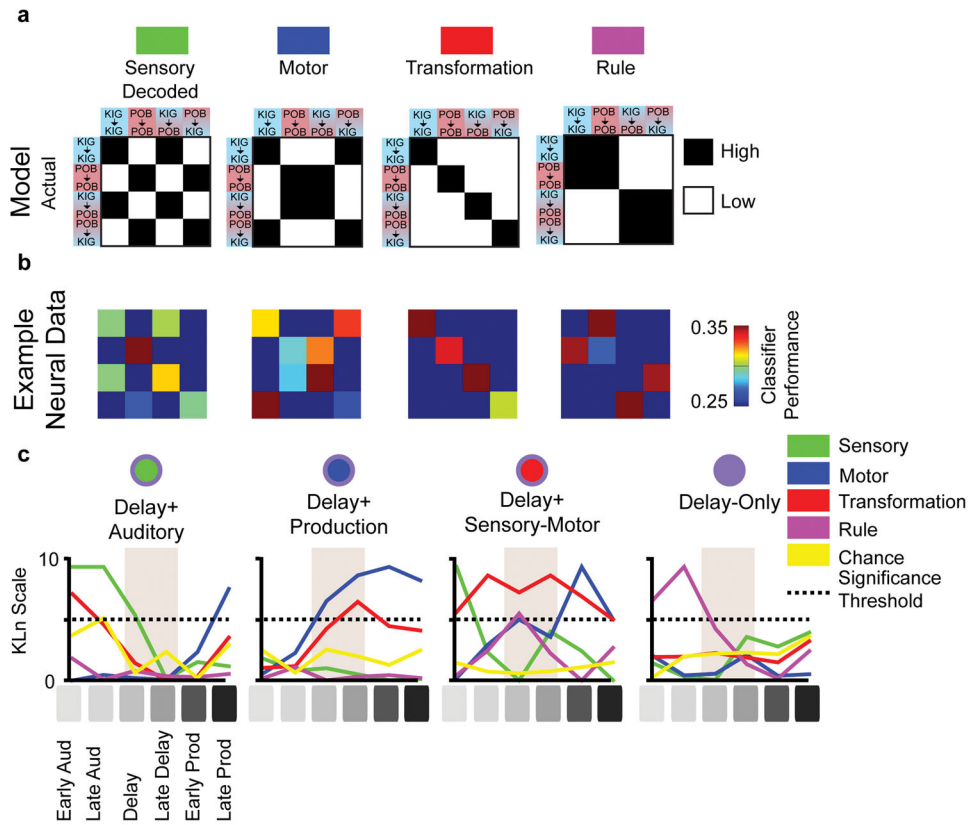
Figure 5. Representation of vWM

a. There are four possible representational models for the content of vWM: A Sensory representation in which the classifier confuses each of the decoded (x-axis) and actual tokens presented (y-axis) with their input sensory equivalent e.g. 'kig' to 'kig' is confused with 'kig' to 'pob' (Left most model). A motor model in which the classifier confuses the output production equivalent, e.g. confusing 'kig' to 'kig' with 'pob' to 'kig' (second model). A sensory-motor model in which all four sensory-motor mapping conditions are decoded separately in a sensory-motor representation (third model), and a Rule model in which the abstract rule is encoded regardless of token (right panel). **b**. Example confusion matrices show all four response classes were obtained when applied to the sub-processes of vWM. **c.** The Strength of each model was assessed for each electrode response category in each temporal epoch using an FDR corrected normalized KL index (KLn, one sided p threshold = 0.005). Each response class represents information differently. The two maintenance systems encode sensory and motor representations respectively: Delay +Auditory electrodes encode an auditory representation during the sensory epoch and the delay epoch, and the Delay+Production electrodes encode the motor plan during the delay, followed by the motor output. The manipulation systems displayed very different representations. The Delay+Sensory-Motor electrodes demonstrated a dynamic representation that switched from an auditory, to a transformation, and finally to production representation, linking perception and production representations. The Delay-Only electrodes demonstrated an abstract rule based representation. The onset of the rule response coincides with the switch from a sensory to transformation representation in the Delay
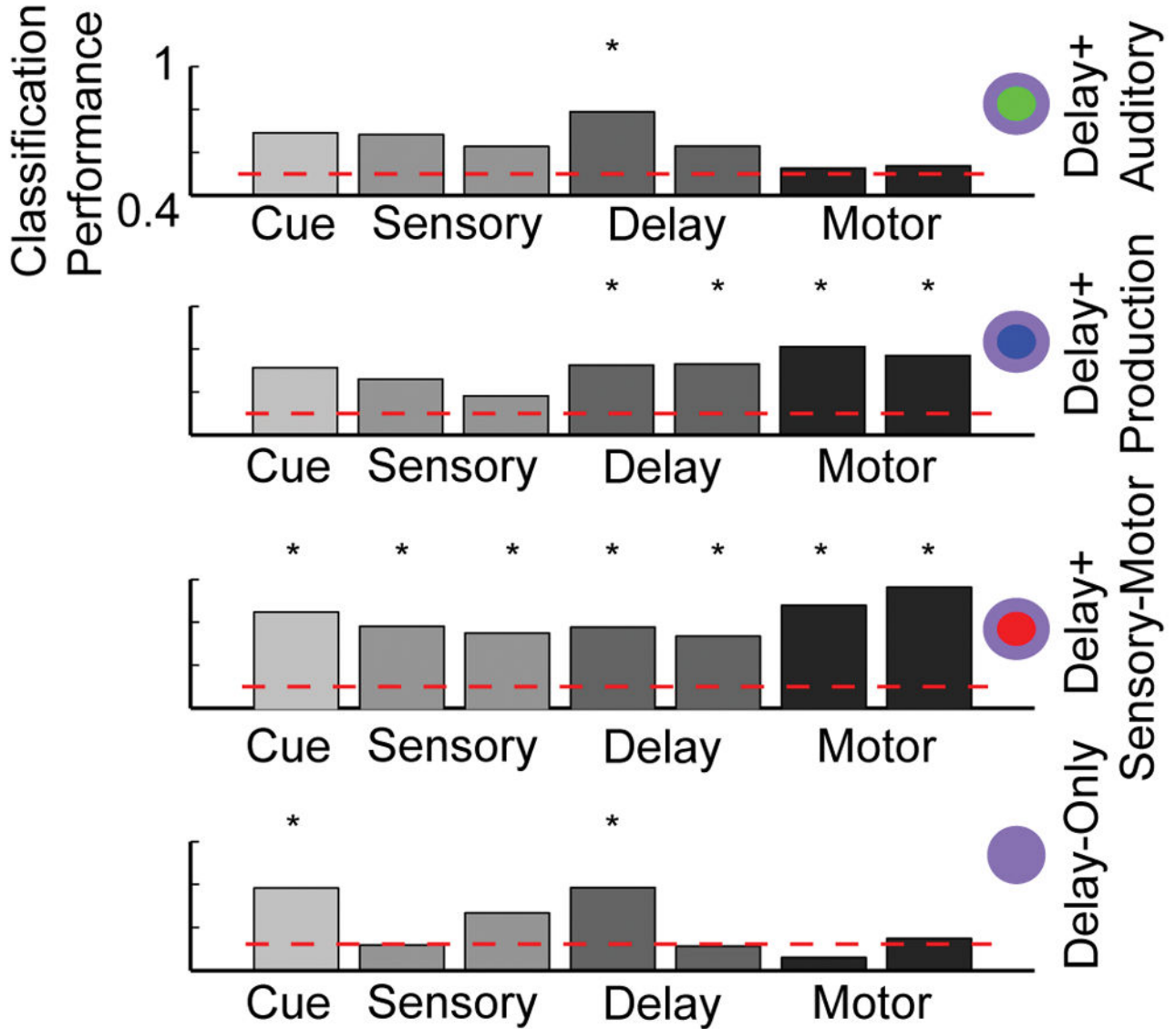
+Sensory-Motor electrodes indicating these two systems work together to manipulate phonological input according to the rule.

Figure 6. Error Analysis

During the delay epoch, all four response classes differentiated between trials with a correct utterance and trials with no utterance (chance is 0.5 – dotted red lines). **a)** Delay+Auditory responses did not significantly differentiate error trials during the sensory epoch. **b)** Delay +Production responses significantly differentiated error trials during the late delay, and the production epochs. **c)** Delay+Sensory-Motor responses significantly distinguished error trials during all task epochs. **d)** Delay-Only responses significantly distinguished error trials during the cue and delay epochs. These results demonstrate that suggesting that errors are likely due to failures to encode and applying the abstract rule. All values were FDR-corrected with an alpha of 0.05, which resulted in a one sided p threshold of 0.02.