Early neural signatures of visual short-term memory

Yigal Agam a,b, Joo-Seok Hyun c, Jared F. Danker c, Feng Zhou d, Michael J. Kahana e, Robert Sekuler d

a Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, USA
b Department of Psychology, Chung-Ang University, Seoul, Korea
c Department of Psychology, Carnegie Mellon University, USA
d Volen Center for Complex Systems, Brandeis University, USA
e Department of Psychology, University of Pennsylvania, USA

ARTICLE INFO

Article history:
Received 27 May 2008
Received in revised form 24 August 2008
Accepted 11 September 2008
Available online 30 September 2008

ABSTRACT

Visual short-term memory (VSTM) relies on a distributed network including sensory-related, posterior regions of the brain and frontal areas associated with attention and cognitive control. To characterize the fine temporal details of processing within this network, we recorded event-related potentials (ERPs) while human subjects performed a recognition-memory task. The task’s difficulty was graded by varying the perceptual similarity between the items held in memory and the probe used to access memory. The evaluation of VSTM’s contents against a test stimulus produced clear similarity-dependent differences in ERPs as early as 156 ms after probe onset. Posterior recording sites were the first to reflect the difficulty of the analysis, preceding their frontal counterparts by about 50 ms. Our results suggest an initial feed-forward interaction underlying stimulus-memory comparisons, consistent with the idea that visual areas contribute to temporary storage of visual information for use in ongoing tasks. This study provides a first look into early neural activity underlying the processing of visual information in short-term memory.

Introduction

Visual short-term memory (VSTM), a key aspect of human cognition, is thought to be supported by a distributed network of different brain areas, including parts of the prefrontal cortex (PFC) and posterior areas associated with sensory perception. A wealth of neuroimaging data (Courtney et al., 1997; Haxby et al., 2000; Mitchell and Cusack, 2008; Pessoa et al., 2002; Postle et al., 2003; Todd and Marois, 2004) show that activity in many brain regions correlates with memory demands, either when measured by contrasting a memory task with a purely perceptual task or by varying the amount of information held in memory. Unfortunately, the limited temporal resolution of functional neuroimaging precludes the identification of the fine temporal details of this network. While VSTM has been studied using electroencephalography (EEG) and magnetoencephalography (MEG) (Agam and Sekuler, 2007; Croize et al., 2004; Morgan et al., 2008; Ruchkin et al., 1990; Ruchkin et al., 1997; Sreenivasan et al., 2007; Vogel and Machizawa, 2004; Vogel et al., 2005), techniques that afford higher temporal resolution, the very early neural events that underlie access to temporarily stored visual information have eluded full description. Knowing the timing and location of the earliest signatures of VSTM will provide clues about mechanisms fundamental to the operation of this important ability.

We had two objectives in carrying out the study reported here. First, we wanted to determine the latency associated with the brain’s comparison of some visual input against stored visual memories. Second, we wanted to identify the direction of information flow within the brain; specifically, in which regions of the brain was evidence of the evaluation process first seen? These questions are important for understanding the nature of VSTM, but have been difficult to answer. One particular challenge is distinguishing between memory-related effects and those resulting from intrinsic properties of the stimuli used to probe memory. Verbal stimuli, for example, rely on long-term, language-based representations. Those extra-laboratory associations are difficult to control and may vary substantially in scope and topography between subjects and between memory exemplars. Some visual stimuli, such as pictures of familiar objects, may also lead to the use of verbal labels. Moreover, when objects are presented in the context of some larger scene, recognition is likely influenced by the early extraction of the scene’s gist or other context-related information (Biederman et al., 1974). To overcome such complexities in studying VSTM, we used test materials that do not depend on extra-sensory representations and associations (Sekuler and Kahana, 2007). Our stimuli were compound gratings that differed only in their vertical spatial frequency. These stimuli are advantageous in that they resist encoding with verbal labels, are unlikely to evoke long-term memory representations, and can easily be manipulated along a single dimension. With the burden of the task squarely on the processes of short-term memory, we could obtain a reliable neural index of access to VSTM: we manipulated the similarity between the probe stimulus and
the stimuli held in memory, a variable that generates a broad range of recognition-memory performance (Kahana and Sekuler, 2002; Kahana et al., 2007; Sekuler and Kahana, 2008; Zhou et al., 2004). The similarity between memory items is known to be a core component of the neural computations for recognition, regardless of the nature of the items: words, objects, sounds etc. (Sekuler and Kahana, 2008). Therefore, results with our simplified, synthetic stimuli can be generalized to other types of sensory information.

We recorded scalp EEG while subjects performed a Sternberg recognition memory task (Sternberg, 1966). To track the time course of recognition, we compared the ERPs time locked to the stimulus probe under different similarity conditions. Rather than focus on predefined ERP components, we made statistical comparisons among similarity conditions at consecutive time points relative to probe onset. This allowed us to identify the time of the earliest differences. This basic method has been used before to study the speed of visual processing (Rousselet et al., 2007; Thorpe et al., 1996), but not in the context of short-term memory. In those studies, the subject had to judge a picture against a fixed criterion (“animal?”). Here, in contrast, subjects had to update the content of memory on every trial. This allowed us to search for the earliest signature of access to working memory, here operationalized as the earliest effect of probe-item similarity.

Materials and methods

Subjects

Subjects were twelve volunteers (all right-handed, 8 female, ages 20–34, who had normal or corrected-to-normal Snellen acuity, and normal contrast sensitivity as measured with Pelli–Robson charts (Pelli and Robson, 1988). All subjects provided written informed consent to a protocol approved by Brandeis University’s Committee for the Protection of Human Subjects.

Stimuli

Stimuli were compound gratings generated by adding vertical and horizontal sinusoidal luminance gratings of equal contrast (Kahana and Sekuler, 2002). Gratings were generated and displayed using Matlab (The Mathworks, Natick, MA) and extensions from the psychophysics and video toolboxes (Brainard, 1997; Pelli, 1997). Stimuli were centered on a 20-inch cathode ray tube display with 800×600 pixels screen resolution and 75 Hz refresh rate. Each stimulus’ mean luminance was matched to the display’s constant, 600 pixels screen resolution and 75 Hz refresh rate. Each 800
definition. Subject’s mean luminance was matched to the display’s constant, 600 pixels screen resolution and 75 Hz refresh rate. Each stimulus produced no potentially distracting luminance transient.

Prior to each subject’s first experimental session, an adaptive psychophysical procedure (Wetherill and Levitt, 1965) determined the subject’s discrimination threshold for spatial frequency. The threshold was defined as the value at which the subject could discriminate successively presented gratings 79% of the time. Expressed as a Weber fraction, the subjects’ mean threshold value was 0.13, with a standard error of 0.06. To take into account differences in the accuracy with which subjects encoded grating’s spatial frequency, the stimuli used to test each subject’s recognition memory were scaled according to his or her own discrimination threshold.

During memory testing, subjects were presented sequentially with two study items (S1 and S2) on each trial, followed by a probe PB. All three stimuli shared a common horizontal frequency, making any differences among S1, S2 and PB dependent solely on their vertical frequencies. Both horizontal and vertical frequencies varied randomly across trials. Horizontal frequencies were randomly drawn from a uniform distribution between 1.0 and 7.0 cycles per degree. The lowest possible vertical frequency was randomly drawn from a uniform distribution between 1.0 and 2.0 cycles per degree, but the actual range of possible vertical frequencies consisted of discrete increments from that minimum, using logarithmic frequency steps based on the individual subject’s discrimination threshold, expressed in units of “just noticeable difference” (JND). In addition, a random phase shift was applied to the vertical frequency of each stimulus. This ensured that subjects could not successfully base memory judgments on residual retinotopic information.

The study stimuli, S1 and S2, always differed from one another by four JNDS. On half of the trials S1’s spatial frequency was higher than S2’s; on half the trials, the reverse was true. Additionally, Target trials, on which PB matched one of the study items, and Lure trials, on which it matched neither study item, were equally probable. On half the Target trials, PB matched S1; on the other half, it matched S2. On Lure trials, PB spatial frequency was drawn from a discrete uniform distribution whose spatial frequencies spanned a range centered on the geometric mean of the two study items. Specifically, this range for possible PB frequencies went from four JNDS below the spatial frequency of the lower frequency study item to four JNDS above the spatial frequency of the higher frequency study item. These equally-probable candidate values for PB spatial frequency on Lure trials were ±6, ±4, and 0 threshold units relative to the geometric mean of that trial’s S1 and S2 frequencies. Expressed in these same coordinates, the study items’ spatial frequencies, as well as PB frequency on Target trials, were ±2 threshold units. For Lure trials, probe frequencies relative to the study items were equally probable and varied randomly.

Procedure

Fig. 1A summarizes the behavioral paradigm. A trial began with a fixation cross at the center of the computer display. The fixation cross was succeeded in turn by S1, S2, and PB, each lasting 750 ms. An inter-stimulus interval (ISI) of 275±75 ms separated fixation, S1 and S2; a 500±75 ms retention interval separated S2 and probe. By pressing designated “Yes” and “No” keys subjects indicated their judgment of whether PB matched one of the study items, or failed to match either of them. An inter-trial interval (ITI) of 3 s separated successive trials. Subjects were instructed to refrain from blinking except during the ITI. In addition, every 20 trials subjects received feedback on their accuracy and response times. Each subject participated in four sessions of 320 trials each, during which both behavioral performance and EEG signals were recorded.

EEG recording

We recorded from 129 electrode sites at 250 Hz using an Electrical Geodesics (Eugene, OR) system. All channels were referenced to the vertex during data acquisition, and were adjusted so that scalp impedances were below 100 kΩ. Data were cleaned of bad channels (chosen by visual inspection), re-referenced to the grand average and reduced to a montage comprising 27 standard electrode locations, two vertical, bipolar channels above and below each eye, and one horizontal, bipolar eye channel, using BESA (MEGIS Software GmbH, Munich). Data were then baseline corrected (100 ms prestimulus), averaged and analyzed using Matlab. Blink artifacts were eliminated by rejecting epochs in which the difference between the maximum and minimum voltage at any EOG channel exceeded 150 µV. This resulted in the removal, on average, of 5% (SD: 4%) of each subject’s trials. Data were high-pass filtered at 0.1 Hz, low-pass filtered at 30 Hz, and notch-filtered at 60 Hz.

Statistical Analysis

To examine the divergence between ERPs for different probe-study item relationships, we performed a repeated-measures ANOVA (with probe frequency, relative to the study items, as the main factor) at each
time sample. Our goal was to identify consecutive periods of time during which the ERPs consistently differed from each other. The appropriate significance thresholds (P-value and number of consecutive samples) were determined by a random permutation procedure: we created 1000 surrogate sets of data by randomly assigning trials to one of the three probe frequency groups, while preserving the temporal and spatial correlations between electrodes (Supplementary Fig. 1A). This provided an empirical estimate of the probability of a Type I statistical error.

Supplementary Fig. 1B shows a flowchart of the random permutation procedure. For this statistical method, the two variables that need to be determined are PV, the P-value threshold for each individual comparison, and N, the number of consecutive samples below PV required to produce F, the false discovery rate (FDR). Either PV or N can be chosen by the experimenter, and the other variable is then determined empirically. For this study, we decided to control PV and searched for a matching N. For each of the 1000 surrogate datasets, we performed the same statistical test (ANOVA) for each electrode and time point, as if we were dealing with the real data. For the data presented here, we assigned PV to be 0.05 and set our target FDR to 1%. Unlike previous implementations of this statistical method, we used random permutations of the data to find the appropriate number of consecutive samples required to declare significance (see Methods). We found significant differences (P<0.05 for at least

Results

As expected, response time (Fig. 1B) and the probability of a “Yes” response (Fig. 1C) both increased as a function of the similarity between the probe and the study items (Kahana et al., 2007; Zhou et al., 2004). In the case of Lure trials, this increase reflected the difficulty in detecting a memory-stimulus mismatch. Because probe frequency strongly influenced task difficulty, differences between ERPs for the three Lure conditions provide an early indication of a comparison between the probe and memory items. As both the behavioral and the ERP data (see Supplementary Figs. 2 and 3) were highly symmetrical around the spatial frequency midway between the two study items (the geometric mean of their respective frequencies), we used three aggregated sets of Lure trials, sorted according to their similarity to the memory items. This aggregation produced three sets of relative probe frequencies: Lure trials with the probe falling either ±6 JNDs, ±4 JNDs, or 0 JNDs from the mid-point.

For each time sample (4 ms), the corresponding data were subjected to a repeated-measures ANOVA; we searched for time periods over which consecutive samples showed significant differences between ERPs (Giard and Peronnet, 1999; Rieger et al., 2005; Rousselet et al., 2007; Rugg et al., 1995; Shams et al., 2001; Thorpe et al., 1996). Unlike previous implementations of this statistical method, we used random permutations of the data to find the appropriate number of consecutive samples required to declare significance (see Methods). We found significant differences (P<0.05 for at least

---

**Fig. 1.** Experimental paradigm and behavioral results. (A) schematic diagram of a trial. Top example is for a Target trial, in which the probe matches one of the study items, and the bottom one is for a Lure trial, in which the probe is different from both study items. (B) response time as a function of the probe spatial frequency with respect to the study items. Solid and dashed lines indicate trials of the opposite frequency order. (C) the probability of a “Yes” response plotted against probe frequency. Note the high symmetry of the results around the central frequency (the geometric mean of the study items’ frequencies). Vertical gray bars indicate the study items’ frequencies. Error bars are within-subject SEM (Loftus and Masson, 1994).
Fig. 2. ERP results. (A) ERP traces for each of the 27 interpolated recording sites (anterior locations at the top). Each color corresponds to a different probe frequency relative to the study items’ frequencies (aggregated across positive and negative distances from the mean). Time 0 is the onset of the probe stimulus. Vertical dashed lines denote the time of the first significant difference between the ERPs (at $P < 0.05$), for locations that met the significance criterion. Colored stripes above each panel indicate the result of the ANOVA at every time point. $P$-values above 0.05 are not included. Number of trials for each condition (red, green, blue; mean±SD): 242±11.38, 244.2±9.9, 121.8±6.2. (B) Topographical maps of the $P$-values at time points separated by 25 ms. Red regions indicate strong modulation of the ERPs by probe frequency, and blue regions indicate little or no modulation. (C) Onset times of ERP differences. The histogram shows the time at which ERPs started to diverge ($P < 0.05$) at ten posterior (red) and ten anterior (blue) locations. (D) ERP traces for aggregated sets of anterior and posterior electrodes. Colored stripes represent significance values (top right color scale); note the earlier significance for the posterior ERPs.
Our results suggest that information in short-term memory exerts a very early influence on the evaluation of a subsequent visual test stimulus. In fact, such evaluation adds no substantial delay to the speed of processing when compared to other ERP studies whose paradigms impose no demand on short-term memory. Those studies (Rousselet et al., 2007; Thorpe et al., 1996) used a categorization task, in which each trial’s test stimulus was compared to the same stable criterion that could be held in memory over many trials (e.g., an “animal” template in long-term memory). As subjects saw and categorized just a single stimulus on each trial, some preparatory memory processing was likely done prior to stimulus presentation in order to establish the categories that would be judged. In contrast, our task required that the contents of memory and the decision criterion had to be updated on every trial and held in VSTM until the probe presentation. In addition, it has been suggested that the differences seen in those studies at around the same latency arise from differences in the physical features of the stimuli (Johnson and Olshausen, 2003); our design overcomes this confound.

It is by now well accepted that prefrontal brain regions and posterior, “sensory” cortex both contribute to short-term memory. Visual selectivity has been observed not only in posterior regions of the visual system, but also in PFC (Everling et al., 2006; Funahashi et al., 1990). Similarly, both regions show delay-period activity during working memory tasks (Freedman et al., 2003; Funahashi et al., 1990; Haxby et al., 2000; Pessoa et al., 2002; Postle et al., 2003; Rainer et al., 1998; Zaksas and Pasternak, 2006), suggesting they are both important for temporary maintenance of information. The role of each component in this distributed network, however, is unsettled. The results of the present study imply that the comparison process in VSTM involves a significant role for early flow of activation in a posterior–anterior direction. A few caveats, however, should be noted: first, 156 ms is long enough to allow feedback connections from prefrontal cortex to more posterior areas (Foxe and Simpson, 2002); for example, ERP experiments have shown that attention modulated visual processing as early as 100 ms poststimulus (Hillyard et al., 1998). Second, ERP measurements provide only coarse information about the localization of activity in specific brain areas, which makes it difficult to know the sources of our ERPs with precision. Moreover, a single dipole may have generated both posterior and anterior differences, in which case the earlier posterior difference could be explained by higher sensitivity in those electrodes. More precise spatial identification, then, would likely require invasive recordings. With these caveats in mind, we suggest that the earlier posterior differences, as well as the early onset of the ERP differences, support the hypothesis that the actual storage of mnemonic visual information relies on brain areas that are also used for visual perception (Postle, 2006). In this view, the role of PFC may be to distribute attention between the representations of the remembered items and the representations activated by the presence of the visual stimulus.

Acknowledgment

Supported by NIH grant MH68404.

Appendix A. Supplementary data


References
