Stimulus-Specific Delay Activity in Human Primary Visual Cortex

John T. Serences, Edward F. Ester, Edward K. Vogel, and Edward Awh

University of California, San Diego, and University of Oregon

ABSTRACT—Working memory (WM) involves maintaining information in an on-line state. One emerging view is that information in WM is maintained via sensory recruitment, such that information is stored via sustained activity in the sensory areas that encode the to-be-remembered information. Using functional magnetic resonance imaging, we observed that key sensory regions such as primary visual cortex (V1) showed little evidence of sustained increases in mean activation during a WM delay period, though such amplitude increases have typically been used to determine whether a region is involved in on-line maintenance. However, a multivoxel pattern analysis of delay-period activity revealed a sustained pattern of activation in V1 that represented only the intentionally stored feature of a multifeature object. Moreover, the pattern of delay activity was qualitatively similar to that observed during the discrimination of sensory stimuli, suggesting that WM representations in V1 are reasonable “copies” of those evoked during pure sensory processing.
remembered either an orientation or a color for 10 s (Haxby et al.,
2001; Haynes & Rees, 2005; Kamitani & Tong, 2005, 2006;
Norman, Polyn, Detre, & Haxby, 2006; Peelen & Downing, 2007;
Serences & Boynton, 2007a, 2007b). The observation of featurespecific activation patterns in V1 suggests that sensory mecha-
nisms are recruited to support the storage of information in WM and
furthermore indicates that subjects have top-down control over
which features of a multifeature stimulus are stored.

methods

Observers
Ten neurologically intact observers participated in a single 2-hr
scanning session. All observers gave written informed consent
in accord with the requirements of the institutional review board
at the University of Oregon. Data from 3 observers were dis-
carded because of technical problems or voluntary withdrawal
from the study. Each observer was trained in the experimental
task for approximately 1.5 hr prior to scanning, to set sample-
test disparities to threshold (see Staircase Procedure).

Staircase Procedure
To encourage a narrow focus of attention on only the relevant
feature, and to discourage verbal-labeling strategies, we titrated
task difficulty for each observer in a separate behavioral testing
session (7–8 blocks of each feature condition). The task was
identical to that just described, and the sample-test disparity
was adjusted independently for each feature dimension until a
criterion level of performance (approximately 75% correct) was
reached. The resulting orientation and color disparities were
determined to use the sample-test disparity for each feature
during scanning.

fMRI Data Acquisition and Analysis
Scanning was performed using a 3-T Siemens Allegra system at
the Robert and Beverly Lewis Center for Neuroimaging at the
University of Oregon. Anatomical images were acquired using a
spoiled-gradient-recalled T1-weighted sequence that yielded
images with a 1-mm³ resolution. Whole-brain echo-planar im-
ages (EPIs) were acquired in 33 transverse slices (2,000-ms
repetition time, 30-ms echo time, 90° flip angle, 64 × 64 matrix,
192-mm field of view, 1.86; Brain Innovations, Maastricht, The
Netherlands) and in one of two orientations (45° or 135°, plus or minus an additional offset randomly
selected from a range of ±10°) and in one of two colors (middle
red or middle green, plus or minus an offset randomly selected
from a range of ±10% saturation). The addition of random jitter
around the canonical features discouraged the use of verbal
labels, which would not have been precise enough to support
accurate performance in this task. For some observers (n = 2),
the spatial frequency of the stimulus varied randomly across
trials (0.75–1.25 cycles/deg); for the others, the spatial fre-
quency was fixed at 1 cycle/deg. The sample stimulus flickered
on and off at 5 Hz and was presented for a total of 1 s, followed by
a 10-s blank retention interval. Next, observers were shown a
test stimulus (again flickering at 5 Hz for a total of 1 s) that either
was identical to or mismatched the sample along one feature
dimension (i.e., either the orientation or the color was slightly
different). Observers then used a custom-made button box to
make a two-alternative forced-choice response regarding whether
the sample and test stimuli matched. The sample and test
stimuli differed on 50% of trials. A 10-s intertrial interval fol-
lowed the offset of each test stimulus.

At the beginning of each block of trials (or scan), observers
were informed that the sample and test stimuli would differ along
only one feature dimension (color or orientation). Thus, re-
member-orientation and remember-color trials were run in
separate scans. Each scan contained eight trials, and each ob-
server completed seven or eight scans in each memory condi-
tion. Feedback (percentage correct) was given after each scan.

Behavioral Task
Stimuli were rendered on a light-gray background and displayed
via a rear-mounted projector (see Fig. 1). Observers were in-
structed to maintain fixation on a central square (subtending 1°
visual angle from a viewing distance of 58 cm) that was present
for the duration of each scan. At the beginning of each trial, ob-
servers were shown the sample, a Gabor stimulus (radius = 13°)
with a small circular aperture (2°) cut around the fixation square.
The sample stimulus was rendered in one of two orientations
(45° or 135°, plus or minus an additional offset randomly
selected from a range of ±10°) and in one of two colors (middle
red or middle green, plus or minus an offset randomly selected
from a range of ±10% saturation). The addition of random jitter
around the canonical features discouraged the use of verbal
labels, which would not have been precise enough to support
accurate performance in this task. For some observers (n = 2),
the spatial frequency of the stimulus varied randomly across
trials (0.75–1.25 cycles/deg); for the others, the spatial fre-

fig. 1. The behavioral paradigm. On each trial, a sample stimulus that
flickered at 5 Hz was presented for 1 s; observers were instructed to re-
member either the exact orientation or the exact color of this sample over
the following 10-s delay period. After the delay, a test stimulus was pre-
sewed, also for 1 s. The task was to indicate with a button press whether
or not the test stimulus matched the sample stimulus on the indicated
dimension. An exaggerated orientation-mismatch trial is depicted here
for demonstration purposes. The test stimulus was followed by a 10-s
intertrial interval (ITI).
(Engel et al., 1994; Sereno et al., 1995). This procedure was used to identify V1, a region known to respond to both color and orientation (e.g., Johnson, Hawken, & Shapley, 2001; Leventhal, Thompson, Liu, Zhou, & Ault, 1995; Sincich & Horton, 2005; Solomon & Lennie, 2007), as well as ventral (V2, V3, hV4) and dorsal (V2, V3, V3a) visual areas.

Voxel Selection
Independent functional localizer scans were used to identify regions of occipital visual cortex that responded to the spatial position occupied by the stimulus aperture in the main experiment. Colored Gabor stimuli identical to those used in the WM task were cycled on and off at 5 Hz for 10 s on each trial and followed by a 10-s passive fixation epoch. At the beginning of each localizer run, observers were instructed to attend to either the color or the orientation of the stimulus for the duration of that run. Color and orientation runs were presented in alternation. During each trial, two or three target events (a change in the value of the relevant feature dimension for 100 ms) occurred, and observers pressed a button whenever they detected a target event. Each localizer run contained 12 trials, and observers completed either one (n = 4) or two (n = 3) runs per attended feature dimension. We then used a general linear model (GLM) to identify voxels within each visual area that responded more strongly during epochs of stimulation than during epochs of passive fixation. The single regressor in the GLM was created by convolving a boxcar model of the stimulus protocol with a gamma function (Boynton, Engel, Glover, & Heeger, 1996). In each visual area, all voxels that passed a statistical threshold of p < .05 (corrected for multiple comparisons using the false-discovery-rate algorithm in BrainVoyager) were retained for further analysis.

Multivoxel Pattern Analysis
For the MVPA, we first extracted the raw time series from each voxel within each region of interest during a time period extending from 4 s to 10 s after the presentation of each sample stimulus. These time series were then normalized using a z transform on a scan-by-scan basis. Temporal epochs from all but one scan were extracted to form a “training” data set for the classification analysis; data from the remaining scan were de-extended from 4 s to 10 s after the presentation of each sample stimulus on each trial from the test scan (classification of color and classification of orientation were carried out separately, so chance for all comparisons was 50%).

This procedure was repeated using a hold-one-scan-out cross-validation approach, so that data from every scan were used as a test set in turn. The SVM’s overall classification accuracy for each observer was then defined as the average classification accuracy across all seven or eight permutations of holding one scan out for use as a test set (depending on the number of scans the subject completed). Classification accuracy was averaged across corresponding regions of interest in the left and right hemispheres because no significant differences were observed between left and right visual areas.

RESULTS

Behavioral Performance
Behavioral performance is summarized in Table 1. During scanning, discrimination thresholds did not differ between the two subtypes of either feature dimension (45° vs. 135°, red vs. green), and overall accuracy was maintained at the level predetermined by the staircase procedure.

Analysis of WM-Related Activation in Visual Cortex
Our data figures (Figs. 2, 3, and 4) are based on the 62 most responsive voxels within each visual area because this was the minimum number of voxels with significant activity across subjects and visual areas; however, our main conclusions are robust even when more or fewer voxels are included in the analysis.

The goal of this study was to use fMRI and MVPA to determine whether, as predicted by the sensory-recruitment hypothesis, there are stimulus-specific modulations in early regions of visual cortex while an observer is remembering a specific orientation or color. Although we examined the response properties of several visual areas in occipital cortex (V1, V2v, V3v, hV4, V2d, V3d, V3a), we focused on V1 because this region contains neurons that are selective for both orientation and color (Johnson et al., 2001; Leventhal et al., 1995; Sincich & Horton, 2005; Solomon &...
Lennie, 2007), making it the ideal region to test for orientation- and color-selective modulations during the storage of information in WM. Independent functional localizer scans were used to identify the 62 V1 voxels in each observer that were most selective for the retinotopic position occupied by the stimulus aperture in the WM task (see the Method section for our voxel-selection logic).

Before performing the MVPA, we compared the mean amplitude of the BOLD response in V1 (collapsed across all 62 voxels) during the delay period of the WM task with the mean amplitude of the BOLD response in a corresponding temporal epoch following the test stimulus (see Figs. 2a and 2b). This comparison controlled for low-level sensory factors, as the sample and test stimuli were essentially identical. Even though response amplitudes were slightly higher 10 s postsample than 10 s posttest, overall activation levels during the WM delay period were statistically indistinguishable from activation levels during the corresponding epoch following the test stimulus on both remember-orientation trials and remember-color trials.

Two-way repeated measures analyses of variance (ANOVAs) with delay type (WM vs. ITI) and time point (four levels, from 4 s through 10 s) as factors showed no reliable main effect of delay type, $F(1, 6) = 0.4, p = 0.55, \eta^2 = 0.06$, and $F(1, 6) = 0.27, p = 0.62, \eta^2 = 0.04$, respectively. We examined activation 4 s through 10 s poststimulus because this epoch should reflect activation associated with the retention of information in WM. No other retinotopically organized region showed sustained amplitude increases related to storing information in WM.

1No extrastriate area that we identified showed a heightened response while observers were remembering orientation or color (over a temporal window extending from 4 s through 10 s poststimulus, all $p$s > .2 for the main effect of WM). We do not rule out the possibility that some modest delay-period activity might be observed if enough subjects were scanned. However, our data do demonstrate that sustained WM-related modulations are relatively weak in primary and extrastriate visual areas (see also Offen et al., in press).
Because comparing activation during the WM delay period with activation during a “passive” ITI following the presentation of the test stimulus is potentially problematic, we also plotted the mean time series for 24 s following the presentation of each sample stimulus (as in Offen et al., in press). Activation levels fell back to baseline approximately 8 s after the onset of the sample stimulus (see Figs. 2c and 2d), in contrast to the sustained amplitude increases often observed in parietal and frontal cortex across the delay period in a WM task (e.g., see Fig. 1 in D’Esposito, 2007). Thus, at least in our study, V1 did not exhibit a robust sustained amplitude increase that is often associated with WM maintenance.

It is important to note that the lack of amplitude changes in V1 does not rule out the possibility that stimulus-specific patterns of
activation are involved in maintaining information in WM. Therefore, we used MVPA to determine if V1 exhibits a stimulus-specific activation pattern during the delay period, consistent with the predictions of the sensory-recruitment hypothesis. The MVPA approach is based on the assumption that some subregions of visual cortex contain submillimeter columns of neurons that are selective for different stimulus features, such as orientation. In comparison, fMRI voxels are large (~3 mm$^3$); however, if slightly more neurons within a voxel prefer a particular orientation than prefer other orientations, then that voxel may exhibit a weak but detectable response bias (Kamitani & Tong, 2005). By examining the distributed voxel-by-voxel activation pattern across a visual area such as V1, one can make inferences about changes in the underlying population profile, and pattern-classification algorithms can be used to predict the specific feature that an observer is viewing, attending, or (in our case) remembering (Haynes & Rees, 2005; Kamitani & Tong, 2005, 2006; Norman et al., 2006; Peelen & Downing, 2007; Serences & Boynton, 2007a, 2007b).

We therefore examined activation patterns in V1 during the delay period of the WM task (4–10 s following sample onset) to determine if information about the remembered feature was being actively represented. As predicted by the sensory-recruitment hypothesis, when observers were remembering the orientation of the sample stimulus, activation patterns in V1 discriminated stimulus orientation, but not stimulus color (see Fig. 3a). The complementary pattern was observed when observers were instructed to remember the color of the stimulus. A two-way repeated measures ANOVA with memory instruction (remember orientation vs. remember color) and stimulus feature (classify orientation vs. classify color) as factors yielded a significant interaction, $F(1, 6) = 21.4, p < .005, \eta^2 = .78$. In contrast, activation patterns associated with a corresponding temporal epoch following the test stimulus—which was physically identical and required a challenging discrimination without storage—did not support above-chance classification accuracy (see Fig. 3b), $F(1, 6) = 0.14, p = .72, \eta^2 = .02$, so that there was a three-way interaction of delay type (WM vs. ITI), memory instruction, and stimulus feature, $F(1, 6) = 6.9, p < .05, \eta^2 = .54$. Given that the test stimulus evoked a BOLD response whose amplitude was statistically indistinguishable from that evoked by the sample stimulus (see Fig. 2), these data suggest that the active discrimination of the test stimulus for 1 s was not sufficient to drive above-chance classification accuracy. Thus, we conclude that the stimulus-specific pattern of delay activity in V1 was a direct consequence of active maintenance in WM. Activation patterns in other retinotopically organized visual areas did not consistently discriminate the remembered feature of the sample stimulus (see Table 2 for classification accuracies).

The three-way interaction illustrated in Figure 3 was also significant when 80 V1 voxels were used to perform the classification, $F(1, 6) = 6.9, p < .05, \eta^2 = .53$. Thus, a similar pattern of classification accuracy is observed even when more than 62 voxels are considered in the analysis. Qualitatively similar results were obtained for pattern sizes ranging from 40 to 100 voxels as well. When a linear discriminant classifier based on the Mahalanobis distance between activation patterns was used to compute classification accuracy, the three-way interaction was again significant, $F(1, 6) = 10.8, p < .025, \eta^2 = .64$. Thus, the results were not idiosyncratically dependent on the use of an SVM. At first glance, it is striking that classification accuracy for orientation was below chance when participants were remembering stimulus color (see Fig. 3a). However, this effect was not robust across all activation pattern sizes, and we never observed below-chance color classification accuracy when participants were remembering orientation.

Together, these analyses suggest that sustained stimulus-specific patterns in V1 reflect active storage in WM, and are not a passive consequence of the attentive encoding of the sample stimulus. Although both the sample and the test stimuli required attentive processing, significant classification accuracy was obtained only during the WM delay period following the sample stimulus. These differences in classification accuracy cannot easily be explained by differences in general arousal or effort related to task demands, as the overall amplitude of the evoked BOLD response was roughly equivalent for the test and the sample stimuli. However, to provide additional support for sustained feature-selective modulations during the WM delay period, we repeated the classification analyses with and without data from the last time point in the delay period (i.e., 10 s after the onset of the sample; see Figs. 2a and 2b). If there were a sustained WM-related activation pattern, then data from the last time point in the delay period would contribute to classification accuracy. By contrast, if feature-selective activation patterns were not sustained across the entire delay period, then adding data from the last time point would not improve classification accuracy (and might even impair classification accuracy if the activation patterns associated with the last time point were dominated by noise).

As Figure 4 shows, classification of the remembered stimulus was significantly better when activation patterns associated with

---

**TABLE 2**

<table>
<thead>
<tr>
<th>Region</th>
<th>Orientation classification accuracy</th>
<th>Color classification accuracy</th>
<th>Orientation classification accuracy</th>
<th>Color classification accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>V2v</td>
<td>.431 (.050)</td>
<td>.543 (.079)</td>
<td>.515 (.043)</td>
<td>.422 (.053)</td>
</tr>
<tr>
<td>V3v</td>
<td>.469 (.063)</td>
<td>.545 (.036)</td>
<td>.534 (.053)</td>
<td>.412 (.064)</td>
</tr>
<tr>
<td>hV4</td>
<td>.426 (.085)</td>
<td>.578 (.035)</td>
<td>.388 (.048)</td>
<td>.471 (.040)</td>
</tr>
<tr>
<td>V2d</td>
<td>.507 (.071)</td>
<td>.552 (.060)</td>
<td>.437 (.073)</td>
<td>.469 (.065)</td>
</tr>
<tr>
<td>V3d</td>
<td>.563 (.055)</td>
<td>.571 (.053)</td>
<td>.453 (.055)</td>
<td>.413 (.054)</td>
</tr>
<tr>
<td>V3a</td>
<td>.595 (.078)</td>
<td>.461 (.046)</td>
<td>.487 (.077)</td>
<td>.526 (.092)</td>
</tr>
</tbody>
</table>

*Note: Standard errors of the means are given in parentheses.*
the last time point in the delay period were included. However, the addition of this last time point in the delay period did not alter classification accuracy for the nonremembered feature (see Fig. 4a). Thus, the mere addition of more data did not necessarily improve the performance of the pattern classifier. A two-way ANOVA with memory instruction (orientation vs. color) and time bin (4 s through 8 s vs. 4 s through 10 s) as factors confirmed that adding information from the last time point in the delay period (10 s poststimulus) selectively enhanced classification accuracy for the remembered feature, $F(1, 6) = 6.7, p < .05, \eta^2 = .53$. Furthermore, although including data from 10 s postsample clearly improved classification accuracy for the remembered stimulus feature, including data from 10 s posttest (at the end of the ITI) had little effect on classification of the test stimulus (see Fig. 4b). The two-way interaction of delay type and time bin was significant, $F(1, 6) = 6.2, p < .05, \eta^2 = .51$.

Finally, we repeated the analysis after removing data from the peak of the stimulus-evoked BOLD response (4 s; see Fig. 2) and used only data collected 6 s to 10 s poststimulus to classify the remembered feature attribute. Even when data from the peak were excluded, classification accuracy for the remembered feature was significantly higher than classification accuracy for the nonremembered feature, collapsed across remember-orientation and remember-color trials (.584 vs. .474), $t(6) = 2.9, p_{rep} > .05$. These control analyses support our conclusion that the interaction depicted in Figure 3a reflects the on-line maintenance of information in WM, rather than the aftereffects of a phasic sensory response.

Although our data suggest that maintaining information in WM gives rise to sustained feature-selective activation patterns in V1, a stronger prediction of the sensory-recruitment hypothesis is that the pattern of activation during the delay period will literally mimic the pattern of activation evoked during sensory processing of the same stimulus. To test this prediction, we trained a classification algorithm using data from the functional localizer scans that were initially used to identify visually responsive voxels in each subregion of occipital cortex (see Method). The stimuli used in the localizer scans were identical to those used in the WM study, except that they were presented continuously for 10 s instead of only 1 s, and WM was not required during the localizer tasks. Each subject completed two to four of these localizer scans (in half of the scans, they attended orientation and ignored color; in the other half, they attended color and ignored orientation). One SVM was trained using data from attend-orientation localizer scans, and another was trained using data from attend-color localizer scans. These SVMs were then used to predict the orientation or color that subjects were remembering on each trial during the main WM task. Collapsed across remember-orientation and remember-color trials, classification accuracy was $.607 (SEM = .42), t(6) = 2.5, p_{rep} > .05$. This above-chance classification accuracy demonstrates that the V1 activation pattern that is sustained during WM resembles the sensory-evoked response that is observed during sensory processing alone.

**DISCUSSION**

These results demonstrate that the maintenance of information in visual WM elicits stimulus-specific activation patterns in the same regions of visual cortex that encode the to-be-remembered sensory information. These activation patterns were specifically tied to the delay period, when active rehearsal in WM was required. Classification accuracy was not above chance following the test stimulus, which involved identical bottom-up stimulation and discrimination of the same stimulus dimension, but no WM load. In addition, the sustained activation patterns observed during the delay period were similar to patterns evoked by the continuous presentation of identical sensory stimuli, which suggests that early feature-selective visual areas are recruited to maintain a “copy” of remembered stimulus attributes, as opposed to a more abstract or categorical representation. In addition to providing these empirical results, this study demonstrates that MVPA is a valuable tool for answering questions about the neural mechanisms that mediate the storage of specific stimulus values in WM.

Finally, these findings are also relevant to the claim that the capacity of visual WM is determined by the number of individuated objects that have to be stored, rather than the total amount of visual detail contained within those items (Awh, Barton, & Vogel, 2007; Irwin, 1992; Luck & Vogel, 1997; Woodman & Vogel, 2008; Xu & Chun, 2006; Zhang & Luck, 2008). For example, Luck and Vogel (1997) showed that capacity estimates for objects defined by a single feature (e.g., color or orientation) were equivalent to capacity estimates for multifeatured objects (e.g., colored oriented lines). This suggests that capacity is determined by the number of objects that are stored, rather than by the total information load. Alternatively, other researchers have proposed that an obligatory set of core features, including attributes such as color and orientation, is maintained regardless of the observer’s intentions (Alvarez & Cavanagh, 2004). This hypothesis might explain Luck and Vogel’s observation of equivalent capacity estimates for single and multifeature objects if all the possible features were obligatorily stored even when only a single feature was relevant. However, our results (Fig. 3) reveal that early sensory areas selectively represent only behaviorally relevant features during a WM delay period and suggest that observers have top-down control over which features are stored (see also Olivers, Meijer, & Theeuwes, 2006; Woodman & Vogel, 2008).

**Acknowledgments**—This study was supported by start-up funds to J.T.S. from the University of California, Irvine; by a National Institutes of Health fellowship to E.F.E.; by National Science Foundation Grant 0617681 to E.K.V.; and by National Institutes of Health Grant R01 MH077105-01A2 to E.A. J.T.S. and E.A. conceived the experiment; J.T.S., E.A., E.F.E., and E.K.V. designed the experimental paradigm; E.F.E. collected the data; J.T.S. and E.F.E. analyzed the data; and J.T.S., E.A., E.F.E., and E.K.V. wrote the article.
REFERENCES


