Visual short-term memory: Activity supporting encoding and maintenance in retinotopic visual cortex

Markus H. Sneve a,*, Dag Alnæs a, Tor Endestad a, Mark W. Greenlee b, Svein Magnussen a

a Center for the Study of Human Cognition, Department of Psychology, University of Oslo, Norway
b Department of Experimental Psychology, University of Regensburg, Germany

A R T I C L E  I N F O
Article history:
Accepted 24 June 2012
Available online 6 July 2012

Keywords:
fMRI
Low-level vision
Multi-voxel pattern analysis
Sensory recruitment
Working memory

A B S T R A C T
Recent studies have demonstrated that retinotopic cortex maintains information about visual stimuli during retention intervals. However, the process by which transient stimulus-evoked sensory responses are transformed into enduring memory representations is unknown. Here, using fMRI and short-term visual memory tasks optimized for univariate and multivariate analysis approaches, we report differential involvement of human retinotopic areas during memory encoding of the low-level visual feature orientation. All visual areas show weaker responses when memory encoding processes are interrupted, possibly due to effects in orientation-sensitive primary visual cortex (V1) propagating across extrastriate areas. Furthermore, intermediate areas in both dorsal (V3a/b) and ventral (LO1/2) streams are significantly more active during memory encoding compared with non-memory (active and passive) processing of the same stimulus material. These effects in intermediate visual cortex are also observed during memory encoding of a different stimulus feature (spatial frequency), suggesting that these areas are involved in encoding processes on a higher level of representation. Using pattern-classification techniques to probe the representational content in visual cortex during delay periods, we further demonstrate that simply initiating memory encoding is not sufficient to produce long-lasting memory traces. Rather, active maintenance appears to underlie the observed memory-specific patterns of information in retinotopic cortex.

© 2012 Elsevier Inc. All rights reserved.

Introduction

Visual short-term memory (VSTM) is a cognitive process subserved by a set of neural processes responsible for the encoding, maintenance, and retrieval of task-relevant information (Cohen et al., 1997; Courtney et al., 1997; Pessoa et al., 2002; Xu and Chun, 2006). In a typical VSTM task, the encoding phase involves the presentation of visual stimuli from which an observer has to remember one or several features. The perceptual analysis of these sample stimuli recruits areas of the visual cortex, and current theories about the neural underpinnings of VSTM suggest that the maintenance of stimulus information, no longer available as sensory input, is supported by the same cortical areas (D’Esposito, 2007; Magnussen, 2000; Pasternak and Greenlee, 2005; Postle, 2006: Ranganath, 2006). Support for this ‘sensory recruitment’ theory of VSTM has been provided by recent studies investigating retention of elementary visual stimulus features. Using functional magnetic resonance imaging (fMRI), Harrison and Tong (2009), and Serences et al. (2009) demonstrated that details about remembered visual features (orientation and color) can be decoded from V1 during the short-term retention period using pattern classification techniques. Similarly, Sneve et al. (2011) demonstrated that cross-channel inhibition between different exemplars of a low-level visual feature (spatial frequency) occurs in V1, even when one of the competing stimuli had been presented several seconds earlier and was maintained as a memory representation.

Several studies have found neural correlates of VSTM encoding in parietal and frontal brain regions (e.g., Pessoa et al., 2002; Roth et al., 2006; Rypma and D’Esposito, 1999; Todd and Marois, 2004; Todd et al., 2011; Zanto et al., 2011), and the observed recruitment of retinotopic visual cortex during memory maintenance is likely a result of modulatory input from these higher-order areas. However, the transition from sensory processing, which is assumed to occur independent of task demands, to maintained memory representations in retinotopic cortex is poorly understood. Some fMRI studies point to a role of intermediate visual areas in the encoding of stimuli into VSTM, including the dorsal occipital cortex/V3a (Pessoa et al., 2002), and the lateral occipital complex (Xu and Chun, 2006), an object-selective part of the extrastriate cortex, late in the ventral pathway hierarchy (LOC; Goodale and Milner, 1992). The posterior LOC contains at least two retinotopic areas (Larsson and Heeger, 2006; Sayres and Grill-Spector, 2008), including one area which resembles early visual areas in showing selective responses to low-level visual features (LO1; Larsson et al., 2006). Whether these retinotopic lateral occipital areas are involved in the encoding of such features into memory representations is currently unknown.
It has been demonstrated that VSTM consolidation\(^1\) of simple items may be fulfilled within the first hundred milliseconds following sample offset when the task requires retention of stimuli for brief periods of time (e.g., Vogel et al., 2006). Feature information seems particularly fragile during this early memory interval, as it is vulnerable to masking stimuli (Pasternak and Zakas, 2003), and TMS to visual cortex (van de Ven et al., 2012) during the first 200 ms of the retention period. However, psychophysical evidence suggests that retention of low-level feature information for longer intervals engages visual cortex to a different degree during the first few seconds following stimulus presentation compared to later phases of the delay interval (Magnussen et al., 1998). In line with this observation, Sligte et al. (2008, 2009) demonstrated that the strength of VSTM representations can be modulated through retro-cueing up to 4 s into the retention interval. Memory encoding of visual information may thus involve visual cortex in a time-limited manner, separable from processes associated with sustained maintenance. Furthermore, in tasks with sufficiently long delay periods, the initial consolidation processes may be prolonged, or succeeded by supplementary encoding operations.

In the current fMRI-study, we have investigated the role of retinotopic cortex in the different component processes of a delayed orientation discrimination task with long (6–14 s) retention intervals. Our main goal was to elucidate the role individual visual areas play in establishing VSTM representations, and to test the hypothesis that encoding operations persist following the initial and brief consolidation process. We developed an experimental design that, within the general linear model (GLM) framework, allows the dissociation of encoding processes from the retention and retrieval phases. A common finding when adopting such a univariate fMRI analysis approach to VSTM tasks is large and consistent responses to the stimulus components, when adopting such a univariate fMRI analysis approach to VSTM encoding/maintenance on subsequent operations in visual cortex. If retention in visual cortex is the result of active maintenance operations, we should be able to decode sample orientation from the delay period when subjects have been cued to prepare for a memory test. However, in trials where subjects are cued to prepare for a discrimination test, we would expect poor classification performance, since the sample orientation has lost its behavioral relevance.

**Materials and methods**

**Participants and scanning sessions**

Six subjects (age 21–28; one female) took part in the study, which was approved by the Regional Committee for Medical and Health Research Ethics (South-East Norway). All subjects gave informed written consent to participate. Five of the participants were naïve with respect to the purpose of the study; the remaining participant was one of the authors (DA). Participants were thoroughly trained (several hours over two days) on the experimental task in a standard laboratory setting prior to experimental data collection. In the scanner, the subjects participated in three sessions of estimating task discrimination thresholds (see below, Threshold estimation). The main experiments comprised 10 scanning sessions per participant (8 sessions for one participant), distributed over several days (see below, Experimental procedure). Finally, we mapped the participants’ individual retinotopic maps and regions of interest (ROI) in two separate scanning sessions (see below, ROI localizers).

**Stimulus presentation**

All stimuli were generated in Matlab (MathWorks, Natick, MA) using Psychophysics Toolbox functions (Brainard, 1997; Pelli, 1997). Visual stimuli were back-projected on a screen inside the scanner by use of a F20 sx + DLP® digital projector (Projectiondesign, Fredrikstad, Norway) at a display resolution of 1400 × 1050. Participants viewed the screen at a total path length of 65 cm through a front surface mirror attached to the head coil. Auditory stimuli were presented to the participants’ headphones through the scanner’s intercom system. A trigger pulse from the scanner synchronized the onset of a trial to the beginning of the acquisition of a fMRI volume. Participants produced their responses using a MR-compatible subject response collection system (ResponseGrip®, NordicNeuroLab, Bergen, Norway).

**Experimental procedure, late-cueing experiment**

The trial sequence of the late-cueing experiment is presented in Fig. 1A. All visual stimuli were Gabor patterns, which are defined as a 2D patch of sinusoidal grating with luminance contrast tapered by a 2D circular-symmetric Gaussian function that subtended 20 degrees of visual angle. The phase of the sinusoid varied randomly between trials. The sinusoid had a maximum Michelson’s contrast of 0.6, and was tapered with a Gaussian kernel with a standard deviation of 2.5°. The stimulus’ full width at half maximum contrast was reached at a radius of 3.3° from center with this configuration. To avoid strong foveal activations, the center of the Gabor grating was masked by a background colored

---

\(^1\) We here use the term consolidation to describe the much studied, early phase (<200 ms) of transforming a perceptual representation into a memory representation (Vogel et al., 2006). Memory encoding is a term more used in the fMRI-literature, and refers to the full set of processes that occurs following a memory sample that can be dissociated from perceptual or maintenance-related processes.
Fig. 1. Experimental design and stimuli. (A) Late-cueing experiment. Dark gray tone symbol and frame around test stimulus represent memory discrimination trials. Light gray tone frame represents stimulus discrimination trials. Partial trials, in which the test stimulus never appeared, are not depicted in the figure. (B) Early-cueing experiment. Similar color conventions as for the late-cueing experiment. Note that the color frames around the stimuli were not present in the experiment, and are included only to facilitate references to other figures. Note also that the test differences have been exaggerated in this figure.

At the start of each trial, a white fixation dot (0.1° in diameter) was displayed at the center of the gray screen. Participants were instructed to fixate this dot as long as it was present. After 2 s, the sample stimulus appeared for 0.5 s in an annulus around fixation. The sample was presented with a spatial frequency of 1.5 or 3 cycles per degree (cpd); jittered from trial to trial with a random value in the range of ±0.1 cpd, and with an orientation centered on 45° or 135° (jittered with a random value in the range of ±5°). This jittering was applied to avoid the formation of permanent long-term representation of the stimuli and discourage the participants from using categorical or verbal strategies in the encoding of the sample (Lages and Treisman, 1998). Independent of these large changes in stimulus properties between trials, the participant’s task was always to remember the precise orientation of the sample grating. After a delay interval of 6–14 s (exponential distribution over 5 discrete intervals, mean duration in seconds = 7.9, STD = 2.4), the participant was presented with a tone cue (500 or 1000 Hz, duration 0.25 s). In 75% of the trials, the tone signaled that the upcoming task would require a comparison of the test stimulus and the memory representation of the sample (memory-based discrimination). In 25% of the trials, the tone instructed the subject to prepare for a difficult orientation-discrimination task for which the remembered information was irrelevant (stimulus-based discrimination). The cueing roles of the two tones were counterbalanced across subjects. Following the tone cue, there was short delay of 1–3 s (uniform distribution) before, in 75% of the trials, the cued test stimulus appeared for 0.5 seconds. In memory-based discrimination trials, the test stimulus (memory test) was identical to the sample grating, except for a slight clockwise or counterclockwise shift in orientation, corresponding to the individually estimated 75% memory discrimination threshold (see below, Threshold estimation). In stimulus-based discrimination trials, the spatial frequency of the test stimulus (discrimination test) was identical to the sample grating, but its orientation would now be shifted clockwise or counterclockwise relative to vertical. This shift corresponded to the individually estimated 75% stimulus discrimination threshold, meaning that the two tasks were approximately similar in their difficulty. The participants knew that, for both tasks, both directions of change (clockwise/counterclockwise relative to sample orientation or relative to vertical) occurred with equal probability, and their task was to indicate the direction of this shift with a button press. In 25% of the trials, no test stimulus followed after the tone cue. These partial trials were included to allow the disentangling of the BOLD responses to the tone cues from the responses to the test stimuli (Ollinger et al., 2001; Serences, 2004). The disappearance of the fixation point, either after a response had been provided in full trials, or when the available response time had expired (3 s after the onset/supposed onset of the test stimulus in full trials/partial trials), signaled the end of a trial. After the offset of the test stimulus there was an inter-trial interval of minimum 10 s to allow the hemodynamic response to return to baseline. Finally, the fixation point appeared again and signaled the start of a new trial.

Each run (duration: 13 min) of the experiment consisted of 32 trials, which was preceded and followed by a baseline measurement period of 20 s. A scanning session consisted of two runs, thus each participant completed 14 runs of the late-cueing experiment (one participant only completed 12 runs). Due to the presence of 25% partial trials, this produced 252 (216 for one participant) full memory-based discrimination trials and 84 (72) full stimulus-based discrimination trials per participant, out of a total of 448 (384) trials. To ensure that the participants always actively encoded the sample orientation into memory, even though this representation only was probed in ~56% of the trials, we made two precautions: 1) During training, partial trials were not included in the experimental structure, thus the memory representation was probed in 75% of the trials when the participant learned the task. 2) Thresholds were estimated before the participants were informed about the presence of partial trials in the experiment. They were also informed that they would be paid a bonus after the completion of the experiment if their accuracy on the memory-based discrimination did not fall below the expected 75% level. As can be seen from the behavioral results, all participants performed at the expected level during the experiment.

The different spatial frequency/orientation combinations and the direction of change during test were counterbalanced and presented in a randomized order. Similarly, the different delay durations (sample-cue/cue-test/cue-partial trial) and the different trial types were counterbalanced and spread equally over the entire experiment in a quasi-randomized fashion. The participants could thus not know in advance when a given trial would end or what kind of trial it would be.

Experimental procedure, early-cueing experiment

The trial sequence of the early-cueing experiment is presented in Fig. 1B. Its experimental structure was in essence a mirror-image of the late-cueing experiment: tone cues appeared 1–3 s after the sample (uniform distribution), and was followed by a delay interval of 6–14 s (exponential distribution) before the cued test stimulus appeared. The experimental details were otherwise identical to the late-cueing experiment, with a few exceptions: To allow investigation of potential sensory adaptation effects, the orientation of the discrimination test stimulus was not shifted around vertical as in the late-cueing experiment, but around the same diagonal as the sample stimulus (degree of shift: individually estimated 75% discrimination threshold). The participants’ task during the discrimination test was to indicate whether the orientation of the grating was closest to the vertical or the horizontal meridian. In 25% of the trials, no tone cue appeared after the sample. These partial trials...
were signaled by the disappearance of the fixation cross 3.25 s after the sample, and were followed by a long inter-trial interval.

Each participant completed 6 runs of the early-cueing experiment. This produced 108 memory-based discrimination trials and 36 stimulus-based discrimination trials out of a total of 192 trials.

Passive viewing trials

In addition to the main experiment, five of the participants completed three runs of passive viewing trials in which they were instructed to attend to the stimuli without performing any tasks. These trials were identical to the late-cueing experiment trials, with the exception that the tone cues did not have any predictive value (either test stimulus could follow after a cue), and the different cue types/test stimuli occurred with 50% probability. Furthermore, to ensure that the participants did not silently perform any task on the “test” gratings, the “memory test” stimulus had the same orientation as the sample in a trial, and the “discrimination test” stimulus was always vertical. The passive viewing runs produced 96 measurements of the sample stimulus, and 36 measurements of each of the test stimuli. Each tone stimulus was presented 48 times.

Delayed spatial frequency discrimination

To investigate whether the reported findings were specific to memory for orientation information, four of the participants and two additional subjects were tested on a version of the paradigm probing memory for spatial frequency. In these trials, the delay interval was held constant at 12 s and a memory test stimulus always appeared after the delay. Since there was no ambiguity about the task in this experiment, no tone cues were played before the presentation of the test stimulus. The experiment details were otherwise similar to the main experiments, with the exception that stimulus orientation was held constant in a trial, and spatial frequency between sample and test differed at the individually estimated 75% threshold level. Participants’ task was to indicate whether the spatial frequency of the test stimulus was higher or lower than that of the sample. Each participant completed three runs of the delayed spatial frequency discrimination experiment, producing 99 trial measurements in total per participant.

Threshold estimation

Individual 75% discrimination thresholds were estimated separately for the two tasks in the experiment, using an adaptive maximum likelihood procedure (QUEST; Watson and Pelli, 1983). The trial structure used to probe memory-based discrimination thresholds was similar to the main experiment, but the sample-cue interval was held constant at 7.9 s (corresponding to the mean duration of this interval in the main experiment). To estimate stimulus-based discrimination thresholds, the trial structure was modified to only contain tone cue and test stimulus. For both tasks, the QUEST-algorithm varied the orientation of the test stimulus around a reference orientation (sample orientation in a trial/vertical), aiming for the difference that produced discrimination accuracy of 75% Individual average threshold levels (estimated from 3 × 32 trials on each of the tasks) were used as orientation differences in the main experiment.

ROI localizers

Standard flickering checkerboard stimuli at maximum contrast (rotating wedges/expanding ring, 10 Hz flickering rate) were used in the retinotopic mapping session (Slotnick and Yantis, 2003). For the ROI localizer session, the stimulus was a flickering (10 Hz) radial checkerboard annulus, presented around fixation and scaled following the linear cortical magnification factor (Qiu et al., 2006). The annulus always had an inner radius of 1°, while the outer radius changed between 3.3° (corresponding to the full width of the Gabor stimuli in the experiment at half maximum contrast) and 5.5°. For the current study, only data from the small-diameter localizer were used to define ROIs. The two localizer stimuli were each presented 24 times over 3 runs—in separate epochs of 15 s duration and interleaved with 15 s rest periods. Both during the retinotopic mapping and the ROI localizer session, the participants were instructed to detect and count sudden changes in the brightness of the central fixation dot. Such changes happened at random intervals, and on average every 15th second.

fMRI methods

Imaging was performed with a Philips Achieva 3 Tesla whole body MR unit equipped with an 8-channel Philips SENSE head coil (Philips Medical Systems, Best, the Netherlands). The functional imaging parameters were equivalent across all scanning sessions: 28 transversally oriented slices (no gap) were measured using a BOLD-sensitive T2*-weighted echo-planar imaging sequence (repetition time (TR), 1500 ms; echo time (TE), 35 ms; flip angle, 74°; voxel size, 3 × 3 × 3 mm; field of view (FOV): 192 × 192 mm; interleaved acquisition). The slices were oriented to cover occipital cortex, parietal cortex, and dorsal frontal cortex, as well as the thalamus. At the start of each fMRI run, 6 dummy volumes were collected to avoid T1-saturation effects in the analyzed data. Anatomical T1-weighted images consisting of 180 sagittally oriented slices were obtained using a turbo field echo pulse sequence (TR, 8.125 ms; TE, 3.72 ms; flip angle 8°; voxel size 1 × 1 × 1 mm; FOV, 256 × 256 mm). We collected an anatomical volume every scanning session to facilitate between-session co-registration and to create a low-noise average anatomical volume for each participant to improve segmentation results.

Imaging data was pre-processed and analyzed using BrainVoyager QX software (version 2.3, Brain Innovation, Maastricht, The Netherlands), custom MatLab scripts, and the NeuroElf toolbox for MatLab (version 0.9c, www.neuroelf.net). To achieve optimal segmentation results, each participant’s individual T1-weighted images were corrected for spatial intensity inhomogeneities, co-registered, and averaged together to produce a single high-resolution anatomical volume for each participant. The white-gray matter boundaries of these volumes were estimated and bridges were removed. Based on the white-matter segment of each hemisphere, 3D-meshes of the cortical surfaces were created. The functional images were time and motion corrected, and co-registered (rigid body) against the individual whole-brain anatomical volume. Because ROIs were precisely localized for each participant, no spatial smoothing was applied, nor was the imaging data transformed into normalized space.

Visual field maps were separated based on the polar angle and eccentricity retinotopic maps. Phase encoded maps were computed using a linear cross-correlation analysis and projected on the corresponding inflated cortical surface. The borders of early visual areas V1, V2, V3, and V4, the dorsal area V3a/b, and the lateral areas LO1-2, were then drawn manually, following well-established guidelines (Larsson and Heeger, 2006; Wandell et al., 2007). The model representing the ROI localizer was specified using two regressors, each representing the onsets of one of the two differently sized flickering annuli. The regressors were modeled with durations of 15 s and convolved with a two-gamma model of the hemodynamic response function (HRF, time to response peak: 5 s; time to undershoot peak: 15 s). Low-frequency drifts were removed using a temporal high-pass filter (cutoff, 0.0125 Hz), and intrinsic autocorrelations were modeled. One set of ROIs in visual cortex was created, based on voxels within the defined visual areas that responded significantly to the small-diameter checkerboard annulus, corresponding to the high-contrast part of the experimental stimuli. The test was false discovery rate (FDR) corrected at p < .001.
For the experimental data, intensity time courses from all voxels within a ROI were high-pass filtered (cutoff, 0.008 Hz), normalized into percent signal change from baseline, and averaged. The resulting single time series from each ROI were then investigated with a set of univariate models. The statistical significance of the measured responses was estimated in group analyses, treating the different participants as random samples. Since the statistical analyses were performed on single time courses, no corrections for multiple comparisons were required.

**Univariate data analysis**

The first model investigating the late-cueing experiment consisted of six regressors representing the three stimulus events (sample, memory test, and discrimination test), the two tone cues, and the retention interval. The durations of the stimulus and tone cue predictors corresponded to their respective presentation lengths during the experiment, while the retention interval was defined as the period between the offset of the sample stimulus and the onset of the tone cue (varying from 6 to 14 s). Since the coarse stimulus properties varied from trial to trial (the sample grating was either left-tilted or right-tilted, and its spatial frequency was either “low” (-1.5 cpd) or “high” (3 cpd)), and these featural differences may produce different BOLD-responses, especially in early visual areas, we created a second version of the model in which each stimulus predictor was specified across orientations and spatial frequencies. This model consisted of 13 regressors, and let us evaluate the effect of these task-irrelevant stimulus changes between trials. For both models, the regressors were convolved with a two-gamma HRF and corrected for serial correlations using a pre-whitening approach.

A set of deconvolution models aimed at estimating the actual hemodynamic responses produced by the individual stimuli in the two experiments. Each stimulus (sample, test stimulus, and tone cue) was modeled by 13 stick predictors, covering the temporal extent of the expected response (one predictor per volume). After finding differential effects of the base spatial frequency on the responses produced by the visual stimuli, the sample and test stimuli were separated into low and high spatial frequency trials to increase model fit. The initial model thus consisted of 104 regressors. Additional models were also investigated, separating sample and memory test stimuli in the late-cueing experiment based on the number of preceding similarly oriented stimuli, and separating sample regressors based on the succeeding cue for the early-cueing experiment (see Results section). The peak responses from the resulting deconvolved BOLD time series were used as estimates of response strengths to the different stimulus types. The point in time of these peak responses varied from participant to participant (4.5–6 s), most likely due to individual differences in the hemodynamic responses (Aguirre et al., 1998).

For visualization purposes, event-related averages were computed for each ROI, separated over trial types and delay intervals. Event-related averages for partial trials were computed separately. Baseline was always calculated as the average of the intensity values at the onset of the sample stimulus and the two preceding volumes.

**Multivariate data analysis, late-cueing experiment**

The task-irrelevant manipulation of the categorical orientation (left versus right) and spatial frequency (high versus low) of the sample stimulus between runs, allowed us to investigate the dataset using multi voxel pattern analysis (MVPA), as implemented in BrainVoyager QX. For each participant, the voxel time courses in each ROI were z-normalized and concatenated across experimental runs. Next, an experimental protocol was set up, sorting trials across base orientation, base spatial frequency, and delay intervals. A linear support vector machine (SVM) was trained in discriminating activity patterns across voxels in a ROI that matched one orientation (or spatial frequency) over the other. The training set always consisted of trials with 6 s delay interval, constituting 50% (224 trials) of the total number of trials, and the SVM was trained on the mean value of the third and fourth volume after sample onset (corresponding to 3–6 s after sample onset). The generalization performance of the classifier was optimized using a 10-fold cross validation of different values of the penalty parameter C on the training set, in effect finding the penalty parameter that maximizes the SVM separation margin, while minimizing the amount of classification errors (Chang and Lin, 2001). The classifier was then tested on the remaining part of the dataset, i.e., trials with a delay interval of 8 s or more between the sample and the tone cue. To investigate how decoding accuracy developed throughout the delay period, we tested the classifier on the mean value of the first and second volume (0–3 s), the third and fourth volume (3–6 s), the fifth and sixth volume (6–9 s), and the seventh and eighth volume (9–12 s) following sample onset. Since this last test bin potentially could be affected by responses to tone cues occurring at 8 s, only trials with a delay interval of 10 s or more (112 trials) were used as test input data during the decoding of the 9–12 s interval.

Training of the orientation classifier and orientation decoding was performed separately for high and low spatial frequency trials, while spatial frequency training and decoding was performed separately for left- and right-tilt trials. Furthermore, decoding was performed separately on the first and the second half of the test input data. The resulting four decoding accuracy scores per interval per ROI were averaged within each participant, and later across subjects. One-sample t-tests were used to establish which intervals produced decoding accuracies significantly above chance-level (50%).

**Multivariate data analysis, early-cueing experiment**

Due to the lower number of trials in the early-cueing experiment, we adopted the approach used by Serences et al. (2009) for the MVPA investigation of this dataset. First, we extracted and z-normalized, run by run, the voxel time courses from the period 7.5–12 s following sample onset for each ROI. To avoid the inclusion of BOLD-activity related to the presentation of test stimuli, only trials with sufficiently long delay intervals, as well as partial trials (that were followed by a long inter-trial interval, and in which the test stimulus never appeared), were included in the analysis. Next, for each participant, we trained the SVM on discriminating orientation patterns from the extracted epochs in a hold-one-run-out cross-validation approach, until all runs had been used as test sets. The training and test sets were thus always independent of each other. The resulting six decoding accuracy scores were averaged within each participant, and decoding above chance-level was established using t-tests.

Memory test trials and discrimination test trials were analyzed separately. After excluding trials with short delay intervals, the remaining number of memory test trials that went into the analysis was 48 per participant. Due to the lower number of sufficiently long discrimination test trials in the experiment, the second analysis was run on discrimination test trials and partial trials combined (64 trials per participant). Note that the partial trials were similar to the discrimination test trials in that the absence of a cue, and the disappearance of the fixation point, signaled that no further memory encoding/maintenance was necessary in a trial.

**Results**

**Behavioral data**

Individual 75% discrimination thresholds and corresponding accuracy scores observed during the experiments on the two discrimination tasks are summarized in Supplementary Table 1. The average accuracy on memory discrimination trials was 75.7% (SD = 4.1%) in the late-cueing experiment, and 80.0% (SD = 3.73%) in the early-cueing experiment. The average accuracy on stimulus discrimination trials was 70.7% (SD =
2.7%), and 72.7% (SD=5.5%) in the late- and early-cueing experiments respectively. A repeated measures ANOVA revealed no main effect of delay interval length on accuracy during memory discrimination trials (late-cueing experiment: F(4,20) = 0.79, p = 0.54, early-cueing experiment: p = 0.73), or during stimulus discrimination trials (late-cueing: p = 0.34, early-cueing: p = 0.50). Furthermore, performance on the memory discrimination trials did not depend on whether the sample stimulus was left- or right-tilted (late-cueing: F(1,5) = 1.476, p = 0.28; early-cueing: p = 0.71), or whether its base spatial frequency was ~3 cpd or ~1.5 cpd (late-cueing: p = 0.64; early-cueing: p = 0.73).

**Univariate data analyses**

Fig. 2 shows a representative example of the developing activation time series across visual areas after the onset of the sample stimulus in the late-cueing experiment. In line with earlier studies of delayed discrimination of orientation (Harrison and Tong, 2009; Offen et al., 2009; Serences et al., 2009), the observed drop to baseline was confirmed in all visual areas by a GLM. As expected, all visual stimuli (sample stimulus/memory test stimulus/discrimination test stimulus) produced highly significant main effects in all ROIs (t(5)>5.1, p<0.004).

A second GLM was set up to reveal any effects from the coarse task-irrelevant stimulus changes across trials. Here, the sample and the test events in memory discrimination trials were sorted into four predictors each (~1.5 cpd/~3 cpd spatial frequency vs. ~45°/~135° base orientation), while the test events in stimulus discrimination trials were separated over spatial frequency only into two predictors. The estimated beta values for the sample predictors for each participant were used as input into a within-subject repeated measures ANOVA with two factors (orientation and spatial frequency), each consisting of two levels. There was no main effect of base orientation in any of the ROIs (F(1,5)<3.2, p>0.13), however the spatial frequency of the sample stimulus had a significant effect on the resulting response in all visual areas (F(1,5)>8.4, p<0.034), consistent with earlier observations in retinotopic cortex (Singh et al., 2000). The remaining deconvolution analyses were therefore set up with two sets of predictors for each of the grating stimuli, representing low and high spatial frequency trials separately. Due to the counterbalancing of jittered delay intervals and partial trials across trial types, we still achieved a sufficient complex distribution of inter-stimulus intervals to allow reliable disentangling of the individual BOLD responses.

In the late-cueing experiment, participants were always instructed to encode and maintain orientation information about a sample stimulus. Towards the end of a delay interval, they were cued to either perform a memory discrimination, or to perform an orientation discrimination task of similar difficulty but for which the remembered information was irrelevant. All stimulus events in a trial involved sensory processing of orientation information; however, the sample stimulus also had to be encoded into a memory representation. The deconvolved event-related responses to the different visual stimulus events in the late-cueing experiment are shown in Fig. 3A. We extracted the peak value of each individually estimated HRF and submitted these to a within-subject repeated measures ANOVA with two factors (spatial frequency and stimulus type) with two (~1.5 cpd/~3 cpd) and three (sample stimulus/memory test stimulus/discrimination test stimulus) levels, respectively. As is evident from Fig. 3A, we found main effects of stimulus type in V3a/b, LO1, and LO2 (F(2,10)>20.25, p<0.001), but no significant effects in any of the other visual areas. V3 did however show a trend towards an effect (p = 0.08). We did not find any main effects of the spatial frequency manipulation, but there was a significant interaction between stimulus type and spatial frequency in V3a/b and LO2 (F(2,10)>4.62, p<0.038), and a trend towards an interaction in LO1 (p = 0.07). Due to this interaction, we ran a one-factor (stimulus type only) repeated measures ANOVA on high and low spatial frequency trials separately, finding similar effects of stimulus type in V3a/b, LO1, and LO2 (p<0.01) independent of spatial frequency, and still no effects in any of the other visual areas. Finally, a set of paired-samples t-tests was performed to elucidate the factors behind the significant main effects. Since the effects of stimulus type were robust across spatial frequencies, we ran the tests on collapsed data. The results of the pairwise comparisons confirmed the visual evidence from Fig. 3A: the sample stimulus evokes significantly more pronounced activity than both the memory test stimulus (t(5)>4.67, p<0.01) and the discrimination test stimulus (t(5)>3.29, p<0.05) in V3a/b, LO1, and LO2 (two-tailed p-values, Bonferroni corrected for two directional tests). These results suggest that the neural populations in V3a/b and LO1/2 that are involved in sensory processing of the experimental stimuli show stronger and/or more extensive responses during memory encoding.

To investigate if the observed memory encoding processes in visual cortex led to prolonged recruitment and persisted after the sample stimulus had disappeared from the screen, we set up a GLM that investigated the early-cueing experiment, and the consequence of receiving information early in the delay interval that no further memory encoding is required. The GLM contained three separate sample stimulus regressors, depending on whether it was followed by a memory test cue (“remember”), discrimination test cue (“forget”), or no cue (partial trials). The different test stimuli were included as separate regressors to increase the model fit. A repeated measures ANOVA with two factors (visual area and sample-cue combination) comparing the peak values from the responses following the different sample conditions, produced main effects of sample-cue combination (F(1,5)=10.59, p<0.03) and visual area (F(6,30)>3.81, p<0.01), but no interaction (p=0.93). As can be seen in Fig. 4, the discrimination test cue (appearing 1–3 s after sample offset) led to weaker BOLD responses throughout retinotopic cortex following the sample stimulus. Since the task, until the cue appeared, was always to remember the sample orientation, and memory test trials were three times as likely as discrimination test trials, we find it likely...
that participants initiated encoding/maintenance processes before cue onset. Thus, the weaker response due to the discrimination test cue suggests that the ongoing process associated with the memory task is interrupted. We did not have sufficient trials in the discrimination test condition to separate trials based on the sample-cue interval; however, all trials were identical until cue onset, thus the observed interruption indicates that memory-related processing lasted at least 1 s after sample-offset, since this was the shortest sample-cue interval.

Neither of the two tone cues produced responses resembling canonical HRFs in the early visual areas V1–V4, according to the initial GLM (both experiments: $t(5) < 0.93$, $p > 0.39$). However, V3a/b, LO1 and LO2 did show significant effects or strong trends towards responding to the presentation of auditory cues indicating the upcoming task in the late-cueing experiment ($t(5) > 2.43$, $p < 0.06$). The responses to the two cue types did however not differ. Examples of responses to the two cue types are shown in Supplementary Fig. 1.

Fig. 3. Deconvolved responses to individual visual stimulus types. (A) Late-cueing experiment. The estimated response time courses are shown superimposed, revealing a significantly larger response to the sample stimulus compared to the test stimuli in LO1 (also representative for V3a/b and LO2, indicated with asterisks in the bar-plot). The bar-plot to the right shows extracted peak values to the three stimulus events from the remaining areas. The curves are averages of all six participants' individually estimated time courses. See Supplementary Fig. 3 for deconvolved time course figures from the remaining visual areas. (B) Early-cueing experiment. Similar effects are observed when the tone cues are presented early in the delay period. See Supplementary Fig. 4 for deconvolved responses from the remaining visual areas. Error bars in all figures represent SEM, corrected to not include variability associated with between-subject differences (Cousineau, 2005; Loftus and Masson, 1994).

Fig. 4. Effect of cue type on the sample response in the early-cueing experiment. The peak sample response is consistently weaker when followed by a cue indicating that no further memory encoding/retention is necessary (light gray). The apparent higher activity levels following the discrimination test cue is likely due to the lack of memory maintenance requirements in these trials, since maintenance is known to produce a drop to baseline following the initial encoding period (Offen et al., 2009). The bar-plot shows peak values from the different visual areas. Error bars represent SEM, corrected to exclude variability associated with between-subject differences.
Alternative explanations

An alternative mechanism that could explain the memory encoding effects in V3a/b and LO1/2 is repetition suppression and/or orientation-selective adaptation, observed as attenuated responses in visual areas to repetitions of stimuli, and larger repetition suppression effects to stimuli with similar orientations (e.g., Weigelt et al., 2008). We ran a set of additional analyses to investigate this possibility. First, we reasoned that if repetition suppression could explain the observed difference in response amplitudes between sample and memory test stimuli, this attenuation should be even more pronounced when a stimulus had been preceded by several gratings of similar orientation (particularly in the LOC: Sayres and Grill-Spector, 2006; Konen and Kastner, 2008). The average inter-stimulus and inter-trial intervals were of comparable lengths (~10 s and ~12 s, respectively), thus if one disregard the tasks associated with the different stimulus events, an experimental run can be viewed as a set of regularly spaced presentations of oriented stimuli. We set up a GLM in which sample and memory test stimuli were sorted into different regressors based on the number of preceding stimuli with similar orientation (left-/right-tilted). A sample stimulus could be preceded by 0, 2 or 4 (or more) same-oriented stimuli: zero if it came after an orthogonal memory test stimulus, a vertical discrimination test stimulus, or a partial trial; two if it came after a memory test trial with similarly oriented stimuli; four if it came after two or more similar memory test trials. Likewise, memory test stimuli could be preceded by 1 or 3 (or more) same-oriented stimuli: one (the sample) if the preceding trial contained orthogonally oriented stimuli or was a discrimination test trial/partial trial; three if it came after one or more memory test trials with similarly oriented stimuli. Tone cues and discrimination test stimuli (which had orientations around vertical) were modeled as separate regressors. The results from this analysis are presented in Fig. 6. ANOVAs confirmed that there was no amplitude differences in dependence on the number of repetitions, neither for the sample stimuli ($F(2,10) = 0.7, p > 0.8$) nor for the test stimuli ($F(1,5) = 2.37, p > 0.18$). Thus, we do not find indications of repetition suppression effects across trials, i.e., simply as a result of successive effortful processing of similarly oriented stimuli.

We further investigated if the observed memory encoding effects could result from within-trial, task-related, adaptation effects. Usually, to achieve sensory adaptation with briefly presented stimuli, the interval between adaptor and test stimulus has to be sufficiently short (<3 s; Boynton and Finney, 2003; Weigelt et al., 2012). The average sample-to-test ISI used in the late-cueing experiment was ~9.9 s, and stimuli were only presented for 0.5 s. However, since the task was to maintain sample information throughout the delay interval, and neurons in retinotopic cortex are thought to remain active during such maintenance periods (Harrison and Tong, 2009), the interval between neural representations of sample and test stimuli could in effect be short, and the effective adaptation period long. One prediction from this interpretation is that adaptation effects should disappear when participants are not required to maintain orientation information. In the early-cueing experiment, a cue could appear shortly after sample offset and inform the participant that no active maintenance was necessary to perform the upcoming discrimination task. Thus, in this experiment, the discrimination test cue effectively signaled the start of a new trial, unrelated to the memory task (and the decoding results show that no, or only negligible, orientation information is stored in retinotopic cortex when maintenance is not required). When comparing the peak values from the deconvolved event-related responses to the different visual stimulus events in the early-cueing experiment (Fig. 3B), we found similar results as for the late-cueing experiment. A within-subject repeated measures ANOVA produced main effects of stimulus type in V3a/b, LO1, and LO2 ($F(2,10) > 5.05, p < 0.03$), and a trend towards an effect in area V3 ($p < 0.08$). Importantly, the response to the sample stimulus was still stronger than the response to the discrimination test in the same areas ($t(5) > 3.46, p < 0.05$), showing that the suggested memory encoding effects are unlikely to be a result of sensory adaptation processes within a trial.

Multivariate analyses

Average decoding performance for the task-relevant feature orientation and the task-irrelevant feature spatial frequency in the late-cueing...
An initial repeated measures ANOVA with cue-condition and visual area showed that this information was irrelevant for the upcoming task, we decoded the role of retinotopic cortex in the different component processes contributing to the STM in visual cortex (Bau mann et al., 2008; Esteve et al., 2009; Serences et al., 2009; Offen et al., 2009; Harrison and Tong, 2009; Serences et al., 2009; Sneve et al., 2011), we here used a paradigm that was optimized for disentangling the unique responses associated with the different requirements of a trial. From our results, intermediate visual areas, both in the dorsal (V3a/b) and ventral (LO1/2) stream (Goodale and Milner, 1992), seem to play a particular role during memory encoding; a process which continues to take place for seconds after the stimulus-to-be-remembered has disappeared from the visual scene. Following this initial memory encoding phase, the average activity, as measured by the percentage signal change in each condition, is a necessary component.

**Passive viewing experiment**

The analysis of the passive viewing runs confirmed that the observed experimental effects were not a result of task-independent interactions between stimuli in a trial: no differences were found in any areas between the estimated peak responses to the different stimulus types when they were not associated with a task (one-factor repeated measures ANOVA, F(2,8) < 1.5, p > 0.28). In light of a recent study showing that repetition suppression is found also when attention is directed away from the stimuli (Larsson and Smith, 2012), we would expect to see indications of suppression also during passive trials if such stimulus interactions took place under the trial specifics of our paradigm.

A second ANOVA with two factors: experiment (late-cueing experiment / passive viewing) and stimulus type (sample / memory test / discrimination test), revealed a significant main effect of experiment type in visual areas V1–V4 and V3a/b (F(1,4) > 9.3, p < 0.04). As is evident from Fig. 7 (V1 being representative for V2–V4), this effect represents the lower absolute strength of responses produced during passive viewing compared with activity during experimental trials in these areas. No such effect was observed in LO1 and LO2, but unsurprisingly the analysis of these areas and V3a/b reflected a significant interaction between experiment type and stimulus type (F(2,8) > 6.5, p < 0.03), due to the strong response to the sample during memory encoding in the experimental trials only (Fig. 7).

**Spatial frequency experiment**

The results from the additional experiment, in which memory for spatial frequency rather than orientation was probed, paralleled the findings from the main experiments. Due to the constant ISI used in the delayed spatial frequency discrimination task, a deconvolution analysis approach was inappropriate. However, when comparing the peak values for the responses associated with the sample and the memory test stimulus, after baseline correcting the recorded signal following a similar routine as described above (Materials and methods section), we observed a significantly stronger response to the sample in V3a/b, LO1, and LO2 ((t(5)) > 3.64, p < 0.02), but no differences in V1–V4 (p > 0.12, two-tailed p-values). Event-related averages showing the group-averaged trial responses in the spatial frequency experiment are depicted in Fig. 8.

**Discussion**

The current study was conducted to enhance our understanding of the role of retinotopic cortex in the different component processes contributing to performance in a VSTM task. Unlike other recent fMRI studies of STM in visual cortex (Baumann et al., 2008; Ester et al., 2009; Harrison and Tong, 2009; Offen et al., 2009; Serences et al., 2009; Sneve et al., 2011), we here used a paradigm that was optimized for disentangling the unique responses associated with the different requirements of a trial. From our results, intermediate visual areas, both in the dorsal (V3a/b) and ventral (LO1/2) stream (Goodale and Milner, 1992), seem to play a particular role during memory encoding; a process which continues to take place for seconds after the stimulus-to-be-remembered has disappeared from the visual scene. Following this initial memory encoding phase, the average activity, as measured by the percentage signal change in each condition, is a necessary component.
measured using BOLD-fMRI, returns to baseline levels in cortical areas coding for stimulus position. However, distributed patterns of activity within these regions (except for LO2) still contain information about the remembered stimulus-feature, and this information persists until it is required to perform a mnemonic task. If the sample details turn out to be behaviorally irrelevant, the activation patterns necessary to infer qualities about the encoded feature fade or disappear, indicating that these patterns of information reflect effortful maintenance operations.

**Encoding phase**

The process of encoding visual information into STM representations has commonly been ascribed to higher-level areas in the frontal/pre-frontal and parietal cortex (e.g., D’Esposito, 2007). Furthermore, recent fMRI evidence suggests that the memory encoding phase can be functionally dissociated from maintenance operations in parts of frontal cortex (inferior frontal junction (IFJ); Todd et al., 2011), and posterior sensory cortex is modulated by IFJ in a top-down fashion when sensory input has to be established as memory representations (Zanto et al., 2011). A similar dissociation between memory encoding and maintenance has been demonstrated psychophysically with tasks involving retention of low-level visual features known to be processed by retinotopic cortex: choice reaction times on delayed discrimination tasks remain constant at short (<3 s) sample-to-test intervals, but increase as the interval gets longer, indicating a transition in representational character following successful encoding (Magnussen, 2000; Magnussen et al., 1998; see also Lalonde and Chaudhuri, 2002). Here, we report two different lines of evidence for the involvement of retinotopic cortex in memory encoding — both likely to reflect modulations through recurrent connections with higher-level areas.

First, in the early-cueing experiment, we find a general effect of interrupted memory encoding across visual areas, leading to weaker responses compared to when fulfilled encoding is required (Fig. 5). Whether this is due to effects in the highly orientation-sensitive V1 propagating across extrastriate cortex, or reflects unique processing in LO2 (Fig. 3A, while the passive viewing results are averages of the individually estimated time courses of the five participants who took part in the passive viewing experiment. The sample response in V3a/b and LO1/2 is only boosted relative to discrimination test when involved in a memory task.

![Passive viewing control experiment](image)

**Fig. 7.** Passive viewing control experiment. Comparison of deconvolved responses to individual visual stimulus types across experiments (solid lines: late-cueing experiment; dotted lines: passive viewing). The response differences between experiments shown in V1 are also representative for visual areas V2–V4, and LO2 is representative for LO1. The curves representing the late-cueing experiment are similar to the ones shown in Fig. 3A, while the passive viewing results are averages of the individually estimated time courses of the five participants who took part in the passive viewing experiment. The sample response in V3a/b and LO1/2 is only boosted relative to discrimination test when involved in a memory task.

![Event-related averages of trial activity in the delayed spatial frequency discrimination experiment](image)

**Fig. 8.** Event-related averages of trial activity in the delayed spatial frequency discrimination experiment. The curves show baseline corrected responses following the presentation of the sample stimulus, averaged across 6 participants. The memory test stimulus appeared after a memory interval of 12 s, and was followed by a 12-s inter-trial interval (stimulus onsets are indicated by gray vertical lines). Note that no tone cues, nor any stimulus discrimination test, were presented in this control experiment. The bar-plot to the right shows extracted peak values to the three stimulus events from the remaining areas. No significant differences in peak responses were observed in V1–V4, V3a/b and LO1–2 showed similar effects as for the delayed orientation discrimination experiments: a significantly stronger response to the sample compared to the test stimulus (indicated with asterisks). Error bars represent ± SEM across participants.
every area, is uncertain. In any case, the involvement of V1 suggests that memory encoding recruits neurons at the earliest levels of cortical sensory processing, and that the temporal window for memory encoding outlasts the stimulation period. The fact that interruption in the current study occurred 1 s or more into the retention interval, supports the hypothesis that ongoing processes associated with memory encoding persist past initial consolidation operations, which are suggested to last only a few hundred milliseconds (Vogel et al., 2006). Two recent reports shed some light on the potential mechanisms behind this prolonged encoding period. Sargent et al. (2011) demonstrated that postcuing of a briefly presented display improved reportability of the cued item when the cue followed 200 ms after stimulus offset, and found a neural correlate to this enhancement in early visual areas. A cue after 1 s also increased the fMRI-response to the cued item, but this increase in activity was not associated with any improvement reportability. This BOLD effect is similar to what has been observed when subjects are instructed to “refresh” a recently viewed visual stimulus (Johnson et al., 2007). In a VSTM task with long maintenance intervals, however, Sligte et al. (2008, 2009) showed how cues presented up to 4 s after sample offset can make items that otherwise would have been forgotten, accessible for report. The cues were most efficient when presented early (1 s) in the retention interval, indicating that the memory trace exists in a gradually less fluid state following consolidation. The memory representation thus seems particulary malleable the first 200 ms after sample offset, however up until a certain point in the retention interval the representation can be further strengthened/refreshed through continued readout, and this process can be observed as an increased BOLD-response.

The second observed encoding effect is the relative boost in activity that intermediate visual areas (V3a/b; LO1/2) show during sample encoding compared to when they are involved in effortful or passive sensory processing of identical stimuli. This modulation was observed on top of the effect of interrupted encoding, thus we are ignorant to whether it reflects recruitment due to the initial and brief consolidation phase, or the suggested readout processes. In any case, the fact that we found similar encoding effects in the same areas also when a different feature, spatial frequency, had to be remembered, points toward a more general role of these areas in the encoding of visual objects into VSTM. Earlier fMRI studies have found indications of the involvement of LOC and areas likely to contain V3a during memory encoding of visual arrays and objects (Pessoa et al., 2002; Xu and Chun, 2006). LO1/2 are posterior parts of LOC, the first area specialized for object perception in the ventral pathway (Grill-Spector et al., 2001). They are suggested to be organized hierarchically, with LO2 showing less sensitivity for low-level features, but more selectivity for objects than LO1 (Larsson and Heeger, 2006). We have not investigated other parts of LOC in this study, but we speculate that the effects we observe in LO1/2 are retinotopic components of object-based VSTM encoding. Similarly, although V3a is located in the dorsal stream, and not traditionally associated with object perception and categorization, it has also been shown to be responsive to objects (Konen and Kastner, 2008).

As discussed in Results section, we are well aware that an alternative mechanism that could be endorsed to explain these effects is repetition suppression (i.e., the observed differential responses in intermediate visual areas could actually be due to attenuated responses to the discrimination test control stimuli rather than increased responses to the sample stimuli). We tested this alternative explanation specifically without finding any indications that such mechanisms in fact are underlyong the observed effects. Earlier reported findings render this explanation yet more unlikely when compared to our pattern of effects: all experiments that find short-term adaptation effects for orientation in visual cortex also find similar or larger attenuation in V4 than in V3a (Fang et al., 2005; Mohr et al., 2009; for similar observations on an object-level, see Konen and Kastner, 2008) or LO-areas (Kourtzi and Hubel, 2005). Such a sample/discrimination-test difference is, however, non-existent in V4 in our experiments. Furthermore, it has been demonstrated that a 45° orientation shift, similar to the difference in orientation between the sample and the discrimination test stimulus in our late-cueing experiment, is equivalent to presenting a new object to LOC (Murray and Wojciulik, 2004) — that is, no suppression effects are observed in LOC when successive similar stimuli are tilted relative to each other by this amount.

It could perhaps be argued that memory encoding comprises an attentionally more demanding task than the baseline task (discrimination test), and that the observed effects in V3a/b and LO1/2 can be attributed to general attentional processes. We find this explanation unlikely, due to the frequent observation of attentional modulations of the fMRI responses also in early visual areas (Baluch and Itti, 2011; Gandhi et al., 1999; Li et al., 2008; McMains et al., 2007). When comparing the results from the memory experiment with the passive viewing trials, the relatively stronger responses observed in V1–V4 during the main experiment are likely to be a result of such attentional enhancement, due to the necessity of active processing of the stimuli when they are associated with a task. Interestingly, in our data, LO1/2 do not seem to be affected by the differences in task demands between the passive viewing trials and the experiment, with the exception of the strong response to the sample, reflecting the memory encoding effect. This lack of attentional modulation in LO1/2 between experiments is somewhat surprising, since a recent fMRI study, investigating the effect of spatial attention on responses in visual cortex during retinotopic mapping, found reliable increases in signal from lateral occipital areas when subjects attended to the location of the checkerboard stimuli (Bressler and Silver, 2010). Nevertheless, the observation of similar responses in LO1/2 to all stimuli, passively or actively processed, except for the memory sample, strengthens our interpretation of this effect as specific to memory encoding. V3a/b, on the other hand, show both increased activity during active processing and an enhanced response to the memory sample during the experiment. FMRI studies of human subjects have reported attentional modulations of stimulus-related responses in V3a (Li et al., 2008; Sylvester et al., 2009). As with LO1/2, however, the response to the sample in V3a/b is similar to the other items when viewed passively, and only enhanced when associated with a memory encoding task.

Retention and retrieval

Replicating the findings of Harrison and Tong (2009), we observed that all investigated areas, with the exception of LO2 (and V3a/b in some participants), held orientation information specific to the remembered stimulus throughout the delay interval in the late-cueing experiment. The MVPA approach furthermore demonstrated that only task-relevant featural aspects of the stimulus were maintained in these areas — spatial frequency decoding was only successful for a short period after sample presentation, and thus probably reflects a sensory trace of the irrelevant spatial frequency information inherent in the slow BOLD response to the stimulus. The observation of orientation-specific retention in LO1, but not LO2, is in accordance with studies on stimulus-selectivity of these areas, showing that adaptation to orientation takes place in LO1, but not LO2 (Larsson et al., 2006). This observation is in line with our interpretation that the observed memory encoding effects in LOC (and perhaps V3a/b) reflect feature-independent processing, and takes place on the object-level of representation. The MVPA results from the early-cueing experiment further strengthen the interpretation that successful decoding during delay intervals, where no stimulus is present, reflects active maintenance of featural information. Throughout orientation-sensitive visual cortex, above-chance classification accuracy was achieved when participants had reason to keep the memory trace alive. However, when participants were cued to prepare for a memory-unrelated task, the classifier tended to perform at chance when it was applied to the ensuing delay period.
We hypothesized that the contrast between the responses to the tone cues towards the end of the maintenance interval in the late-cueing experiment would provide us with information about retrieval processes in retinotopic cortex. The monkey homologue of LOC; inferior temporal cortex (ITC), has been associated with several memory operations, including the reinstatement / strengthening of memory representations towards the end of the retention interval (Meyers et al., 2008; Woloszyn and Sheinberg, 2009). Both tone cues in the late-cueing experiment evoked responses in the same visual areas showing a memory encoding effect (V3a/b, L01/2). However, we did not find any differential responses in these areas based on the predictive content of the cues; the strength and shape of the observed BOLD response was independent of whether the participants were instructed to prepare a memory-based discrimination (suggesting they should retrieve the memory representation) or a stimulus-based discrimination (suggesting they should drop the remembered information). The finding that auditory cues can produce increased BOLD responses in visual areas has been reported previously (Silver et al., 2007; Sylvester et al., 2007), and probably reflects allocation of spatial attention to the expected upcoming stimulus. We do not find these effects in early visual areas in the current study — most likely since no spatial shifts of attention were required in our task, as it has been shown that attentional allocation affects LOC, but not V1–V3 when the same position is cued throughout an experiment (Murray, 2008; Murray and He, 2006).

Concluding remarks

Successful decoding during the delay period indicates that information about the remembered orientation is present in the measured neural populations. However, on a voxel-level, activity drops to baseline levels throughout retinotopic cortex, probably reflecting that the average activity levels in the underlying populations of neurons are close to null. This apparent contradiction could be a result of inhibitory processes between neurons coding for task-relevant and task-irrelevant features (Magnussen et al., 1991). Activity levels in populations coding for dimensional values other than the one to remember represent noise, and the system solving the task would thus benefit from the silencing of these neurons (Serenes et al., 2009). From our results it seems like such noise (and fMRI signal) cancellising processes take time to build up, since univariate effects clearly are visible during the early delay interval.

Recent theories about encoding and retention of VSTM representations suggest that the mechanisms underlying sustained attention to objects or features also are the driving forces behind maintained sensory recruitment during memory operations (Chun and Johnson, 2011; Gazzaley and Nobre, 2011). However, maintained attention to low-level featural information produces sustained activity levels in visual cortex, but memory storage does not (e.g., Offen et al., 2009). Whether this indicates that selective attention to internal perceptual representations (Chun et al., 2011) and VSTM in fact are qualitatively different processes, or that they only differ in the degree of sharpening of task-relevant neural populations, is to our knowledge unknown. In any case, the hypothesis that inhibitory processes cause the observed drop to baseline during memory intervals has yet to be empirically tested.

Acknowledgments

We would like to thank Kartik Sreenivasan and Torgeir Moberget for helpful comments on the manuscript. This work was supported by a grant from the University of Oslo to Center for the Study of Human Cognition.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2012.06.053.

References

Murray, S.O., 2008. The effects of spatial attention in early human visual cortex are stimulus independent. J. Vis. 8, 2, 1–2.11.


