Feedback signals can explain spike-silent working memory traces in visual cortex

Krause, Noa noa.krause@esi-frankfurt.de
Ernst-Struengmann Institute for Neuroscience, Deutschordenstrasse 46, 60528 Frankfurt am Main, Germany
Rademaker, Rosanne rosanne.rademaker@esi-frankfurt.de
Ernst-Struengmann Institute for Neuroscience, Deutschordenstrasse 46, 60528 Frankfurt am Main, Germany

Abstract

Storage of visual memories is known to engage early visual cortex (EVC), where mnemonic representations are presumed to be spike-silent. Possibly, such spike-silent representations rely on short-term synaptic plasticity (STSP) traces laid down by previous sensory inputs. However, STSP cannot account for selection of information from working memory on the basis of external information, and is not robust against visual distraction. As such, it fails to explain data from neuroimaging studies in humans, which shows that representations recovered from EVC are mnemonic in nature, that is, only task-relevant information can be recovered. Additionally, these traces return even after temporary quenching from visual distraction. We show that feedback projections from more anterior cortical sites, known to reflect memories via sustained spiking, can explain these findings that STSP cannot account for.

Keywords: visual working memory; neural mechanisms; STSP; feedback; early visual cortex

Introduction

Early visual cortex (EVC) has repeatedly been implicated in visual working memory (VWM; Harrison & Tong, 2009; Serences et al., 2009). Memory representations appear to be stored in spike-silent formats but can be recovered from EVC via neural perturbations (Wolff et al., 2017). This phenomenon can be explained with short-term synaptic plasticity (STSP; Pals et al., 2020; Mongillo et al., 2008), which posits that visual stimuli leave a slowly decaying trace of potentiation in EVC that can be reactivated using perturbations.

Here, we argue that STSP is a passive mechanism that fails to account for two hallmarks of VWM: The ability to flexibly select from within the memory store, and the ability to resist distraction. We demonstrate these shortcomings in two neural network simulations and show that an alternative model based on feedback signaling from frontal areas (that encode VWM contents in persistent spiking activity; Mendoza-Halliday, 2014) to EVC can explain spike-silent storage and reactivation via perturbations, as well as flexible selection from VWM and distractor robustness.

Methods

We simulated competing neural networks equipped with STSP or feedback to test their performance as VWM mechanisms. Both models consisted of an EVC and a frontal module, each comprised of two circuits (one per cortical hemisphere; Figure 1a). All circuits were modeled using the same ring-shaped architecture (Wimmer et al., 2014; Compte et al., 2000) consisting of interconnected excitatory (E) and inhibitory (I) neural populations (N=180). E-neurons were placed on a ring according to their preferred orientation. Recurrent connectivity was structured by a von Mises ($\kappa = 5$), such that a neuron's excitation to other E-neurons diminished as a function of their difference in orientation preference. Connections from E- to I-, I- to I-, and I- to E-neurons were uniform (all-to-all). EVC circuits were tuned to respond transiently, frontal circuits were tuned to produce attractor dynamics.

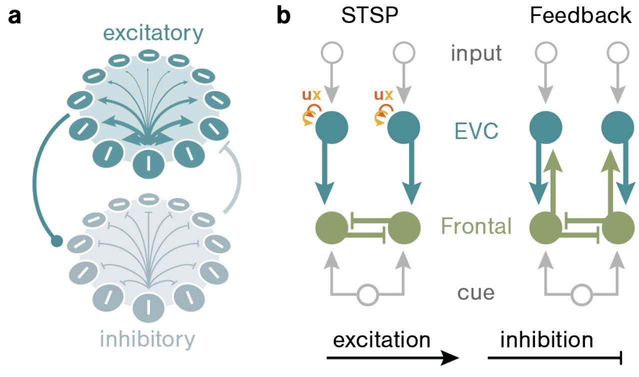

Figure 1: **a** Circuit architecture. **b** Model architectures.

Input to the model consisted of Gabor patches of varying orientations, which were multiplied with EVC neurons' receptive fields. After encoding, EVC neurons projected information to the frontal module, where information was maintained in persistent activity. The crucial difference between models was the mechanism by which silent neural traces were maintained in EVC (Figure 1b): In the STSP model, the information is maintained locally via changes in synaptic connectivity on the basis of recent firing history. In the feedback model, the frontal module projects back to EVC via a feedback connection providing continuous weak excitatory drive to neurons in EVC.

Results

Feedback model can explain spike-silent storage

To first test that both STSP and feedback mechanisms can lead to spike-silent storage in EVC in the absence of external drive, we presented both models with a Gabor patch and perturbed EVC neurons after a short delay. We replicate findings that STSP involves spike-silent representations (decoding decays to baseline after stimulus offset) that can be reactivated using perturbations (temporary return of decoding; Mongillo et al., 2008; Pals et al., 2020) and additionally show that feedback signals to EVC could account for this effect as well (Figure 2).

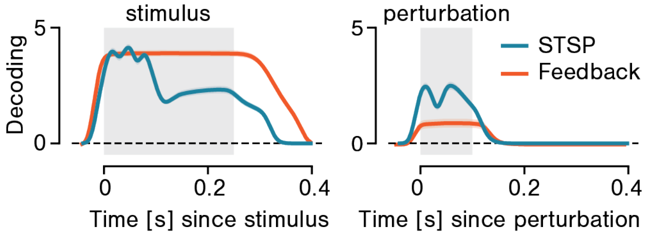

Figure 2: Decoding during stimulus and perturbation period.

Feedback model can explain selection from VWM

To test selection in VWM, we simulated a retro-cue VWM task in which the models remembered two simultaneously presented orientations until a retro-cue marked only one as relevant (Figure 3a). The cue comprised of a uniform current to all frontal E-neurons of the 'cued' circuit (representing the cued orientation). Via mutual inhibition between left and right frontal circuits, activity is deleted in the uncued circuit. In the STSP model, the cue does not propagate back to EVC due to the lack of feedback connection. Without this connection, the model cannot account for selection from VWM (Figure 3b). However, keeping the feedback connection intact, and instead inducing EVC-like tuning in frontal circuits, also did not lead to selection of the cued orientation. Finally, applying the cue directly to the EVC circuit representing the cued orientation weakened the cued representation instead of selecting it.

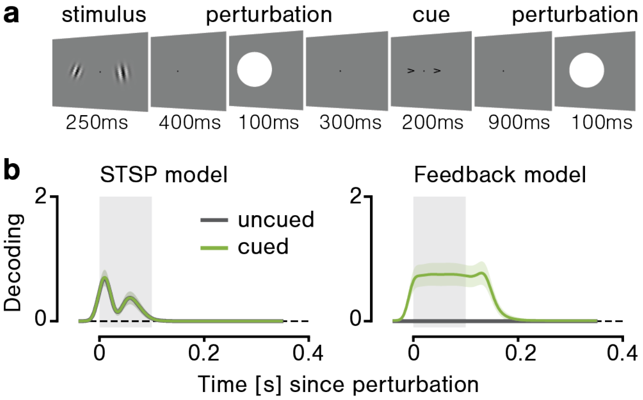

Figure 3: **a** Task-design. **b** Decoding cued and uncued stimuli from EVC in response to a perturbation shown after selection.

In contrast, the feedback model performs in line with experimental results (Wolff et al., 2017). After deletion of the uncued representation from its frontal circuit, feedback from this circuit terminates, and the perturbation reveals only the cued item in the EVC module.

Feedback model is distractor-resistant

To test for resistance against visual distraction, we simulated a task with two targets and two distractors, each followed by a perturbation (Figure 4a), expecting the models to remember only the targets. In the STSP model, targets and distractors both instilled neural traces in EVC, and distractor traces were even stronger than target traces at the time of the second perturbation (Figure 4b). This suggests that the strength of the STSP trace is governed by stimulus recency, not stimulus relevance.

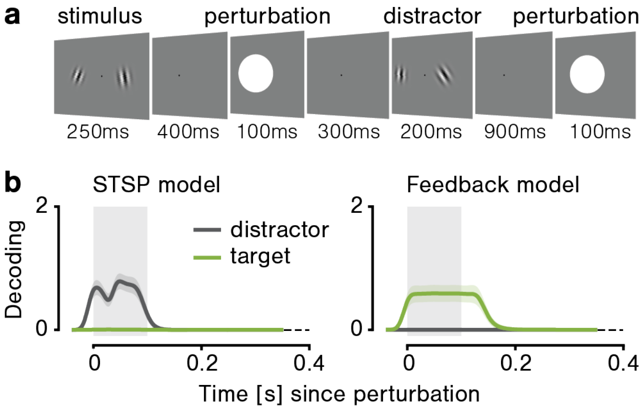

Figure 4: **a** Task-design. **b** Target and distractor decoding in response to second perturbation from EVC.

In contrast, the ring attractor used in the frontal module of the feedback model displays some robustness against distraction (already discussed by Compte et al., 2000). Once a stimulus is encoded, it is difficult for new information to penetrate the circuit. In line with this, the distractor was never encoded into the frontal circuits, which therefore did not project information about the distractor back to EVC. From the second perturbation, only information about the target, not the distractor, can be read out from EVC.

Discussion

We demonstrate that STSP is unlikely to be the sole mechanism underlying VWM traces in EVC, given that it cannot easily account for flexible selection and deletion of information from VWM and is not robust against distraction. In contrast, a model that includes sustained suprathreshold VWM maintenance in a higher-level area like frontal cortex allows more flexibility for selection and competition among information in VWM. Our work exemplifies one of many possible mechanism by which that can be achieved, namely, maintenance in a ring attractor with mutual inhibition between individuated circuits maintaining VWM information. Irrespective of the exact storage mechanism, we show that sub-threshold feedback signals from higher-level cortical areas can account for the spike-silent EVC storage and reactivation through perturbation reported in human neuroimaging studies.

References

Compte, A., Brunel, N., Goldman-Rakic, P. S., & Wang, X.-J. (2000, September). Synaptic Mechanisms and Network Dynamics Underlying Spatial Working Memory in a Cortical Network Model. *Cerebral Cortex*, *10*(9), 910–923. doi: 10.1093/cercor/10.9.910

Harrison, S. A., & Tong, F. (2009, April). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–635. doi: 10.1038/nature07832

Mendoza-Halliday, T. S. . M.-T. J., Diego. (2014, April). Sharp emergence of feature-selective sustained activity along the dorsal visual pathway. *Nature Neuroscience*(7238), 1255-1262. doi: 10.1038/nn.3785

Mongillo, G., Barak, O., & Tsodyks, M. (2008, March). Synaptic Theory of Working Memory. *Science*, *319*(5869), 1543–1546.

Pals, M., Stewart, T. C., Akyürek, E. G., & Borst, J. P. (2020). A functional spiking-neuron model of activity-silent working memory in humans based on calcium-mediated short-term synaptic plasticity. *PLoS Computational Biology*, *16*(6). doi: 10.1371/journal.pcbi.1007936

Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009, February). Stimulus-Specific Delay Activity in Human Primary Visual Cortex. *Psychol Sci*, *20*(2), 207–214. doi: 10.1111/j.1467-9280.2009.02276.x

Wimmer, K., Nykamp, D. Q., Constantinidis, C., & Compte, A. (2014, March). Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nature Neuroscience*, *17*(3), 431–439. doi: 10.1038/nn.3645

Wolff, M. J., Jochim, J., Akyürek, E. G., & Stokes, M. G. (2017, June). Dynamic hidden states underlying working-memory-guided behavior. *Nature Neuroscience*, *20*(6), 864–871. doi: 10.1038/nn.4546