

Neuronal population coding of parametric working memory

Omri Barak^{1*}, Misha Tsodyks¹⁺, Ranulfo Romo²⁺

¹ Department of Neurobiology, Weizmann Institute of Science, Rehovot, Israel

² Instituto de Fisiología Celular-Neurociencias, Universidad Nacional Autónoma de México, 04510 México, D.F., México.

* Current address: Center for Neurobiology and Behavior, Columbia University College of Physicians and Surgeons, New York, NY 10032, USA

⁺ Corresponding authors (misha@weizmann.ac.il and rromo@ifc.unam.mx)

- Abbreviated title: Population coding of working memory
- Number of figures and tables : 5 figures, no tables.
- Contents of supplemental material (if applicable): Supplemental text and figures.
- Number of pages: 24 (including title page)
- Number of words for Abstract (133), Introduction (452), and Discussion (1236)
- Six keywords : Prefrontal cortex, single neuron recordings, remapping, delay activity, synaptic facilitation, monkeys
- Acknowledgements: We thank B. Blumenfeld, C. Brody, S. Fusi and L. Abbott for helpful comments on the manuscript. O.B. thanks Ranulfo Romo and all members of the Romo's lab for the heartwarming hospitality and helpful discussions on his visits there. O.B. is supported by the Azrieli Foundation. M.T. is supported by the Israeli Science Foundation. The research of R.R. was partially supported by an international Research Scholars Award from the Howard Hughes Medical Institute and grants from CONACYT and DGAPA-UNAM.

Abstract

Comparing two sequentially presented stimuli is a widely used experimental paradigm for studying working memory. The delay activity of many single neurons in the prefrontal cortex (PFC) of monkeys was found to be stimulus-specific, however, population dynamics of stimulus representation has not been elucidated. We analyzed the population state of a large number of PFC neurons during a somatosensory discrimination task. Using the tuning curves of the neurons, we derived a compact characterization of the population state. Stimulus representation by the population was found to degrade after stimulus termination, and emerge in a different form towards the end of the delay. Specifically, the tuning properties of neurons were found to change during the task. We suggest a mechanism whereby information about the stimulus is contained in activity dependent synaptic facilitation of recurrent connections.

Introduction

Working memory refers to the ability to maintain and manipulate items for short periods of time (Baddeley and Hitch, 1974). A particular example is the sequential comparison of two vibrational stimuli separated by a delay period of a few seconds (Hernandez et al., 1997), which forces the subject to maintain an analog value in memory (frequency of vibration). Single cell activity in the prefrontal cortex (PFC) during this task was extensively analyzed (Miller et al., 2003; Brody et al., 2003; Romo et al., 1999). It was found that the firing rate of many of the cells either increases or decreases monotonically as the stimulus frequency increases. This positive or negative tuning was found both for the firing rates during the stimulus and for various points in time during the delay period following the stimulus offset. Importantly, while throughout the delay there is always a certain subpopulation of neurons that are tuned to the stimulus, this subpopulation is not constant as is evidenced by the repertoire of temporal evolutions of single cell tuning properties (Jun et al., 2010).

The diversity of single neuron behaviors indicates that population encoding of stimuli in working memory exhibits nontrivial evolution during various stages of the task. We believe that elucidating how the population activity states represent the stimulus and how this representation evolves during the task is a crucial step to understand the neuronal mechanisms of working memory. In this contribution, we analyze a large data set of single-neuron recordings, in order to characterize the population encoding of stimulus during the delay period of a working memory task. We derive a compact description of the population state, specified by a small number of parameters. This compact description allows one to follow the dynamics of population encoding during the task. Surprisingly, we find that population state consistently reflects the vibrational frequency during the stimulus and towards the end of the delay, but less so during other parts of the delay. Moreover, we observed very substantial transformation of the population states between the stimulus and the end of the delay, indicating two almost independent PFC representations of frequency. These findings challenge the standard assumption that information in the working memory is solely encoded in spiking activity of dedicated neuronal populations. In our recent contribution (Mongillo et al., 2008) we proposed that information about the stimulus is maintained in the facilitation level of the synaptic

connections. To illustrate how this idea could be adapted to memory of analog variables, such as frequency, we propose a simplified circuit model, which is based on the assumption that stimulus frequency is encoded in synaptic facilitation during stimulation, and is subsequently extracted into the population activity state during the delay period of the task.

Materials and Methods

Discrimination task

Two monkeys (*Macaca Mulatta*) were trained to perform the vibrotactile task as depicted in Figure 1 (Hernandez et al., 1997). Briefly, the monkey places the tip of its finger below a mechanical probe. A trial begins with the probe indenting the finger (Probe Down – PD), followed by the monkey grasping a metal bar with his other hand (Key Down – KD) to signal readiness. After a variable delay of 1500-3000 ms, the first stimulus is applied for 500 ms, followed by a 3000 ms delay and the second stimulus. The monkey then releases the bar (Key Up – KU) and presses one of two buttons to indicate whether the first stimulus frequency was larger than the second one or vice versa (Push Button – PB). The stimuli are sinusoidal vibrations of the probe with frequencies of 10-34Hz which are in the flutter range (Mountcastle et al., 1990). The amplitude is adjusted to maintain constant subjective amplitude across frequencies (Hernandez et al., 1997; Mountcastle et al., 1990).

It might be argued that for some stimulus conditions (e.g. 10 vs. 18 Hz) the monkey reaches a decision upon receipt of the first stimulus, and does not engage memory circuits. A detailed argument against this interpretation appears in the supplementary material.

During part of the recordings, visual control trials were also performed. In these trials the vibrotactile stimuli were as in the discrimination task, but a light indicating the correct response was active from the Probe Down until the Probe Up.

Recording Sessions and Sites

Neurophysiological recordings began after training was complete with an array of seven independently movable microelectrodes (2-3 $M\Omega$) (Romo et al., 1999; Hernandez et al., 1997) separated from each other by 500 μm and inserted in parallel into the inferior convexity of the PFC of the right hemisphere of both monkeys, and also of the left hemisphere of one of the monkeys. Recording sites changed from session to session. After advancing the electrodes into the cortex, window discriminators were used to isolate one neuron per electrode while the monkey was performing the task.

Data Analysis

Offline analyses and statistical tests were performed using custom software written for Matlab (The Mathworks, Natick, MA). Trials were excluded from analysis if more than 1% of the inter spike intervals (ISI) were less than 1msec. Neurons for which more than 10% of the trials met this criterion were removed. In order to exclude neurons which did not respond at all, but retain those which responded only in a specific part of the trials, mean firing rates for each of the following epochs were calculated: The entire trial; Probe down to stimulus; First stimulus; Delay period; Second stimulus; End of second stimulus to end of trial. Neurons in which all mean firing rates were below 1Hz were removed. We then took the remaining neurons which had at least 20 trials, and split their trials into blocks of 10 (each block covers all stimulus conditions). Stability of the recording was assessed by performing a t-test on the firing rates of the first vs. the second half of the blocks. Neurons for which the p value was less than 5% were removed. This procedure identified 899 neurons - 26 neurons from the right hemisphere of one monkey, and 782 (91) neurons from the right (left) hemispheres of the second monkey respectively.

Firing rates of all neurons were calculated for each stimulus frequency in non overlapping 100 ms bins, by averaging the number of spikes emitted by the neuron over all trials with this frequency. The vector of all 899 firing rates at a given time bin is defined as the instantaneous population state. All figures show binned data without any smoothing.

The correlations between population states (Figure 2) are sensitive to the number of trials used to calculate average firing rates. In order to control for this effect we randomly removed trials until each experimental condition had the same total number of trials sampled (5194 trials per condition).

Labeling cells according to their tuning

In order to attain a compact description of the population state, cells were characterized according to their stimulus tuning in the following way: The firing rate of the cell during the second bin of the first stimulus, r , was fitted to the actual stimulus frequency, f , according to a linear equation:

$$r = a_0 + a_1 f \tag{1}$$

The a_1 values from this linear regression were called significant if the linear regression had a p-value of less than 0.05. Note that this value was used even for cells in which a sigmoidal function might have given a better fit (Brody et al., 2003; Jun et al., 2010).

Early, late and persistent neurons

In the Discussion we compare our current findings with a previous classification of neurons in the literature of this task, namely: early late and persistent neurons (Romo et al., 1999). To this end we defined significant time bins for each neuron as those 100 msec bins in which the linear fit of the firing rate to stimulus frequency (Eq. 1) was significant with $p < 0.05$. Early neurons were defined by having at least one significant bin in the first second of the delay period and none in the last second. Late neurons had at least one significant bin in the last second and none in the first. Persistent neurons had at least one significant bin in each of the three seconds.

Model

Our simplified model network, illustrated in Figure 5A, consists of two layers each having 3 units, representing populations of neurons that are negatively, positively, or non-significantly tuned to stimulus frequency. The first layer corresponds to the second time bin of stimulus presentation (sensory representation) and the second layer - to the last bin of the delay (memory representation). We further assume that neurons with significant tuning have higher excitability compared to non-tuned neurons. Note that every single neuron is represented in both layers, corresponding to its sensory and memory tuning. So for example a neuron that is positively tuned during the stimulus and not significantly tuned at the end of the delay will be placed in the lower right and upper middle populations. Connections in the network are going from sensory level populations to corresponding populations in the memory layer with the same tuning, i.e. we assume that neurons with positive sensory tuning project selectively to neurons with positive memory tuning, etc. Note that even though the connectivity pattern appears to be feedforward, the underlying network is recurrent since the same neurons are represented

in both layers as explained above. During the stimulus presentation, populations in the sensory layer receive an external input according to their tuning (see Figure 3):

$$\begin{aligned} I_{s,+} &= C + A(f - f_{\min}) \\ I_{s,0} &= C \\ I_{s,-} &= C + A(f_{\max} - f) \end{aligned} \quad (2)$$

where f is the stimulus frequency, A is its amplitude and C is a non specific background input. Parameters f_{\min} and f_{\max} denote the frequencies for which the positive and negative branches correspondingly become horizontal, i.e. they define the range of frequencies represented by the population. The activity of the tuned neurons is further increased by factor $1 + \Delta$ corresponding to their increased excitability. The memory layer receives these firing rates as input multiplied by an initial weight w_0 . We model synaptic augmentation by setting the synaptic weights from the sensory to the memory representation to be proportional to the activity of the sensory populations during the stimulus:

$$\begin{aligned} W_{++} &= w_0 \gamma (C + A(f - f_{\min})) (1 + \Delta) \\ W_{00} &= w_0 \gamma (C) \\ W_{--} &= w_0 \gamma (C + A(f_{\max} - f)) (1 + \Delta) \end{aligned} \quad (3)$$

We define the slopes of the V-shaped activity profile as the difference between the activity of the tuned and the untuned populations. During the readout, a steadily increasing uniform input I is applied to the sensory layer, which in turn activates the memory representation in a frequency-dependent manner due to the previous facilitation of synaptic weights. This results in the following slopes of activity profile in the memory representation for given input amplitude I :

$$\begin{aligned} \alpha_+ &= w_0 \gamma I \left[A(f - f_{\min}) (1 + \Delta)^2 + C(2\Delta + \Delta^2) \right] \\ \alpha_- &= w_0 \gamma I \left[A(f_{\max} - f) (1 + \Delta)^2 + C(2\Delta + \Delta^2) \right] \end{aligned} \quad (4)$$

where α_+, α_- are the positive and negative slopes. As can be seen from this expression, growing input results in gradually increasing slopes of the V-shaped profile, as illustrated in Fig. 5B,C of the results. The slopes of the V shaped profile for sensory and memory representations are plotted in Figure 5B,C using the following parameters:

$\Delta = 0.3, \gamma = 1.5, A = 0.7, f_{\min} = 0, f_{\max} = 40, f = 34, C = 10$. The uniform external input I grows linearly from 0.2 to 2 during the delay period.

Results

Two monkeys (*Macaca Mulatta*) were trained in the vibrotactile discrimination task until their psychophysical thresholds were stable (Hernandez et al., 1997; Mountcastle et al., 1990). The time course of an individual trial is schematized in Figure 1A. Briefly, the monkey had to remember the frequency of the first stimulus for 3 seconds, and then compare it to the frequency of the second one. After training, we recorded 899 neurons from the inferior convexity of the prefrontal cortex (PFC). The neurons were selected according to the quality of recording (see Materials and Methods), regardless of their response properties. In this paper, we focus on the presentation of the first stimulus and the delay period.

We begin our analysis by considering the mean firing rate across the entire recorded population, across all different stimuli (Figure 1C, blue). An interesting observation is that the mean firing rate decreases after the stimulus offset relative to the pre stimulus level, and gradually ramps up during the delay. Note that in control trials (see Materials and Methods), where the correct choice was indicated by a light and the monkey did not have to memorize the stimulus, the post stimulus activity decrease is less pronounced and there is no ramping towards the end of the delay (Figure 1C, green).

Population coding of frequency

Averaging over the entire population across all conditions hides the great variety of single cell responses to different stimuli (Romo et al., 1999; Brody et al., 2003). We therefore computed the mean firing rate of each cell, obtaining a vector of 899 firing rates for each stimulus frequency and each time bin, which we call an instantaneous *population state* (see Materials and Methods). In our analysis we calculate the mean firing rates by averaging across different trials, while the monkey's brain has to reach decisions in single trials. We assume that in the brain, there are many groups of neurons with similar response properties, of which we only sample a small fraction. By averaging across the activity of neurons within each group, the brain could then have a single trial access to the mean firing rates used in our analysis (Wilson and Cowan, 1972). For the sake of clarity, we stress that all the analysis refers to the first stimulus frequency.

Having defined the population state, we examine how well it reflects the stimulus frequency by calculating the correlation coefficient between the population states

corresponding to every pair of frequencies, for each time bin along the trial. If different frequencies reliably drive the population to different states, we expect the correlation to monotonically decrease with frequency difference. Figure 2A shows average correlation coefficients grouped according to the difference in the stimulus frequencies, Δf (i.e. the yellow curve shows the average of the two pairs with a 20Hz difference: 10Hz, 30Hz and 14Hz, 34Hz). At the first 100 ms of the stimulus, correlations sharply increase to a high value of almost 0.9 for all pairs of frequencies, implying that the stimulus initially brings the population close to a fixed state regardless of the precise frequency used. At the second stimulus bin, the curves separate according to Δf , indicating a sensitivity of the population state to the stimulus. The separation between the curves decreases to a minimum around 600 ms into the delay, suggesting that at this time bin the stimulus frequency is not consistently mapped into a population state. Surprisingly, frequency sensitivity of the population state then steadily increases till the very end of the delay period.

In order to characterize the dynamics of the population state during the non monotonic changes in stimulus sensitivity, we calculate the cross correlation between the population states at different times for a given condition. Figure 2B shows the average of the crosscorrelation coefficient across all conditions, with the reference time being either the second time bin of the stimulus presentation (blue curve) or the last time bin of the delay period (green curve). The results show that while the population encodes the stimulus at both of these time points, the corresponding population states are different: correlation coefficient between them does not exceed the pre-stimulus level. In the following we call these population states *stimulus* and *memory* representations respectively. The graphs also reveal that the stimulus representation decays quickly while the stimulus is still present, whereas the memory representation slowly builds up towards the end of the delay period.

Sensory and memory representations

The above results show that the stimulus frequency is encoded in the population state in a highly dynamic manner. In order to gain insight into the mechanisms which generate this encoding, we search for a reduced parameterization of the population state. Commonly, this is achieved by labeling neurons according to their tuning curves (e.g. preferred

orientation of V1 neurons). Neurons across the sensory pathway, beginning with secondary somatosensory cortex (S2), respond to increasing stimulus frequency with a monotonic increase or decrease of their firing rate (Romo et al., 2002). Following Romo et al, we fitted each cell's firing rate after the presentation of the stimulus as a linear function of the stimulus frequency f :

$$r(f) = a_0 + a_1 f . \tag{5}$$

This equation does not necessarily mean that the cell's response function is perfectly linear, but is simply a first order characterization of monotonically responding cells. It allows us to label each neuron by its a_1 value, obtained at the 2nd bin of the stimulus (where the separation between frequencies is maximal, see Fig 2). Importantly, all the neurons in the PFC are divided into two roughly equal populations of positively ($a_1 > 0$) and negatively ($a_1 < 0$) tuned neurons. We can now use these labels to characterize the population state by plotting r vs. a_1 . If a_0 and a_1 coefficients were uncorrelated across the population, Eq. (5) would predict that single neuron firing rates should scatter around a straight line with slope f . Figures 3A,B,C show examples of population states for $f = 10, 22$ and 34 Hz with r taken from the same time bin, using all cells with a_1 coefficients significantly different from zero (Corresponding plots for all frequencies are in the supplementary information, Fig. S1). Surprisingly, the figures reveal a rotated V shaped profile, characterized with positive and negative branches with positive and negative slopes respectively. The existence of the negative branch means that for a given frequency, neurons with progressively more negative a_1 coefficients respond *stronger*, in apparent contradiction to Eq. 5. The contradiction is resolved however if for neurons with negative tuning, their a_0 coefficients are negatively correlated with a_1 , which is indeed the case (see Fig. S3).

The above results indicate that the population state can be compactly characterized by two parameters – the slopes of the positive and negative branches that together encode f . Figure 3D shows the dependency of branch slopes on the stimulus frequency. Note that reading out the slopes of the V shape is simply a weighted linear summation of the neurons' activity, where the weight for each neuron is proportional to its a_1 label,

making this representation highly accessible to downstream neurons (see Supplementary material, Eqs 2,3).

We now utilize this compact representation to examine the dynamics of the population state throughout the delay period more closely. The low cross correlation values in Figure 2B suggest that the a_1 values extracted during the stimulus will not aptly describe the population state at the end of the delay period. Indeed, Figure 3E shows that under this representation the activity profile becomes almost symmetric and does not change with f . The correlation analysis of Fig. 2A, however, suggests that the population state encodes the stimulus frequency at this point, perhaps using a different representation. Indeed, sorting the cells according to a_1 values obtained at the end of the delay (Figure 3F) restores the rotated V shapes similar to Figure 3A-C, albeit with a smaller range of a_1 values. The cell labels have thus been “remapped” between the stimulus and delay periods. Some of the cells are only tuned to f in one of the epochs (significant a_1 value: 233 in the stimulus, 158 in the end of delay), but even the a_1 values of those neurons which have a significant a_1 in both epochs (59, expected from independence: 41) are only weakly correlated (correlation coefficient 0.29, 35% change sign of a_1).

This result confirms our conclusion that the same neurons represent the stimulus in two coexisting representations, which we called *sensory* and *memory* representations. The population undergoes a remapping from sensory to memory representation between the stimulus and the end of the delay period. In order to characterize the dynamics of remapping, we plot the evolution of positive and negative branch slopes for a given stimulus frequency (34Hz) throughout the delay period under the two representations. Figure 4 shows that the sensory representation of the stimulus quickly decays to a symmetric activity profile, while the memory activity profile gradually rotates to the direction of the stimulus frequency, again in agreement with Fig. 4B above. We thus see that the gradual ramping up of firing rate during the delay period (Figure 1C) is not a simple scaling of the firing rates, but rather a redistribution which improves the representation of the stimulus towards the end of the delay period.

A candidate neural mechanism for dynamic representation of stimulus frequency

There is currently no comprehensive theoretical understanding of the underlying mechanisms of working memory of continuous variables (stimulus frequency in the current study) that would be compatible with the experimental results. Existing models rely on the idea of a *line attractor* (Seung, 1996; Miller et al., 2003; Machens et al., 2005; Miller and Wang, 2006) whereby the network can be in infinitely many stable activity states each representing a different stimulus frequency. Such models, however, are structurally unstable and thus require a high degree of tuning (Seung, 1996; Koulakov et al., 2002). Moreover, the temporal evolution of neural firing rates cannot be accounted by line attractor models (Brody et al., 2003).

Line attractor networks result in a gradual degradation of the information representation due to accumulations of the effects of noise over time (Seung, 1996). Our analysis presented above indicates, however, that the stimulus encoding in the firing rates decays after stimulus offset and then gradually improves during the delay period. Such non-monotonic encoding dynamics may indicate that information about the stimulus is not exclusively maintained in a spiking form, but also in the facilitation level of the synaptic connections (Mongillo et al., 2008, Deco et al., 2010). Here we illustrate this idea with a schematic neural circuit.

Figure 5A shows a highly reduced network model of the PFC population, where we divide the neurons in six populations, according to their tuning properties – positive, negative and unselective, in both sensory and memory representations as described above (see Methods for more details). For the sake of illustration, we consider the two representations (sensory and memory) as two separate layers of the network. However, it is important to remember that PFC neurons belong to both sensory and memory representations, so they are divided between the corresponding populations in both layers in an almost independent manner. Furthermore, connections in the network are going from sensory level populations to corresponding populations in the memory layer with the same tuning, i.e. we assume that neurons with positive sensory tuning project selectively to neurons with positive memory tuning, etc.

During the stimulus presentation with a particular frequency, populations in the sensory layer receive and external input according to their tuning, resulting in the corresponding V shaped activity profile (branch slopes are shown in Fig. 5B for $f=34\text{Hz}$). Connections from the sensory to the memory layers are then facilitated, such that their baseline weights are scaled with a facilitation factor proportional to the activity of corresponding pre-synaptic populations (see Methods). During the delay period, we assume that a gradually increasing uniform input is applied to the sensory layer – effectively reading out the facilitation state into the firing rates of the memory representation (see also Deco et al., 2010). The crucial result is that due to the prior facilitation of the connections, uniform activation in the sensory layer evokes the V shaped activity profile in the memory layer (corresponding branch slopes are shown in Fig. 5C), i.e. neurons are now tuned to stimulus frequency according to their memory labels and not sensory ones (see Materials and Methods for the derivation of this plot). Modeling results presented in Fig. 5B,C are broadly consistent with the corresponding experimental data (Fig. 4A,B).

Finally, we briefly discuss a possible mechanism for the decision phase of the task in the framework of our model. When the second stimulus arrives, the memory representation is tuned to the first stimulus frequency, f_1 , while the sensory representation is driven by the second stimulus frequency f_2 . If the two representations compete with each other via inhibition between corresponding branches, then the outcome of this competition can be used to solve the task. In this scheme, for trials with $f_1 > f_2$ the positive memory neurons will be more active than positive sensory neurons, and negative memory neurons will be less active than negative sensory neurons (vice versa for $f_1 < f_2$). We test this prediction by fitting the firing rate of each cell during the second stimulus as $r = b_0 + b_1(f_1 - f_2)$. Positive b_1 values indicate “Yes” cells which fire at a higher rate when $f_1 > f_2$, while negative values indicate “No” cells. Indeed, there is a tendency for Yes neurons to be negatively tuned to f_1 in the sensory representation (correlation coefficient -0.44), and positively tuned to f_1 in the memory representation (correlation coefficient 0.33). The negative correlation between b_1 and sensory- a_1 is inconsistent with Machens et al. (2005). Note that similar correlations are predicted also in the model of Miller and Wang (2006).

Discussion

We studied the dynamics of stimulus representation in working memory. We focused on population state analysis, taking advantage of the large dataset of PFC single neuron recordings in monkeys performing delayed memory task. The population analysis revealed two surprising outcomes. First, the quality of stimulus encoding by the population degrades after the offset of the stimulus, and gradually recovers throughout the delay period. Second, the encoding of the stimulus after this recovery is quite different than that during the stimulus itself. Non-monotonic dynamics of stimulus encoding and the remapping of stimulus representation during the delay period pose a challenge to previously suggested models of working memory of stimulus frequency (Machens et al., 2005; Miller et al., 2003; Miller and Wang, 2006). Two specific forms of remapping were introduced in previous models. Machens et al (2005) describe a sign flip between the sensory representation of the first and second stimulus, which does not account for the random remapping observed in the data. Miller et al. (2006) suggest two distinct populations – *memory neurons* which maintain sign from stimulus to delay and inhibit *comparison neurons* which flip sign from stimulus to delay. While pooling over these two groups might produce a remapping similar to the one we observed, the firing rate dynamics do not match this model. Specifically, when we split the neurons into two groups according to whether they flipped sign from the stimulus to the delay, we found that both groups ramp up during the delay period in an indistinguishable way, without any sign of inhibition from the *memory* to the *comparison* populations (not shown). Furthermore, both models rely on persistent firing to maintain the stimulus information and are hence expected to display a gradual deterioration of stimulus representation during the delay period, opposite to the observed trend in the data.

To study stimulus encoding in more detail, we propose a compact representation of the population state in the form of a V-shaped activity profile characterized by the slopes of two branches. Using the slopes allows for estimation of stimulus frequency from the instantaneous population state (see Figure 3D), via a simple linear readout of the cells' activities. This approach is similar to the extraction of population vectors in the motor cortex, where the firing rates of the neurons are first combined in a vectorial manner in order to estimate the direction of movement (Georgopoulos et al., 1986). Using this representation, we showed that a V shaped profile in the memory representation slowly

emerges during the delay period, thus improving stimulus encoding. These results raise the possibility that other neural processes besides the firing rates are involved in storing the stimulus attributes. Following our recent proposal for synaptic facilitation based working memory of discrete items (Mongillo et al., 2008), we suggest a putative mechanism in which the frequency of the stimulus is reflected in the graded facilitation profile of the connections between sensory and memory representations. The facilitation can be subsequently extracted into the firing rates by increasing excitatory input applied uniformly to the entire network, which could correspond to the ramping up of firing rates observed during the delay period. For this model to be compatible with experimental delay durations, facilitation has to be stable for the entire delay period. Some forms of facilitation, such as synaptic augmentation, that are driven by high-frequency bursts of presynaptic firing, can persist for periods of many seconds (Zucker and Regehr, 2002; Hempel et al., 2000; Wang et al., 2006). We therefore propose that the firing rates in the stimulus period, which can reach values of more than 100 Hz (Fig. 3A-C) are sufficient to trigger synaptic augmentation. The model that we analyzed in this contribution is aimed to illustrate the basic mechanism of facilitation-dependent representation of stimulus frequency by PFC activity. We therefore simplified the exposition of the model by ignoring the short-term dynamics of facilitation during the delay period. We assume that lower-rate activity of neurons during the delay does not significantly change the facilitation level of the corresponding connections, but this issue will have to be studied more carefully in future extensions of the model.

The two representations we described (sensory and memory) are reminiscent of the early and late neurons reported in (Romo et al., 1999), where neurons were classified according to the segments of the delay period in which they were stimulus selective. The sensory representation, however, is defined according to the stimulus period and not the early delay period. In our data, 47 of the 183 early neurons (25%) were also sensory neurons and none of them were memory neurons. Of the 191 late neurons, 51 (27%) were memory, 44 (23%) were sensory and 11 (6%) were both. The categories are thus only weakly correlated, and we therefore did not use this earlier classification in our analysis. .

Previous work from the Romo's lab also identified persistent neurons, which were significantly tuned to the stimulus throughout the entire delay period (Romo et al., 1999).

It might be argued that the memory is carried by this sub-population, and the non-monotonic encoding observed in our analysis is an artifact of considering other irrelevant neurons. In order to test this possibility we identified persistent neurons (20% of the population, see Materials and Methods) and analyzed them separately. The firing rate of this group varied throughout the delay period, in agreement with previous studies (Brody et al., 2003). Moreover, the firing rate dynamics of this group are similar to those of the general population (Fig. S5), making it unlikely that they form an isolated group that carries working memory.

A recent model proposed a continuous attractor in which neurons are tuned both for stimulus frequency and for time since the stimulus termination (Singh and Eliasmith, 2006). In this model, a_i values of the neurons are expected to vary linearly with time, which could explain the remapping between the stimulus and the end of the delay period. However, such a linear change in a_i s is inconsistent with the low correlation between the population state at the stimulus and the state following the stimulus offset (Figure 2B).

Information about a stimulus travels through many stations in the brain, undergoing various transformations along the way. The stimulus is initially coded in positive monotonic primary somatosensory cortex (S1) neurons. S2 already displays negatively tuned neurons, and some early delay activity. Similar response properties to those reported in PFC have been also found in medial premotor cortex (Hernández et al., 2002) and ventral premotor cortex (Romo et al., 2004), using the same task and suggesting a contribution of these two cortical areas in maintaining sensory information during working memory. At the end of the chain M1 neurons code the decision made by the monkey. The surprising observation in the PFC is that the same neurons seem to participate in different stages of the computation, and their tuning at each stage is largely independent of other stages. PFC neurons are used for many tasks, and it is thus not likely that every stage of every task will employ a dedicated group of neurons. It is probably the common case that a given neuron participates in various stages of different tasks. Furthermore, overlapping representations provide an opportunity for interaction between the different stages of the task. If both the memory and the sensory representations reside within the same neural pool, then mutual inhibition between the representations could serve to solve the decision task.

Acknowledgments

We thank B. Blumenfeld, C. Brody, S. Fusi and L. Abbott for helpful comments on the manuscript. O.B. thanks Ranulfo Romo and all members of the Romo's lab for the heartwarming hospitality and helpful discussions on his visits there. O.B. is supported by the Azrieli Foundation. M.T. is supported by the Israeli Science Foundation. The research of R.R. was partially supported by an international Research Scholars Award from the Howard Hughes Medical Institute and grants from CONACYT and DGAPA-UNAM.

References

- Baddeley A, Hitch G (1974) Working Memory In G. Bower, ed. Recent advances in learning and motivation New York: Academic Press, p. 47-90.
- Brody CD, Hernandez A, Zainos A, Romo R (2003) Timing and neural encoding of somatosensory parametric working memory in macaque prefrontal cortex. *Cerebral cortex* 13:1196-1207
- Deco G, Rolls ET, Romo R (2010) Synaptic dynamics and decision making. *Proc Natl Acad Sci U S A* Available at: <http://www.ncbi.nlm.nih.gov/ezproxy.cul.columbia.edu/pubmed/20360555> [Accessed April 10, 2010].
- Georgopoulos AP, Schwartz AB, Kettner RE (1986) Neuronal population coding of movement direction. *Science* 233:1416-1419
- Hempel CM, Hartman KH, Wang X, Turrigiano GG, Nelson SB (2000) Multiple Forms of Short-Term Plasticity at Excitatory Synapses in Rat Medial Prefrontal Cortex. *J Neurophysiol* 83:3031-3041
- Hernandez A, Salinas E, Garcia R, Romo R (1997) Discrimination in the sense of flutter: new psychophysical measurements in monkeys. *Journal of Neuroscience* 17:6391-6400
- Hernández A, Zainos A, Romo R (2002) Temporal evolution of a decision-making process in medial premotor cortex. *Neuron* 33:959-972
- Jun JK, Miller P, Hernández A, Zainos A, Lemus L, Brody CD, Romo R (2010) Heterogenous population coding of a short-term memory and decision task. *J. Neurosci* 30:916-929
- Koulakov AA, Raghavachari S, Kepecs A, Lisman JE (2002) Model for a robust neural integrator. *Nat. Neurosci* 5:775-782
- Machens CK, Romo R, Brody CD (2005) Flexible control of mutual inhibition: a neural

model of two-interval discrimination. American Association for the Advancement of Science.

Miller P, Brody CD, Romo R, Wang XJ (2003) A recurrent network model of somatosensory parametric working memory in the prefrontal cortex. *Cerebral Cortex* 13:1208-1218

Miller P, Wang XJ (2006) Inhibitory control by an integral feedback signal in prefrontal cortex: A model of discrimination between sequential stimuli. *Proceedings of the National Academy of Sciences* 103:201-206

Mongillo G, Barak O, Tsodyks M (2008) Synaptic Theory of Working Memory. *Science* 319:1543-1546

Mountcastle VB, Steinmetz MA, Romo R (1990) Frequency discrimination in the sense of flutter: psychophysical measurements correlated with postcentral events in behaving monkeys. *Journal of Neuroscience* 10:3032-3044

Romo R, Brody CD, Hernández A, Lemus L (1999) Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature* 399:470-473

Romo R, Hernández A, Zainos A, Lemus L, Brody CD (2002) Neuronal correlates of decision-making in secondary somatosensory cortex. *nature neuroscience* 5:1217-1225

Romo R, Hernández A, Zainos A (2004) Neuronal correlates of a perceptual decision in ventral premotor cortex. *Neuron* 41:165-173

Seung HS (1996) How the brain keeps the eyes still. *National Acad Sciences*.

Singh R, Eliasmith C (2006) Higher-dimensional neurons explain the tuning and dynamics of working memory cells. *Journal of Neuroscience* 26:3667-3678

Wang Y, Markram H, Goodman PH, Berger TK, Ma J, Goldman-Rakic PS (2006) Heterogeneity in the pyramidal network of the medial prefrontal cortex. *Nature neuroscience* 9:534-542

Wilson HR, Cowan JD (1972) Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J* 12:1-24

Zucker RS, Regehr WG (2002) Short-term synaptic plasticity. *Annu. Rev. Physiol* 64:355-405

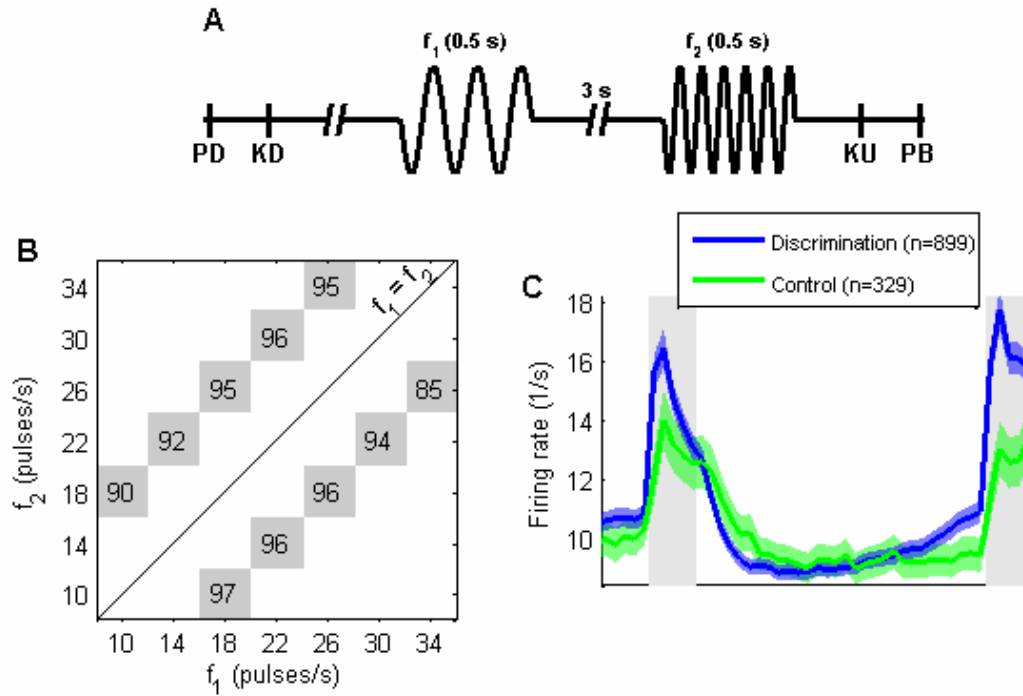


Figure 1: Behavioral task, stimulus sets and firing rate. **A** Schematic diagram of the behavioral task. A mechanical probe goes down to indent the monkey's finger (PD). The monkey signals readiness by grasping a key with his free hand (KD). After a random delay of 1500-3000 ms, the first stimulus begins followed by a 3000 ms delay and a second stimulus. After the second stimulus the monkey releases his free hand (KU), and pushes one of two buttons (PB) to indicate whether $f_1 > f_2$ or vice versa. **B** The pairs of f_1, f_2 used in the task. Numbers indicate percent of correct trials in each condition. **C** Mean firing rate of the entire recorded PFC population, across all stimulus conditions (blue). The green curve shows a subset of the neurons which were tested with the light control condition. The colored shading denotes error bars. Throughout the paper, gray shaded areas denote stimulus periods, and all data is binned with 100ms non overlapping bins without any smoothing.

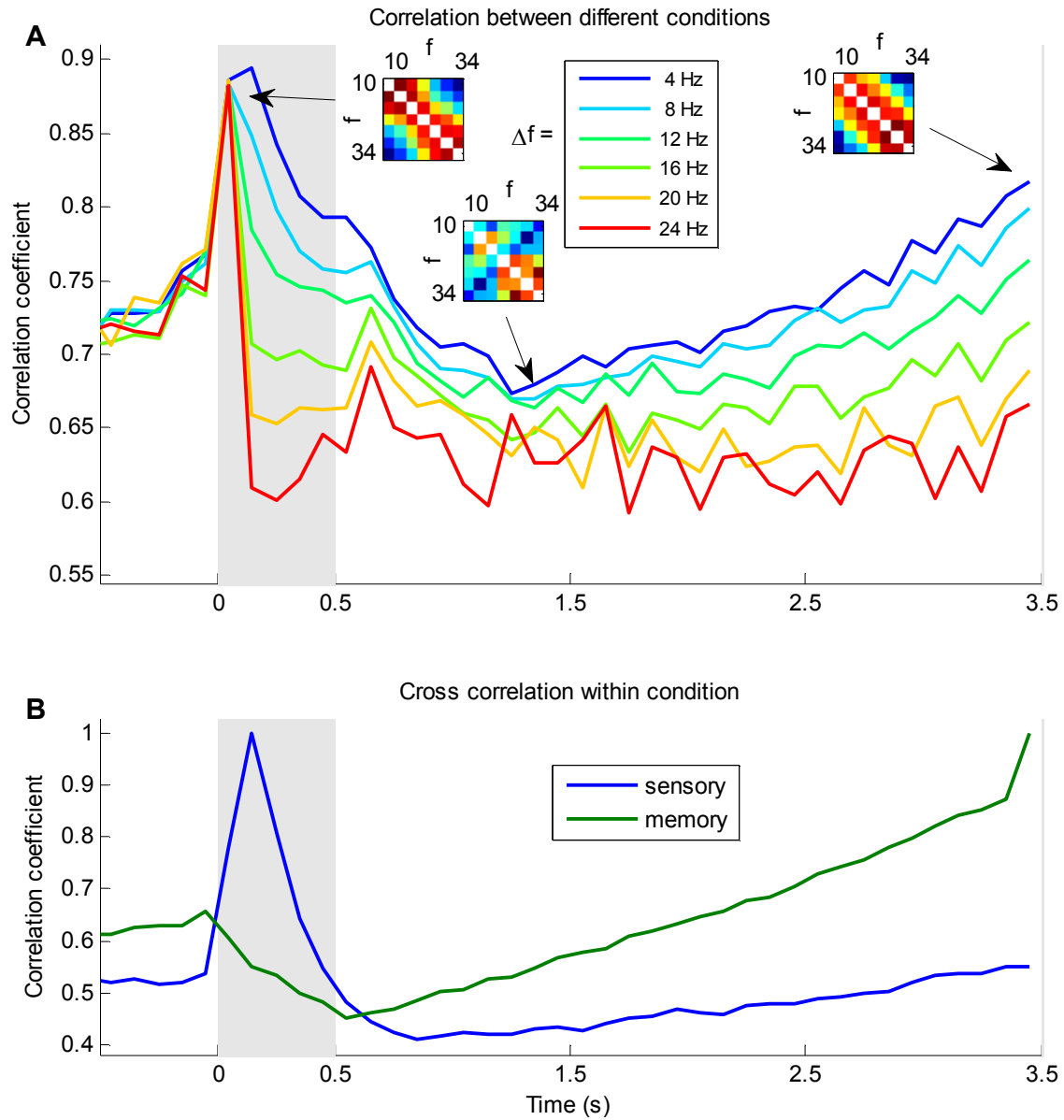


Figure 2: Correlation between different population states as a function of time. **A** Correlation within time bin, across conditions. Each curve is the average of all pairs of stimuli with a fixed difference between their stimulus frequencies. Insets show correlation coefficients of all pairs of frequencies for three selected points in time. Color in the insets varies from blue to red representing from left to right the ranges [0.61,0.92], [0.61,0.74] and [0.67,0.83]. **B** Correlation across time within conditions. The blue (green) curve shows the correlation of the population state during the stimulus (at the end of delay) to the state throughout the delay period.

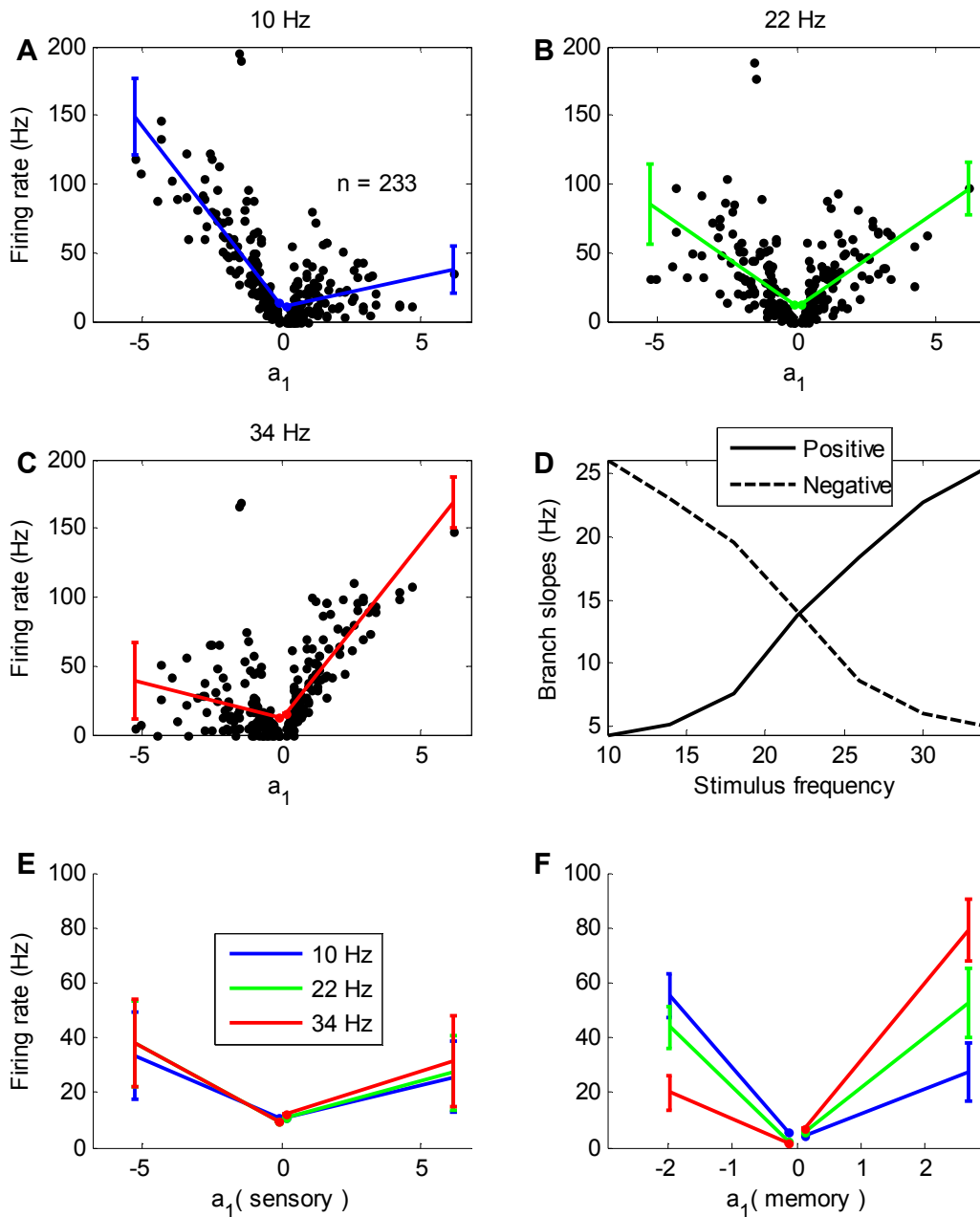


Figure 3: Rate as a function of a_1 . **A,B,C** Firing rates of 233 neurons with significant a_1 values during 10, 22 and 34 Hz stimulation. Neurons are sorted according to a_1 from the stimulus period (sensory). The solid lines show a linear fit to the data, with a confidence interval for the extreme points. **D** The slopes of the lines for different stimulus frequencies. **E,F** The firing rate at the end of the delay period using either the sensory (E) or the memory (F) representation for three stimulus frequencies.

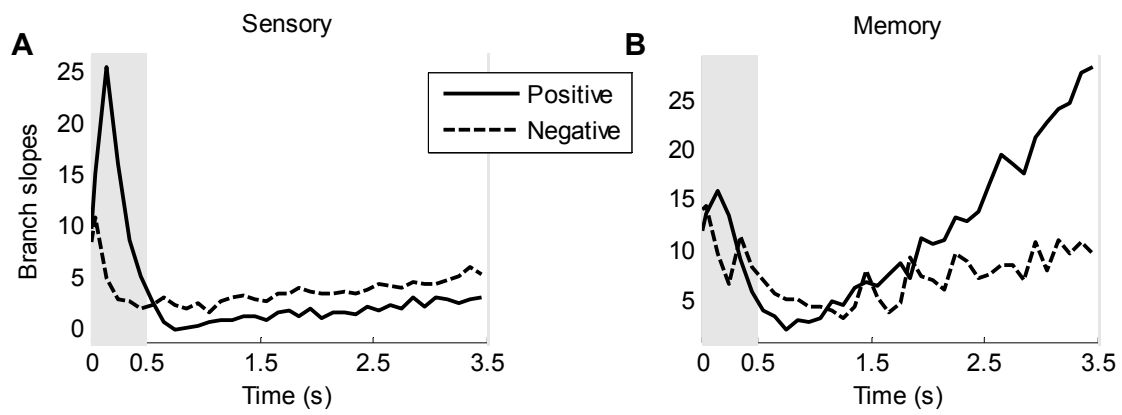


Figure 4: Dynamics of the population state. Branch slopes for a 34Hz stimulus plotted from the stimulus till the end of delay. **A** Sensory representation: a_1 labels extracted from the stimulation period. **B** Memory representation: a_1 labels extracted from the end of the delay period.

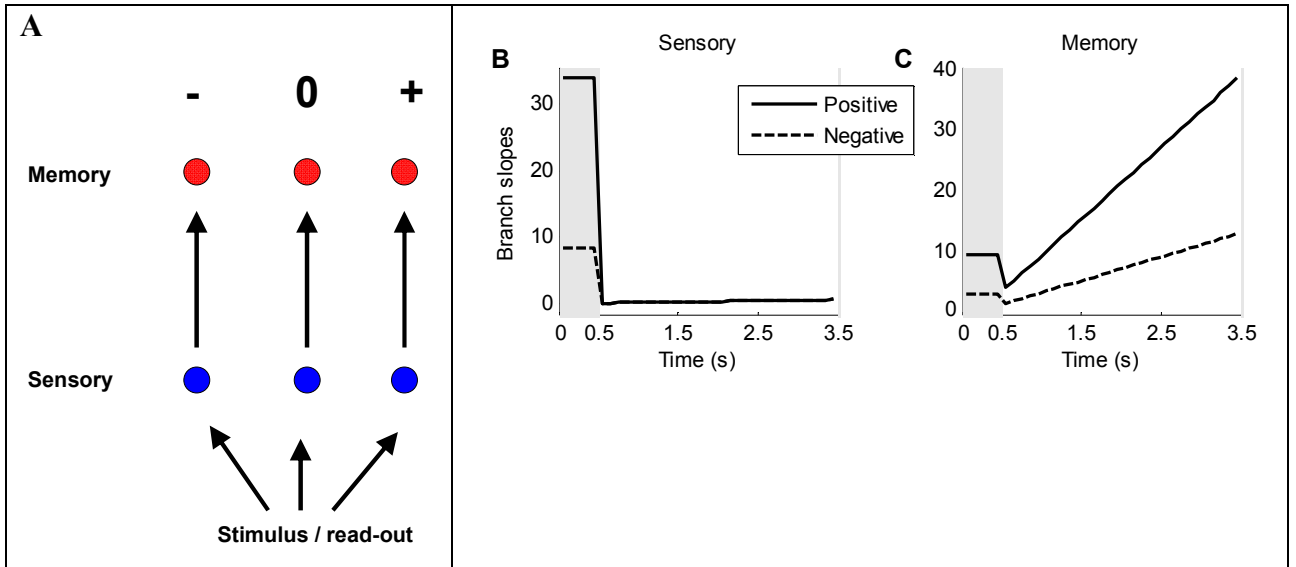


Figure 5: Model for facilitation based stimulus memory. **A**: Neurons were grouped according to their sensory and memory tuning, with facilitating connections from the sensory to the memory representation. **B,C**: Branch slopes of the V shaped activity profile in the sensory and memory representation, respectively, following the stimulus with $f=34\text{Hz}$.