Investigating Mechanisms of Maintaining Multiple Items in Short-Term Memory Through Recurrent Neural Networks

Xinyu Wei, Chaihyun Lee, Nicolas Y. Masse
Freedman Laboratory, Computational Neuroscience, University of Chicago, Chicago, IL

Key words: **recurrent neural network; short-term memory; sequential items**

Introduction

Short-term memory is indispensable for a variety of tasks in our daily lives. Recent *in vivo* studies suggest that short-term memory is maintained by persistent neural activities[1-2] [Fig.1]. While many studies have examined how single items are encoded and maintained in short-term memory, much less is known about how multiple items in a sequence are encoded, largely due to the difficulties in training animals and recording neural data from these tasks. In response to these difficulties, we investigated short-term memory mechanisms for maintaining multiple items by training a biologically-realistic Recurrent Neural Network (RNN) on sequence memory tasks. The current approach allows us to derive putative mechanisms for brains. Furthermore, by understanding how interference between multiple items affects the networks' memory task performance, we may be able to design a novel neural network with enhanced short-term memory capacity.

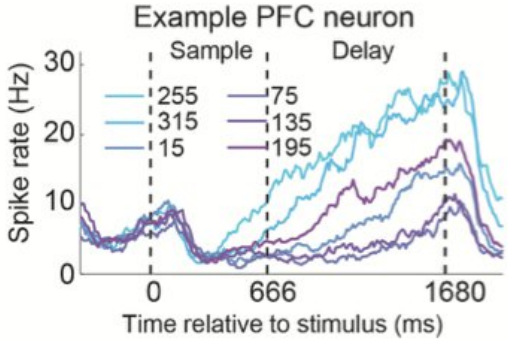

Figure 1. Persistent neural activity during delay period (from Masse et al. 2017)
This figure shows the neural activity of an example neuron recorded from prefrontal cortex (PFC) during a short-term memory task. The subject was trained to indicate whether a sample and test motion direction stimulus, separated by a delay period (between 666 ms and 1680 ms), matched. The six colored curves represent the mean spike rate of the neuron for each of the six sample motion directions (angle indicated in degrees). It was observed that the identity of the sample stimulus is encoded in the neural activity of this example neuron throughout the delay period. This is consistent with previous results that information in short-term memory are maintained through persistent neural activities[7] (Mongillo et al. 2008).

Materials and Methods

Biologically-inspired Recurrent Neural Network (RNN)

We used a biologically-inspired Recurrent Neural Network (RNN) to examine the putative neural mechanisms underlying short-term memory. In a typical neural network, neurons can exhibit positive (excitatory) or negative (inhibitory) activities without constraint. However, in a biological setting, there is a strict division between excitatory and inhibitory neurons. Therefore, we set 80% of the neurons to be excitatory and 20% to be inhibitory, consistent with what is known *in vivo*.

Furthermore, the firing rate of a neuron in the brain is relatively low either due to the metabolic costs[4] (Laughlin, 1998), or possibly to facilitate information read-out from the neural activities[3,9] (Hawkins, 2016 & Olshausen, 2004). Therefore, we added a punishment to the high neural activities of the network. The network would learn to solve problems with relatively low neural activities.

Most importantly, synaptic connections between the recurrently connected neurons were modulated by synaptic plasticity in order to mimic the ways in which brain might be encoding information for various tasks. Recent research suggests that short-term synaptic plasticity, which is maintained through short-term changes in the neural network, is critical for maintaining information during short-term memory tasks. The short-term synaptic plasticity in brain alters synaptic

efficacy based on presynaptic activity mainly through two mechanisms: presynaptic activity typically increases residual calcium concentration in the presynaptic terminal (short-term facilitation) and decreases available neurotransmitters (short-term depression). In our neural network, we incorporated two terms that respectively represented the calcium concentration and number of available neurotransmitters for each neuron. The synaptic efficacy was made proportional to the product of the two terms.

Task

In the task, after a short initial fixation period of 200 ms, we presented a sequence of motion directions pulses to the network, each lasting for 200 ms and followed by a delay period of 200 ms [Fig.3]. During the delay, no stimulus was present. The last motion direction was followed by a long delay period of 500 ms. From the start of the task to the end of the long delay period, the network should always fixate. The long delay was followed by a sequence of test periods during which the network is cued to recall the motion directions in the order of their presentation. Each test period was cued by one of the eight response cues to indicate which stimulus the network should recall. The response cue for each motion direction would last for 200 ms, and was followed by a 200 ms short delay period before the next cue. When the response cue was on, the network was trained to output a one-hot vector with one at the index of the correct direction.

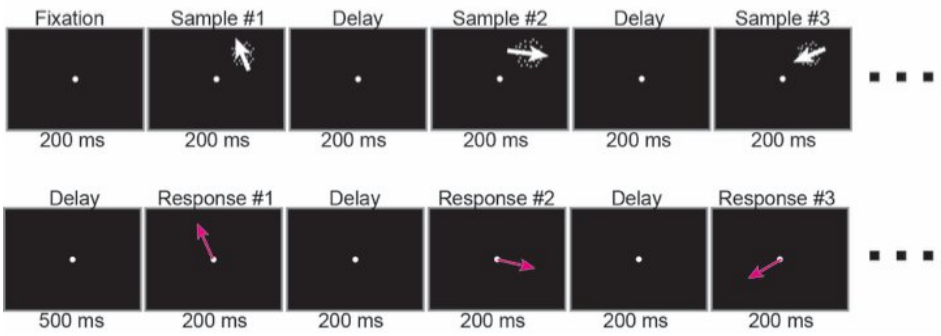

Figure 3. Task
Schematic diagram of the task in which the network is shown a sequence of stimulus chosen from the eight possible motion directions and asked to recall the directions in the same order later on.

Network configuration

The input layer of the model consisted of 24 motion-tuned neurons for each of the four receptive fields, two fixation-tuned neurons, and one cue-tuned neuron for each pulse presented during the task [Fig.3]. The activities of the motion-tuned neurons were determined based on the motion stimulus that was presented as well as the neurons' preferred directions. The fixation-tuned neurons fired when the network should be fixating - the stimuli presentation period and the delay period. The cue-tuned neuron for each motion directions fired during the presentation of that specific motion direction and when the network needed to recall that direction. The input neurons projected to 100 hidden neurons, with 80 being excitatory and 20 being inhibitory. The output was a one-hot vector with a length of nine, with eight of them representing the eight motion directions respectively and one representing fixation.

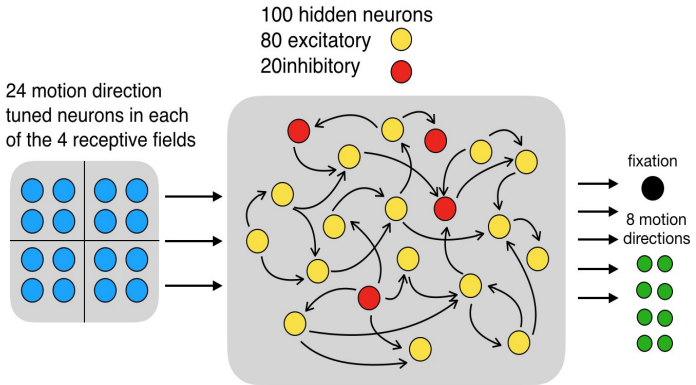

Figure 3. Overall schematic diagram of a neural network with EI network

Training method

To examine the effect of the sequence size on the network's behavior, we trained the network on trials with different number of motion stimuli. Connection weights and biases were trained using stochastic gradient descent to minimize a loss function comprised of two terms: 1) the cross-entropy between the expected output and actual response of the output neurons and 2) the mean firing rate of the hidden layer neurons. We started from training the network on a sequence of three motion directions, and increased the number of motion directions until the network's learning performance could not reach the accuracy of 85% any more. Throughout training, the accuracy for each motion stimulus and the neural and synaptic activities were recorded to be used for analysis. In order to make the network flexible to various lengths of delay period, we made the short delay periods between stimuli and response cues vary at each iteration of the training, but no longer than 500 ms.

Results

Accuracy

Throughout the test period of the task, we determined the network's recall of the stimuli based on the activities of its output neurons. The motion direction that the neuron with the highest activity represented was the network's recall. The accuracy for each pulse was calculated by the ratio of the number of times the network correctly recall and the total number of trials. The overall task accuracy was calculated by averaging all the pulse accuracies in the sequence. For trials with less than six stimuli, the network was able to achieve 90% accuracy in recalling all stimuli after 38,000 iterations. For tasks with more than six total motion stimuli, the network required much more number of iterations to learn.

Encoding and maintenance of information via neural activity and synaptic efficacy

To investigate how the network maintained information of multiple stimuli, we decoded the neuronal activity and synaptic efficacy of the hidden neurons. We used a linear support vector machine (SVM) to classify each sample motion directions from a population of neuronal activities or synaptic efficacies. The decoding accuracy was calculated as the percentage of the motion directions that can be correctly predicted by the SVM classifier. The decoding accuracy represented the amount of information about the stimuli that can be extracted from the hidden neurons.

For both tasks that involved only a few stimuli (3 sample stimuli) and tasks that involved more stimuli (8 sample stimuli), information about the stimuli could be extracted well from the synaptic efficacy during the delay period, in which the stimuli were no longer presented. However, the information about the stimuli was no longer stored in the neuronal activity during the delay period [Fig. 5]. This contrast between the encoding and maintenance behaviors of neural activity and synaptic efficacy aligns with previous research[5] (Masse et al., 2018) that highlighted the importance of synaptic plasticity in maintaining short-term information.

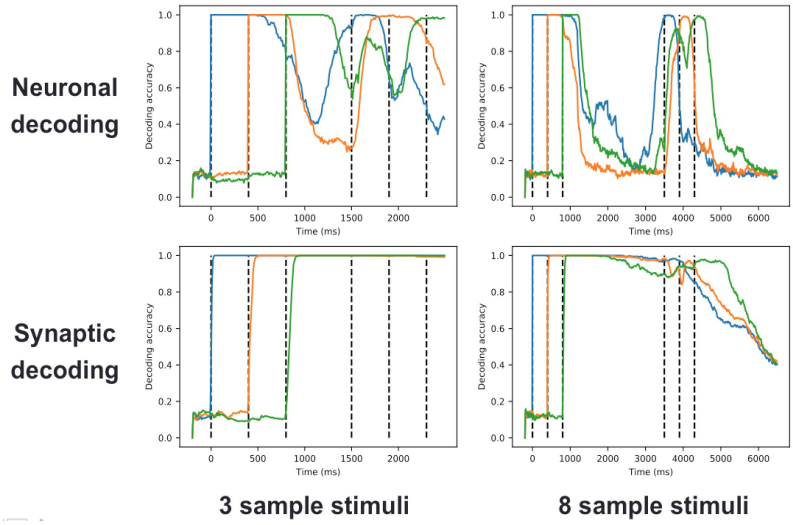

Figure 5. Decoding accuracy of neuronal activity and synaptic efficacy of the hidden neurons for the first three stimuli
Biologically-inspired RNN were trained on tasks that require the network to maintain various number of sequentially presented stimuli. This figure shows the decoding accuracy of hidden neurons' neuronal activity and synaptic efficacy for the first three motion direction pulses in one task where only three stimuli were presented (left panels) and another task where eight stimuli were presented (right panels).

Sparse encoding

One possible way of maintaining multiple stimuli in short-term memory could be sparse encoding. By having a selective number of hidden neurons encoding and maintaining information about a specific stimulus, the network may be able to minimize the interference from other stimuli in the sequence and successfully keep the memory for each stimulus intact.

In order to analyze the encoding specificity of the hidden neurons, we calculated the number of neurons that are robustly encoding two different motion stimuli. We calculated the proportion of explained variance (PEV) of each neuron's neuronal activity and synaptic efficacy throughout the trial. The PEV measured the amount of explained variance a linear model related neuronal activity or synaptic efficacy to the motion direction. Thus, higher PEV meant a stronger relationship between the stimulus identity and the neuronal activity or synaptic efficacy. We defined neurons with PEV values larger than 0.25 at the end of the long delay period as the ones robustly encoding the motion stimulus. A significantly larger number of neurons were found selective for one single stimulus by synaptic efficacy than by neuronal activity [Fig. 6].

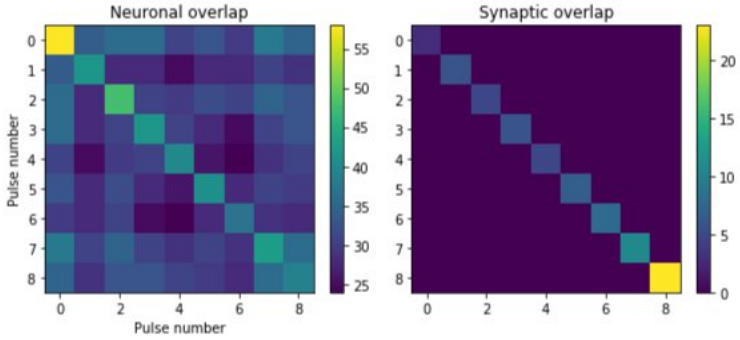

Figure 6. Specificity in neuronal and synaptic encoding
These are heat maps illustrating the number of hidden neurons that are responsible for encoding two different motion stimuli (one indicated on the x axis and another indicated on the y axis). The two heat maps were generated by counting the number of neurons that have PEV values larger than 0.25, calculated from neuronal activity (left) and synaptic efficacy (right) respectively at the end of the long delay period. The diagonal positions show the number of neurons encoding for one specific stimulus.

Observation of the Primacy-and-Recency Effect

While training the recurrent neural network to encode multiple sequentially presented items in its memory, we observed an interesting phenomenon in which the accuracy for the first and last couple items were higher than that of the ones occurred in the middle [Fig.7]. This phenomenon is analogous to the "Primacy Effect" (remembering items at the beginning of the list better compared to the rest) and the "Recency Effect" (remembering items that are presented near or at the end of the list better than the ones before) that are observed in human experiments[6,8] (Murdock, 1962 & Mayo, 1964). A potential reason for the Primacy-and-Recency Effect could be that the network had more difficulties projecting strong inhibitory signal to the hidden neurons when it encoded items in the middle of the sequence. With less inhibitory signal, the neurons encoding for the middle stimuli would suffer from more interference from other stimuli.

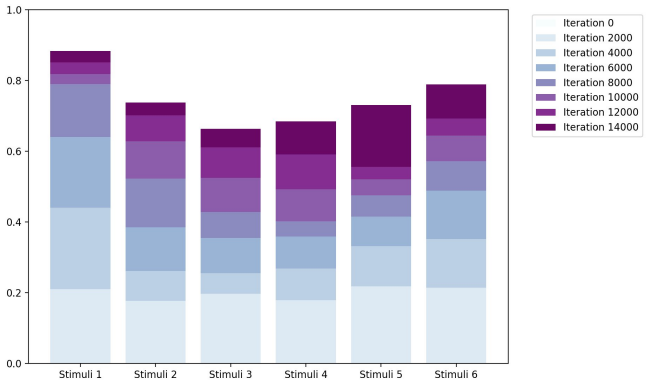

Figure 7. Recall accuracy for each stimulus demonstrates the Primacy-and-Recency Effects
In a task with 6 motion stimuli, the network performed better for the first and last few stimuli compared to the ones in the middle of the sequence. This figure illustrates such trend at different points of the training.

Potential mechanism for reducing interference

To understand the strategy taken by the network to reduce interference between stimuli, we examined the currents flow between five neurons that most strongly encodes a specific stimulus. The five neurons having highest PEVs at the end of the long delay period were selected. Then we calculated the currents into these neurons following the formula: presynaptic neural activities * corresponding connection weights * synaptic efficacy. These currents were separated into activities from excitatory neurons, inhibitory neurons, motion tuned neurons, and cue neurons. We were able to identify a neuronal circuitry in which the neurons responsible for encoding each stimulus had high excitatory inputs soon followed by a strong inhibitory current [Fig.8]. Having a strong inhibitory current inhibited any neuronal activity after the stimulus presentation in these neurons, keeping the information for the encoded stimulus intact. This explained how the network was able to minimize the interference from proceeding or subsequent stimuli when encoding each stimulus.

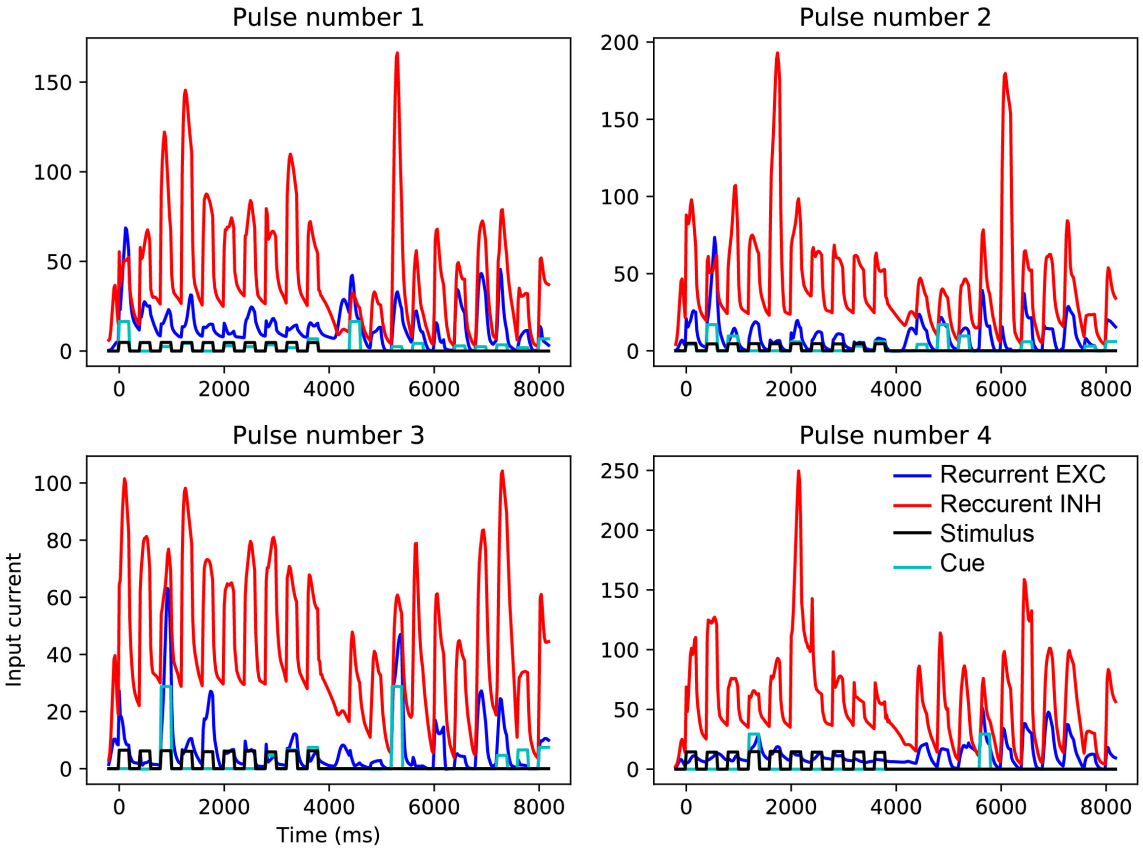

Figure 8. Currents flow between hidden neurons for the first four motion stimuli

The currents flowing into the top five neurons responsible for encoding each motion stimulus were plotted. The figure shows input currents to the top five neurons of each pulse stimulus based on four sources of the incoming activities: excitatory neurons ("Recurrent EXC"), inhibitory neurons ("Recurrent INH"), motion-tuned neurons ("Stimulus"), and cue-tuned neurons ("Cue").

Discussion

The current results suggest that the network encodes multiple sequentially presented stimuli by sparsely storing information for each stimulus in the synaptic efficacy of hidden neurons. With the understanding of how the network uses excitatory and inhibitory currents to gate the information getting into each neuron to minimize interference, we are on the way of designing novel neural networks that further strengthen such gating mechanism. This could potentially alleviate the network's deficiency in memorizing items presented in the middle, as well as improve its memory capacity to store more items. Such enhancement in short-term memory capacity would be beneficial for more complex tasks that require the network to keep track of multiple events, similar to what humans need to do in lots of real-life situations.

References

1. Chafee, M. V. & Goldman-Rakic, P. S. Matching patterns of activity in primate prefrontal area 8a and parietal area 7ip neurons during a spatial working memory task. *J. Neurophysiol.* **79**, 2919–40 (1998).
2. Funahashi, S., Bruce, C. J. & Goldman-Rakic, P. S. Mnemonic Coding of Visual Space in the Monkey's Dorsolateral Prefrontal Cortex. *JOURNALOFNEUROPHYSIOLOGY* **6**, (1989).
3. Hawkins, J. & Ahmad, S. Why Neurons Have Thousands of Synapses, A Theory of Sequence Memory in Neocortex. *Front. Neural Circuits* **10**, 23 (2016).
4. Laughlin, S. B., de Ruyter van Steveninck, R. R. & Anderson, J. C. The metabolic cost of neural information. *Nat. Neurosci.* **1**, 36–41 (1998).
5. Masse, N. Y., Yang, G. R., Song, H. F., Wang, X., & Freedman, D. J. Circuit mechanisms for the maintenance and manipulation of information in working memory. doi:10.1101/305714 (2018).
6. Mayo, C. W., & Crockett, W. H. Cognitive complexity and primacy-recency effects in impression formation. *The Journal of Abnormal and Social Psychology*, *68*(3), 335 (1964).
7. Mongillo, G., Barak, O., & Tsodyks, M. Synaptic theory of working memory. *Science*, *319*(5869), 1543-1546 (2008).
8. Murdock Jr, B. B. The serial position effect of free recall. *Journal of experimental psychology*, ***64*(5)**, 482 (1962).
9. Olshausen, B. & Field, D. Sparse coding of sensory inputs. *Curr. Opin. Neurobiol.* **14**, 481– 1932 487 (2004).