

Modeling Hippocampal-Cortical Dynamics During Event Processing

Qihong Lu (q lu@princeton.edu)

Uri Hasson (hasson@princeton.edu)

Kenneth A. Norman (knorman@princeton.edu)

Department of Psychology and Princeton Neuroscience Institute, Princeton University

Abstract

In this study, we present a two-component neural network model that describes how cortex and hippocampus jointly process event sequences. Cortex is modeled as a recurrent neural network, whose behavior is governed by actively maintained parameters specifying features of the current event. Hippocampus can take “snapshots” of sets of actively maintained parameters and then retrieve these parameter sets in response to partial cues. With functional alignment methods, we qualitatively captured patterns of inter-subject correlation (ISC) from a recent human neuroimaging study. Specifically, we observed enhanced ISC when hippocampus retrieved stored parameters relating to the current event and fed these into the cortex. These results support our formalization of how hippocampus and cortex collaboratively process events, and provide a proof-of-concept demonstration of our computational modeling framework for group-level “brain coupling” phenomena.

Keywords: episodic retrieval; event cognition; neural net

Introduction

Although the hippocampus is critical for episodic retrieval, the processing of coherent natural dynamic stimuli can be done (to some degree) in the absence of a functioning hippocampus. For example, it is known that patients with hippocampal damage can carry on a coherent conversation (Duff, Hengst, Tranel, & Cohen, 2006; Milner, Corkin, & Teuber, 1968; Scoville & Milner, 1957). We hypothesize that this occurs because cortex can actively maintain key parameters of the current event (the “event model”; Radvansky and Zacks, 2011) on its own. These parameters, in conjunction with learned weights, allow cortex to predict upcoming states of the world (e.g., actively maintaining that Bob is the barista, coupled with learned weights encoding how baristas usually act, is enough to predict how Bob will act). This system works so long as there are no interruptions that disrupt storage of actively maintained parameters. If an interruption occurs, hippocampus is needed to retrieve these parameters from long-term storage.

A recent experiment simulated this phenomenon with normal human subjects (J. Chen et al., 2016). In the study, three different groups of participants watched a movie, divided into two parts. The Recent Memory (RM) group watched the two parts without any interruption; the Distant Memory (DM) group watched the two parts with a one-day gap in between; the No Memory (NM) group watched the second part of the movie only. The NM group mimics the experience of hippocampal patients. The one-day gap in the DM simulates an inter-

ruption, which is known to be disruptive for hippocampal patients. Empirically, stronger hippocampal-cortical interactions were observed among participants in the DM group, suggesting stronger episodic reinstatement compared to other groups (J. Chen et al., 2016). We hypothesize that this occurred because i) the one-day gap cleared out the parameter buffer, and ii) in this situation, hippocampus was called upon to retrieve missing parameters once the movie resumed.

In the present study, we propose i) a computational formalization of the functions of hippocampus and the cortex during event processing and episodic retrieval; ii) a method of modeling inter-subject correlation (Hasson, Malach, & Heeger, 2010). We qualitatively captured several findings. For example, during the processing of the second half of the movie, the DM-RM inter-group ISC gradually converged with the RM-RM inter-group ISC, suggesting that the interpretation of ongoing events in the DM group gradually converged with the interpretation of these events in the RM group (J. Chen et al., 2016).

Model

In this section, we state the high-level formalization of the functions of the cortex and hippocampus during event processing. Then we describe the structure of the stimulus environment and the inner workings of the model in more detail.

Model overview

The cortex is implemented by a standard long short-term memory (LSTM) network (Hochreiter & Schmidhuber, 1997) with noisy neurons. The LSTM layer receives information from two layers: 1) a representation of the current (observable) state, and 2) a buffer that actively maintains the values of parameters that specify features of the current situation (for expediency, the buffer is set up with intrinsic active maintenance, such that units maintain their activity levels unless directly reset to a new value). As described below, the environment is set up such that the mapping between the current state and the next state is dependent on the parameters stored in the buffer (i.e., you need to know certain parameter values to predict certain state transitions; for example, you need to know whether a coffee shop accepts cash to predict what will happen if you try to pay in cash). In the current (early-stage) version of the model, the hippocampus is not modeled explicitly – rather, it is simulated by allowing the network to “fill in” the full set of parameters for a given event if it has previously been exposed to these parameters (future versions of the model will simulate the hippocampal network directly).

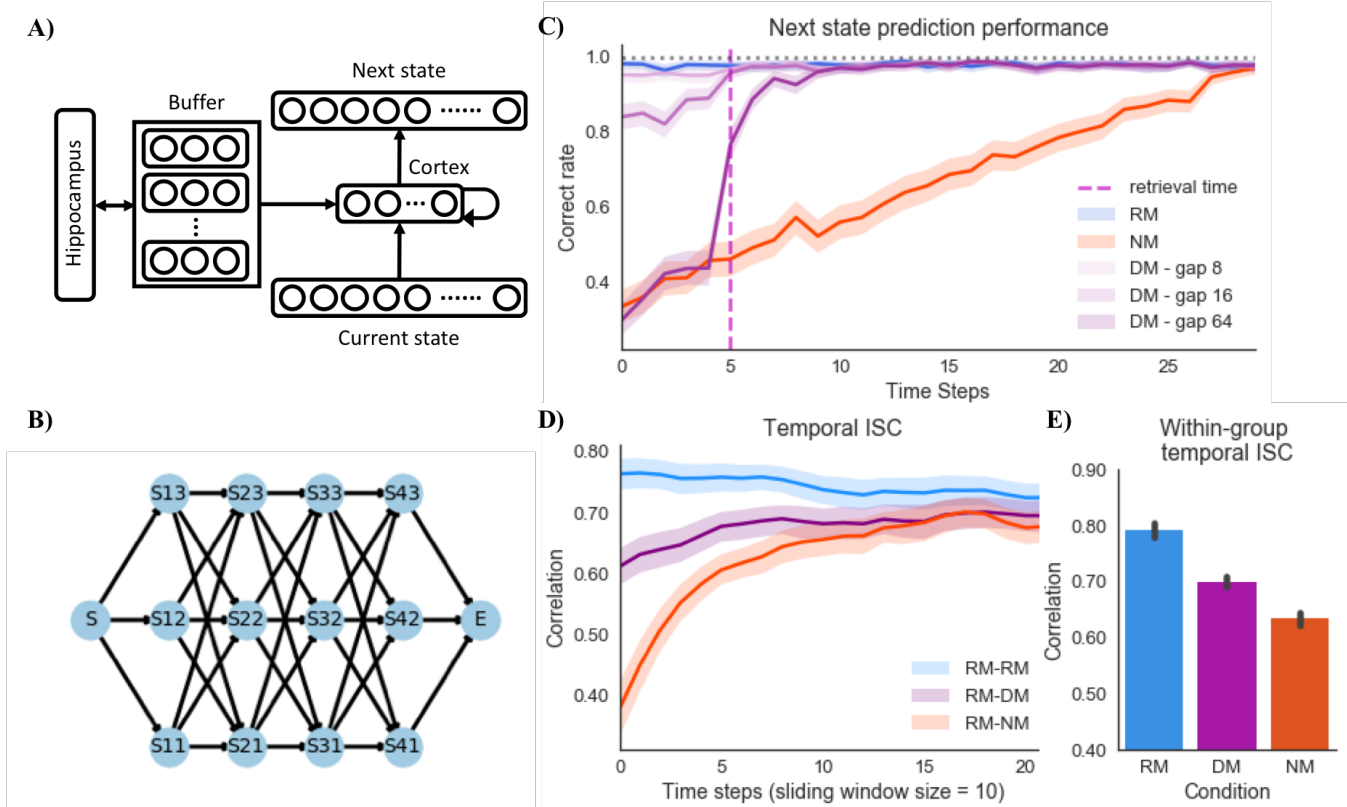


Figure 1: **A)** The model architecture. The cortex is modeled as a standard LSTM. Parameters specifying properties of the current event are actively maintained in the buffer, which feeds into the LSTM layer. Hippocampus is not directly simulated as a neural network; rather, hippocampal retrieval is simulated by filling in missing parameter values in the buffer. The model is trained to predict the next state. **B)** A shortened event graph with length $T = 4$ and branching factor $K = 3$. S_{tk} denotes the k^{th} event at time t . We used $T = 30$ and $K = 3$ for our simulation. **C)** Next state prediction accuracy during the second event sequence. The gap length (for the DM group) is the number of noise vectors fed to the network between the two event sequences. We chose length 64 for the ISC analysis. **D)** Inter-group sliding-window inter-subject correlation (ISC). Initially, the RM group is more correlated with other members of that group than with the DM group or the NM group. When hippocampal retrieval occurs for the DM group (at time 5), this pulls the DM group closer to the RM group. The ISC value at time t is computed using the hidden states from time t to $t + 10$. **E)** Within-group ISC for the entire time course (the second event sequence). *Error bars/bands for all plots are 99% bootstrapped confidence intervals with 10,000 bootstrapped samples.

Stimulus representation

The model is trained on synthetic event sequences, generated by a graph (Figure 1A). The graph has T layers of nodes, where T is the length of the event sequence. Each node, representing an *event*, has K branches, which represent K possible next events. All nodes have the same branching factor. To obtain graphs with inhomogeneous branching factors, we can set the probability of some branches to zero, so we think the graph is sufficiently general. In our simulation, we chose $T = 30$ and $K = 3$.

To simulate the parameter-dependence of event structure, the transition probabilities out of each layer t in the graph are controlled by a *event parameter*, p_t , which can take K possible values; each of these K values specifies a unique successor state S_{tk} . For example, if $p_4 = 1$, then S_{41} will happen at $t = 4$; if $p_4 = 2$ then S_{42} will happen at $t = 4$. Concretely, p_4 might

represents the event parameter “weather”, which can take any value from {“sunny”, “windy”, ..., “rainy”}. Then $p_4 = 1$ represents “weather” is “sunny” which leads to successor state S_{41} (e.g., putting on sunscreen). The upshot of this scheme is that transitions are completely deterministic, conditioned on knowing the right parameter values. If the model does not know the parameter value that controls a particular transition, it can not predict which of the K successor states will occur next.

At time t , the model is allowed to observe the current state (i.e., where it is on the graph). The model is also informed about a new, randomly chosen parameter value at each time step (sampling without replacement from the set of parameters); this act of parameter observation is simulated by activating the units corresponding to that parameter value in the buffer. Note that the parameter that is revealed on a particular time step may not be relevant on that time step. For example,

on time step 2, the model might learn that $p_4 = 1$ (“weather” is “Sunny”), but that particular fact is not relevant until time step 4. Because the sequence of parameter values is revealed in a random order, the probability of observing or having observed the “right” parameter value at time t is t/T . This mimics the fact that in reality, we typically do not know all the relevant information to make the correct prediction.

In addition to observing parameter values, the model is also able to retrieve stored sets of parameter values from the hippocampus; as noted above, this is not directly simulated in the initial version of the model – rather, we simply reinstate these parameter values at moments when we deem hippocampal retrieval to have taken place. We refer to the fully-specified set of parameter values as an “episode”.

We simulated the two parts of a movie (J. Chen et al., 2016) by generating two random sequences using the same set of parameter values, which ensures the transition structure in the second event sequence is consistent with the episode learned from in the first sequence. This design choice implicitly assumes that two parts of the movie are consistent (i.e., the world is relatively stationary over time).

Experiment conditions

To simulate the three experiment conditions in the fMRI experiment (J. Chen et al., 2016), we experimentally controlled the action of the hippocampus for all time steps. During the second half of the movie, since the Recent Memory (RM) group was not interrupted, they should have access to almost all relevant event parameters. Therefore, we revealed all true parameter values for all time steps to the models in the RM group. For the No Memory (NM) group, the values of the event parameters need to be discovered as the event unfolds. Since the model observes parameter values in a random order, it is often the case that NM models have not yet learned the parameter required to predict the next state. The Distant Memory (DM) group has stored an episodic memory pertaining to the ongoing event. Initially, the experience of the DM group is similar to NM. We simulated episodic retrieval by activating all of the true parameters in the buffer on the 5th time point. We simulated the one-day gap in the DM condition by feeding a sequence of noise patterns in between the two event sequences to corrupt the cortical representation.

Training and testing

We trained 30 networks on 5000 training examples (from $3^{30} \times 30!$ possibilities). Each example contains two subsequences (length T) with identical event parameter values. The model’s training experience was meant to capture all three conditions: During the first half of each training example, the model started with an empty buffer and the parameters were revealed over time (as in NM); during the latter half of each training example, the model was given a chance to learn with all parameters specified (as in RM and also DM, after episodic retrieval occurred). To ensure the model had experience being interrupted, we inserted a random length noise sequence between the two subsequences. The training sets for different

models were generated independently, so different models do not share identical experience, though the generative model for the world is consistent across individuals.

All networks were tested on the same held-out set, which can be viewed as the “movie” in the fMRI experiment (J. Chen et al., 2016). We assigned 10 models to each condition.

Results

Behavior results: “stop-and-ask”

The next state prediction performance during the second event sequence is shown in Figure 1C. This task is similar to the stop-and-ask experiment for humans (J. Chen et al., 2016). For the RM condition, since all parameter values were revealed in the buffer and the cortex contains information from the first event sequence, the prediction performance among subjects in the RM condition is optimal from the very beginning. The models in the NM condition need to learn the true event parameter values as the event sequence unfolds, so accuracy linearly increases from the chance level to the optimal level. For the DM group, the prediction performance peaks when these networks retrieve the episode at $t = 5$.

For the DM group, initial performance in the second sequence depends on the length of the one-day gap, operationalized as the number of noise vectors inserted between the sequences (Figure 1C). A short noise sequence (e.g., length = 8) had minimal impact on the information maintained in the cortex. For the ISC analysis, we chose length 64 noise, which is sufficient to corrupt all information in the cortex.

Neuroimaging results: inter-subject correlation

Typically, inter-subject correlation (ISC) is computed by taking the fMRI timecourse from a particular brain area in each subject and correlating these timecourses; here, we computed ISC based on the hidden representation in the LSTM (Figure 1D) – see “analysis detail” below. The resulting pattern is qualitatively similar to the empirical fMRI result (see Figure 6 by J. Chen et al., 2016). During the “second half of the movie” in the simulation, the ISC between the two subgroups of the RM group (or RM-RM ISC for short) was the highest for all time windows. The RM-DM ISC started out lower than the RM-RM ISC but converged after the “movie” began. Overall, the RM-DM ISC was higher than the RM-NM ISC. These results indicate that the DM group’s representation of the event sequence gradually converged with the RM group’s representation as the event sequence unfolded.

ISC analysis detail The analysis procedure is designed to closely mimic the human fMRI data analysis pipeline (J. Chen et al., 2016). The procedure involves an hyperalignment step (P.-H. Chen et al., 2015; Haxby et al., 2011) followed by a sliding window ISC. We first functionally aligned the hidden representation across networks using the shared response model (SRM) (P.-H. Chen et al., 2015) – a generalization of hyperalignment. Hyperalignment is necessary for comparing hidden unit activities across networks, since the representational geometry of a trained neural network might be arbitrary.

ily oriented relative to hidden unit activity space. Conceptually, SRM factorizes the time course matrix for each model as a subject-specific transformation and a lower dimensional shared time course. We have previously shown that SRM is an effective method for functionally aligning neural networks (Lu, Ramadge, Norman, & Hasson, 2018). We fitted SRM on the training set, then we used the learned SRM to transform the held-out neural activities.

In the second step, we analyzed these functionally-aligned held-out activities with ISC (J. Chen et al., 2016; Hasson et al., 2010). For sliding-window ISC (In Figure 1D), at time t , given two groups of subjects, we extract their hidden units' time courses from time t to time $t + 10$, and compute the Pearson correlation between these two time courses. Repeating this procedure for all time steps traces out a curve that indicates the degree of similarity of internal representations between groups. For the inter-group ISC analysis, we first evenly divided the 10 subjects in each condition into two subgroups. For example, the 10 RM subjects were divided into two subgroups of five subjects, RM1 and RM2. Then we computed the sliding-window ISC between RM1 and RM2 (blue), RM1 and DM2 (purple), RM1 and NM2 (orange) (Figure 1D).

Finally, we did within-group ISC for the entire time course (for the second subsequence) (Figure 1E). We found that RM-RM ISC is the highest, followed by DM-DM ISC and finally NM-NM ISC. This pattern reflects the ordering of the overall uncertainty level across the three conditions. Uncertainty arises whenever the relevant parameter value has not been observed, because the empirical distribution over all possible next states is uniform. The models in the NM group had the highest degree of uncertainty, which made them more sensitive to noise, so the use of noisy neurons had the larger impact on NM models, relative to other two conditions. Consequently, the ISC of the internal representation across subjects within the NM group was the weakest.

Conclusion and future directions

With a simple neural network model, we formalized how the cortex and hippocampus jointly process event sequences. Although the cortex can predict event sequences by actively maintaining parameters of the current event (which serve to disambiguate the network's predictions), any interruption can corrupt this actively-maintained information. Hippocampus can recover the information by reinstating a previously stored episode, which contains a list of event parameters. With functional alignment, we were able to qualitatively capture some group-level brain-coupling patterns from a recent fMRI study (J. Chen et al., 2016): When the delayed-memory networks retrieved memories, making their buffer content more similar to the recent-memory group, we observed enhanced inter-group ISC. These results validate our model formalization and provide a proof-of-concept demonstration of our computational framework for further ISC modeling.

The present study is our initial attempt to understand cortical-hippocampal interaction during naturalistic event pro-

cessing. In the future, we plan build a more detailed hippocampal model to study the computational principles of episodic retrieval and cortical-hippocampal interaction. For the training environment, we plan to explore richer event transition structure, more realistic parameter-dependency and naturalistic event schemas.

Acknowledgments

This work was supported by a Multi-University Research Initiative grant to KAN and UH (ONR/DoD N00014-17-1-2961). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Office of Naval Research or the U.S. Department of Defense. We thank Daniel Greenidge, Andre Beukers and Silvy Collin for insightful discussion.

References

- Chen, J., Honey, C. J., Simony, E., Arcaro, M. J., Norman, K. A., & Hasson, U. (2016). Accessing Real-Life episodic information from minutes versus hours earlier modulates hippocampal and High-Order cortical dynamics. *Cereb. Cortex*, 26(8), 3428–3441.
- Chen, P.-H., Chen, J., Yeshurun, Y., Hasson, U., Haxby, J., & Ramadge, P. J. (2015). A Reduced-Dimension fMRI shared response model. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in neural information processing systems 28* (pp. 460–468). Curran Associates, Inc.
- Duff, M. C., Hengst, J., Tranel, D., & Cohen, N. J. (2006). Development of shared information in communication despite hippocampal amnesia. *Nat. Neurosci.*, 9(1), 140–146.
- Hasson, U., Malach, R., & Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends Cogn. Sci.*, 14(1), 40–48.
- Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., ... Ramadge, P. J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, 72(2), 404–416.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.*, 9(8), 1735–1780.
- Lu, Q., Ramadge, P., Norman, K. A., & Hasson, U. (2018). Measuring representational similarity across neural networks. *Poster to be presented at the 40th Annual Meeting of the Cognitive Science Society*, Madison, Wisconsin.
- Milner, B., Corkin, S., & Teuber, H.-L. (1968). Further analysis of the hippocampal amnesic syndrome: 14-year follow-up study of H.M. *Neuropsychologia*, 6(3), 215–234.
- Radvansky, G. A., & Zacks, J. M. (2011). Event perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(6), 608–620.
- Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatry*, 20(1), 11–21.