Published in final edited form as:

IEEE Trans Neural Syst Rehabil Eng. 2018 February ; 26(2): 272–280. doi:10.1109/TNSRE.2016.2604423.

Sparse Large-Scale Nonlinear Dynamical Modeling of Human Hippocampus for Memory Prostheses

Dong Song^a, Brian S. Robinson^a, Robert E. Hampson^b, Vasilis Z. Marmarelis^a, Sam A. Deadwyler^b, and Theodore W. Berger^a

^aDepartment of Biomedical Engineering, University of Southern California, Los Angeles, CA 90089 USA

^bDepartment of Physiology & Pharmacology, Wake Forest University, School of Medicine, Winston-Salem, NC 27157 USA

Abstract

In order to build hippocampal prostheses for restoring memory functions, we build sparse multi-input, multi-output (MIMO) nonlinear dynamical models of the human hippocampus. Spike trains are recorded from hippocampal CA3 and CA1 regions of epileptic patients performing a variety of memory-dependent delayed match-to-sample (DMS) tasks. Using CA3 and CA1 spike trains as inputs and outputs respectively, sparse generalized Laguerre-Volterra models are estimated with group lasso and local coordinate descent methods to capture the nonlinear dynamics underlying the CA3-CA1 spike train transformations. These models can accurately predict the CA1 spike trains based on the ongoing CA3 spike trains during multiple memory events, e.g., sample presentation, sample response, match presentation and match response, of the DMS task, and thus will serve as the computational basis of human hippocampal memory prostheses.

Keywords

hippocampus; spatio-temporal pattern; spike; memory

I. Introduction

We define hippocampal prosthesis as a closed-loop system that bypasses damaged hippocampal region(s) to restore or enhance memory functions (Fig. 1) [1]. Hippocampus is a brain region responsible for the formation of new long-term episodic memories [2]. It receives signals carrying short-term memories from neocortices and transforms them with its feedforward pathways into signals that can be stored as long-term memories back in neocortices. If a hippocampal region is damaged due to diseases or injuries, new long-term memories cannot be formed although short-term memories remain intact [3]. A hippocampal memory prosthesis is designed to bi-directionally communicate with the hippocampus or

Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

Correspondence and proofs should be sent to: Dong Song, 403 Hedco Neuroscience Building, University of Southern California, Los Angeles, CA 90089, Tel: 213-740-8063, Fax: 213-740-5687, dsong@usc.edu.

surrounding cortical regions by recording input signals from an upstream region (e.g., CA3), and stimulating output signals back to a downstream region (e.g., CA1). If this input-output transformation sufficiently mimics the input-output transformation performed by the intact hippocampus, damaged hippocampal region would be bypassed with the reinstated hippocampal signals and long-term memory functions would be restored.

Hippocampal memory prosthesis is a specific form of cortical prostheses. Different from sensory prostheses where the inputs are external sensory (e.g., visual or auditory) signals and the outputs are internal electrical stimulations to the brain or its peripherals [4, 5], or motor prostheses where the inputs are internal motor cortical signals (e.g., M1 spike trains) and the outputs are external signals such as movements or muscle activations [6–9], a cortical prosthesis, by contrast, uses internal brain signals, i.e., ensemble neural activities or spike trains, which carry highly processed sensory and motor signals, as both inputs and outputs (Fig. 1). A computational model that can accurately replicate the transformations from input spike trains to output spike trains then becomes essential for building such a cortical prosthesis. This model can be formulated as a multiple-input, multiple-output (MIMO), point-process, nonlinear dynamical model of spike train transformation [10, 11].

For the past decades, our team has been working on developing hippocampal memory prostheses in various animal preparations. In our previous studies, we have shown that, using a closed-loop system based on nonlinear dynamical input-output models, we can restore the hippocampal signal transmissions and transformations in (a) hippocampal slices of rodents [12], (b) intact hippocampus of behaving rodents [13–15], and (c) intact hippocampus of behaving nonhuman primates (NHPs) [16, 17]. In the latter two cases, it has been shown that electrical stimulation of a downstream hippocampal region (i.e., CA1) predicted from an upstream hippocampal region (i.e., CA3) by the MIMO nonlinear dynamical model of the hippocampal CA3-CA1 can restore and even enhance memory functions. Following these studies, we have moved on to the most critical step towards our ultimate goal: developing hippocampal memory prostheses for human subjects [18]. In this paper, we report our recent progress on building MIMO nonlinear dynamical models of the human hippocampus using spike train data recorded from epileptic patients performing long-term memory dependent tasks.

One of the main challenges of developing human hippocampal prostheses is the significantly larger model scales. Human memory functions are much more complex than those of rodents or NHPs. In order to replicate those functions, large numbers of hippocampal neurons encoding more memory information need to be recorded, and the potentially more complex transformations between their spiking activities need to be modeled. For example, the human CA3-CA1 datasets analyzed in this study can include approximately a hundred hippocampal neurons, and this number is expected to increase rapidly in near future due to the fast advance of recording techniques. The resulting human CA3-CA1 MIMO model contains a much larger number of model coefficients than those of rodents or NHPs, and thus poses serious challenge in model estimation. To solve this problem, we have developed a new sparse MIMO model estimation approach utilizing a group regularized likelihood formulation implemented with a highly scalable and robust local coordinate descent (LCD) technique. This approach solves the large-scale model estimate problem by optimizing the

MIMO model structure with a sparse representation of the model coefficients. With LCD, coefficient estimation does not rely on the maximum-likelihood method, which becomes difficult or even impossible in large-scale models. Using this method, we have built compact MIMO model that can highly accurately predict the CA1 (output) spatio-temporal patterns of spikes based on the ongoing CA3 (input) spatio-temporal patterns of spikes on a single-trial basis. Such MIMO models will be further used to drive the electrical stimulation to the hippocampal CA1 region for restoring and enhancing memory functions of the patients.

II. Methods

A. Human Subject Recording Procedures

Adult patients ($n = 3$) suffering from pharmacologically refractory epilepsy are surgically implanted with FDA-approved hippocampal electrodes capable of field potential (macro-) and single-unit (micro-) recordings (Ad-Tech Medical Instrumentation Corporation, Racine, WI) for localization of seizures. All procedures performed in human patients were reviewed and approved by Wake Forest University Health Sciences IRB (protocol IRB00023148, approved 4/4/13, continuing review approval 4/4/16). Inclusion in this study is voluntary and consented to separately from the surgical procedure. All study participants undergo appropriate clinical epilepsy screening evaluations. A frameless BrainLab Cranial Navigation System (BrainLab North America, Westchester, IL) is used to plan and guide electrode entry points, stereotaxic electrode trajectories and targets within the CA3 and CA1 subregions of each hippocampus. Electrode localization is confirmed using postoperative MRI. Single unit neural activities (i.e., spike trains) are recorded and isolated using the Blackrock Cervello Elite electrophysiological recording system with a raw data acquisition frequency at 30k samples/sec without filtering, and a spike sorting frequency at 30k samples/sec with 500–5,000 Hz bandpass filtering.

B. Cognitive and Behavioral Tasks

Cognitive and behavioral experiments consist of visual object and spatial position oriented delayed match-to-sample (DMS) tasks (Fig. 2). Object and position trials start with circle and square cues presented on the touchscreen, respectively. In the Sample Phase, an object is presented in a specific position of the screen (Sample Presentation) and the patient touches the object to form a Sample Response event. In the Match Phase of an object trial, the patient needs to choose and touch the correct object that he has seen in the previous Sample Phase among distractors to generate a correct Match Response; in the Match Phase of a position trial, patients need to choose and touch the correct position where he has seen the object in the previous Sample Phase among other positions to generate a correct Match Response. Memory functions are evaluated with the percentage of correct responses during a DMS session that consists of approximately 100 trials. Objects are either CANTAB (Cambridge Neuropsychological Test Automated Battery) non-verbalizable images [19] or clip-art images. All MIMO models described in this paper are estimated and validated using CA3 and CA1 data recorded during the DMS tasks. Patients are not recorded when they are not performing the task.

C. Sparse Large-Scale MIMO Nonlinear Dynamical Model

1) MIMO Nonlinear Dynamical Model—We have formulated a nonlinear dynamical MIMO model for the development of hippocampal memory prostheses [10, 11, 20, 21]. In this approach, the identification of spatio-temporal pattern transformations from the input region to the output region is formulated as the estimation of a MIMO model that can be decomposed into a series of multiple-input, single-output (MISO) spiking neuron models as (Fig. 3):

$$w = u(k, x) + a(h, y) + \varepsilon(\sigma) \quad (1)$$

$$y = \begin{cases} 0 & \text{when } w < \theta \\ 1 & \text{when } w \geq \theta \end{cases} \quad (2)$$

$$u(t) = k_0 + \sum_{n=1}^N \sum_{\tau=0}^M k_1^{(n)}(\tau) x_n(t-\tau) + \sum_{n=1}^N \sum_{\tau_1=0}^M \sum_{\tau_2=0}^M k_2^{(n)}(\tau_1, \tau_2) x_n(t-\tau_1) x_n(t-\tau_2) \quad (3)$$

$$a(t) = \sum_{\tau=1}^M h_1(\tau) y(t-\tau) + \sum_{\tau_1=1}^M \sum_{\tau_2=1}^M h_2(\tau_1, \tau_2) y(t-\tau_1) y(t-\tau_2) \quad (4)$$

Variables x and y are input (e.g., CA3) and output (e.g., CA1) spike trains. k are feedforward Volterra kernels describing the mapping from x to the post-synaptic potential u . h are feedback Volterra kernels describing the transformation from preceding y to u . Zeroth-order kernel k_0 models the input-independent baseline firing rate. First-order feedforward kernels $k_1^{(n)}$ describe the linear relation between the n^{th} input x_n and u , as functions of the time intervals τ between the past time and the present time. Second-order feedforward kernels $k_2^{(n)}$ describe the nonlinear interaction between pairs of spikes in the n^{th} input x_n as they jointly affect u , in addition to their individual first-order effects. First-order feedback kernel h_1 and second-order feedback kernel h_2 can be interpreted similarly by treating preceding y as an extra input. N is the number of inputs. M is the system memory length. The total synaptic potential u are added with a feedback after-potential a involving feedback kernels h and preceding output y , and a Gaussian noise ε with standard deviation σ , to form the pre-threshold potential w . When w crosses threshold θ , an action potential is generated. For simplicity, higher-order kernels or cross-kernels are not included in this study.

As shown in Equation 3 and 4, Volterra kernels essentially express the systems nonlinear dynamics in a linear form by pre-structuring the nonlinearity in multiplications between their input variables. In addition, the joint effect of the pre-threshold Gaussian noise and the threshold is equivalent to a sigmoidal normal cumulative distribution function and a

Bernoulli process (Fig. 3), which is equivalent to a *probit* link function in a generalized linear model [22]. Therefore, this model is termed generalized Volterra model (GVM) [20, 21].

2) Laguerre Expansion of Volterra Kernels—To reduce the model complexity, we expand both feedforward and feedback Volterra kernels with Laguerre basis functions b (Fig. 3, bottom):

$$u(t) = c_0 + \sum_{n=1}^N \sum_{j=1}^J c_1^{(n)}(j) \nu_j^{(n)}(t) + \sum_{n=1}^N \sum_{j_1=1}^J \sum_{j_2=1}^{j_1} c_2^{(n)}(j_1, j_2) \nu_{j_1}^{(n)}(t) \nu_{j_2}^{(n)}(t) \quad (5)$$

$$a(t) = \sum_{j=1}^J c_1^h(j) \nu_j^h(t) + \sum_{j_1=1}^J \sum_{j_2=1}^{j_1} c_2^h(j_1, j_2) \nu_{j_1}^h(t) \nu_{j_2}^h(t) \quad (6)$$

where $\nu_j^{(n)}(t) = \sum_{\tau=0}^M b_j(\tau) x_n(t - \tau)$ and $\nu_j^h(t) = \sum_{\tau=1}^M b_j(\tau) y(t - \tau)$; $c_1^{(n)}$, $c_2^{(n)}$, c_1^h , and c_2^h are the sought Laguerre expansion coefficients of $k_1^{(n)}$, $k_2^{(n)}$, h_1 , and h_2 , respectively; c_0 is simply equal to k_0 ; J is the number of basis functions. The formula of obtaining discrete Laguerre basis functions are provided in a previous publication [21]. Since Laguerre basis consists of orthonormal functions with exponentially decaying shapes (Fig. 4, left), it can effectively fit a variety of temporal processes with a small number of basis functions. In this study, 3–4 Laguerre functions are used; M is equal to 500, reflecting a 1 sec memory length with a 2 msec temporal resolution.

3) Sparse Model Estimation with Group Lasso—To yield sparse representation of the model, model coefficients are calculated with a group regularized estimation method [21, 23, 24]. In this approach, model coefficients are grouped with respect to each input and each model order. Since two model orders, i.e., first-order and second-order are included in this study, there are a total of $2N+2$ groups for the N inputs and one output (Equation 7). With a group lasso formulation, the coefficients are selected and estimated simultaneously at the group level by minimizing a target function S consisting of the negative log-likelihood $-l$ and a grouped penalty term involving the summation of L_2 -norm of each group (Fig. 4, middle).

$$S(c) = -l(c) + \lambda \left(\sum_{n=1}^N \|c_1^{(n)}(j)\|_2^1 + \sum_{n=1}^N \|c_2^{(n)}(j_1, j_2)\|_2^1 + \|c_1^h(j)\|_2^1 + \|c_2^h(j_1, j_2)\|_2^1 \right) \quad (7)$$

$$l(c) = \sum_{t=1}^T [y(t) \log P(t) + (1 - y(t)) \log(1 - P(t))] \quad (8)$$

$$P(c) = \Phi(u(t) + a(t)) \quad (9)$$

where Φ is the normal cumulative distribution function that maps the summation of post-synaptic potential u and after-potential a into the output firing probability P . It is essentially a sigmoidal function that transforms $u + a$ into a value between 0 and 1. T is the data length; λ is the sparsity parameter.

4) Implementing Group Lasso with Local Coordinate Descent—In this study, group lasso estimation is implemented with a LCD method [25, 26], in which the model coefficients are updated one by one along fixed descent directions with line search to minimize the target function (Fig. 4, right). Since the computational cost increases only linearly with the number of coefficients (i.e., model scale), LCD can be reliably and efficiently applied to solve very large-scale model estimation problems.

In LCD, optimization of model coefficients c is performed by iterating between (a) making a quadratic approximation of the log-likelihood l at the current estimated linear predictor at each step, and (b) individually updating the estimate of each coefficient. The local quadratic approximation of l requires the calculation of second derivative vector w of l with respect to the current estimate of the linear predictor, i.e., $w(t) = -2l(t) / \Phi^{-1}(\tilde{P}(t))^2$. The calculation of $w(t)$ for the *probit* link function is given in a previous paper [21]. In LCD, w is recalculated after updating all individual coefficient estimates. The model residual value $\tilde{r}$ is used to accelerate estimation and updated after each coefficient estimate.

For simplicity, in the following we express each group of coefficients

$c_1^{(n)}(j)$, $c_2^{(n)}(j_1, j_2)$, $c_1^h(j)$, and $c_2^h(j_1, j_2)$ as a vector c_q with individual elements c_{qp} where $q = 1, 2, \dots, 2N+2$ and $p = 1, 2, \dots, P_q$. P_q is the total number of coefficients in group q . It is apparent that P_q is equal to J and $J(J+1)/2$ in first-order and second-order groups,

respectively. Similarly, all corresponding convolution vectors $\nu_j^{(n)}(t)$ and $\nu_j^h(t)$, and their element-wise products, $\nu_{j_1}^{(n)}(t)\nu_{j_2}^{(n)}(t)$ and $\nu_{j_1}^h(t)\nu_{j_2}^h(t)$, are expressed as v_{qp} ; the collection of all vectors within a group is expressed as matrix V_q , which are all combined to form V . The concatenation of c_0 and all c_{qp} is denoted as c .

The algorithm for estimating $\tilde{c}$ is described below:

1. Start with a set of the initial values of $\tilde{c}$
2. Calculate w with Equation 10
3. $\tilde{r} = y - \Phi(V\tilde{c})$

$$4. \quad \tilde{c}_0 \leftarrow \nu_0^T W \tilde{r} / \nu_0^T W \nu_0 + \tilde{c}_0$$

$$5. \quad \tilde{r} \leftarrow \tilde{r} - (\tilde{c}_0 - \tilde{c}_0^*) \nu_0$$

6. For each group $\tilde{c}_q$

a. If

$$\frac{1}{T} \|V_q^T \tilde{r} + V_q^T V_q c_q\| < \sqrt{P_q} \lambda$$

$$i. \quad \tilde{r} \leftarrow \tilde{r} + \tilde{c}_{qp} V_q$$

$$ii. \quad \text{Set } \tilde{c}_q = 0$$

b. Otherwise, for each $\tilde{c}_{qp}$

i.

$$\tilde{c}_{qp} \leftarrow \frac{\frac{1}{T} \nu_{qp}^T W \tilde{r} + \frac{1}{T} \nu_{qp}^T W \nu_{qp} \tilde{c}_{qp}}{\frac{1}{T} \nu_{qp}^T W \nu_{qp} + \lambda_i \sqrt{P_q} / (\|\tilde{c}_q\| + \delta)}$$

ii.

$$\tilde{r} \leftarrow \tilde{r} - (\tilde{c}_{qp} - \tilde{c}_{qp}^*) \nu_{qp}$$

7. Repeat (2) through (6) until convergence

In the above algorithm, * denotes the estimate from the previous iteration; W is a diagonal matrix with elements w ; δ is a small value added to prevent division by zero. In addition, each column of V must be standardized to have zero mean and unit variance before estimation.

5) Optimization of the Sparsity Parameter—The relative importance of the likelihood and the penalty term is controlled by the sparsity parameter λ (Equation 7). A larger value of λ yields sparser estimation of the coefficients, i.e., more groups of model coefficients are deselected or set to zeros. In this study, λ is optimized with a 10-fold cross-validation (CV) procedure.

First, the minimal value of λ that yields complete sparsity is calculated as:

$$\lambda_{\max} = \max_q \frac{1}{T} \frac{\|\nu_q^T \tilde{r}\|}{\sqrt{P_q}}$$

where $\tilde{r}$ is calculated with the c_0 only model, i.e., zeroth-order model. A series of 100 λ values that are logarithmically spaced between $\lambda_{\max}$ and $0.01 \lambda_{\max}$ are used to estimate the

sparse models and determine the selection path. The λ value that yields the smallest out-of-sample negative log-likelihood is selected as the optimal λ .

6) Model Validation and Prediction—As described previously [10, 27, 28], model goodness-of-fit can be evaluated with a Kolmogorov-Smirnov (KS) test based on the time-rescaling theorem (Fig. 5) in means of the maximal distance between the KS plot and the 45-degree diagonal line. In this study, we use normalized KS-score, i.e., the ratio between the maximal distance between the KS plot and diagonal (Fig. 5, variable a) to the distance between the 95% confidence bound and the diagonal (Fig. 5, variable b), as the final measure. If the normalized KS score is below 1, the KS plot is within the bounds and the model is considered accurate (Fig. 5, right).

To predict y , u is calculated with inputs x and the estimated feedforward kernels. This forms the deterministic part of pre-threshold potential w . A Gaussian random sequence with the estimated standard deviation is then generated and added to u to render w stochastic. At each time t , if w crosses threshold θ , a spike is generated, i.e., $y(t)$ is set to one, and a feedback process a is triggered and added to the future values of w . This is equivalent to transforming $u + a$ into the firing probability θ and then generate 1 and 0 with a Bernoulli random process (Fig. 3). The calculation then moves on to time $t+1$ with updated w until it reaches the end of the data length.

III. Results

Using methods described in the previous section, we have built human hippocampal CA3-CA1 models with spike trains recorded from epileptic patients performing different forms of the memory-dependent DMS task, i.e., CANTAB and Clip-Art tasks. Depending on the specific surgery procedure, electrode placement, and condition of each patient, various numbers of neurons are recorded. In this paper, we present one set of results from a patient (i.e., Patient A, CANTAB task, unilateral recordings from anterior and posterior hippocampus) with a relatively large number of neurons, and two sets of results from two other patients (i.e., Patient B, CANTAB task, unilateral recordings from anterior hippocampus; Patient C, Clip-Art task, bilateral recordings from anterior hippocampus) with relatively small number of neurons for comparison. Continuous recordings of CA3 and CA1 spike trains are used to estimate and validate the MIMO models. The data lengths of the three patients are 1000, 1000, and 1751 seconds, respectively. To avoid overfitting, only out-of-sample results are shown.

A. Sparse MISO Model of Human Hippocampal CA3-CA1

We have recorded the largest number of neurons (48 CA3 neurons and 49 CA1 neurons) among the three patients from Patient A. The estimated MIMO model thus contains 49 48-input, single-output models. Figure 6 shows the first-order and second-order Volterra kernels (feedforward and feedback) of one representative MISO model.

First-order kernels k_1 are 1-D vectors quantifying the first-order causal relationships between each input (or previous output in the case of feedback kernels h_1) and the output as functions of the time intervals (Fig. 6, top). Second-order kernels k_2 are 2-D matrices

describing the second-order joint effects of pairs of input (or previous output in h_2) spikes on the output in addition to their individual first-order effects (Fig. 6, bottom). In this specific model, among the 48 inputs, 14 inputs show significant first-order kernels, and 16 inputs show significant second-order kernels. It is shown that CA3 neurons are sparsely connected to the CA1 neuron with various kernel shapes and memory lengths. The CA1 neuron has a significant feedback component containing both first-order and second-order kernels.

The goodness-of-fit of this MISO model is evaluated with KS scores. A zeroth-order model, which contains only the mean firing rate of the output, is used as the control. KS scores of zeroth-order, second-order non-sparse, and second-order sparse models are 3.57, 1.71, and 0.99, respectively. This results show that although the non-sparse second-order model can capture a fairly large amount of the system dynamics, it is the sparse second-order model that provides the most accurate prediction among the three models.

B. Validating MIMO Models of Human Hippocampal CA3-CA1

MIMO models are formed by concatenating MISO models of each patient. Since cross-validations are used to obtain optimal sparse models, out-of-sample negative log-likelihood values of sparse models are always no larger than those of non-sparse models, and those of non-sparse models are always no larger than those of zeroth-order models. In fact, some non-sparse models even show infinite negative log-likelihood, indicating serious overfitting with the full set of possible coefficients. Sparse models, on the other hand, always reduce out-of-sample negative log-likelihood from their corresponding zeroth-order models. These results are not shown in this paper for simplicity. Instead, model goodness-of-fits are evaluated with an independent measure: out-of-sample normalized KS-scores of each MISO model (Fig. 7). In Patient A, out of 49 MISO models, 22 (44.9%) sparse and 7 (14.3%) non-sparse MISO models show a KS plot within the 95% confidence bounds. Sparse models significantly out-perform their corresponding non-sparse models in 44 (89.8%) outputs. These results show that sparse models are required for modeling large-scale datasets.

In comparison, Patient B has only 9 inputs and 18 outputs. Out of 18 MISO models, 6 (33.3%) sparse and 5 (27.8%) non-sparse MISO models are within the 95% confidence bounds. Sparse models out-perform their corresponding non-sparse models in 13 (72.2%) outputs. In two outputs, sparse models perform the same as non-sparse models since all inputs are selected in sparse models and make them identical to their corresponding non-sparse models. Patient C has 20 inputs and 23 outputs. Out of 23 MISO models, 10 (43.4%) sparse and 9 (39.1%) non-sparse MISO models are within the 95% confidence bounds. Sparse models out-perform their corresponding non-sparse models in 13 (56.5%) outputs. These results show that in smaller-scale models, the improvements of sparse models over non-sparse models are less significant and the overall performances of MIMO models are not as good as in the larger-scale model, i.e., Patient A, presumably due to the much smaller number of input neurons.

C. Predicting CA1 Spatial-Temporal Patterns with Sparse MIMO Model

Most importantly, CA1 spatio-temporal patterns are predicted from the CA3 spatio-temporal patterns using the estimated sparse MIMO CA3-CA1 models. Figure 8 illustrates CA1

predictions in Patient A. Two 100s long segments containing multiple DMS trials are shown. To facilitate visualization, neurons are ordered descending with respectively their mean firing rates. It is evident that this sparse MIMO model can highly accurately predict the CA1 spatio-temporal patterns on a single-trial basis. The prediction captures both the global trend and fine details of the CA1 patterns. The correlation coefficients between the actual and predicted patterns is 0.78.

We further plot the spatio-temporal patterns of CA1 activities around Sample Presentation and Sample Response events for Patient A (Fig. 9). CA1 neurons consistently exhibit decreased activities before Sample Response events and increased activities afterwards. Despite this general trend, the CA1 patterns also show high degree of variations across different trials. These different patterns presumably encode different objects and object positions that the patient has seen during DMS trials. It is clearly shown that the sparse MIMO model can faithfully replicate these single-trial CA1 patterns during these critical memory formation periods.

CA1 predictions in Patient B are similarly plotted in Figure 10. It shown that the sparse MIMO model can still predict the CA1 patterns with a high degree of accuracy, despite the much smaller number of input and output neurons. However, some fine details are missed, especially in the second segment. Clip-Art tasks, instead of CANTAB tasks are used in Patient C. Variations of firing are significantly weaker in this dataset compared to those of CANTAB tasks (Fig. 11). However, the MIMO model is still able to capture these variations, especially in the high activity ranges.

IV. Discussions

In this paper, we present the first sets of sparse MIMO nonlinear dynamical models of the human hippocampal CA3-CA1 using spike trains recorded from epileptic patients. Since the numbers of neurons recorded in human (which can reach approximately 100) are significantly larger than those recorded in rodents or NHPs (typically 20 – 40 per animal), estimation of the subsequently much larger number of model coefficients becomes a real challenge. In order to solve this large-scale modeling problem, we have improved our MIMO modeling methodology by utilizing sparse representations of MIMO models estimated with a group lasso model selection/estimation method, which is implemented with a more flexible and scalable LCD-based technique. Results show that the sparse MIMO models can accurately predict CA1 spatio-temporal patterns from the ongoing CA3 spatio-temporal patterns on a single-trial basis.

Predicting downstream hippocampal (e.g., CA1) signals from upstream hippocampal (e.g., CA3) signals using a computational model is the first important step towards to our ultimate goal. To build a viable hippocampal memory prosthesis, we need to solve several additional critical issues. First, we need to make sure that the predicted output signals are not only adequately similar to the actual output signals, as we have shown in this paper, but also contain sufficient amount of memory-related information for the downstream region to interpret. In order to do this, we need to combine the MIMO model, which restores output signals from input signal, with a memory decoding model that can interpret both input and

output spatio-temporal patterns as memories. For example, we may first build a memory decoding model for the hippocampal CA3 patterns to assess what memories are encoded, and then build another memory decoding model for the hippocampal CA1 patterns predicted by the MIMO model, to quantify how much of these memories are maintained in the restored signals. We have formulated this modeling framework and successfully applied it in our rodent studies [29], where much simpler memories (i.e., Left or Right locations in the delayed nonmatch-to-sample task) are involved. Human studies, on the other hand, involves much more complex memories with multiple categories, attributes and associations. More powerful and sophisticated modeling strategies are required for building successful memory decoding models for human.

The next critical step, of course, will be to stimulate electrically the downstream hippocampal region with the predicted output signals to restore or enhance memory functions. We have developed a closed-loop system for performing this experiment. In addition to the recording electrodes described in this paper, this system also includes a designated computer for running near real-time MIMO prediction, and a stimulator for delivering electrical pulses to the stimulating electrodes. The experiments will be performed as follows: (a) in the recording session, hippocampal CA3 and CA1 spike trains will be recorded during DMS trials. The difficulty of the DMS task will be adjusted to the level that the patient will make a significant amount of error responses, e.g., 60–70% correct. This metrics is used to assess the memory function of the patient and provides a control for improvement. (b) A Sparse MIMO model will be built using the CA3 and CA1 activities recorded in this session, preferentially those of the correct response trials where the spatio-temporal patterns presumably contain stronger encodings of memories. (c) In the stimulation session, CA3 will be recorded and CA1 patterns will be predicted by the sparse MIMO model. The predicted CA1 patterns will be used to stimulate the CA1 region in real time as the CA3 patterns unfold. The memory performance will be evaluated and compared with that of the control condition. Such experiments are expected to be performed in near future with multiple patients.

Due to experimental limitations, all of our current MIMO models are specific to one memory task, i.e., DMS involving either CANTAB or Clip-Art images. Although the DMS task already contains important memory processes such as memory encoding (during the Sample Response event) and memory retrieval (during the Match Response event) and our results have shown that the MIMO model can accurately predict the CA1 activities during the whole task, it is still unknown how well the MIMO model can be generalized to other memory tasks. Another important future study is to collect data from multiple memory tasks involving richer sets of memory contents and build multi-task MIMO models to more comprehensively test the generality of our modeling approach. Nonetheless, the sparse MIMO nonlinear dynamical modeling study presented in this paper has represented a critical step towards building clinically viable hippocampal memory prostheses.

Acknowledgments

This work was supported by the Defense Advanced Research Projects Agency (DARPA N66001-14-C-4016).

References

1. Berger TW, Ahuja A, Courellis SH, Deadwyler SA, Erinjippurath G, Gerhardt GA, et al. Restoring lost cognitive function. *IEEE Eng Med Biol Mag.* Sep-Oct;2005 24:30–44.
2. Eichenbaum H. The hippocampus and mechanisms of declarative memory. *Behavioural Brain Research.* 1999; 103:123–33. [PubMed: 10513581]
3. Milner, B. Memory and the medial temporal regions of the brain. In: Pribram, KH., Broadbent, DE., editors. *Biology of Memory.* N.Y.: Academic; 1970. p. 29-50.
4. Humayun MS, de Juan E Jr, Weiland JD, Dagnelie G, Katona S, Greenberg R, et al. Pattern electrical stimulation of the human retina. *Vision Research.* 1999; 39:2569–76. [PubMed: 10396625]
5. Loeb GE. Cochlear prosthetics. *Annual Review of Neuroscience.* 1990; 13:357–71.
6. Donoghue JP. Connecting cortex to machines: recent advances in brain interfaces. *Nature Neuroscience.* 2002; 5(Suppl):1085–8. [PubMed: 12403992]
7. Schwartz AB. Cortical neural prosthetics. *Annu Rev Neurosci.* 2004; 27:487–507. [PubMed: 15217341]
8. Nicolelis MA. Brain-machine interfaces to restore motor function and probe neural circuits. *Nature Reviews Neuroscience.* 2003; 4:417–22. [PubMed: 12728268]
9. Song D, Hendrickson P, Marmarelis VZ, Aguayo J, He J, Loeb GE, et al. Predicting EMG with generalized Volterra kernel model. *Conf Proc IEEE Eng Med Biol Soc.* 2008; 2008:201–4. [PubMed: 19162628]
10. Song D, Chan RH, Marmarelis VZ, Hampson RE, Deadwyler SA, Berger TW. Nonlinear dynamic modeling of spike train transformations for hippocampal-cortical prostheses. *IEEE Trans Biomed Eng.* Jun.2007 54:1053–66. [PubMed: 17554824]
11. Song, D., Berger, TW. Identification of Nonlinear Dynamics in Neural Population Activity. In: Oweiss, KG., editor. *Statistical Signal Processing for Neuroscience and Neurotechnology.* Boston: McGraw-Hill/Irwin; 2009.
12. Hsiao MC, Song D, Berger TW. Nonlinear dynamical model based control of in vitro hippocampal output. *Front Neural Circuits.* 2013; 7:20. [PubMed: 23429994]
13. Berger TW, Hampson RE, Song D, Goonawardena A, Marmarelis VZ, Deadwyler SA. A cortical neural prosthesis for restoring and enhancing memory. *Journal of Neural Engineering.* Aug.2011 8:046017. [PubMed: 21677369]
14. Berger TW, Song D, Chan RHM, Marmarelis VZ, LaCoss J, Wills J, et al. A Hippocampal Cognitive Prosthesis: Multi-Input, Multi-Output Nonlinear Modeling and VLSI Implementation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering.* Mar.2012 20:198–211. [PubMed: 22438335]
15. Hampson RE, Song D, Chan RHM, Sweatt AJ, Riley MR, Gerhardt GA, et al. A Nonlinear Model for Hippocampal Cognitive Prosthesis: Memory Facilitation by Hippocampal Ensemble Stimulation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering.* Mar.2012 20:184–197. [PubMed: 22438334]
16. Hampson RE, Song D, Opris I, Santos LM, Shin DC, Gerhardt GA, et al. Facilitation of memory encoding in primate hippocampus by a neuroprosthesis that promotes task-specific neural firing. *J Neural Eng.* Dec.2013 10:066013. [PubMed: 24216292]
17. Song D, Opris I, Chan RH, Marmarelis VZ, Hampson RE, Deadwyler SA, et al. Functional connectivity between Layer 2/3 and Layer 5 neurons in prefrontal cortex of nonhuman primates during a delayed match-to-sample task. *Conf Proc IEEE Eng Med Biol Soc.* 2012; 2012:2555–8. [PubMed: 23366446]
18. Song D, Robinson BS, Marmarelis VZ, Hampson RE, Deadwyler SA, Berger TW. Sparse generalized Volterra model of human hippocampal spike train transformation for memory prostheses. *Proceedings of the IEEE EMBS Conference.* 2015; 2015:3961–4.
19. Sahakian BJ, Morris RG, Evenden JL, Heald A, Levy R, Philpot M, et al. A comparative study of visuospatial memory and learning in Alzheimer-type dementia and Parkinson's disease. *Brain.* Jun; 1988 111(Pt 3):695–718. [PubMed: 3382917]

20. Song D, Chan RH, Marmarelis VZ, Hampson RE, Deadwyler SA, Berger TW. Nonlinear modeling of neural population dynamics for hippocampal prostheses. *Neural Networks*. Nov.2009 22:1340–51. [PubMed: 19501484]
21. Song D, Wang H, Tu CY, Marmarelis VZ, Hampson RE, Deadwyler SA, et al. Identification of sparse neural functional connectivity using penalized likelihood estimation and basis functions. *J Comput Neurosci*. Dec.2013 35:335–57. [PubMed: 23674048]
22. McCullagh, P., Nelder, JA. *Generalized Linear Models*. 2. Boca Raton, FL: Chapman & Hall/CRC; 1989.
23. Song, D., Wang, H., Berger, TW. Estimating sparse Volterra models using group L1-regularization; *Proceedings of the 32nd Annual International Conference of the IEEE EMBS*; 2010. p. 4128-4131.
24. Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society Series B-Statistical Methodology*. 2006; 68:49–67.
25. Breheny P, Huang J. Penalized methods for bi-level variable selection. *Stat Interface*. Jul 1.2009 2:369–380. [PubMed: 20640242]
26. Robinson BS, Song D, Marmarelis VZ, Hampson RE, Deadwyler SA, Berger TW. Estimation of a large-scale generalized Volterra model for neural ensembles with group lasso and local coordinate descent. *Proceedings of the IEEE EMBS Conference*. 2015; 2015:2526–9.
27. Brown EN, Barbieri R, Ventura V, Kass RE, Frank LM. The time-rescaling theorem and its application to neural spike train data analysis. *Neural Computation*. 2002; 14:325–46. [PubMed: 11802915]
28. Haslinger R, Pipa G, Brown E. Discrete Time Rescaling Theorem: Determining Goodness of Fit for Discrete Time Statistical Models of Neural Spiking. *Neural Computation*. Oct.2010 22:2477–2506. [PubMed: 20608868]
29. Song D, Harway M, Marmarelis VZ, Hampson RE, Deadwyler SA, Berger TW. Extraction and restoration of hippocampal spatial memories with non-linear dynamical modeling. *Front Syst Neurosci*. 2014; 8:97. [PubMed: 24904318]

Figure 1.

Modeling hippocampal memory function as a transformation of spatio-temporal patterns of spikes. Top: a conceptual illustration of the hippocampal memory prosthesis for restoring memory functions. Bottom: a hippocampal memory prosthesis bypasses a damaged region by reinstating output spike trains from input spike trains using a multiple-input, multiple-output (MIMO) model.

Figure 2.

Delayed match-to-sample (DMS) tasks for human patients. DMS tasks contain both Object Trials and Spatial Trials, in which either objects or locations of objects during Sample Phases need to be remembered to give correct Match Responses during Match Phases. In CANTAB tasks, objects are non-verbalizable geometric patterns. In Clip-Art tasks, objects are verbalizable cartoon images.

Figure 3.

A MIMO nonlinear dynamical model consists of a series of multiple-input, single-output (MISO) models of spiking neurons that are equivalent to generalized Laguerre-Volterra models.

Laguerre Basis Functions

Group Lasso

Local Coordinate Descent

Figure 4.

Sparse MIMO model estimation are achieved by combining Laguerre basis function expansion, group lasso model selection, and local coordinate descent methods.

Figure 5.

Evaluating the model goodness-of-fit with a Kolmogorov-Smirnov test based on the time-rescaling theorem. Left: time-rescaling of the spike train intervals. Right: the normalized KS-score (i.e., a/b) is used as the final measure of model accuracy. Solid blue line: KS plot of the model; solid black line: 45-degree diagonal; dashed black lines: 95% confidence bounds.

Figure 6.

Patient A: First-order and second-order feedforward kernels (first 48 plots in each figure; k_1 and k_2) and feedback kernels (last subplot in each figure; h_1 and h_2) of one MISO model of the sparse MIMO CA3-CA1 model of Patient A. Note that, in both first-order and second-order kernels, only a subset of inputs have significant effects (i.e., non-zero kernels) on this specific output. All kernels are normalized for visualization.

Figure 7.

Normalized KS-scores of MIMO models of Patient A. Top panel: KS-scores of second-order sparse and non-sparse MISO models. Zeroth-order model are plotted as control. Dashed line: 95% confidence bound; Middle panel: improvements of KS-scores of sparse and non-sparse models over zeroth-order models; Bottom panel: improvements of KS-scores of sparse models over non-sparse models.

Figure 8.

Patient A, CANTAB task: predicting CA1 spatio-temporal patterns from CA3 spatio-temporal patterns with the sparse MIMO model. SP: sample presentation; SR: sample response; MP: match presentation; MR: match response.

Figure 9.

Patient A, CANTAB task, hippocampal CA1 perievent activity patterns (2s before SP to 2s after SR) are accurately replicated by the sparse MIMO model.

Figure 10.

Patient B, CANTAB task: predicting CA1 spatio-temporal patterns from CA3 spatio-temporal patterns with the sparse MIMO model.

Figure 11.

Patient C, Clip-Art task: predicting CA1 spatio-temporal patterns from CA3 spatio-temporal patterns with the sparse MIMO model.