Interpreting Neural Activity Through Linear and Nonlinear Models for Brain Machine Interfaces

Justin C. Sanchez[1], Deniz Erdogmus[2], Yadunandana Rao[2], Sung-Phil Kim[2],
Miguel Nicolelis[3], Johan Wessberg[4], Jose C. Principe[2]

[1]Department of Biomedical Engineering, University of Florida, Gainesville, FL
[2]Department of Electrical Engineering, University of Florida, Gainesville, FL
[3]Department of Neurobiology, Duke University, Durham, NC
[4]Department of Physiology, Göteborg University, Göteborg, Sweden

Abstract—**Brain machine interface (BMI) design can be achieved by training linear and nonlinear models with simultaneously recorded cortical neural activity and behavior (typically the hand position of a primate). We propose the use of optimized BMI models for analyzing neural activity to assess the role of individual neurons and cortical areas in generating the performed movement. Two models (linear-feedforward and nonlinear-feedback) are trained to predict the hand position of a primate from neural recordings in a reaching task. Qualitative and quantitative investigation of the effect of neurons and their corresponding cortical areas through both models yields conclusions consistent with neurophysiologic knowledge. In addition, this analysis revealed the role of these areas and the importance of the neurons in terms of BMI design.**

Keywords—**Brain machine interface, linear/nonlinear models, recurrent neural network, analysis of neural activity**

I. INTRODUCTION

In a landmark article, Nicolelis and Wessberg showed that linear and nonlinear models can approximate the hand trajectory of a primate performing a reaching task using only the firing patterns of an ensemble of cortical neurons [2]. While they and others showed the feasibility of using optimal input-output models for Brain Machine Interfaces (BMIs), the primary focus was to find the best model for accurate trajectory reconstructions [2-9]. For a reaching task not all models perform equally due to differences in ability to exploit trends in neural activity [10-13].

Here, we shift the focus of BMI studies to investigating the interpretations we can make about neural activity from these trained models. All of the proposed models have the ability to encode and store the fundamental timing relationships between neural inputs and hand trajectory in the model parameters [1, 14-16]. A natural next step in BMI development is to hypothesize that these models can be studied in a principled way to extract neurophysiologic trends in the neural recordings. By analyzing the model parameters in a signal-processing context, we can ascertain importance to individual neurons and extract relationships between cortical regions and the desired behavior. Note that this approach to neuronal analysis contrasts with traditional neuroscience methods that implement data-driven reasoning under extremely well controlled experimental paradigms. However, we feel that this methodology seems ill prepared to attack the study of interactions between large populations of neurons used in BMI experiments. Consequently, we propose the use of signal processing constructs to deduce interpretations of neural activity.

The success of neural analysis through models depends on having well trained models that encode the mapping from neural activity to hand kinematics. We must be aware that the model could bias the interpretation due to abstract level of modeling, and difficulties in determining a reasonable fit, model order, and topology. For these reasons, the validity of this approach must be tested before continuing this type of analysis. By investigating how the choice of the model topology affects the interpretation of neural activity, we can test and validate the aptness of this approach.

With these considerations, we compare interpretations of neural activity through two radically different models: linear feedforward Wiener filter (WF) and nonlinear dynamic recurrent multilayer perceptron (RMLP). The WF is extensively used in the BMI literature [2, 7]and the RMLP produced the best results for a reaching task thus far [10-13].

In this paper we train both models to predict the hand trajectory of a behaving primate using various combinations of cortical activity from the primary motor, premotor and posterior parietal cortices. This investigation will show how each model identifies which cortical regions are involved with the production of the hand trajectory. Secondly we will perform a sensitivity analysis on each model to rank the neurons for importance and determine if this importance ranking is influenced by the model selection.

II. METHODOLOGY

A. *Model Topologies*

Figs. 1 and 2 show the topology of WF and RMLP that are used in our studies. For the WF, the output is a weighted linear combination of neuronal inputs $\mathbf{x}$ (10 most recent values) given by (1) [2]. The optimal MSE solution is given by (2), where $\mathbf{d}$ is the hand trajectory[1].

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t) \tag{1}$$

$$\mathbf{W} = \mathbf{R}^{-1}\mathbf{P} = E(\mathbf{x}^T\mathbf{x})^{-1}E(\mathbf{x}^T\mathbf{d}) \tag{2}$$

---

[1] For ill-conditioned matrices, the inversion might assign large weights to neurons with few firing counts. To prevent this, we implement the ridge-regression regularization technique [1] to improve the condition number.
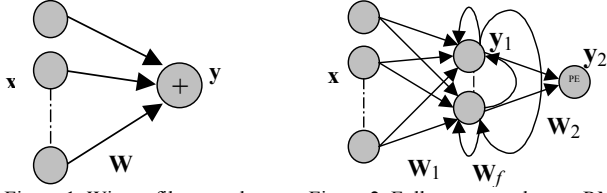
Figure 1. Wiener filter topology.
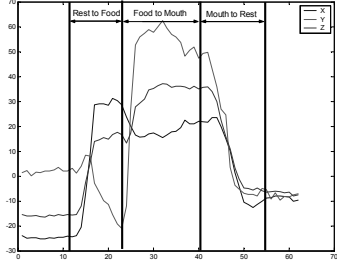

Figure 2. Fully connected, state RMLP.

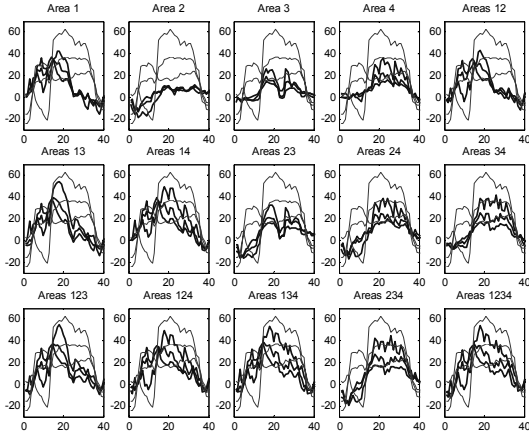Figure 3. One movement segmented into rest/food, food/mouth, and mouth/rest motions.

Figure 4. **Wiener filter** - Testing output X, Y, and Z trajectories (bold) for one desired movement (light) from fifteen Wiener filters trained with neuronal firing counts from all combinations of four cortical areas.
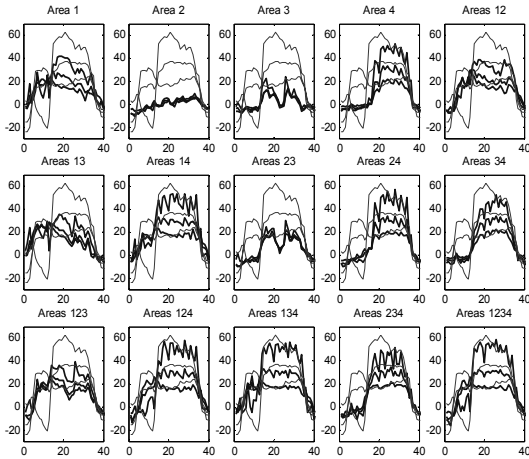

Figure 5. **RMLP** - Testing output X, Y, and Z trajectories (bold) for one desired movement (light) from fifteen RMLPs trained with neuronal firing counts from all combinations of four cortical areas

The RMLP uses only the instantaneous neural activity to compute each output. The hidden layer consists of 5 *tanh* processing elements (PEs) that are fully connected to each other with a feedback matrix. The state vector of the hidden layer in (3) is a nonlinear function of the linear combination of input and previous state. The feedback of the state creates memory and allows representations on multiple timescales. The output layer has 3 linear PEs (for X, Y, Z coordinates) and produces the output as in (4). Each hidden PE is a nonlinear adaptive basis for the output that projects the high dimensional neuronal data. These projections are then linearly combined to form the outputs (position predictions) of the RMLP. Optimal weights are determined by minimizing MSE using BPTT [14].

$$\mathbf{y}_1(t) = f(\mathbf{W}_1\mathbf{x}(t) + \mathbf{W}_f\mathbf{y}_1(t-1) + \mathbf{b}_1) \qquad (3)$$

$$\mathbf{y}_2(t) = \mathbf{W}_2\mathbf{y}_1(t) + \mathbf{b}_2 \qquad (4)$$

B. *Cortical Contributions*

We are interested in the cortical regions that contribute to the triad of movements defined in Fig. 3. Each reaching movement is segmented to three reaches: rest/food, food/mouth, and mouth/rest. By training both WF and RMLP using combinations of neurons from different cortical areas, and observing the network outputs, we can build a set of constructs to compare with established neurophysiologic principles. Both models are trained using multichannel neuronal firing times from up to 104 cells and hand trajectories that were collected synchronously from owl monkeys (Aotus trivirgatus). The firings of neurons are binned in 100ms non-overlapping windows. The neural recordings are collected by 64 electrodes implanted in four cortical regions (posterior parietal (PP) - Area 1, primary motor (M1) - Area 2, Area 4, and premotor dorsal (PMd) - Area 3) each receiving 16 electrodes [17].

Using approximately 33mins of neural activity from all combinations of the base set of cortical regions, we trained 15 WFs and RMLPs. After training, the weights of each topology were fixed and 5mins of novel data were presented to produce hand position predictions. For this short period of testing, there is no noticeable degradation in any of the models. This procedure was repeated on two separate datasets from the same primate that were collected on consecutive days, as well as a third dataset from a second primate. The results presented here are only for the first recording session and primate, however, the results are consistent among all trials.

The X, Y, Z network outputs (bold) and the actual hand coordinates for one sample movement are plotted in Figs. 4-5 for each network. Even though the RMLP outperforms the WF in capturing the trajectory peaks (subplots Area 1234), we observe that the trends of both topologies are consistent. For example in both topologies, Area 1 captured rest/food, but showed a poor fit in food/mouth. Area 2 does not display any correlation to this desired trajectory even though
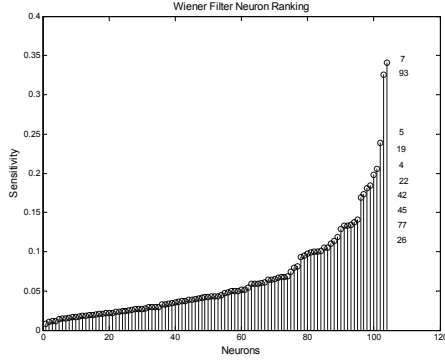
Figure 6. **Wiener filter** - Neuronal sensitivities sorted from minimum to maximum for a movement. The ten highest sensitivities are labeled with the corresponding neuron.
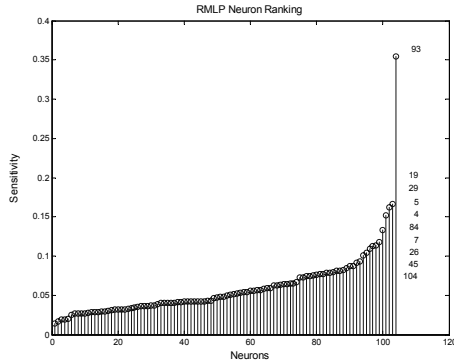

Figure 7. **RMLP** - Neuronal sensitivities sorted from minimum to maximum for a movement. The ten highest sensitivities are labeled with the corresponding neuron.

neuronal firing in this region is nonzero. Sharp changes in the model output appear in movement transitions for the network trained with Area 3. Area 4 accurately captures the food/mouth and mouth/rest regions, but misses the beginning of movement. Both the WF and RMLP display the following trends in the hand trajectory reconstruction:

- Area 1 is necessary to capture rest/food.
- Area 2 is not crucial in trajectory reconstruction.
- Area 3 relates to sharp transitions in trajectory.
- Area 4 is necessary to capture food/mouth.
- Combining multiple areas (e.g., 1, 3, 4) reduce magnitude of fluctuations in trajectory predictions.

C. *Sensitivity Analysis – Selecting Important Neurons*

With the trained weights of both WF and RMLP topologies, we have a tool to identify the neurons that affect the output most. A sensitivity analysis, using the Jacobian of the output vector with respect to the input vector, tells how each neuron's spike counts affect the output given the data of the training set. The procedure for deriving the sensitivity for a feedforward topology is an application of the chain rule [18]. For the case of the WF, differentiating the output with respect to the input yields $\partial\mathbf{y}(t)/\partial\mathbf{x}(t)=\mathbf{W}$. Hence, a neuron's importance is determined by its corresponding weight value. In this paper we consider the absolute values of the weights

averaged over the three output dimensions and the ten delays per neuron. With this procedure a single value indicating the importance of the neuron is obtained.

Since our RMLP model displays dependencies over time that result from feedback in the hidden layer, we must modify the procedure presented in [18]. Starting at each time t we compute the sensitivities to instantaneous inputs as well as previous inputs. In the RMLP we apply the chain rule as shown in (9)-(12), where $\mathbf{D}_t$ is the derivative of the hidden layer nonlinearity evaluated at the operating point shown in (10) using the input sample at time t. The dependencies to previous inputs introduced by the feedback are handled similar to the transition from (9) to (11) to obtain the general form in (12). Experimentally we determined that the effect of an input decays to zero over a window of 20 samples. At each time t the absolute-sensitivity of the output with respect to the inputs is represented as the averages of the absolute values of the sensitivities over the 20-sample window and the three output coordinates. Again, a single *importance* value for each neuron is obtained through the RMLP.

For both the WF and the RMLP, the sorted neuronal sensitivities for the ensemble of neurons plotted in Figs. 6 and 7 show a model-independent trend in the sensitivity profiles. An initially sharp decrease from maximum indicates that only a few neurons are required for outlining the movements in both topologies. Of the 104 neurons, 7 of the 10 highest-ranking neurons are common for the WF and RMLP. The most important neurons are primarily distributed over cortical areas 1 and 4. This finding that the output is less sensitive to the neurons in regions 2 and 3 is in agreement with the results in Section II.B.

$$\frac{\partial\mathbf{y}_2(t)}{\partial\mathbf{x}(t)} = \mathbf{W}_2^T\mathbf{D}_t\mathbf{W}_1^T \tag{9}$$

$$\mathbf{D} = diag\left(f'(z_1^1) \quad f'(z_1^2) \quad \cdots \quad f'(z_1^n)\right) \tag{10}$$

$$\frac{\partial\mathbf{y}_2(t)}{\partial\mathbf{x}(t-1)} = \mathbf{W}_2^T\mathbf{D}_t\mathbf{W}_f^T\mathbf{D}_{t-1}\mathbf{W}_1^T \tag{11}$$

$$\frac{\partial\mathbf{y}_2(t)}{\partial\mathbf{x}(t-\Delta)} = \mathbf{W}_2^T\mathbf{D}_t\left(\prod_{i=1}^{\Delta}\mathbf{W}_f^T\mathbf{D}_{t-i}\right)\mathbf{W}_1^T \tag{12}$$

V. Conclusion

We investigated the possibility of analyzing neural data from a BMI design perspective by considering the relative contributions of individual cortical regions and single neurons to the construction of hand trajectories through optimally trained models. It is encouraging that the qualitative interpretations of neural activity is independent of the model topology, even with two distinct topologies (linear-feedforward vs nonlinear-feedback). This builds confidence in our conclusions about neurophysiology drawn from signal processing techniques. The interpretations

obtained by the model analyses corroborate the view of broad tuning of the motor cortex, that is, the spatio-temporal encoding of the motor information is such that only a minute population of 100+ neurons is enough to enable a relatively precise mapping spike counts to hand movements. But the broad tuning seems to be limited to local organization. It seems that different cortical areas are required to track the different parts of the reaching task. The PP was controlling the reach to the food and the M1 was controlling the reach from food to mouth. To a lesser extent, the PMd was controlling the transitions in the movement. This indicates that the electrodes should be placed strategically throughout the motor cortex to capture vital information. If electrodes are not placed in a cortical region important for a part of the movement, the trajectory cannot be reconstructed well.

Moreover, only a handful of neurons in these respective cortical areas seem to be highly correlated to the hand trajectory. We have experimentally verified this observation by computing the sensitivities through both WF and RMLP topologies. Using these two methods of analysis, we found 7 of the 10 most important neurons for each model are common. Hence, one may be lead to think that BMIs can be constructed from just a few electrodes, but our data shows otherwise. Although, in 100+ neurons only 10 are indispensable to produce *good* mappings, we can only find these neurons after the analysis is done through a model. This means that finding good neurons for a large repertoire of tasks during surgery will be virtually impossible. So at this point the best strategy seems to sample the cortex with as many neurons as technically feasible.

This cortical analysis motivates appropriate questions about the roles of cortical regions in voluntary movements so that we can compare our observations with experimental neuroscience. Using well controlled experiments, the posterior parietal cortex has been associated with motor imagery [19], visual/tactile manipulation of objects [20], and spatial coordinate transformations [21]. In our alternative analysis approach, we repeatedly identified the PP as the active area during the rest/food reach, which is a task that may involve all of the mentioned PP associations. This could be an additional example confirming the role of PP.

Our last comment is on the scalability of this analysis approach. So far, we only utilized data from a reaching task. Therefore, it remains to be seen how the important neurons change from task to task. Due to the different importance of areas in the various segments of the reaching task, we anticipate that important neurons will vary for different movements. If this is the case, two problems arise. First, a given sampling of neurons may provide a limited repertoire of reconstructable movements. Second, a network trained for all possible movements may perform worse than one trained just for one movement. This is particularly the case for linear mappers, but it will affect to a certain extent the nonlinear models; how much remains to be seen.

REFERENCES

[1] S. Haykin, *Adaptive filter theory*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall International, 1996.

[2] J. Wessberg, C. R. Stambaugh, J. D. Kralik, P. D. Beck, M. Laubach, J. K. Chapin, J. Kim, S. J. Biggs, M. A. Srinivasan, and M. Nicolelis, "Real-time prediction of hand trajectory by ensembles of cortical neurons in primates," *Nature*, vol. 408, pp. 361-365, 2000.

[3] J. K. Chapin, K. A. Moxon, R. S. Markowitz, M. A. Nicolelis, "Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex," *Nature Neuroscience*, vol. 2, pp. 664-670, 1999.

[4] Y. Gao, M. J. Black, E. Bienenstock, W. Wu, and J. P. Donoghue, "A Quantitative Comparison of Linear and Non-linear Models of Motor Cortical Activity for the Encoding and Decoding of Arm Motions," IEEE EMBS CNE, Capri, Italy, 2003.

[5] A. P. Georgopoulos, A. B. Schwartz, and R. E. Kettner, "Neuronal Population Coding of Movement Direction," *Science*, vol. 233, pp. 1416-1419, 1986.

[6] A. B. Schwartz, D. M. Taylor, and S. I. H. Tillery, "Extraction algorithms for cortical control of arm prosthetics," *Current Opinion in Neurobiology*, vol. 11, pp. 701-708, 2001.

[7] M. D. Serruya, N. G. Hatsopoulos, L. Paninski, M. R. Fellows, and J. P. Donoghue, "Brain-machine interface: Instant neural control of a movement signal," *Nature*, vol. 416, pp. 141-142, 2002.

[8] D. M. Taylor, S. I. H. Tillery, and A. B. Schwartz, "Direct cortical control of 3D neuroprosthetic devices," *Science*, vol. 296, pp. 1829-1832, 2002.

[9] W. Wu, M. J. Black, Y. Gao, E. Bienenstock, M. Serruya, and J. P. Donoghue, "Inferring hand motion from multi-cell recordings in motor cortex using a Kalman filter," SAB, 66-73, Edinburgh, Scotland, 2002.

[10] J. C. Sanchez, D. Erdogmus, J. C. Principe, J. Wessberg, and M. Nicolelis, "A Comparison Between Nonlinear Mappings and Linear State Estimation to Model the Relation from Motor Cortical Neuronal Firing to Hand Movements," SAB, 59-65, Edinburgh, Scotland, 2002.

[11] J. C. Sanchez, S. P. Kim, D. Erdogmus, Y. N. Rao, J. C. Principe, J. Wessberg, and M. Nicolelis, "Input-Output Mapping Performance of Linear and Nonlinear Models for Estimating Hand Trajectories from Cortical Neuronal Firing Patterns," NNSP, pp. 139-148, Martigny, Switzerland, 2002.

[12] S. P. Kim, J. C. Sanchez, D. Erdogmus, Y. Rao, J. C. Principe, and M. Nicolelis, "Modeling the Relation from Motor Cortical Neuronal Firing to Hand Movements Using Competitive Linear Filters and a MLP," IJCNN, Portland, Oregon, 2003.

[13] S. P. Kim, J. C. Sanchez, D. Erdogmus, Y. N. Rao, J. C. Principe, M. Nicolelis, "Divide-and-Conquer Approach for Brain Machine Interfaces: Nonlinear Mixture of Competitive Linear Models," *Neural Networks*, 2003.

[14] S. Haykin, *Neural networks*, New York: Macmillan, 1994.

[15] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications*. Cambridge, MA: MIT Press, 1949.

[16] G. Orr and K.-R. Müller, *Neural networks: tricks of the trade*, vol. 1524. New York: Springer, 1998.

[17] M. A. Nicolelis, A. A. Ghazanfar, B. M. Faggin, S. Votaw, and L. M. Oliveira, "Reconstructing the engram: simultaneous, multisite, many single neuron recordings," *Neuron*, vol. 18, pp. 529-537, 1997.

[18] L. Fu, T. Chen, "Sensitivity Analysis for Input Vector in Multilayer Feedforward Neural Networks," ICNN, pp. 215-218, San Francisco, CA, 1993.

[19] D. J. Crammond, "Motor imagery: never in your wildest dream," *Trends in Neurosciences*, vol. 20, pp. 54-57, 1997.

[20] Kupfermann, "Localization of Higher Cognitive and Affective Functions: The Association Cortices," in *Principles of Neural Science*, E. R. Kandel, J. H. Schwartz, and J. T. M, Eds., 3 ed. Norwalk, Conn: Appleton & Lange, 1991, pp. 823-838.

[21] R. A. Andersen, L. H. Snyder, D. C. Bradley, J. Xing, "Multimodal Representation of Space in the Posterior Parietal Cortex and its Use in Planning Movements," *Annu. Rev. Neurosci*, vol. 20, pp. 303-330, 1997.