Towards Predicting Persistent Activity of Neurons by Statistical and Fractal Dimension-Based Features

Panagiotis C. Petrantonakis, Athanasia Papoutsi, and Panayiota Poirazi

Abstract—Persistent activity is the prolongation of neuronal firing that outlasts the presentation of a stimulus and has been recorded during the execution of working memory tasks in several cortical regions. The emergence of persistent activity is stimulus-specific: not all inputs lead to persistent firing, only ‘preferred’ ones. However, the features of a stimulus or the stimulus-induced response that determine whether it will ignite persistent activity remain unknown. In this paper, we propose various statistical and fractal dimension-based features derived from the activity of a detailed biophysical Prefrontal Cortex microcircuit model, for the efficient classification of the upcoming Persistent or Non-Persistent-activity state. Moreover, by introducing a novel majority voting classification framework we manage to achieve classification rates up to 92.5%, suggesting that selected features carry important predictive information that may be read out by the brain in order to identify ‘preferred’ vs. ‘no-preferred’ stimuli.

I. INTRODUCTION

Recent studies [1], [2] have revealed the existence of small groups or clusters of neurons in several cortical regions. Neurons within such clusters are strongly interconnected and, thus, co-activated during behavioral tasks. Such types of self-organized assemblies have also been detected in the Prefrontal Cortex (PFC), and it has been argued that they play a key role in crucial operations associated with complex behaviors, such as working memory and decision making [3].

Although little is known about the functional contribution of such clusters of neurons to high level cognitive tasks, it is evidenced that PFC neuronal assemblies support the spontaneous emergence of Up and Down states [4]. The Up- and Down- phenomenon is linked to the expression of persistent neuronal activity, which, in turn, stands for the cellular-level expression of a working memory process. Thus, the way persistent activity is driven by the properties of a PFC microcircuit and intra- or inter-connectivity of it, defines an intriguing research direction.

In [5], the properties of persistent activity expression are investigated in a detailed biophysical PFC microcircuit model. Here, various parameters of the microcircuit’s structure and activity, prior to the expression of persistent firing, are utilized in order to seek for features that incorporate predictive information of the upcoming state, namely the emergence of Persistent (P) or Non-Persistent activity (NP). Besides the feature extraction task based on statistical moments and Fractal Dimension (FD) estimation, a classification task is also developed. Support Vector Machines are used in order to facilitate the discrimination between the two different states. Moreover, in an attempt to deal with the large number of features compared to the number of observations provided for the two classes (states), a novel Majority Voting (MV) classification framework is introduced where various SVM classifiers are ranked according to a similarity score before the MV process. Thus, apart from the contribution of this work to the derivation of PFC microcircuits’ features that drive persistent activity, a new framework of MV classification is proposed.

The rest of this paper is structured as follows. In the next section, methodological aspects regarding dataset construction as well as feature extraction and classification framework formulation are presented. Section III presents the results and discusses several issues raised by the whole approach. Finally, section IV concludes the paper.

II. MATERIALS AND METHODS

A. Prefrontal Cortex Microcircuit

The PFC layer V microcircuit whose activity features are used in this work, was constructed in [5] and consists of 7 pyramidal neurons and 2 interneurons. The morphologically simplified models were developed using the NEURON simulation environment [6]. The pyramidal neuron model consists of a soma, a basal, a proximal and a distal dendritic compartment and includes modeling equations for 15 types of ionic mechanisms previously reported in [7], [8], known to be present in these neurons, as well as modeling equations for the regulation of intracellular calcium concentration. The interneuron model is a basket-type fast spiking (FS) neocortical interneuron, known to innervate the soma and the proximal dendritic compartments of pyramidal neurons in the PFC. The interneuron model used in the present study was obtained from ModelDB (accession number 82784).

The electrophysiological profile (active and passive properties), of both neuronal types was validated against experimental data. The neurons were fully connected through recurrent connections. The connectivity between the model neurons of the network (number of synapses/location/delays) was also based on experimental anatomical and electrophysiological data. Importantly, each pyramidal neuron was connected to another pyramidal neuron through 5 synaptic contacts [9] and also received 1 autapse [10]. Moreover, the kinetics of fast (iAMPA) and slow (iNMDA) synaptic excitation and fast (iGABA_A) and slow (iGABA_B)
synaptic inhibition were validated against electrophysiological data. This microcircuit contains several random processes, including poisson-like membrane noise, and experimentally observed fluctuations in synaptic delays. This ensemble of neurons was used to assess the induction of persistent activity, the cellular correlate of working memory. Persistent activity is the prolonged post stimulus activity in the absence of incoming stimuli. In the present study persistent trials were considered trials that spiking activity outlasted a 5-second period.

B. Dataset

The whole analysis implemented in this work is based on data produced after recurrent simulations of the PFC microcircuit. Finally, 90 cases were isolated corresponding to 45 expressions of persistent activity (P) and 45 of non-persistent (NP). Persistency or non-persistency was documented based on the activity of a single neuron in the microcircuit, as cluster dynamics are adequately represented by each individual neuron. From these 90 cases, one structural parameter of the microcircuit and four activity-based ones were taken under consideration. The structural parameter was the randomized connection of pyramidal to pyramidal neurons in terms of the distance of the connection from the soma, normalized in the range [0, 1]. The activity-based parameters are the estimated total AMPA, iGABA_A, iNMDA, and iGABA_B currents measured in a single neuron. These currents were measured over a 500ms-window before the expression of the state to be predicted (P or NP), sampled with a 0.1ms-sampling period resulting in 5000 samples for each current. Finally, it must be noted that not all receptors are activated from the beginning of the 500ms-duration and thus initial missing values were zero-padded to have 5000 sample values for all currents.

C. Feature Extraction

For the feature vector construction, statistical as well as fractal dimension (FD)-based approaches were used. For the data referring to the structural parameters of the PFC microcircuit, i.e., the distance from the soma of the pyramidal-to-pyramidal connection, two statistical moments were estimated. The mean, \( \mu \), and the skewness, \( \gamma \):

\[
\gamma = E \left[ \left( \frac{x - \mu}{\sigma} \right)^3 \right] \tag{1}
\]

were estimated for all pyramidal-to-pyramidal connection distances of the PFC microcircuit (in (1) \( E(\cdot) \) is the expectation operator whereas \( \mu \) and \( \sigma \) are the mean and standard deviation values, respectively). As evidenced in [7], expression of persistent activity depends on whether distal or proximal connections are used and, hence, mean and skewness will capture this dependence if existed. Fig. 1 (a) and (b) show the boxplots of the mean and skewness values, respectively, for the two distinct states (P and NP).

For the activity-based data, i.e., the currents corresponding to different receptors, AMPA, GABA_A, NMDA, and GABA_B, FD was estimated within a 1000-sample window. Taking into account the increased complexity of the neuronal dynamics while/during exhibiting persistent firing, a FD measure is expected to get higher values during the P versus the NP case. The FD measure was estimated using the Higuchi Algorithm (HA) [11]. To adjust the \( k_{max} \) parameter for the HA, the same procedure followed in [12] was adopted here, resulting in a value of \( k_{max} = 309 \) for a 1000-sample window. Thus, for each one of the four currents analyzed, five FD values were obtained, as each current signal consisted of five non-overlapping 1000-sample windows. In Fig. 2 the mean of the
FD values along with the standard deviation error is depicted for all four currents and for all five windows. As was expected, P state’s FD tend to be slightly higher than the NP state for any type of current, confirming the increased complexity in the neuronal response before entering the persisted state.

Finally a new signal, namely Balanced Excitatory Inhibition (BEI) [5] was constructed from the other four currents-signals according to the rule:

$$BEI = \frac{\text{AMP} + \text{NMDA}}{GABA_A + GABA_B}$$

(2)

and the FD measure was also estimated. As noted in II.B there were zero-padded values that resulted in non-definable BEI signal values. Thus, the FD value was estimated only from the second to fifth window of the BEI signal. The final feature vector was structured as:

$$FV = [\mu, \gamma, \text{AMP1}_{1-5}^{\text{FD}}, \text{GABA1}_{1-5}^{\text{FD}}, \text{NMDA1}_{1-5}^{\text{FD}}, \text{GABA1}_{1-5}^{\text{FD}}, \text{BEI}_{1-5}^{\text{FD}}]$$

(3)

were the subscription refers to the windows where the FD value was estimated. Hence the feature vector, $FV$, contained 26 features.

D. Classifier

A Support Vector Machine (SVM) classifier was used to discriminate between the Persistent and Non-Persistent observations. As a kernel function that projects the data to high dimensional feature space, the gaussian Radial Basis Function (RBF) is chosen:

$$K(x_i, x_j) = e^{-\frac{1}{2\sigma^2}\|x_i - x_j\|^2}$$

(4)

where $x_i$ and $x_j$ are two different feature vectors. For the best tuning of the SVM classifier, i.e., the adjustment of $\sigma$ in (4) and the penalty parameter for the optimization of the SVM margins, $c$, the procedure proposed in [13] was also adopted here. This procedure corresponds to a “grid search” for both parameters $\sigma$ and $c$ via a cross-validation framework (for details on cross-validation see paragraph II.G).

E. Majority Voting

As described in paragraph II.B, the dataset consisted of 45 observations for each class. Given that the feature vector $FV$ contains 26-features, the features-to-observation ratio is big. To deal with this issue we adopted a Majority Voting (MV) framework for the classification of the two different states. For each pair of features, that is $\binom{26}{2} = 325$ pairs, in order to decrease the dimensionality of the feature space, for each individual classifier, as much as possible, one SVM classifier was tuned and trained (see II.G). Subsequently, the final classification of a specified observation was obtained via a MV process among the 325 classifiers. It should be noted however, that this approach dramatically increases the number of classifiers to be tuned and trained leading to a severe computational burden despite the limited feature space for each classifier.

F. Similarity Measure Between two Classifiers

To deal with the computational issue of increased classifiers’ number, we introduce a new similarity measure between two classifiers. This similarity leads to the determination of a ranking that in turn provides a selection criterion for the SVM units to be taken into account during the MV process.

Suppose a small set of $n$ observations $X: x_i, i = 1, \ldots, n$ where the first $a$ observations, i.e., $x_i, i = 1, \ldots, a(< n)$,
correspond to class 0 and the rest \( b \) observations, i.e., \( x_i, i = a + 1, \ldots, a + b (= n) \) correspond to class 1. Then assume the binary vectors \( O_{S_1} \) and \( O_{S_2} \), where each element of these vectors corresponds to the outcome of the subscripted classifiers, \( S_1 \) and \( S_2 \), for the respective observation as in set \( X \). An ideal outcome vector (target outcome, \( O_{\text{target}} \)) would be the one with its \( a \) first elements assigned to 0 and the rest \( b \) elements assigned to 1.

Now we define a similarity measure between two classifiers, \( S_i \) and \( S_j \), as:

\[
S_{ij} = 2 - \left( \frac{\sum_{k=1}^{a} \left[ 2O_{S_i}(k) - O_{S_j}(k) = 0 \right]}{a} + \frac{\sum_{k=a+1}^{b} \left[ 2O_{S_i}(k) - O_{S_j}(k) - 1 = 0 \right]}{b} \right)
\]

where \([\cdot] \) is the Iverson bracket. The \( S_{ij} \) measure is a distance-like measure that performs a similarity comparison between the outcomes of two classifiers and between these outcomes and the target outcome \( O_{\text{target}} \) at the same time. It is obvious from (5) that the value range of \( S_{ij} \) is \([0, 2]\) with 0 corresponding to maximum similarity between two classifiers, \( S_i \) and \( S_j \).

G. Implementation Issues

For the implementation of the methodology aspects previously described, the dataset was separated in three distinct subsets. The first one is the parameter Adjustment (\( A \)) set that is reserved for the tuning of the SVM classifiers in a cross-validation manner. The second one is used for the extraction of the ranking (\( R \) set) of the classifiers according to the \( S_{ij} \) measure and the last one is reserved for the Evaluation (\( E \) set) stage. All three sets consisted of 15 observations per class.

Throughout this paper, whenever a cross-validation task is considered, the 70% of the corresponding dataset is used for training and the rest 30% for testing. The whole training-testing process is realized for 100 iterations and finally the mean classification rate, \( \bar{C} \), is estimated across all iterations.

To determine the ranking of classifiers according to the similarity measure introduced in II.F, we used a hierarchical clustering approach with the clustering criterion being the \( S_{ij} \) measure. In each iteration of the cross-validation process, a clustering procedure was performed and after 100 iterations, the classifiers were ranked according to the frequency they belonged to a low proximity cluster \( G^t \). This cluster is the one the incorporates pairs of classifiers with \( S_{ij} < t \). In this work we set \( t = 1 \).

III. RESULTS

Firstly we examine the case where all 26 features are used for the classification task. By using \( A \) set for parameter adjustment, the best values obtained for \( \sigma \) and \( c \) are 120 and 10000 respectively, providing with \( \bar{C} = 81.36\% \). For the \( E \) set, the mean classification rate with the adjusted SVM parameters was \( \bar{C} = 80.29\% \). Nevertheless, when using both \( R \) and \( E \) sets for evaluation, the mean classification rate is \( \bar{C} = 84.60\% \), slightly higher than the tuning phase, probably due to the augmentation of the training set with more observations.

On the other hand, the classification performance of each one of the 325 classifiers, developed for all possible pairs of 26 features, was also checked. In Fig. 3 the mean classification rate \( \bar{C} \) is presented for \( A, R, \) and \( E \) sets, Fig. 3 (a), (b), and (c), respectively. The serial number of each classifier in horizontal axis corresponds to the outcome of each pair of the 26 features in (3), that is, classifier #1 is developed with the features \([\mu, \gamma]\), classifier #2 with features
TABLE I

<table>
<thead>
<tr>
<th>C (%) VALUES FOR SETS A, R, AND E USING DIFFERENT NUMBER OF PCA FEATURES (# OF FEATURES / % OF VARIANCE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 / 76.43</td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>89.96</td>
</tr>
<tr>
<td>83.90</td>
</tr>
<tr>
<td>82.14</td>
</tr>
</tbody>
</table>

[μ, AMPA<sup>FD</sup>], etc. In all three subfigures in Fig. 3, the maximum classification rates along with the # of the classifier with which they were achieved are highlighted. It is notable that many pair-wise feature combinations exhibit better performance than the full feature set. This is probably due to the large features-to-observations ratio when using the whole dataset, as mentioned in the previous section. It should be stressed out that within the range of classifiers #26-49 a plateau in performance (C) is observed. This plateau corresponds to all combinations of feature y with all other features, denoting the crucial role of that feature in discriminating the two different neuronal states, P or NP.

Moreover, in Fig. 4, the results from the MV realization are depicted. Fig. 4(a) shows the results of MV applied to E set, when ranking the classifiers according to the performance of the pair-wise classification in R set (see Fig. 3(b)). In Fig. 4(b) however, classifiers are ranked according to the procedure described in section II.G in R set, and the C values correspond to the MV outcome on the E set. Furthermore, Fig. 4(c) shows the mean and standard deviation of C, estimated from 50 random permutations (rankings) of the classifiers revealing the outlying behavior of the proposed similarity measure-based ranking, especially for the case where few classifiers are used. The serial number (#) of the classifiers do not represent the combinatorial outcome as in Fig. 3 but, instead, represent the #1-first classifiers, according to the specified ranking, that was used to extract the MV classification rate C. As evident in Fig. 4, the C rate when using the MV approach is significantly increased compared to the case where the whole feature set is used (C = 80.29%). Moreover when the similarity-based ranking is used, instead of the benchmark, performance-based ranking, not only the C rate is increased (92.5% instead of 90.9%) but also, the number (#) of classifiers used to accomplish maximum performance with MV is significantly decreased (3 instead of 181), a goal for which the s<sub>fj</sub> measure was initially developed. It should be noted that all three classifiers that accomplished maximum performance via MV in Fig. 4 (b) include feature y. Particularly, the features used for each classifier were [y, μ], [y, AMPA<sup>FD</sup>] and [y, AMPA<sup>FD</sup>]. Thus, finally, four features and three classifiers are exploited for the best classification performance to be achieved.

Despite the fact that, the whole similarity measure-based MV approach ends up like a dimensionality reduction methodology for the feature space, the initial intention was just to rearrange the rank of the classifiers, according to a similarity measure, in a way that the MV phase exhibit high performance with as few classifiers as possible; this is not equivalent with the feature-space dimensionality reduction. For instance, it is possible to use the whole feature vector, i.e., 26 features, by using only 13 classifiers for the MV process, in which case the similarity measure-based ranking would had also succeeded (from 181 classifiers in Fig. 4(a) to 13 classifiers in the hypothesized case). Nevertheless, in order to define a benchmark case for the dimensionality reduction task we implemented a Principal Component Analysis (PCA)-based feature selection task. In Table I, the results for the A, R, and E sets, for different feature sizes, corresponding to different percentage of variance of the initial feature set, are presented. For the evaluation set E, the best performance, 85.21%, is achieved using only three features after the PCA analysis, which represent 82.93% of the initial variance. On the other hand the proposed approach exhibits C rate of 92.5% using four features. The performance of the proposed approach is even more prevalent when four features are also selected after the PCA analysis.

IV. CONCLUSION

In this work we presented a feature set and a classification framework in order to identify PFC microcircuit properties that lead to the expression of persistent activity. Our results provide the first steps toward the challenging task of inferring the characteristics of network activity and/or dynamics that determine when a given input signal will induce persistent firing in neuronal microcircuits. Despite the promising discriminating capabilities of both statistical and FD-based features, extended evaluation of the results is needed in order to establish a concrete connection between those features and the firing mode of a neuron.

REFERENCES

