35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Feature Extraction and Unsupervised Classification of Neural Population Reward Signals for Reinforcement Based BMI

Noeline W. Prins *Student Member, IEEE*, Shijia Geng, *Student Member, IEEE,* Eric A. Pohlmeyer, Babak Mahmoudi and Justin C. Sanchez, *Senior Member, IEEE*

***Abstract*—New reinforcement based paradigms for building adaptive decoders for Brain-Machine Interfaces involve using feedback directly from the brain. In this work, we investigated neuromodulation in the Nucleus Accumbens (reward center) during a multi-target reaching task and investigated how to extract a reinforcing or non-reinforcing signal that could be used to adapt a BMI decoder. One of the challenges in brain-driven adaptation is how to translate biological neuromodulation into a single binary signal from the distributed representation of the neural population, which may encode many aspects of reward.**

To extract these signals, feature analysis and clustering were used to identify timing and coding properties of a user's neuromodulation related to reward perception. First, Principal Component Analysis (PCA) of reward related neural signals was used to extract variance in the firing and the optimum time correlation between the neural signal and the reward phase of the task. Next, k-means clustering was used to separate data into two classes.

I. Introduction

Brain-Machine Interfaces (BMIs) have shown great potential to restore movement function for amputees and for people living with paralysis through the control of external devices or through functional electrode stimulation (FES) [1-5]. The design of neural decoders to translate brain activity into behavior is typically trained in a supervised manner with either real or inferred kinematic signals. In cases of severe paralysis or amputation, it may not be possible to collect these signals as a desired response. Therefore, there is a need to develop other methods of acquiring training signals and using them to adapt neural decoders.

As an alternative to supervised learning which is being tested in subjects living with paralysis [6, 7], Reinforcement Learning (RL) provides a method of biological and computational learning that does not depend on specific known outcomes but rather performance outcomes. [8, 9]. Using this approach, we have developed a new method of decoding that is based on actor-critic RL [10-12]. In this approach, the actor is driven by motor neural inputs and translates them into behavioral actions. The role of the critic is to adapt the actor based on experience. The only feedback the critic should provide is the appropriateness or the value of the chosen action; in this case if the action selected was correct or incorrect. This feedback signal can be obtained by the external environment or from the brain itself.

* This work was supported by DARPA REPAIR project N66001-10-C-2008.

N. W. Prins, S. Geng, E. A. Pohlmeyer, B. Mahmoudi, and J. C. Sanchez are with the Department of Biomedical Engineering, University of Miami, Coral Gables, FL 33146 USA (e-mail: jcsanchez@miami.edu).

Obtaining reward information from the brain has a variety of challenges associated with it. Much research has gone into identifying reward centers in the brain [13-15]. Of these centers, the Nucleus Accumbens (NAcc) is a main component in the ventral striatum and plays a key role in the linking of reward to motor behavior [16]. If signals from this structure are to be used to adapt BMI decoders, a first step is to determine how to preprocess and extract reward signals from it.

The nature of neural representation, especially reward activation, is complex. The timing, type, magnitude, and expectation of reward can also affect the related neuromodulaton [14]. For RLBMIs, three main aspects of reward are important: differentiation between rewarding and non-rewarding targets, the timing of modulation related to these conditions, and how to extract features in the neuronal firing that signal these conditions [15, 17, 18]. In this work, we seek to investigate preprocessing methodologies for extracting reward signals from NAcc for BMIs. The approach is to identify major modes of variance through Principal Component Analysis (PCA), reduce dimensionality, and extract relevant features related to reward. Once the modes are identified an unsupervised method is applied to identify rewarding and non-rewarding neural activation.

II. Methodology

A. Neural Recordings

Neural data was acquired while a marmoset monkey (*Callithrix jacchus*) was interacting with a robot in a two-choice decision task. To access deep brain reward signals, a 16-channel tungsten microelectrode array, (Tucker Davis Technologies, FL) was surgically implanted in ventral striatum targeting the NAcc under isoflurane anesthesia and sterile conditions. All surgical and animal care procedures were consistent with the National Research Council Guide for the Care and Use of Laboratory Animals and were approved by the University of Miami Institutional Animal Care and Use Committee.

Neural recordings were sampled at 24,414Hz using Tucker Davis Technologies RZ2 system. Spike sorting of neuronal signals was performed in real-time based on the shape and amplitude of action potential waveforms and using manually set threshold levels. Both multiunit as well as single unit neurons were recorded and used equivalently in all applications. Multiunit signals and single unit signals collectively are referred to here as neuronal signals. During the real-time experiment, 29 neuronal signals from NAcc were isolated and recorded.

978-1-4577-0216-7/13/$26.00 ©2013 IEEE

B. Experimental task

The task studied here was a two-choice decision making task. The monkey was trained to move a robot arm to one of two targets to receive a food reward (Figure 1).

The monkey initiated trials by placing its hand on a touchpad for a random (700-1200msec) hold period. At the onset of the trial, an audio go signal was provided that corresponded to a robot arm moving upwards, out from behind an opaque shield, and presenting its gripper. The gripper held either a desirable (waxworm or marshmallow, 'A' trials) or undesirable (wooden bead, 'B' trials) object. Simultaneously, the A (red) or B (green) spatial target LED corresponding to the type of object in the gripper was illuminated.

Each type of trial required a different action; for A trials, the monkey had to reach a second sensor within 2 second reach time limit and the robot would move to A target; for B trials, it was required to keep its hand motionless on the touchpad for 2.5 seconds and the robot would move to B target. For both A and B trials, if the robot moved to the target indicated by the LED, the monkey was given a food reward. Trials where the animal either did the wrong action or was not interacting with the task were removed from the analysis.

To create robot perturbations that contrast with reward trials, the robot was occasionally overridden and moved in the direction opposite to that of the action commanded by the monkey. These trials where the monkey sees an undesirable action in the environment (evoking negative response in the brain) were considered 'catch' trials. There were 24% and 35% catch trials for A and B trials respectively. The trials where the robot moved to the intended target and the animal received a food rewards were called 'standard' trials.

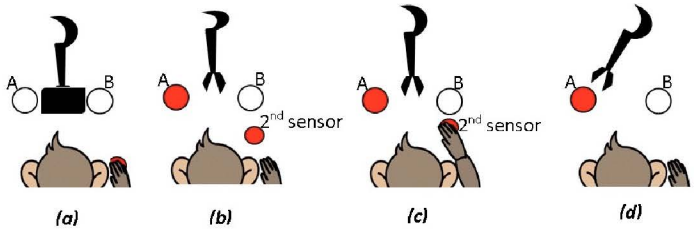

Figure 1 'A' trials: (a) Animal initiates trial (b) Robot comes out from opaque screen and reveals gripper, target LED lights come on, second sensor light comes on (c) Animal makes arm movement and triggers second sensor (d) Robot moves to target 'A'.

C. Feature Analysis

A and B trials were considered separately for this analysis. The purpose was to separate the standard trials from catch trials. Figure 2 shows a timeline of a trial. All analysis was performed relative to the beginning of the robot movement (RM) time (which began when the second sensor was triggered in A trials or at the end of the hold period in B trials).

The analysis done was using a 0.5 second sliding window (0.1 second overlap) with the sum of firing rate within the given window of each of the 29 neuronal signals as the feature space. This goal was to find the optimal window that correlated with the robot moving to or away from the desired target.

Next, PCA was used for feature analysis. PCA is a widely used technique in neuroscience because it can exploit the high variance of neural data [19-21]. PCA also gives the direction of maximal variance, which helps in extracting relevant features and in dimensionality reduction, which is helpful in BMI applications.

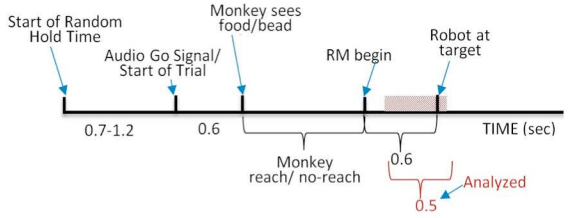

Figure 2 Timeline for the trial (in black). Red shows the 0.5sec window that was focused on for classifying standard and catch trials from the NAcc (0.2-0.7sec after RM)

D. Clustering and Labeling of PCA Data

After features were extracted using PCA, the signals then needed to be classified into reinforcing and non-reinforcing classes. k-means clustering, an unsupervised method, was used for clustering the data. The only prior knowledge required was the number of clusters. In this application the number of clusters is already known to be two (reinforcing or non-reinforcing).

k-means is used to classify n objects of input space $I\{i_1 i_2 \dots i_n\}$, each having measurements on p variables $i_j\{x_{j1}x_{j2} \dots x_{jp}\}$, into k clusters with cluster centroid $C\ (c_1 c_2 \dots c_k)$. In this case, n = number of trials, p = number of principal components used and k=2. The algorithm was set to start by setting C to an initial value (randomly picked from I). The centroid value for cluster c_k is given by:

$$c_k = \frac{1}{n_k}\sum_{j=1}^{n_k} i_j;\ \forall\ i_j\{x_{j1}x_{j2} \dots x_{jp}\}\epsilon c_k$$

where n_k is the number of objects in k.

Next, clustering is done based on minimizing the cost function which is a measure of the distance between each data point and the centroid. Three different cost functions were used: squared Euclidean distance, sum of absolute differences and one minus the cosine of the included angle between points (treated as vectors). The results are of squared Euclidean distance are presented as the clusters aligned better with this criterion.

For each $i_j \epsilon I$, the squared Euclidean distance (d) between i_j and its centroid, c_k was calculated.

$$d(i_j, c_k) = (i_j - c_k)^2;\ \forall\ i_j\{x_{j1}x_{j2} \dots x_{jp}\}\epsilon c_k, j = 1,2 \dots n$$

The objects of I were moved to the cluster whose centroid was closest, until d was minimum [22].

$$\underset{C}{\operatorname{argmin}}\left(\sum_{j=1}^{k} d(i_j, c_k)\right)$$

The two clusters obtained from k-means clustering were assigned labels (standard and catch) manually and compared against the class labels standard ('+') and catch ('o') categories in the experiment. The classification accuracy was the number of trials correctly classified (True Positive + True Negative) out of the total number of trials.

III. Results

Recordings of 3 consecutive sessions were analyzed individually (S1,S2,S3). We also aggregated the sessions together (S1+S2+S3) to see if there was consistency among the sessions and also to have a higher number of trials.

A. PCA Analysis of Variance

For all the sessions analyzed, the first 9 and 15 principal components accounted for at least 80% and 90% of the variance respectively (Figure 3). After the data was converted using PCA, all combinations of the first 7 principal components were plotted and inspected. The first two principal components contained 48% of the variance and showed best separability. Hence, the first 2 principal components were selected as the features for analysis.

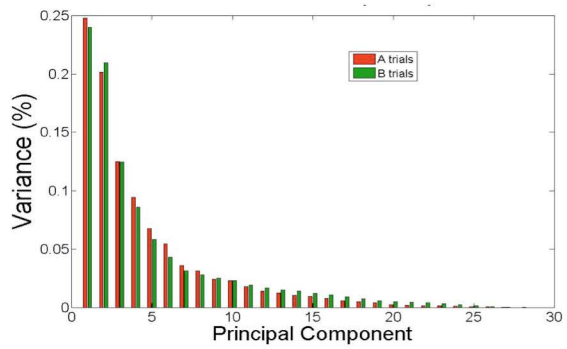

Figure 3 Variance of data relative to RM. Red: 'A' trials . Green: 'B' trials. Window 0.2-0.7sec (S1+S2+S3)

B. Unsupervised Clustering

The data projections for the first and second principal components of the two different trial types (A and B) were used for clustering. k-means was used to partition the PC space into two clusters as seen Figure 4 and Figure 5 (blue and yellow Voronoi diagrams). Next we labeled the trials,'+' for standard and 'o' for catch, and compared the k-means classes against the labels and calculated the resulting classification accuracy.

PCA and k-means analysis of the NAcc firing revealed a difference in the separability between standard and catch trails for the A and B trials. More overlap in the neural representation and clustering was observed for A trials.

Figure 4 and Figure 5 show the clustering for a time widow of 0.2-0.7 sec time following RM, which gave a classification accuracy of 64.1% and 87.5% for A and B trials, respectively (S1+S2+S3). Other time windows (0-0.5, 0.1-0.6, 0.3-0.8, 0.4-0.9 and 0.5-1 sec) were tested as well, with 0.1-0.6 sec and 0.2-0.7 sec showing the greatest separability. Table-I gives the differences in clustering performance, PCA representation and timing. The highest classification for the 0.1-0.6 sec window was 81% and 90.5% for A (S2) and B (S3) trials respectively. For the 0.2-0.7 window it was 90.5% for A trials (S2) and 85.7% for B trials (S3). This difference could be due to variance of the neuronal signals from each session. The baseline for the trials was the window 0.5-0 sec prior to the go signal where the animal was initiating the trial. The PCA showed no pattern between the two categories and performance of k-means clustering was at chance (55.5% for both A and B trials).

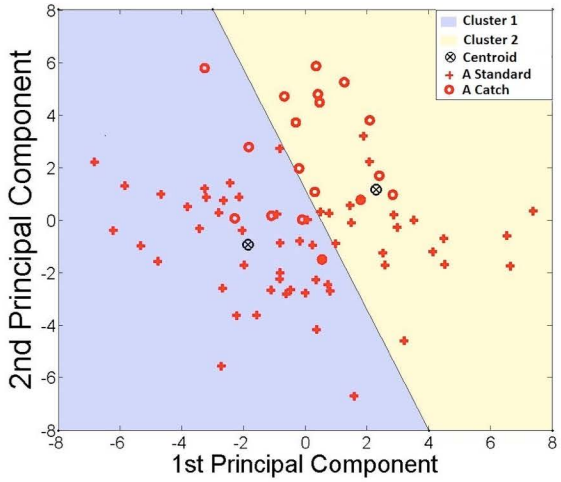

Figure 4 Data clustered in PC space using k-means for 'A' trials. Blue: Cluster 1. Yellow: Cluster 2. '+': standard. 'o': catch and ⊗: cluster centers. Window 0.2-0.7sec

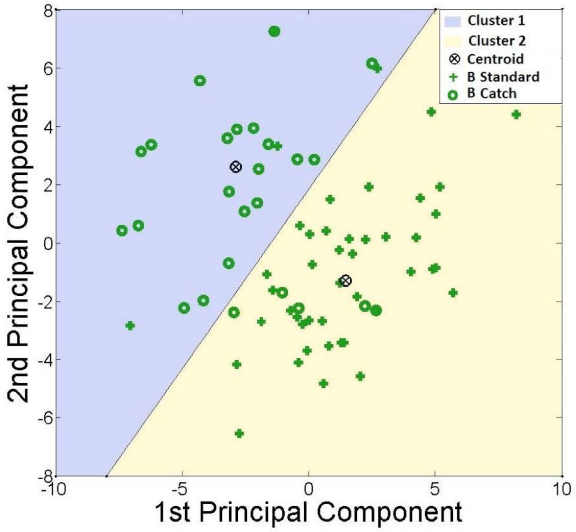

Figure 5 Data clustered in PC space using k-means for 'B' trials. Blue: Cluster 1. Yellow: Cluster 2. '+': standard. 'o': catch and ⊗: cluster centers. Window 0.2-0.7sec

TABLE I. Overall Accuracy of Clustering using k-means[a] for different window sizes relative to the RM

	0.1-0.6sec window		*0.2-0.7sec window*	
	'A' trials	*'B' trials*	*'A' trials*	*'B' trials*
S1	59.3%	66.7%	63.0%	57.1%
S2	81.0%	73.3%	90.5%	83.3%
S3	53.3%	90.5%	80.0%	85.7%
S1+S2+S3	66.7%	58.3%	64.1%	87.5%

a. Criterion: minimize squared Euclidean distance.

IV. Discussion

The purpose of this paper was to find reward representation in the NAcc for the application in reinforcement based decoders in BMI. At present we have developed and used an actor critic decoding paradigm that uses an ideal feedback or feedback from the environment to control a robot arm for a two-choice task [10-12]. The next step is to incorporate the methods in this paper to give a processed biological signal as the feedback.

As a first step to process this biological signal, we tested the separability of NAcc data by projecting to the first two principal components. We also tested the separability by adding the third principal component. The performance did not significantly increase and in some cases, it was reduced beyond that of two principal components. We concluded that two principal components were sufficient for this basic task.

The next step to process this biological signal was to separate the data into two clusters. The k-means algorithm was used for this purpose. It is a basic unsupervised clustering method which required only a few iterations (<10). Another advantage of the approach is that it is fully unsupervised and can be applied to natural environments where no a priori knowledge is known about the targets. However, it has several disadvantages: it cannot handle outliers or deal with overlapping clusters, the clustering is locally minimum and increasing the number of clusters will reduce the training error within a cluster [22].

The cost function used in k-means will affect the performance of the clustering. Even though qualitatively, we observed separation in the standard ('+') and catch ('o') categories which indicated there was a difference in the neuromodulation for the two conditions. But since k-means is a clustering algorithm and not a classifier, it is only interested in optimizing with respect to the distance, not inaccuracies in labeling. When we compared the clusters given by k-means to the standard ('+') and catch ('o') categories, we saw that the k-means did not represent the standard and catch categories as accurately for A trials compared to B trials as there was more overlap in A trials. This suggests that the criterion used (squared Euclidean distance) may not be the best for the information required. The separation of classification could be improved with advanced supervised techniques and we may be able to get higher accuracies however the tradeoff is the need for supervision.

V. References

[1] M. Velliste, S. Perel, M. C. Spalding, A. S. Whitford, and A. B. Schwartz, "Cortical control of a prosthetic arm for self-feeding," *Nature,* vol. 453, pp. 1098-1101, 2008.

[2] C. T. Moritz, S. I. Perlmutter, and E. E. Fetz, "Direct control of paralysed muscles by cortical neurons," *Nature,* vol. 456, pp. 639-642, 2008.

[3] J. M. Carmena, M. A. Lebedev, R. E. Crist, J. E. O'Doherty, D. M. Santucci, D. F. Dimitrov*, et al.*, "Learning to control a brain–machine interface for reaching and grasping by primates," *PLoS biology,* vol. 1, p. e42, 2003.

[4] 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.

[5] J. R. Wolpaw and D. J. McFarland, "Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 101, pp. 17849-17854, 2004.

[6] J. L. Collinger, B. Wodlinger, J. E. Downey, W. Wang, E. C. Tyler-Kabara, D. J. Weber*, et al.*, "High-performance neuroprosthetic control by an individual with tetraplegia," *The Lancet,* 2012.

[7] L. R. Hochberg, D. Bacher, B. Jarosiewicz, N. Y. Masse, J. D. Simeral, J. Vogel*, et al.*, "Reach and grasp by people with tetraplegia using a neurally controlled robotic arm," *Nature,* vol. 485, pp. 372-375, 2012.

[8] B. W. Balleine and A. Dickinson, "Goal-directed instrumental action: contingency and incentive learning and their cortical substrates," *Neuropharmacology,* vol. 37, pp. 407-419, 1998.

[9] Y. Niv, "Reinforcement learning in the brain," *Journal of Mathematical Psychology,* vol. 53, pp. 139-154, 2009.

[10] B. Mahmoudi and J. C. Sanchez, "A symbiotic brain-machine interface through value-based decision making," *PloS one,* vol. 6, p. e14760, 2011.

[11] J. DiGiovanna, B. Mahmoudi, J. Fortes, J. C. Principe, and J. C. Sanchez, "Coadaptive brain–machine interface via reinforcement learning," *Biomedical Engineering, IEEE Transactions on,* vol. 56, pp. 54-64, 2009.

[12] E. A. Pohlmeyer, B. Mahmoudi, G. Shijia, N. Prins, and J. C. Sanchez, "Brain-machine interface control of a robot arm using actor-critic rainforcement learning," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, 2012, pp. 4108-4111.

[13] T. E. Schlaepfer, M. X. Cohen, C. Frick, M. Kosel, D. Brodesser, N. Axmacher*, et al.*, "Deep brain stimulation to reward circuitry alleviates anhedonia in refractory major depression," *Neuropsychopharmacology,* vol. 33, pp. 368-377, 2007.

[14] R. Kawagoe, Y. Takikawa, and O. Hikosaka, "Expectation of reward modulates cognitive signals in the basal ganglia," *Nat Neurosci,* vol. 1, pp. 411-416, 1998.

[15] W. Schultz, P. Dayan, and P. R. Montague, "A neural substrate of prediction and reward," *Science,* vol. 275, pp. 1593-1599, 1997.

[16] C. Pennartz, H. J. Groenewegen, and F. Da Silva, "The nucleus accumbens as a complex of functionally distinct neuronal ensembles: an integration of behavioural, electrophysiological and anatomical data," *Progress in neurobiology,* vol. 42, pp. 719-761, 1994.

[17] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," *Annual review of neuroscience,* vol. 24, pp. 1193-1216, 2001.

[18] P. N. Tobler, C. D. Fiorillo, and W. Schultz, "Adaptive coding of reward value by dopamine neurons," *Science,* vol. 307, pp. 1642-1645, 2005.

[19] D. A. Adamos, E. K. Kosmidis, and G. Theophilidis, "Performance evaluation of PCA-based spike sorting algorithms," *Computer methods and programs in biomedicine,* vol. 91, pp. 232-244, 2008.

[20] P. Jahankhani, V. Kodogiannis, and K. Revett, "EEG signal classification using wavelet feature extraction and neural networks," in *Modern Computing, 2006. JVA'06. IEEE John Vincent Atanasoff 2006 International Symposium on*, 2006, pp. 120-124.

[21] J. K. Chapin and M. Nicolelis, "Principal component analysis of neuronal ensemble activity reveals multidimensional somatosensory representations," *Journal of neuroscience methods,* vol. 94, p. 121, 1999.

[22] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm," *Applied statistics,* pp. 100-108, 1979.