 DOI:10.3217/978-3-85125-260-6-115

Using Multiple Reward Related Signals in the Adaptation of Neuroprosthetic Decoders

S. A. Roset[1], N. W. Prins[1], S. Geng[1], H. F. Gonzalez[1], B. Mahmoudi[1], E. A. Pohlmeyer[1], J. C. Sanchez[1]

[1]*University of Miami, Coral Gables, FL, USA*

Correspondence: J. C. Sanchez, University of Miami, Coral Gables, FL, USA. E-mail: jcsanchez@miarni.edu

Abstract. Using neuroprosthetics during daily living presents new challenges that affect design choices in neural decoders. New classes of architectures for neural decoding that are based on reinforcement learning (RL) are being investigated. RL decoders use experience to help shape and adapt the decoder such that the benefits of tasks are maximized for the user. Since RL is based on reward feedback, reward signals are a critical part of the functionality. In this work, we investigate and compare 4 sources of reward related signals that can provide feedback for RL based decoders: the external environment, error-related potentials (ErrP) in EEG and LFPs, and single neuron activity in the Nucleus Accumbens (NAcc).

Keywords: Reward, Adaptive, Decoder, Reinforcement, Learning

1. Introduction

The use of neuroprosthetics in activities of daily living present new challenges that affect the design choices used in neural decoding. These include the dynamics of working in multiple environments, effects of rehabilitation on neuroplasticity, effects of user learning on performance, dependence on a caregiver to help calibrate the system, and the overall stability of neural signals acquired from the neural interface itself. Our laboratory has been investigating a new class of architectures for neural decoding that are based on reinforcement learning (RL) [Mahmoudi and Sanchez, 2011]. RL decoders are inspired by physiological, computational, and behavioral principles involved in the process of using experience to help shape and adapt performance such that the benefits of tasks are maximized for the user. In simple terms, RL decoders work in the following way. When users generate neural activity that leads to successful use of the neuroprosthetic, the functionality of the decoder is reinforced. Conversely, when users generate neural activity and the neuroprosthetic is unsuccessful, the interface is adapted. This framework leads to continuous interaction between user and neuroprosthetic based on success or failure.

Since RL decoding is based on reward feedback, acquisition and access to these signals is a critical part of the functionality. Typically in he development of living organisms, the signals used to reinforce behaviors during learning come from a wide variety of sources. These include the external environment, other individuals, and the organism's own brain [Holroyd and Coles, 2002]. All could be good candidates for adapting a BCI decoder.

In this work, we investigate 4 sources of reward related signals that can be used to provide feedback for RL based decoders.

2. Material and Methods

Our adaptive neural interface uses actor-critic based RL decoders. As shown in Fig. 1, the actor maps motor related brain activity to intended actions of a prosthetic or assistive device. The actor's weights are initialized randomly and then adjusted after each trial based on feedback from the critic. The critic provides feedback by decoding the user's brain activity or environmental cues to determine if the decoded behavior should be reinforced.

Multiple sources of reward related feedback can be extracted from the environment and the brain. The differences among them are related to: frequency of expression, robustness of their representation, and modality of acquisition. Our lab has been evaluating several common feedback sources, including environmental cues and several types of neural recordings.

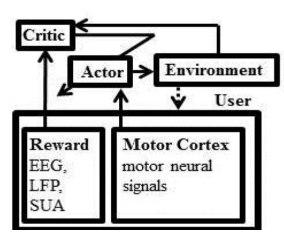

Figure 1. *Reinforcement Learning BCI.*

When reward is from a source other than neural signals from the user's brain, the source is considered part of the external environment. The external environment can include a system with prior knowledge of available actions, another person providing feedback, or the user themselves through feedback such as a button press or eye blink. However, such signals are not always available in the severely paralyzed.

Reward can also come from the electroencephalogram (EEG) in the form of error-related potentials (ErrP). This is typically detected in the 5-10 Hz band from electrode Cz when the user thought an error was committed [Ferrez and Millan, 2008]. By targeting reward centers in the brain such as the Nucleus Accumbens (NAcc), single unit activity (SUA) can also be used to extract reward information. However, a significant challenge is to construct a single reward signal from the distributed representation of the neural population, which may encode many aspects of reward as it is linked to behavior. Local field potentials (LFP) from the NAcc can also produce event related potentials associated with reward.

Our lab is investigating all four signal types as possible sources of reward feedback. To illustrate the use of this class of signals, a support vector machine (SVM) [Muller, Smola et al., 1999] was used to classify human EEG, and LFP and SUA signals from nonhuman primate NAcc. All tested behaviors were a two choice target selection task (33 trials). To preprocess features for classification, the power of the 5-10 Hz band, 1 s window, from Cz EEG was used. Likewise, the power of broadband (1-500 Hz) LFP was used from four electrodes. The vector firing rate of 22 neurons (from 16 electrodes) was used as a feature for the SVM in the case of SUA.

3. Results

Table 1 shows average classification results across two subjects. The classification accuracy of the different sources of reward is comparable: SUA (82%), LFP (75%), and EEG (69%).

***Table 1**. Classification Accuracy of Reward Sources.*

	Environment	EEG-ERN	LFP	SUA
Reward	100%	69%	75%	82%
No-Reward	100%	63%	71%	69%

4. Discussion

Overall, the three neural signals provided similar average classification rates. SUA can provide higher spatial resolution, but determining how they relate to global processing such as reward or no-reward can be a challenge. Since LFP is believed to detect the synaptic activity of many neurons and since reward processing is an integrative procedure less preprocessing of LFP was required to extract reward for similar performance. EEG’s major advantage is non-invasiveness. However, EEG has a lower signal to noise than other sources and ErrPs might not be detectable during rapidly paced tasks. LFP appear to have a similar limitation, while SUA may not. The external environment can give feedback that is completely accurate. However, incorporating it into a system may limit its usability in daily life, since additional inputs from the user, such as muscle movements, would be required.

Acknowledgements

This work was supported by DARPA REPAIR project N66001-1O-C-2008 and DARPA RNR subcontract for project W31P4Q-12-C-0200.

References

Ferrez PW, Millan JdR. Error-Related EEG Potentials Generated During Simulated Brain-Computer Interaction. *IEEE Trans Biomed Eng*, 55(3):923-929, 2008.

Holroyd CB, Coles MGH. The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychol Rev*, 109(4):679-709, 2002.

Mahmoudi B, Sanchez JC. A Symbiotic Brain-Machine Interface through Value-Based Decision Making.*PLoS ONE*, 6(3):e14760, 2011.

Müller KR, Smola A, et al. Using support vector machines for time series prediction. *Advances in kernel methods: support vector learning*. 253, 1999.