33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Optimizing Microstimulation using a Reinforcement Learning Framework

Austin J. Brockmeier, *Student Member, IEEE*, John S. Choi, Marcello M. DiStasio, Joseph T. Francis, and José C. Príncipe, *Fellow, IEEE*

***Abstract*— The ability to provide sensory feedback is desired to enhance the functionality of neuroprosthetics. Somatosensory feedback provides closed-loop control to the motor system, which is lacking in feedforward neuroprosthetics. In the case of existing somatosensory function, a template of the natural response can be used as a template of desired response elicited by electrical microstimulation. In the case of no initial training data, microstimulation parameters that produce responses close to the template must be selected in an online manner. We propose using reinforcement learning as a framework to balance the exploration of the parameter space and the continued selection of promising parameters for further stimulation. This approach avoids an explicit model of the neural response from stimulation. We explore a preliminary architecture—treating the task as a k-armed bandit—using offline data recorded for natural touch and thalamic microstimulation, and we examine the methods efficiency in exploring the parameter space while concentrating on promising parameter forms. The best matching stimulation parameters, from $k = 68$ different forms, are selected by the reinforcement learning algorithm consistently after 334 realizations.**

I. Introduction

Neuroprosthetics are envisioned to both translate neural signals to prosthetic operation and to deliver information back to the nervous system. Electrical microstimulation has been the primary vehicle for the latter. Microstimulation in somatosensory cortical regions has been shown to elicit responses that can be discriminated [1], [2]. In brain-machine interfaces for motor restoration, somatosensory feedback (both tactile and proprioceptive) may be useful to enhance a users performance, where information from sensors would be delivered to the central nervous system via spatio-temporal microstimulation. In general, closed-loop feedback provides context to the user, which may lead to more efficient co-adaptation. However, if the microstimulation feedback causes unnatural neural responses, then it is unclear whether this can convey the proper context. Consequently, our goal is to elicit a near-natural response for tactile or proprioceptive sensation.

In order to provide realistic feedback, the neural response elicited by microstimulation should match the response to a natural stimulus (e.g. cutaneous touch) as closely as possible. However, due to the underlying dynamics the neural response may vary between realizations; thus, multiple realizations may be needed to fairly evaluate the similarity of the response to template of the natural response. Explicitly modeling and inverting a model of the neural response to microstimulation of varying parameters is alluring if the model can be adapted online [3]; however, a general model that encompasses all the parameters of spatio-temporal stimulation may be ill-posed without sufficient training data. If the possible stimulation space has many dimensions it is inefficient to naively sample the space to generate training data, and without sufficient data the model would fail to generalize the complex relationships between neural response and stimulation parameters [4]: making a model-free approach appealing [5].

A human expert is able to efficiently probe both location of stimulation and any stimulation parameters such as the minimum amplitude needed to elicit a neural response. Recent work in stimulation optimization points to the promise of having automated systems to fine tune these parameters in a closed-loop. In [6] the authors use nonlinear regression to optimize the stimulation amplitude at a predetermined electrode position for a given target potential in closed-loop operation on multiple animals. The amplitude is selected given the ongoing local potential at another predetermined electrode position to increase the reliable control of the evoked potentials. The authors in [5] use genetic algorithms to optimize the temporal waveform for deep-brain stimulation on a neural simulator. With microelectrode arrays, there is the possibility to modify both the spatial and temporal parameters of the stimulation. An action-selection approach with binary rewards was used to select spatio-temporal stimulation patterns in [7], but in their *in vitro* study the goal of the stimulation was to adjust the underlying network connectivity through repeated stimulation. A study [8] using simulated neural circuits showed the ability to control firing rates treating the stimulation as MIMO control problem. Recent work described in [3] uses an inverse control architecture to precisely control spiking on simulated neural circuits.

A. Reinforcement Learning

Along similar lines as the genetic algorithms, a reinforcement learning framework may offer advantages for online closed-loop operation without explicit models. Reinforcement learning is both intuitive and mathematically sound. Ultimately, the system must track non-stationarity inherent in neural environments, explore numerous spatio-temporal

This work was supported in part by the University of Florida Graduate School Fellowship and DARPA Contract N66001-10-C-2008.

A. J. Brockmeier and J. C. Príncipe are with the Department of Electrical and Computer Engineering, University of Florida, P.O. Box 116130 NEB 486, Bldg #33, University of Florida, Gainesville, FL 32611 USA. {ajbrockmeier, principe}@cnel.ufl.edu

J. S. Choi, M. M. DiStasio, and J. T. Francis are with the Department of Physiology and Pharmacology, State University of New York Downstate Medical School, Brooklyn, NY 11203 USA. {john.choi, marcello.distasio, joe.francis} @downstate.edu

stimulation parameters, modify the stimulation with regard to current neural activity, and perform well using only stochastic evaluations of a similarity measure.

In this work, we explore a preliminary step toward this goal: naive action selection method for choosing both the electrode(s) (monopole or dipole) from a multi-electrode array and the current amplitude for a single biphasic symmetric rectangular pulse. The objective is to maximize the average cross-correlation between the stimulation evoked potentials and the natural touch template recorded as local field potentials in the somatosensory cortex. As the action selection is agnostic to the ongoing neural activity, this is a k-armed bandit problem in a possibly non-stationary environment (k is the combinations of electrode configurations and amplitudes). We test the performance of reinforcement comparison [9] on an awake resting rat dataset with $k = 68$ total stimulation parameter forms, 4 amplitudes and 17 electrode configures. The method is able to find stimulation parameters that elicit neural responses that near the natural template that was recorded during mechanical thwacking of a forepaw digit.

II. Method

The microstimulation parameter selection algorithm utilizes an action selection function estimated by reinforcement comparison. The underlying reward in the reinforcement learning paradigm is based on the similarity between the natural template and stimulation-elicited neural response.

We test the method on data collected in the somatosensory cortex during microstimulation in the somatosensory region of the thalamus. Stimulation earlier in the somatosensory pathway may be an efficient mechanism in eliciting cortical responses that match a natural template since intracortical stimulation artifacts are avoided.

A. Similarity measure

We evaluate the similarity of single realization of the microstimulation neural response to a template formed from multiple realizations of a natural stimuli. Let the mean natural neural response be denoted $\bar{X} \in \mathcal{X}$ and $Y_{i,a} \in \mathcal{X}$ be the neural response from ith response to microstimulation with parameter form $a \in A$. The set of stimulation parameter forms A is discrete with cardinality denoted $|A|$. The objective is to choose a subset of stimulation forms $A^\star$ such that the neural responses $Y_{1,a}, Y_{2,a}, \ldots$ of form $a \in A^\star$ are closer to the natural neural response than the remaining stimulation forms $A \setminus A^\star$. Closeness is judged in terms of a dissimilarity measure $d(X, Y)$ that operates in $\mathcal{X}$.

For local field potential data we let $\mathcal{X} = \mathbb{R}^{M \times T}$ be M channels of local field potential values for T consecutive samples surrounding the stimulation or natural touch. We form the dissimilarity measure from the channel-average cross-correlation,

$$d(X,Y) = 1 - \max_{\tau_1,\tau_2} \frac{R_{X,Y}(\tau_1,\tau_2)}{\sqrt{R_{X,X}(\tau_1,\tau_1)R_{Y,Y}(\tau_2,\tau_2)}} \quad (1)$$

$$R_{X,Y}(t,s) = \sum_{i=1}^{M}\sum_{j=1}^{T} X(i,t+j)\cdot Y(i,s+j) \quad (2)$$

Either τ_1 or τ_2 is allowed to be nonzero to account for jitter in the stimulation timing. In addition, samples outside the window size are assumed to be zero.

B. Reinforcement comparison

In general, a state-action policy defines the probability of choosing each stimulation form, at iteration n, given the ongoing neural activity S, i.e. $\Pr\{a_n{=}a|S{=}s\}$. In our current treatment, we do not estimate a state variable and denote this probability $\pi_n(a) = \Pr\{a_n{=}a\}$. Here an 'action' determines the parameters of the microstimulation event of a given iteration, but does not control the relative timing of the microstimulation.

For reinforcement learning we consider the estimated reward to be the negative of average dissimilarity measure. To avoid bias problems we choose each action once to initialize the estimated reward and action values

$$\rho_0(a) = r_0(a) = -d(\bar{X}, Y_{1,a}) \quad \forall a \in A. \quad (3)$$

The probability vector at any iteration n is the normalized action values

$$\pi_n(a) = \frac{\exp \rho_n(a)}{\sum_{i\in A} \exp \rho_n(i)} \quad \forall a \in A. \quad (4)$$

After initialization, at each iteration $n = 1, \ldots$, we draw an action $a_n \in A$ with probability $\pi_{n-1}(a_n)$. The reward for a_n is updated by averaging all of its realizations $Y_{1,a_n}, \ldots, Y_{m+1,a_n}$ where m is the number of iterations a_n was selected after initialization; specifically, m is the cardinality of the set $\big\{j \in \{1,\ldots,n\} : a_j = a_n\big\}$.

$$r_n(a_n) = \frac{-\sum_{i=1}^{m+1} d(\bar{X}, Y_{i,a_n})}{m+1} \quad (5)$$

The equation (6) for updating the action value $\rho(a_n)$ is called reinforcement comparison [9] since it compares each reward to the average reward $\bar{r}$.

$$\rho_n(a_n) = \rho_{n-1}(a_n) + \beta\big(r_n(a_n) - \bar{r}_{n-1}\big) \quad (6)$$
$$\bar{r}_n = \frac{1}{|A|}\sum_{i\in A} r_n(i)$$

C. Action set adjustment

In this reinforcement learning framework, it is simple to adjust A at any iteration: poor performing actions can be removed and new actions can be added to approach a global optimal. New actions are assigned an action value equal to the current maximum. An explicit model trained concurrently may also be able to extrapolate or interpolate new stimulation forms not in the original set.

D. Data collection

Animal procedures were approved by SUNY Downstate Medical Center IACUC and conformed to National Institutes of Health guidelines. A single female Long-Evans rat was implanted with two microarrays (16 contacts each in a 2×8 grid Neuronexus). The electrodes covered somatosensory areas of the cortex, S1, and the VPL nucleus of the thalamus [10]. The neural activity on all channels was recorded using a multi-acquisition processing system (Plexon, Inc.). Multiple channels of both arrays had multiunit activity responsive to cutaneous touch of forepaw digit.

Microstimulation was administered on single electrodes (monopolar) or adjacent pairs (dipoles) of the thalamic array. The stimulation waveforms were *single* symmetric biphasic rectangular current pulses; for monopolar configurations the first phase was negative. Each rectangular pulse was 200μs long and had an amplitude from the set $\{10\mu\text{A}, 20\mu\text{A}, 30\mu\text{A}, 40\mu\text{A}\}$. Inter-stimulus intervals were exponentially distributed with mean interval of 500ms. Stimulus isolation used a custom built switching headstage.

Field potential data from each of the 16 cortical channels was amplified and filtered with an active bandpass filter with cutoffs at 0.7Hz and 8.8kHz and sampled at 20kHz (National Instruments PCI-6701E). Offline stimulation artifact cancellation was performed using recursive least squares to predict the potential using a causal finite impulse response filter on the stimulation pulses; the residual approximates the artifact-less signal. After cancellation, the signal was digitally filtered with a Butterworth band-pass filter with cutoffs at (5Hz, 200Hz) and resampled at 0.8kHz.

During physiological recording, the rat was awake resting in a box with a suspended mesh floor. Manual stimulation was performed on the forepaw digit with a tactor when the digit was not obscured by the mesh.

E. Test set

Each local field potential sample consists of 720 samples (300ms) on the 16 cortical channels (see Fig. 1). Of these samples 72 (30ms) precede the event timestamp (stimulation or thwack onset), and the offsets τ_1, τ_2 from (1) are used to best align the stimulation response to the natural touch template. The results presented are from a singe recording session. The natural template, $\bar{X}$ in (5), was formed from over 100 responses to mechanical *thwacking*, short light touch with tactor, of a forepaw digit while the animal was awake.

The microstimulation set, A, consisted of single biphasic pulse waveforms with 68 parameter forms, 4 amplitudes $\{10\mu\text{A}, 20\mu\text{A}, 30\mu\text{A}, 40\mu\text{A}\}$ and 17 electrode configuration (a combination of single electrodes and adjacent pairs). Each form had 75 realizations in the full dataset. All 75×4 realizations for a given electrode configuration were done consecutively, but the amplitudes were pseudo-randomly arranged.

The quantitative results used offline analysis with 8 Monte Carlo runs through the dataset. In each run the order of realizations for each stimulation was permuted, and the first realization from each of the 68 parameter forms was used to initialize the reward estimates (3). An additional 500 realizations were sampled based on the update equations (4), (5), and (6). The comparison gain in (6) was limited to $\beta = 0.1$ in this analysis. This gain is linked to the cardinality of the parameter set, A; for higher values of β the probability vector will converge quickly, and for low values of β, actions are explored more often before convergence.

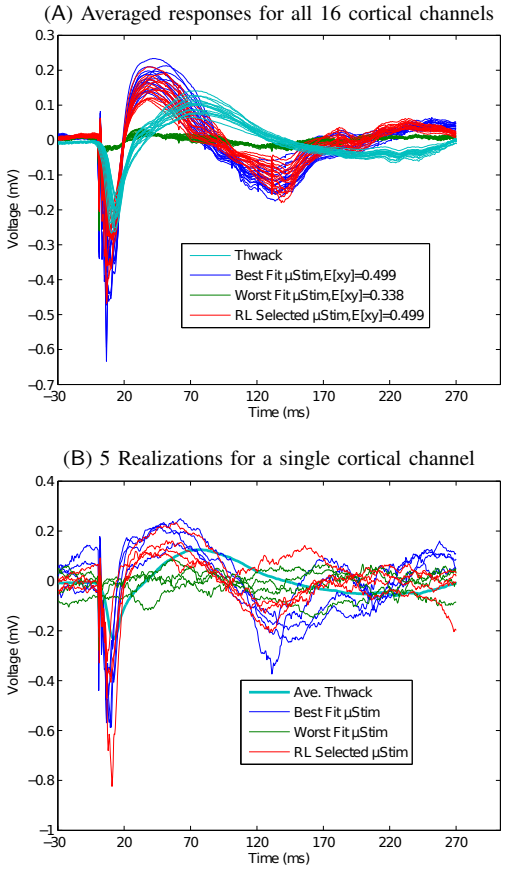

Fig. 1. Local field potential means (A) and realizations (B) for thwacking (mean only) and a subset of microstimulation (after artifact cancellation). For microstimulation the responses are shown for parameters with the best fit response, the worst fit, and the reinforcement learning selected response. Even the best fitting microstimulation evoked potential differs in shape from the natural response; obtaining a better fit requires moving beyond single pulse waveforms.

III. Offline Analysis

To assess the performance of the paradigm the 8 Monte Carlo runs using the action selection algorithm were compared to results obtained using the full dataset. Optimality is assessed using the cross-correlation, $1 - d(X, Y)$ from (1), of all 75 samples from each parameter form. Cross-correlation across the full range of parameter forms is shown in Fig. 2. The relatively low results for cross-correlation show the evoked potentials differ from naturally evoked potentials. This is not surprising since this test set is limited to single pulse waveforms on single or adjacent electrodes; more natural responses may be possible through spatio-temporal patterns.

The selection rates for the top 64 parameter forms are shown in Fig. 3, where color indicates the ranking based on the *full* data. It is clear that after 250 iterations the stimulation forms with the best matching responses are consistently selected with increasing probability. This results shows the promise of the method to explore the space while concentrating on near optimal solutions. The average reward for each stimulation form at the end of the 500 iterations is comparable to the reward after sampling from the full dataset in Fig. 4.

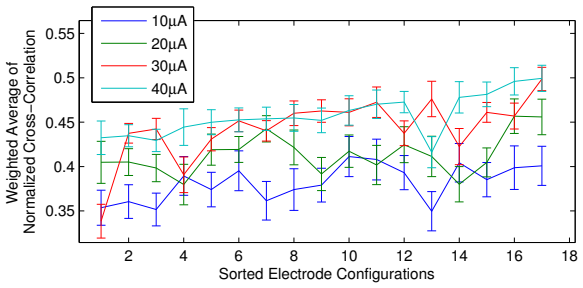

Fig. 2. Normalized cross-correlation across stimulation configurations (different monopoles and dipole pairs of thalamic channels) and stimulation amplitudes. Error bars show standard deviation over 500 bootstrap resampling.

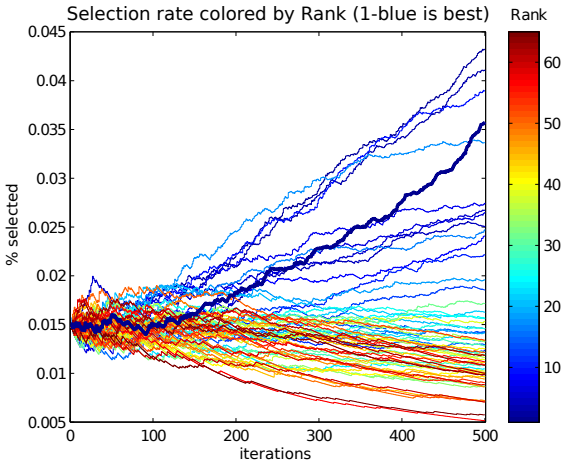

Fig. 3. Selection traces for the top 64 parameter forms. Color indicates rank where smaller is indicated by blue and is better, the best-matching form is bolded.

IV. Conclusion

In this work we propose using reinforcement learning as a framework for online selection of microstimulation parameters to elicit an evoked response close to a natural template. This is a primarily step toward optimization of feedback to the brain for somatosensory neuroprosthetics, which can augment such brain-machine interfaces. Ideally, microstimulation that elicits a near-natural neural response may be well suited to provide contextual feedback in a neuroprosthesis, but for brain-machine interfaces the results must be gauged by behavior experiments or improvements in task performance.

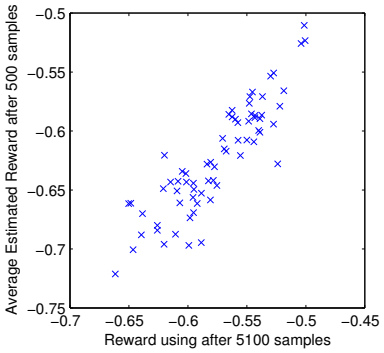

Fig. 4. Reward for each parameter form using all ($5100 = 68 \times 74$) realizations versus average reward after 500 iterations (568 realizations).

We treated stimulation parameter selection as a k-armed bandit problem, whereas future work can couple state estimation into the action selection such as [6] and also spatio-temporal waveform selection. The results from offline analysis of neural data show reinforcement learning algorithms can efficiently sample from promising microstimulation parameter forms, while still exploring the parameter space. The next step is testing this paradigm online with more complex spatio-temporal patterns that may be able to elicit more naturalistic responses.

References

[1] R. Romo, A. Hernández, A. Zainos, and E. Salinas, "Somatosensory discrimination based on cortical microstimulation," *Nature*, vol. 392, no. 6674, pp. 387–390, 1998.

[2] R. Romo, A. Hernndez, A. Zainos, C. D. Brody, and L. Lemus, "Sensing without touching: Psychophysical performance based on cortical microstimulation," *Neuron*, vol. 26, no. 1, pp. 273 – 278, 2000.

[3] L. Li, A. Brockmeier, J. T. Francis, J. C. Sanchez, and J. Principe, "An adaptive inverse controller for somatosensory microstimulation," in *Workshop on Neural Engineering (NER), 2011 5th International IEEE*, Apr. 2011.

[4] S. Butovas and C. Schwarz, "Spatiotemporal effects of microstimulation in rat neocortex: a parametric study using multielectrode recordings," *Journal of neurophysiology*, vol. 90, no. 5, p. 3024, 2003.

[5] X. Feng, B. Greenwald, H. Rabitz, E. Shea-Brown, and R. Kosut, "Toward closed-loop optimization of deep brain stimulation for Parkinson's disease: concepts and lessons from a computational model," *Journal of Neural Engineering*, vol. 4, p. L14, 2007.

[6] D. Brugger, S. Butovas, M. Bogdan, and C. Schwarz, "Real-time adaptive microstimulation increases reliability of electrically evoked cortical potentials," *Biomedical Engineering, IEEE Transactions on*, 2011.

[7] D. Bakkum, Z. Chao, and S. Potter, "Spatio-temporal electrical stimuli shape behavior of an embodied cortical network in a goal-directed learning task," *Journal of neural engineering*, vol. 5, p. 310, 2008.

[8] J. Liu, K. Oweiss, and H. Khalil, "Feedback control of the spatiotemporal firing patterns of neural microcircuits," in *Decision and Control (CDC), 2010 49th IEEE Conference on*, Dec. 2010, pp. 4679 –4684.

[9] R. Sutton and A. Barto, *Reinforcement learning: An introduction*. The MIT press, 1998.

[10] J. Francis, S. Xu, and J. Chapin, "Proprioceptive and cutaneous representations in the rat ventral posterolateral thalamus," *Journal of neurophysiology*, vol. 99, no. 5, p. 2291, 2008.