Information theoretic bounds on the effectiveness of neural prosthetics

Ilan N. Goodman and Don H. Johnson Electrical & Computer Engineering Department, MS380 Rice University Houston, Texas 77005–1892 {igoodman,dhj}@rice.edu

1 Introduction

Recent successes in the study of brain-machine interfaces (Lebedev and Nicolelis, 2006; Hochberg et al., 2006) raise an obvious question: What are the theoretical barriers to the successful use of neural prosthetics in general, and how well can such devices be expected to perform? Neural prosthetics fall into two classes (Nicolelis, 2001). Neural stimulation devices such as cochlear implants and deep-brain stimulators for Parkinson's patients consist of an electrode (or multiple electrodes) that electrically stimulates a population of neurons. On the other hand, neural control devices (e.g., prosthetic limbs and EEG-computer interfaces) rely on summation potentials recorded from multiple neurons to control a mechanical device. How well these prostheses can perform depends on, in the first case, how effectively information from the stimulation device can be conveyed through the population and in the second, how accurately the interface can extract information from a population. We investigate the accuracy of neural prosthetics using an information theoretic approach.

In this framework, neurons (or populations of neurons) are viewed as a communication channel. A stimulus signal S is encoded into a signal X, which then serves as the input to a system for further processing. This system could be a single neuron, a population of neurons, or several layers of neural systems. Regardless, we implicitly assume that this system introduces noise, making it difficult to decode the stimulus. The system produces a spike train output, denoted by Y. Although a stimulus estimate may not explicitly be produced in an actual neural system, the ability of the system to transmit information about the stimulus is characterized by how well the stimulus *could* be estimated from Y. Therefore, we conceptually place an optimal decoder immediately after the channel that produces a stimulus estimate \hat{S} to the best of its ability. The quantity that determines this estimate's quality is the *capacity* of the channel, defined as the maximum mutual information between X and Y over all inputs (defined by their probability distribution) that belong to some constraint class \mathscr{C} :

$$C = \lim_{T \to \infty} \max_{p(X): p(X) \in \mathscr{C}} \frac{1}{T} I(X;Y) \,.$$

Here, T represents the time over which the stimulus is sent and decoded.

Spike trains are modeled as point processes (Johnson, 1996; Snyder, 1975). Mathematically, the point process channel produces a sequence of events that encodes an input signal X_t according to an intensity $\mu(t; \mathcal{H}_t)$, which represents how the instantaneous event rate depends on the input X_t and on the process's history \mathcal{H}_t . Kabanov (1978) derived the capacity of the single point process channel when the minimal and maximal instantaneous rates are constrained according to $\lambda_{\min} \leq \mu(t; \mathcal{H}_t) \leq \lambda_{\max}$. In most cases of interest in neuroscience, the minimum-rate constraint is zero and the resulting capacity formula is

$$C^{(1)} = \frac{\lambda_{\max}}{e \ln 2}$$
 bits/s,

where $C^{(1)}$ denotes the single-neuron capacity. Kabanov's derivation showed that the capacity of *any* point process satisfying the intensity constraint cannot exceed the Poisson process's capacity.

The significance of capacity for stimulus estimation becomes evident when viewed in the context of the *rate-distortion function*. This quantity measures how accurately a source must be encoded to achieve any specified degree of error (Berger, 1971). The quality of \hat{S} is characterized by the distortion (error) $d(S, \hat{S})$ between the signal and its estimate. The average distortion D equals the expected value of the distortion with respect to the joint distribution of the signal and its estimate: $D = \mathsf{E}\left[d(S, \hat{S})\right]$. The rate-distortion function $\mathscr{R}(D)$ is defined to be the minimal mutual information between the signal and its estimate when the average distortion is no greater than D. To characterize how well a specific source S (as described by its probability distribution) can be encoded, the minimization is calculated with respect to all conditional distributions $p(\hat{S}|S)$, regardless of how the signal is encoded, corrupted, and processed.

$$\mathscr{R}(D) = \lim_{T \to \infty} \min_{p(\widehat{S}|S) : \mathsf{E}[d(S,\widehat{S})] \le D} \frac{1}{T} I(S;\widehat{S})$$

The rate-distortion function for any error measure is strictly convex and equals zero at some maximum distortion D_{max} (Berger, 1971). Zero rate means nothing about the signal is encoded and the estimation system simply makes an intelligent guess.

Although rate has units of bits/s, rate-distortion functions can be meaningfully calculated for all cases, even those where no physical bits play a role. Shannon's rate-distortion theory states that if a communication channel or processing system having capacity *C* intervenes between an encoder and decoder, the point at which capacity equals the rate-distortion function defines the minimum achievable distortion D_{\min} : $C = \mathscr{R}(D_{\min})$. Since this result holds for *any* source and *any* distortion function, capacity dictates the ultimate limit to which a communication system can convey information about any source signal regardless of how error is quantified. Because all rate-distortion curves are strictly decreasing and convex, increasing capacity *always* means reducing the smallest achievable average distortion. For example, when the stimulus is a bandlimited Gaussian random process having bandwidth *W*, we find that the minimal mean-squared-error distortion for a single neuron decreases exponentially in its maximal spike rate.

$$D_{\min} = D_{\max} \exp\left\{-\frac{\lambda_{\max}}{e \cdot W}\right\}$$

Consequently, computing the capacity of different neural population models provides an objective comparison of how effectively different structures can convey information. In work under review (Johnson and Goodman, 2007), we derived the capacity for several population structures wherein each neuron's output is modeled as a Poisson process. We extend and interpret that work here to characterize how well neural prostheses can perform.

2 Results

As a baseline, consider the population structure shown in Figure 1a. There, each neuron in a population of size M is independently innervated. Simple manipulations using the properties of mutual information show that the total capacity of this population equals the sum of the individual capacities: $C^{(M)} = \sum_{m} C_m^{(1)}$.

2.1 The Neural Stimulation Channel

Our population model for an electrical stimulation prosthetic is shown in Figure 1b. Here, a single input drives M individual neurons, modeling how a single electrical stimulation affects several neighboring neurons. Somewhat surprisingly, the capacity of this channel is *identical* to the independent input case. Independent Poisson processes driven by a common input act like a single





Poisson process having a rate equal to the sum of the individual rates. Since a Poisson process's capacity is proportional to the maximal rate, the common-input capacity is proportional to population size, just as in the separate-input case.

2.2 The Neural Control Channel

Most neurally controlled prosthetics use aggregated recordings of simultaneous activity of many neurons. To model this situation we sum the outputs of the baseline population to produce a single output, $Y = \sum_m Y_m$, as shown in Figure 1c. Summing the outputs models unsorted extracellular recordings made with a single-electrode and EEG recordings that represent the superposition of many identical neural signals. According to the Data Processing Inequality (Cover and Thomas, 2006), aggregating the outputs can only decrease the mutual information because the identity of which neuron produced a given output spike is lost. However, this result does not provide any detail on how much capacity is reduced.

As Figure 2 shows, the resulting capacity, expressed as a multiple of the single-channel capacity, never exceeds two. Consequently, not separating an aggregate recording into its constituents greatly reduces the information that can be gleaned, with the aggregate information flow being less than what two single-neuron channels can sustain. Since many recording protocols use more than one electrode, we also studied the extent to which using two electrodes could enhance capacity. Here, each electrode sums the activities from a subpopulation that overlaps the other, mimicking the typical situation with a gross electrode recording. We found that while aggregating from two subpopulations substantially increases capacity from the single electrode case, it is still far less than the unaggregated, single-input capacity.

3 Conclusions

The two populations we analyzed—a neural population having a single, common input and an aggregated recording from any population—model two common neural prosthetic situations. The first describes electrical stimulation situations, such as cochlear implants, in which each electrode excites multiple neurons. The second models the use of gross recordings, such as EEGs and unsorted extracellular recordings, to control devices. Our capacity results suggest that electrical stimulation faces no fundamental barriers to creating viable replacement sensory or motor inputs. The only question that remains is finding the most effective stimulation signal. On the other hand, using gross recordings for controlling motion must somehow surmount the limitations that our capacity results indicate must be present. The sharp reduction of channel capacity means that the best-case distortion *must* increase, exponentially in the Gaussian case described previously, and cannot be better than if two neurons' signals alone were used to drive the prosthesis. This capacity barrier can be lifted in three ways. First, using more electrodes does help, but these must be processed properly to extract the available information. Secondly, if the original constituents could be com-



Figure 2: The capacity of the aggregated channel first decreases, then grows slowly with the population size M. The line shows capacity in the case when the entire population is summed together. The shaded area shows the range of capacities that using two summations afforded. The variation depends on the degree of overlap between the equal-sized subpopulations that contributed to the summations. In the common-input scenario that models the electrical stimulation prosthesis, capacity is proportional to M.

pletely untangled from the aggregate using spike sorting techniques, no capacity decrease would occur. Finally, recent work has shown that feedback can increase capacity, though the results do not indicate by how much (Venkataramanan and Pradhan, 2005). Thus, visual feedback may play a crucial role in enabling artificial limb motion guided by neural control.

References

Berger, T. (1971). Rate Distortion Theory. Prentice-Hall, Englewood Cliffs, NJ.

Cover, T. and Thomas, J. (2006). *Elements of Information Theory*. Wiley, New York, second edition.

Hochberg, L. et al. (2006). Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, 442:164–171.

Johnson, D. (1996). Point process models of single-neuron discharges. J. Comp. Neuroscience, 3:275–299.

Johnson, D. and Goodman, I. (2007). Inferring the capacity of the vector poisson channel with a bernoulli model. *Network: Computation in Neural Systems*. Submitted.

Kabanov, Y. (1978). The capacity of a channel of the Poisson type. *Theory Prob. and Applications*, 23:143–147.

Lebedev, M. and Nicolelis, M. (2006). Brain-machine interfaces: past, present and future. *Trends in Neurosciences*, 29:536–546.

Nicolelis, M. (2001). Actions from thoughts. Nature, 409:403-407.

Snyder, D. (1975). Random Point Processes. Wiley, New York.

Venkataramanan, R. and Pradhan, S. (2005). Directed information for communication problems with common side information and delayed feedback/feedforward. *Proc.* 43rd Allerton Conf..