

Neural Prosthetic Systems: Current Problems and Future Directions

Cindy A. Chestek*, John P. Cunningham*, Vikash Gilja, Paul Nuyujukian, Stephen I. Ryu, Krishna V. Shenoy

Abstract—By decoding neural activity into useful behavioral commands, neural prosthetic systems seek to improve the lives of severely disabled human patients. Motor decoding algorithms, which map neural spiking data to control parameters of a device such as a prosthetic arm, have received particular attention in the literature. Here, we highlight several outstanding problems that exist in most current approaches to decode algorithm design. These include two problems that we argue will unlikely result in further dramatic increases in performance, specifically spike sorting and spiking models. We also discuss three issues that have been less examined in the literature, and we argue that addressing these issues may result in dramatic future increases in performance. These include: non-stationarity of recorded waveforms, limitations of a linear mappings between neural activity and movement kinematics, and the low signal to noise ratio of the neural data. We demonstrate these problems with data from 39 experimental sessions with a non-human primate performing reaches and with recent literature. In all, this study suggests that research in cortically-controlled prosthetic systems may require reprioritization to achieve performance that is acceptable for a clinically viable human system.

I. INTRODUCTION

In recent years, advances in neural technologies have enabled the creation of neural prosthetic systems (variously called neural interfaces, brain-machine interfaces, or BMI) that aim to help severely disabled human patients. There are many medical, scientific, and engineering challenges in developing such systems [1]–[5], and all neural prosthetic systems share in common a signal processing backend. This backend takes as input raw voltage waveforms from multi-electrode recordings (or other technologies), and it produces as output a control signal such as kinematic parameters to control a prosthetic arm. Along this signal flow, there are two major steps: first, raw voltage must be separated into spike trains from single or multiple neural units, called “spike sorting”; second, these spike trains must be processed by a decoding algorithm to produce behavioral control signals.

This work was supported by: Burroughs Wellcome Fund Career Award in the Biomedical Sciences, Christopher and Dana Reeve Foundation, HHMI Fellowship (P.N.), Johns Hopkins University Applied Physics Laboratory under the DARPA Revolutionizing Prosthetics program contract N66001-06-C-8005, The McKnight Endowment Fund for Neuroscience, NDSEG Fellowship (V.G.), NIH-NINDS N01-NS-4-2362, NIH-CRCNS-R01, NSF Graduate Research Fellowships (C.A.C., V.G.), Soros Fellowship (P.N.), Stanford’s CIS, SGF (C.A.C., J.P.C.), and Medical Scholars Program (P.N.)

C.A.C. and J.P.C. are with the Dept of Electrical Engineering, Stanford U, Stanford, CA, V.G. is with the Dept of Computer Science, Stanford U, Stanford, CA, S.I.R. is with the Dept of Neurosurgery, Stanford U, Stanford, CA (Present address: Dept of Neurosurgery, Palo Alto Medical Foundation); K.V.S. is with the Dept of Electrical Engineering, Bioengineering, and Neurosciences Program, Stanford U, Stanford, CA USA, with the Department of Electrical {cindyc, jcunnin, gilja, phn, seoulman, shenoy}@stanford.edu

*These authors contributed equally to this work.

Both of these steps have been well studied: spike sorting [6]–[8] and decode algorithms [5], [9]–[23]. These works have delivered important proofs of concept that brain-machine interfaces can translate neural signals into physical commands. However, moving to a clinically viable system will require several significant developments. These developments exist at all stages: in the recording technologies [24], the signal processing backend, and prosthetic end effectors such as robotic arms and computer interfaces [25]. This study introduces several problems in the signal processing domain.

First, the field must better understand how the recorded signals change over time, as there has been much work suggesting various levels of stability in recorded neural activity over time [26], [27] - discussed in Section III-A below. Next, noisy spike trains must be meaningfully processed into neural firing rates or other quantities appropriate for input into decode algorithms; we address this potential problem in Section III-B below. Decode algorithms calculate a mapping between physical behavior and neural activity. We introduce unresolved questions in these models in Section III-C below. Further, a large problem may be fundamental limitations in the data - discussed in Section III-D. These limitations exist due to an insufficient signal to noise ratio in the limited number of neural channels available, as well as model mismatch (e.g., many algorithms assume linear mappings to model nonlinear relationships). Other limitations may also exist in experimental design and algorithmic testing, and we discuss those potential issues in Section III-E. Many of these aspects of BMI performance can interact in complex ways. However, as a starting point in this study, we will address them individually.

II. METHODS

A. Animal Task and Neural Recordings

Animal protocols were approved by the Stanford University Institutional Animal Care and Use Committee. We trained a rhesus monkey (*Macaca mulatta*), monkey L, in a standard reaching paradigm that has been extensively reported elsewhere [25], [28], [29]. We give a short overview here. We implanted a 96-electrode Utah electrode array (Blackrock Microsystems, Salt Lake City, UT) into premotor cortex. The array was implanted 10 months prior to the experiments, showed substantial neural activity, and continued to do so for several months after the experiments¹. The monkey is trained to make instructed reaches to a number of points (28 peripheral targets at 4 radial distances from

¹Previous reports discuss the same monkey. Here we use a newer implant (same technology) and a very similar experimental paradigm.

the central target, uniformly distributed in 7 directions) on a vertical screen. Monkey L begins with his hand on a target at the screen center. After a brief hold time, a peripheral target appears, indicating the goal of his reach. Restrictions on reaction time ensure that the monkey will reach quickly and accurately to the peripheral target, then receiving a juice reward [28]. This experiment was performed on 39 days over a period of 7 weeks. Prior to this time, this monkey had been heavily trained on similar tasks for several years. Here we analyze the first hour of data from each day, with an average of 1655 reaches per dataset.

B. Neural Prosthetic Decoding

In the results that follow, we will demonstrate the quality of decoding neural activity based on different segmentations of the neural data. Accordingly, we need a method with which to decode neural activity into action so that we can compare performance of different signal processing techniques. We describe those methods briefly here, where we refer to blocks of the general signal flow for a BMI, as shown in Fig. 1.

To extract spike trains from raw voltage, neural units were isolated off-line using a PCA-based spike sorting algorithm [30], and quality was assessed by hand using the waveforms and clusters in principle component space. Units were labeled single unit, contaminated single-units (with waveforms from other neurons), and multi-units. For analyses using threshold crossings only, all events that crossed a threshold of three times rms noise were used; more explanation can be found in [29]. All of these threshold crossings were classified as single or multi-units in the full spike sorting analysis.

First, we use a simple maximum likelihood (ML) decoder, as seen in [15], [25]. This method uses training data to build an expectation, for reaches to each of the reach targets, of the number of spikes recorded from each neuron. Given test data, the ML decoder evaluates the likelihood (under a Poisson noise model) and picks the reach condition with the largest value (hence maximum likelihood) as the decoded reach. The percentage of reach conditions correctly decoded is reported as overall performance [25].

ML decoding makes a discrete choice. In some cases, we also want to decode moment-by-moment parameters of the subject's reach. To do this, we use the popular linear decoder (LD), which assumes that movement is a linear combination of recorded neural activity. Using least squares, the movement can be decoded from neural activity, and common metrics such as root-mean-squared-error (RMS) or correlation coefficient can be used to determine the quality of decode [28], [29]. A third common approach is the Kalman filter [31], which stipulates a linear relationship between physical behavior over time and between neural activity and physical behavior [14], [29].

III. RESULTS AND DISCUSSION

Here we discuss the problems we highlighted in the introduction, and we demonstrate these problems in our experimental data and recent publications.

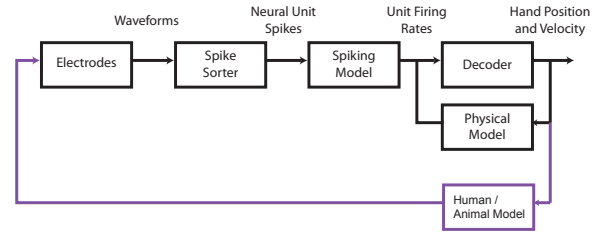


Fig. 1. Block diagram of a typical BMI illustrating potential areas for improvement. The lower feedback loop illustrates aspects of neural adaptation that can be engaged only in closed-loop experiments.

A. Spike Sorting

Spike sorting is a major challenge in neural signal extraction, both for basic neuroscience studies and for neural prosthetic systems [6]–[8]. We discuss here the importance of isolating single neuron activity, and the instability of neural recordings over time.

1) *Single Unit Activity*: When studying the properties of individual cells, it is important to isolate “single units” with accurate spike sorting. Often, recorded neural activity that likely arises from multiple cells is excluded from analysis, despite the fact that such activity typically comprises a substantial portion of recorded neural activity. For example, from a single dataset, we differentiated all neural events into 205 clusters. Of these 205 neural units, only 53 came from well-isolated or somewhat contaminated units (units with clearly differentiated waveforms that were either not adjacent to other waveforms in voltage and PCA space (well isolated) or adjacent but clearly distinct (somewhat contaminated)). The remaining 152 neural units were classified as likely multi-unit. Table I shows the percent correct for a ML decode of reach directions from single unit only and multi-unit activity. Also, adding multi-unit activity to single unit activity increases performance from 74% to 82%. Therefore, it seems clear that multi-units should be included in prosthetic decoders, despite being “unclean” isolations in a basic science sense.

The complexity of the spike sorting process has substantial power implications for integrated circuits that may be used as part of future clinical systems to transmit wireless neural data from the patient [24], [32], [33] since it changes the number of bits required per channel for full waveforms versus threshold crossings. This may be partially alleviated by small process technologies or novel powering methods, but is still likely to be a substantial concern. This raises the question, does full spike sorting produce a large improvement in decode performance over threshold based systems? Table I shows a performance comparison between PCA based spike sorting, using all single and multi-units (third row) and a single threshold per channel (3 times RMS noise, fourth row). While small increases in performance can be important to users, using sorted spikes instead of thresholds produces a surprisingly small improvement of 7%. Also, the threshold number represents a base level performance which could likely be improved by setting the thresholds optimally on a

TABLE I
DECODE PERFORMANCE BY UNIT TYPE.

	Number of Units	Decode Performance	
		ML Decode ¹	Correlation Coeff. ²
Single units ³	53	74%	0.86
Multi-units	152	79%	0.91
All sorted units	205	82%	0.92
Thresholds ⁴	96	75%	0.89

¹ Note that chance ML decode accuracy is 1/7, or 14%.

² Correlation coefficient based on a linear decoder.

³ Includes definite and high-confidence single units.

⁴ Standard thresholding at 3 times RMS noise.

per-channel basis. Further, two thresholds per channel could also substantially make up the difference in performance without requiring full broadband data. The optimal may resemble [34], in which bits of resolution are distributed to channels based on information content.

2) *Waveform Shape Instability*: While there is evidence that neurons themselves maintain stable tuning properties at least over the course of a day [28], there is significant doubt about the stability of the raw voltage recordings of those neurons over the same time periods (due to changing position of the electrodes with respect to the neurons, or similar) [26], [27]. If these recordings are not stable, accurate spike sorting will require additional sophistication to track neural units over the course of minutes, hours, and days [35].

Fig. 2 shows the dramatic effect of this instability on decode performance. We fix a decode algorithm to the population recorded on an array on the first day of recording. We fix both the maximum likelihood parameters and the spike-sorting projections (waveform shapes) across seven weeks. Performance falls precipitously after only a few days, which must be due to changes in the recorded neural activity (the signal processing backend has been held constant). This suggests that nonstationarity will be a substantial problem in future clinical systems. Human systems to date have used daily calibration by skilled technicians [36], but this approach will not economically scale to broad use.

Fig. 3 shows how much a recording can change over the course of a single experimental recording session. While the average change in waveform shape is small (many remain within $\pm 5\%$ of their time zero size) several neurons indeed change their waveforms significantly over just one hour (e.g. red growing by 25% and the blue shrinking by 25%). The average absolute change was 0.3%/min. Looking at the slopes in absolute voltages, the average change was 1 $\mu\text{V}/\text{min}$, but changes above 5 $\mu\text{V}/\text{min}$ were observed on several units. Examination over even greater time spans may reveal even greater excursions. More accurately characterizing these changes requires analysis of multi-week wireless recordings

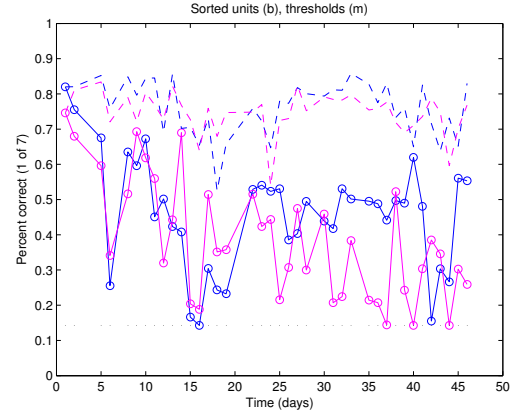


Fig. 2. Decoder performance (ML accuracy) determined by generating spike sorting templates and maximum likelihood coefficients from the first day of the experiment, and applying those models across 7 weeks of similar experiments. Sorted data shown in blue, threshold data shown in magenta; dotted lines indicate data that was re-fit on each experimental day.

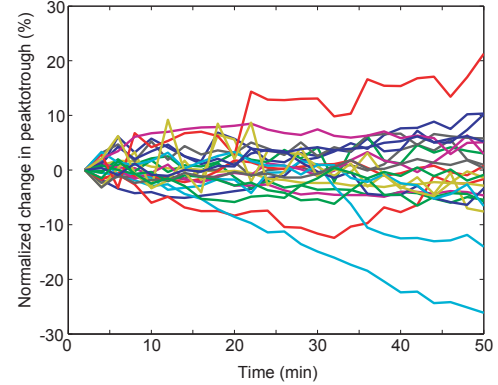


Fig. 3. Waveform change over one hour for 23 example neurons. Change is normalized to the size of the initial waveform.

[37] from additional animals². These waveform changes can and do cause serious spike sorting difficulties. Fig. 4 shows a single unit that remained well isolated over several weeks (uncontaminated units like this are rare). The tuning curves in the second row suggest this is the same neuron, despite substantial waveform changes. More commonly, as shown in the third row, a unit that is initially well isolated disappears over days into multi-unit activity. Since spike-sorting algorithms rely on waveform shape, these instabilities may confound spike sorting significantly over the course of several hours, and certainly across days.

Some aspects of these instabilities might be particular to the experimental preparation considered here. However, these devices were approved for initial human studies [36], are likely to be used for future human work, and are believed to be at least as stable (if not more so) than other multi-electrode technologies due to its ability to move with the brain rather than being secured to the skull. Accordingly,

²It is possible that some of the change in performance over days is due to slight differences in connector impedance. However, this likely makes only a small contribution since the noise across the array was fairly stable (mean 1.1 μVrms , std 0.2 μVrms). Also, there were a few electrodes with highly similar waveforms across days. At the the same time, waveform changes on individual electrodes could be dramatic over only a few hours.

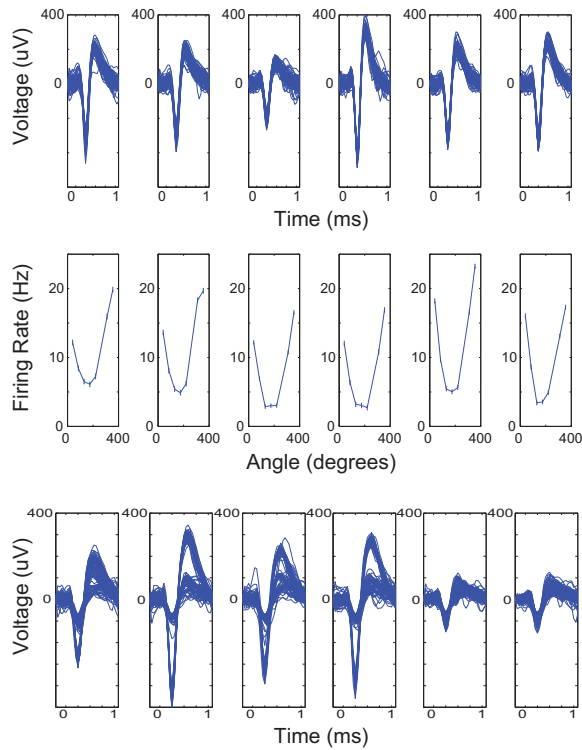


Fig. 4. Example waveforms over days. Panel A shows an isolated unit changing over time but (usually) remaining well isolated. From left to right, top panels show the single unit from days 1, 5, 16, 17, 34, and 45. Panel B shows the tuning pattern across 7 angular directions on those days. Panel C shows a more common example of a waveform shape collapsing into the multi-unit activity. Waveforms crossing threshold are shown for one electrode on six consecutive days.

one might consider different strategies going forward to compensate for these instabilities. This may be substantially easier in neural prosthetics than basic science because it is arguably not important to track single units. Perhaps decode algorithms can be designed by sorting on tuning alone over time or using simply no spike-sorting algorithm whatsoever [20]. In any event, since the effect on performance is high, serious effort to address these instabilities must be committed.

B. Models for Neural Spiking

Spike trains present analytical challenges due to their noisy, spiking nature. A common view is that spikes are generated from a smooth function of time (the firing rate) and that this function carries a significant portion of the neural information (vs. the precise spike timing). If so, decoding neural activity may require accurately estimate firing rate. There has been extensive work in modeling spike trains [5], [38]–[42] and estimating firing rates [43]–[47]. While some decode algorithms average over neural activity in small temporal windows [17], some algorithms use firing rates or use spiking models directly [18]. Spiking models are another source of approximation in BMIs. Though sophisticated firing rate estimation has proven valuable in basic neuroscience, a recent study found minimal differences in prosthetic decode performance using different estimators [29]. Perhaps models for neural spiking, though clearly yet

one more approximation in decoding, may not be a source of major performance gain for future research.

C. The Mapping between Physical Behavior and Neural Spiking

To date, essentially all prosthetic decode algorithms (population vectors, linear decoders, Kalman filters, etc.) have assumed a linear mapping between kinematic parameters of the arm and neural firing rates (or, in some cases, spiking activity directly) [5], [9]–[23]. There are a few potential shortcomings with this linear choice, including the fact that most algorithms ignore meaningful nonlinearities in neural data, and the poor generalization of these models.

1) *Nonlinearities*: There is wide variation in how well the activity of a neuron can be linearly related to a given kinematic parameter, shown in Fig. 5. The top panel shows the average reach speed, and the second panel shows X-position for reaches to 7 out of 28 targets. These represent typical kinematic signals that one would like a linear model to accurately predict. The middle panel shows that some neurons that have a strongly linear relationship with speed given a specific time lag. In general, a subset of neurons may have a strong linear relationship with a given kinematic parameter. However, the remaining two panels show firing rates from four neurons with firing rates that do not have an obvious linear relationship to any parameter. For example, activity in the fourth panel comes from two units with long plateaus of activity that precede and follow movement. The two bottom units show double peaks, that also have no obvious linear transformation to kinematic parameters.

One way linear decoders can cope with non-linearities is to use only units with clear linear relationships and set other coefficients to low values. In our data, linear decoders can come within 10% of the optimal error using between 15-29% of the 53 predominantly single units (for x and y position and velocity). This number was obtained by sorting individual units for correlation with the various parameters, and adding them to the decoder until the error was within 10% of its value for the whole ensemble. To achieve this performance level on all 4 kinematic parameters together, only 49% of the units were required. More than half the units were unused. This underutilization may occur because neurons with nonlinear relationships to behavior provide a source of model mismatch to linear decoders. However, model mismatch is not noise; there is possibly information in these neural units that linear decoding models (like the linear filter, the population vector, and the Kalman filter) are unable to exploit. Future algorithmic designs may offer significant performance improvements by modeling nonlinearities in this mapping. Also, nonlinearities could be introduced at many points in the signal flow shown in Figure 1, not just the mapping considered here.

2) *Generalization*: The ability of a model to generalize to novel conditions is a major concern with any decoding algorithm. Linear algorithms in particular may generalize poorly to novel reaches. For example, in the current dataset, determining an optimum linear filter using 27 out of the 28

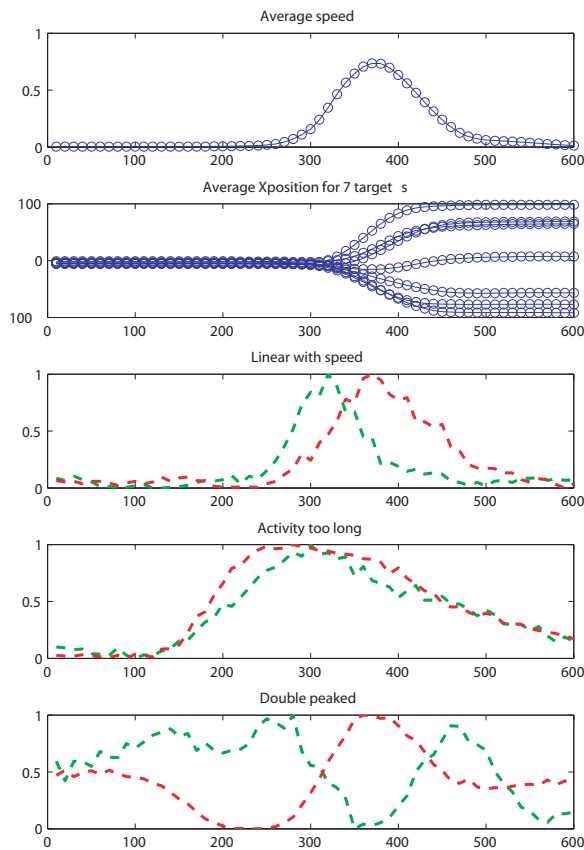


Fig. 5. Comparison of kinematic parameters (blue traces) with neural firing rates (red and green traces) neural (A) Average speed during reaches. (B) Average X-position during reaches to 7 out of 28 targets. (C) Normalized average firing rates from two units with a strong linear relationship to velocity. (D) Two example units whose activity precedes and postcedes neural activity. (E) Two example units with double peaked average firing rates. For both (D) and (E) there is no obvious linear transform between neural activity and any of the kinematic parameters.

targets and testing on reaches to the remaining target resulted in a 4x greater squared error on average than training on a dataset that included reaches to that target. This occurred despite the fact that the training dataset included 3 other examples of reaches to the same angle and 6 other reaches to the same distance. It is notable that many other prosthetic experiments to date have used highly constrained movement tasks which may overestimate the ability of linear models to generalize [17], [48]. While these tests indeed demonstrate useful signal extraction from cortex, they do not test a broad range of behavior. Accordingly, it may be that these constrained experimental settings pose an unrealistic proxy to the eventual user mode.

A real prosthesis user will desire a broad range of potentially novel behaviors. An accurate model mapping physical behavior to neural activity must be able to decode novel reaching conditions. Moving to unconstrained settings in three dimensions, with many other types of reaching - curved, straight, point-to-point, continuous, and more - there are many possible model mismatches. Further, arbitrary movement in three dimensions engages long-studied questions of reference frames and coordinate transforma-

tions [49], which may complicate things further. In short, experimental constraints may not translate to a prosthesis that generalizes to the needs of a human user. Some effort should be made in vetting all BMI developments with a range of experimental control (including very little).

D. Limitations on Precision

There is obviously not arbitrarily large information content in a given number of neural channels. For example, while the output of a continuous linear decoder can exhibit high correlation with the actual hand movement, single trials often decode to erratic reach behavior. Figure 6A shows an example of actual reaches to one of the 28 targets in the center out task. A position-based linear decoder trained on the first half of the dataset predicted reaches in the second half of the data with a correlation coefficient of $R=0.88$, which shows similar performance to other results in the literature [14], [48]. Reaches decoded by the linear model are shown in Fig. 6B. While the average correlation is apparent, the endpoints exhibit a much higher standard deviation (21 mm vs 6 mm) than actual reaches (the red ellipse).

This illustration represents an “offline” linear decode. One might argue that these incorrect trajectories can be corrected using feedback in an “online” BCI experiment. However, online linear models have shown a tendency to move erratically as well [48]. This may place limitations on how closely spaced potential targets can be and whether undesired targets can be avoided. Moving from computer control to the control of a robotic limb would further emphasize this problem.

1) *Models for Physical Behavior*: One weakness of algorithms like the linear decoder (and population vector) is that these algorithms do not have an explicit physical behavior model, and thus all noise in the recorded signal is passed through to the decoded arm trajectory. In contrast to this shortcoming, models such as the Kalman filter [31], which stipulate a model for physical behavior in arm reaches, have been shown to outperform the linear decoder in a variety of cases [14], [48]. This success led to extensions that assume similar models for physical behavior [5], [13], [17], [18], [21]–[23], [50]. Unfortunately, this class of models for physical behavior is inappropriate in some ways for reaching movements.

Specifically, the Kalman filter assumes a linear dynamical system ($\mathbf{x}_t = A\mathbf{x}_{t-1} + v$, where v is some noise). Depending on the matrix A , reaches from this distribution can only converge to the origin, oscillate, or diverge to infinity, which conflicts with the reality that the majority of arm reaches are point-to-point [49]. Fig. 6C shows that this model can decoded reaches that fail to stop. Overall, the performance does not appear substantially less erratic than those from the simple linear decoder. The Kalman filter, like other linear models, fails to infer the correct reach goal and stop precisely. These inadequate physical models have been chosen in large part because of their mathematical tractability. Instead, a model could exploit the deep literature describing how reaches are actually made in human behavior [49].

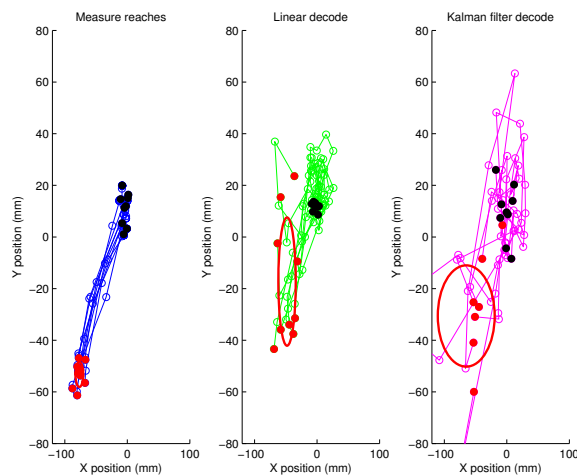


Fig. 6. Panel A shows actual reaches to one of 28 targets measured using an infrared motion tracking system. Panel B shows the linear decode of neural activity during those reaches. Panel C shows a Kalman filter decode of that same activity. Black dots denote the end points of each reach, and the black ellipse denotes the standard deviation in the X and Y direction. Note the red end-point variance ellipse is very small in the first panel.

2) *Data Limitations*: Current recording technologies can record from up to hundreds of individual neurons, which is a tiny fraction of the many millions involved in arm reaching. Accordingly, the field is and will continue to be limited in the amount of information it can record from cortex.

BMI devices using lower SNR sources such as ECoG often try to maximize information throughput by using an “indirect” signal source. For example, imagining something that can be somewhat unrelated to arm movement in order to generate cursor movement [51]. Cortical BMI’s have relatively higher SNR, and can attempt a “direct” decode [3]. For example, [15], [17] decoded movements towards particular targets, but the number and position of targets was small, fixed, and known. [48] demonstrated improved accuracy with humans controlling a computer cursor by using a training paradigm of reaches that moved very slowly.

Much traction might be gained by restricting the space of movements that can be decoded from neural activity. Researchers may consider the field of human motor control (e.g., [49]), where work has shown fundamental constraints on the human reaching system. By similarly constraining the space of movements that can be decoded from neural activity, some performance improvements may be achieved using currently available signal sources. By recognizing that there is not an arbitrary amount of information in the recorded neural activity, the field can begin to ask meaningful questions about what actions we may hope to extract from cortex. Designing decode strategies in this way will be critical in moving towards a clinically viable system.

E. Experimental Limitations

In this final section we introduce another potential issue in current prosthetic design, and we discuss why we think addressing this issue may be a valuable direction for future investigation. As previously noted, experimental constraints do not necessarily translate to a prosthetic device that can

generalize well. For example, decoding success is often determined by how well the decoded arm trajectory matches the true arm movement that was recorded alongside the (possibly synthetic) neural activity. Unfortunately, this “offline” approach neglects potentially important features of a real neural prosthesis, including the prosthetic user’s ability to modify behavioral strategies to improve control of the prosthetic device (via the decode algorithm). In other words, as soon as the prosthesis user sees the prosthetic device act, he/she will bring to bear all his/her behavioral modification strategies to attempt to drive a natural, desired reach. In moving towards a usable prosthesis, experimental paradigms should be tested in this “online” context in order to provide a realistic proxy to clinical use. This feature is noted by the large feedback loop in Fig. 1. The field should investigate the extent to which the subject can (for a given decode algorithm, spike sorting approach, or other signal processing choice) engage feedback mechanisms, learning and adaptation, and other control strategies to improve decode performance.

IV. CONCLUSIONS

Neural prostheses have received much attention in the last decade. In this study, we used 39 neural datasets, from a single monkey making center out reaches day after day, to examine potential areas for future advances. These analyses suggest that areas such as single unit spike sorting and advanced spiking models, while useful to pure neuroscience research, may not provide dramatic performance increases in future BMIs. However, there are three areas that we believe may provide more space for improvement. First, non-stationarity of neural waveforms must be addressed when moving towards long term clinical systems. Second, linear models may not be fully exploiting information available from particularly non-linear neurons. This may also lead to observed difficulties in model generalization. Third, erratic decoded movements cause difficulty in predictably controlling a BMI cursor. This shortcoming could be mitigated by more careful analysis of the neural information content, by limiting the types of reaches based on the information available, and by meaningfully testing algorithmic developments in an online context. In all of these issues, it is of great value for the field to review and compare available methods at each step in the BMI signal path, and to design future studies (both experimental and algorithmic) with those results in mind.

V. ACKNOWLEDGMENTS

We thank M. Risch for veterinary care, D. Haven for technical support, and S. Eisensee for administrative support.

REFERENCES

- [1] M. A. Lebedev and M. A. L. Nicolelis, “Brain-machine interfaces: past, present, and future,” *Trends in Neurosci.*, vol. 29, 2006.
- [2] A. B. Schwartz, “Cortical neural prosthetics,” *Ann. Rev. Neurosci.*, vol. 27, pp. 487–507, 2004.
- [3] J. P. Donoghue, “Bridging the brain to the world: A perspective on neural interface systems,” *Neuron*, vol. 60, pp. 511–521, 2008.
- [4] M. D. Linderman, G. Santhanam, C. T. Kemere, V. Gilja, S. O’Driscoll, B. M. Yu, A. Afshar, S. I. Ryu, K. V. Shenoy, and T. H. Meng, “Signal processing challenges for neural prostheses,” *IEEE Signal Processing Magazine*, vol. 25, pp. 18–28, 2008.

- [5] J. C. Sanchez, J. C. Principe, T. Nishida, R. Bashirullah, J. G. Harris, and J. Fortes, "Technology and signal processing for brain-machine interfaces: The need for beyond the state-of-the-art tools," *IEEE Sig Proc Magazine*, vol. 25, no. 1, pp. 29–40, 2008.
- [6] M. S. Lewicki, "A review of methods for spike sorting: the detection and classification of neural action potentials," *Net Comp Neur Sys*, 1998.
- [7] F. Wood, M. J. Black, C. Vargas-Irwin, M. Fellows, and J. P. Donoghue, "On the variability of manual spike sorting," *IEEE Trans. Biomed. Eng.*, vol. 51, pp. 912–918, 2004.
- [8] F. Wood and M. J. Black, "A nonparametric bayesian alternative to spike sorting," *J Neurosci Methods*, vol. 173, pp. 1–12, 2008.
- [9] A.P. Georgopoulos, A.B. Schwartz, and R.E. Kettner, "Neuronal population coding of movement direction," *Science*, vol. 233, pp. 1416–1419, 1986.
- [10] Y. Gao, M.J. Black, E. Bienenstock, S. Shoham, and J.P. Donoghue, "Probabilistic inference of hand motion from neural activity in motor cortex," *Advances in NIPS*, vol. 14, 2002.
- [11] C.T. Kemere, K.V. Shenoy, and T.H. Meng, "Model-based neural decoding of reaching movements: a maximum likelihood approach," *IEEE Transactions on Biomedical Engineering – Special issue on Brain-Machine Interfaces*, vol. 51, pp. 925–932, 2004.
- [12] A.E. Brockwell, A.L. Rojas, and R.E. Kass, "Recursive Bayesian decoding of motor cortical signals by particle filtering," *J Neurophysiol*, vol. 91, no. 4, pp. 1899–1907, 2004.
- [13] S. Kim, J. C. Sanchez, Y. N. Rao, D. Erdogmus, J. M. Carmena, M. A. Lebedev, M. A. L. Nicolelis, and J. C. Principe, "A comparison of optimal mimo linear and nonlinear models for brain-machine interfaces," *J Neural Engineering*, vol. 3, pp. 145–161, 2006.
- [14] W. Wu, Y. Gao, E. Bienenstock, J.P. Donoghue, and M.J. Black, "Bayesian population decoding of motor cortical activity using a Kalman filter," *Neural Comp*, vol. 18, no. 1, pp. 80–118, Jan. 2006.
- [15] G. Santhanam, S. I. Ryu, B. M. Yu, A. Afshar, and K. V. Shenoy, "A high-performance brain-computer interface," *Nature*, vol. 442, pp. 195–198, 2006.
- [16] G. Shakhnarovich, S. P. Kim, and M. J. Black, "Nonlinear physically-based models for decoding motor-cortical population activity," in *Advances in NIPS 18*. MIT Press, Cambridge, MA, 2006.
- [17] B. M. Yu, C. Kemere, G. Santhanam, A. Afshar, S. I. Ryu, T. H. Meng, M. Sahani, and K. V. Shenoy, "Mixture of trajectory models for neural decoding of goal-directed movements," *J Neurophysiol*, vol. 97, pp. 3763–3780, 2007.
- [18] L. Srinivasan, U. T. Eden, S. J. Mitter, and E. N. Brown, "General purpose filter design for neural prosthetic systems," *J Neurophysiol*, vol. 98, pp. 2456–2475, 2007.
- [19] 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, 2008.
- [20] V. Ventura, "Spike train decoding without spike sorting," *Neural Comp*, vol. 20, pp. 923–963, 2008.
- [21] W. Wu and N. G. Hatsopoulos, "Target-included model and hybrid decoding of stereotyped hand movement in the motor cortex," in *Intl Conf on Biomed Robotics and Biomech*, 2008. (IEEE RAS/EMBS), 2008, pp. 55–60.
- [22] J. E. Kulkarni and L. Paninski, "State-space decoding of goal-directed movements," *IEEE Sig Proc Magazine*, vol. 25, no. 1, pp. 78–86, 2008.
- [23] J. DiGiovanna, B. Mahmoudi, J. Fortes, J.C. Principe, and J.C. Sanchez, "Coadaptive brain-machine interface via reinforcement learning," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 1, pp. 54–64, 2009.
- [24] K.D. Wise, D.J. Anderson, J.F. Hetke, D.R. Kipke, and K. Najafi, "Wireless implantable microsystems: High-density electronic interfaces to the nervous system," *Proc IEEE*, vol. 92, pp. 76–97, 2004.
- [25] J. P. Cunningham, B. M. Yu, V. Gilja, S. I. Ryu, and K. V. Shenoy, "Toward optimal target placement for neural prosthetic devices," *J Neurophysiol*, vol. 100, no. 6, pp. 3445–3457, 2008.
- [26] G. Santhanam, M. D. Linderman, V. Gilja, A. Afshar, S. I. Ryu, T. H. Meng, and K. V. Shenoy, "HermesB: A continuous neural recording system for freely behaving primates," *IEEE Trans. in Biomed. Eng.*, vol. 54, pp. 2037–2050, 2007.
- [27] S. Suner, M. R. Fellows, C. Vargas-Irwin, G. K. Nakata, and J. P. Donoghue, "Reliability of signals from a chronically implanted silicon-based electrode array in non-human primate primary motor cortex," *IEEE Trans. Neural Sys. Rehabil. Eng.*, pp. 524–541, 2005.
- [28] C.A. Chestek, A.P. Batista, G. Santhanam, B.M. Yu, A. Afshar, J.P. Cunningham, V. Gilja, S.I. Ryu, M.M. Churchland, and K.V. Shenoy, "Single-neuron stability during repeated reaching in macaque premotor cortex," *J Neurosci*, vol. 27, pp. 10742–10750, 2007.
- [29] J. P. Cunningham, V. Gilja, S. I. Ryu, and K. V. Shenoy, "Methods for estimating neural firing rates, and their application to brain-machine interfaces," *Neural Networks*, 2009, In Press.
- [30] M. Sahani, *Latent Variable Models for Neural Data Analysis*, Ph.D. thesis, California Institute of Technology, 1999.
- [31] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Engineering*, vol. 82, pp. 35–45, 1960.
- [32] R. R. Harrison, P. T. Watkins, R. J. Kier, R. O. Lovejoy, D. J. Black, B. Greger, and F. Solzbacher, "A low-power integrated circuit for a wireless 100-electrode neural recording system," *IEEE J of Solid State Circuits*, vol. 42, pp. 123–133, 2007.
- [33] Z. S. Zumsteg, C. Kemere, S. O'Driscoll, G. Santhanam, R. E. Ahmed, K. V. Shenoy, and T. H. Meng, "Power feasibility of implantable digital spike sorting circuits for neural prosthetic systems," *IEEE TNSRE*, vol. 13, pp. 272–279, 2005.
- [34] S. O'Driscoll, T. H. Meng, K. V. Shenoy, and C. Kemere, "Neurons to silicon: Implantable prosthesis processor," *Int. Solid State Circuits Conference (ISSCC)*, pp. 552–553, 2006.
- [35] A. Bar-Hillel, A. Spiro, and E. Stark, "Spike sorting: Bayesian clustering of non-stationary data," *Journal of Neuroscience Methods*, vol. 157, pp. 303–316, 2006.
- [36] L. R. Hochberg, M. D. Serruya, G. M. Friehs, J. A. Mukand, M. Saleh, A. H. Caplan, A. Branner, D. Chen, R. D. Penn, and J. P. Donoghue, "Neuronal ensemble control of prosthetic devices by a human with tetraplegia," *Nature*, vol. 442, pp. 164–171, 2006.
- [37] C. A. Chestek*, V. Gilja*, P. Nuyujukian, R. Kier, F. Solzbacher, S. I. Ryu, R. R. Harrison, and K. V. Shenoy, "HermesC: Low-power wireless neural recording system for freely moving primates," *IEEE Trans. in Neural. Sys. and Rehabil. Eng. (In Press)*, 2009.
- [38] D. Johnson, "Point process models of single-neuron discharges," *J Comput Neurosci*, vol. 3, pp. 275–299, 1996.
- [39] R. Barbieri, M.C. Quirk, L.M. Frank, M.A. Wilson, and E.N. Brown, "Construction and analysis of non-Poisson stimulus-response models of neural spiking activity," *J Neurosci Methods*, pp. 25–37, 2001.
- [40] V. Ventura, R. Carta, R.E. Kass, S.N. Gettner, and C.R. Olson, "Statistical analysis of temporal evolution in single-neuron firing rates," *Biostatistics*, vol. 3, no. 1, pp. 1–20, 2002.
- [41] R.E. Kass and V. Ventura, "A spike-train probability model," *Neural Comp*, vol. 14, pp. 5–15, 2003.
- [42] W. Truccolo, U.T. Eden, M.R. Fellows, J.P. Donoghue, and E.N. Brown, "A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects," *J Neurophysiol*, vol. 93, pp. 1074–1089, 2004.
- [43] H. Shimazaki and S. Shinomoto, "Kernel width optimization in the spike-rate estimation," *Neural Coding, Montevideo, Uruguay*, 2007.
- [44] J. P. Cunningham, B. M. Yu, K. V. Shenoy, and M. Sahani, "Inferring neural firing rates from spike trains using Gaussian processes," in *Advances in NIPS 20*, J. Platt, D. Koller, Y. Singer, and S. Roweis, Eds. MIT Press, Cambridge, MA, 2008.
- [45] I. DiMatteo, C.R. Genovese, and R.E. Kass, "Bayesian curve-fitting with free-knot splines," *Biometrika*, vol. 88, pp. 1055–1071, 2001.
- [46] R.E. Kass, V. Ventura, and E.N. Brown, "Statistical issues in the analysis of neuronal data," *J. Neurophysiol*, vol. 94, pp. 8–25, 2005.
- [47] D. Endres, M. Oram, J. Schindelin, and P. Foldiak, "Bayesian binning beats approximate alternatives: estimating peri-stimulus time histograms," *Advances in NIPS*, vol. 20, 2008.
- [48] S. Kim, J. D. Simeral, L. R. Hochberg, J. P. Donoghue, and M. J. Black, "Neural control of computer cursor velocity by decoding motor cortical spiking activity in humans with tetraplegia," *J Neural Engineering*, vol. 5, pp. 455–476, 2008.
- [49] R. Shadmehr and S. P. Wise, *The Computational Neurobiology of Pointing and Reaching*, MIT Press, Cambridge, 2005.
- [50] W. Wu, J. E. Kulkarni, N. G. Hatsopoulos, and L. Paninski, "Neural decoding of goal-directed movements using a linear state-space model with hidden states," *COSYNE Abstract*, 2008.
- [51] E. C. Leuthardt, K. J. Miller, G. Schalk, R. P. N. Rao, and J. G. Ojemann, "Electrocorticography-based brain computer interface - the seattle experience," *IEEE. Trans. on Neur. Sys. Rehabil. Eng.*, vol. 14, pp. 194–198, 2006.