PAPER

Performance sustaining intracortical neural prostheses

To cite this article: Paul Nuyujukian et al 2014 J. Neural Eng. 11 066003

View the <u>article online</u> for updates and enhancements.

You may also like

- A high performing brain-machine interface driven by low-frequency local field potentials alone and together with spikes Sergey D Stavisky, Jonathan C Kao, Paul Nuyujukian et al.
- Intention estimation in brain—machine interfaces
 Joline M Fan, Paul Nuyujukian, Jonathan C Kao et al.
- Feedback control policies employed by people using intracortical brain-computer interfaces
 Francis R Willett, Chethan Pandarinath,

Beata Jarosiewicz et al.



J. Neural Eng. 11 (2014) 066003 (13pp)

doi:10.1088/1741-2560/11/6/066003

Performance sustaining intracortical neural prostheses

Paul Nuyujukian 1,2,3 , Jonathan C Kao 4 , Joline M Fan 1,5 , Sergey D Stavisky 6 , Stephen I Ryu 4,7 and Krishna V Shenoy 1,4,8,9

E-mail: shenoy@stanford.edu

Received 24 March 2014, revised 29 June 2014 Accepted for publication 10 July 2014 Published 13 October 2014

Abstract

Objective. Neural prostheses, or brain-machine interfaces, aim to restore efficient communication and movement ability to those suffering from paralysis. A major challenge these systems face is robust performance, particularly with aging signal sources. The aim in this study was to develop a neural prosthesis that could sustain high performance in spite of signal instability while still minimizing retraining time. Approach. We trained two rhesus macaques implanted with intracortical microelectrode arrays 1–4 years prior to this study to acquire targets with a neurally-controlled cursor. We measured their performance via achieved bitrate (bits per second, bps). This task was repeated over contiguous days to evaluate the sustained performance across time. Main results. We found that in the monkey with a younger (i.e., two year old) implant and better signal quality, a fixed decoder could sustain performance for a month at a rate of 4 bps, the highest achieved communication rate reported to date. This fixed decoder was evaluated across 22 months and experienced a performance decline at a rate of 0.24 bps yr⁻¹. In the monkey with the older (i.e., 3.5 year old) implant and poorer signal quality, a fixed decoder could not sustain performance for more than a few days. Nevertheless, performance in this monkey was maintained for two weeks without requiring additional online retraining time by utilizing prior days' experimental data. Upon analysis of the changes in channel tuning, we found that this stability appeared partially attributable to the cancelling-out of neural tuning fluctuations when projected to two-dimensional cursor movements. Significance. The findings in this study (1) document the highest-performing communication neural prosthesis in monkeys, (2) confirm and extend prior reports of the stability of fixed decoders, and (3) demonstrate a protocol for system stability under conditions where fixed decoders would otherwise fail. These improvements to decoder stability are important for minimizing training time and should make neural prostheses more practical to use.

S Online supplementary data available from stacks.iop.org/jne/11/066003/mmedia

Keywords: neural prosthesis, brain-machine interface, robustness, high performance, rhesus macaque, longevity

1

¹ Department of Bioengineering, Stanford University, Stanford, CA

² School of Medicine, Stanford University, Stanford, CA

³ Department of Neurosurgery, Stanford University, Stanford, CA

⁴ Department of Electrical Engineering, Stanford University, Stanford, CA

⁵ School of Medicine, University of California San Francisco, San Francisco, CA

⁶ Neurosciences Program, Stanford University, Stanford, CA

⁷ Neurosurgery Department, Palo Alto Medical Foundation, Palo Alto, CA, USA

⁸ Department of Neurobiology, Stanford University, Stanford, CA

⁹ Stanford Neurosciences Institute, Stanford University, Stanford, CA

1. Introduction

Neural prostheses translates brain activity into control signals for guiding assistive devices such as robotic arms and computer cursors. Several reports of neural prostheses have shown promising proof-of-concept demonstrations [1–11], but clinically-useful performance remains a challenge despite recent advances [12-14]. Most approaches have relied on systems requiring daily retraining [15], which is a user-level burden often requiring expert assistance. Prior reports have demonstrated robust performance with fixed decoders. One study utilized fixed decoders trained from an electrocorticographic (ECoG) array over five days [16]. Another offline study found decoder stability with ECoG data across months [17]. Similar stability has been reported with hand selected single-unit recordings over 19 days [18]. An alternative approach to using spiking information may be the recording of local field potential (LFP), which could remain after the loss of spiking information [19-22]. Finally, a recent report detailed decoder stability with LFP and spike thresholds, approaching a year in some cases [23]. These findings, while encouraging, assume that decoder performance can be sustained with fixed parameters. With high quality signal sources, or low frequency signal sources such as LFPs [20], this may be a reasonable assumption given the stability of singleunits and their correlated behavior across hours to days [24]. However, particularly when using aging, years-old electrode arrays, fixed decoders cannot always sustain performance across time. In this study, we sought a framework that sustained high performance in the case where fixed parameter decoders failed without resorting to retraining the decoder from scratch. Such a system would maximize the time during which the prosthesis maintained high performance, while minimizing time spent during retraining.

2. Methods

2.1. Experimental setup

All procedures and experiments were approved by the Stanford University Institutional Animal Care and Use Committee (IACUC). Experiments were conducted with adult male rhesus macaques (L and J), implanted with 96 electrode Utah arrays (Blackrock Microsystems Salt Lake City, UT) using standard neurosurgical techniques. Electrode arrays were implanted in arm motor regions of primary motor cortex (M1) and dorsal premotor cortex (PMd), as estimated visually from local anatomical landmarks [5]. Monkey L was implanted with a single array at the M1/PMd border on 22-01-2008, whereas Monkey J had two arrays implanted, one in M1 and the other in PMd on 2009-08-24. Except for the daily retraining experiments, where in Monkey J only the M1 array was used, all available electrode channels of the arrays were fully utilized: 96 in Monkey L and 192 in Monkey J. The monkeys were trained to make point-to-point reaches in a two dimensional (2D) plane with a virtual cursor controlled by the contralateral arm or by a neurally-driven decoder, as diagrammed in figure 1(a). The neural prosthetic decoder was trained using the ReFIT-KF algorithm [13]. This algorithm utilizes a two-stage training process, starting with arm-controlled training data. That training data is then utilized to build a decoder that is run online and serves as a second set of training data. From that second set of training data, the final decoder is constructed, resulting in control that is well aligned with the monkey's intention [25].

The ipsilateral arm was restrained at all times in these experiments. The contralateral arm, on the other hand, was left unbound and typically continued to move even during neurally-controlled sessions. This approach was preferred because it minimized any behavioral, and thus neural, differences between the training set (arm control sessions) and the testing set (neural prosthesis session). This animal model was selected because we believe it more closely mimicked the neural state a human user would employ when controlling a neural prosthesis in a clinical setting [26]. This model has been previously demonstrated comparable performance to the dual arm-restrained animal model, where the contralateral arm is restrained [13].

The virtual cursor and targets were presented in a 3D environment (MSMS, MDDF, USC, Los Angeles, CA) [27]. Hand position was measured with an infrared reflective bead tracking system (Polaris, Northern Digital, Ontario, Canada) polling at 60 Hz. Neural data were initially processed by the Cerebus recording system (Blackrock Microsystems Salt Lake City, UT) and were available to the behavioral control system within 5 ± 1 ms. Spike counts were collected by applying a single negative threshold, set to $-4.5 \times \text{root}$ mean square of the voltage of the spike band of each neural channel using the Cerebus software's thresholding [12]. Behavioral control and neural decode were run on separate PCs using the Simulink/xPC platform (Mathworks, Natick, MA) with communication latencies of 3 ms. This system enabled millisecond-timing precision for all computations. Visual presentation was provided via two LCD monitors with refresh rates at 120 Hz, yielding frame updates within 7 ± 4 ms. Two mirrors, setup as a Wheatstone stereograph, visually fused the monitors into a single 3D percept for the monkey, although all cursor motion was constrained to two dimensions [28]. Datasets are referenced by a monkey and date format where a dataset from Monkey J would be JYYMMDD.

2.2. Grid task

A keyboard-like grid task was used to measure performance of the system. The goal of the task was to navigate the cursor onto the green target and hold it within the acceptance region for a required hold time. This task was intentionally designed to be similar to a cursor-based keyboard interface that a human subject could use when controlling a neural prosthetic cursor. The task uniformly divided the 24×24 cm workspace into a 5×5 grid of non-overlapping 4.8×4.8 cm square acceptance regions each containing a potential target in yellow. Within this workspace, the cursor was always in the acceptance region of a target. Targets were prompted one at a time in a random-with-replacement fashion, indicated to the

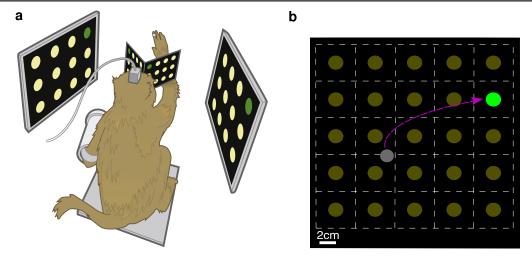


Figure 1. Experimental setup. (a) Diagram of monkey in experimental rig interacting with virtual workspace. (b) Diagram of grid task configured for 25 targets. Dashed white lines are drawn for illustration purposes to denote the boundaries of the acceptance regions (4.8 cm tiles) and were not shown to the monkey. The gray cursor was under neural control and the task was to navigate and hold the cursor over the green target for 450 ms.

monkey by illuminating the target in green. Selections were made after the cursor was held over a single target for 450 ms and the monkey had 5 s within which to acquire the target before failing that trial. After a selection was made, a lockout was employed for 200 ms, during which no hold time was accumulated. This lockout helped prevent accidental selections from the cursor staying still while the monkey scanned for the next target. An incorrect selection occurred when the cursor was held over a non-prompted target for the 450 ms hold time. This task design (i.e., number of targets and hold time) was used becuase it represents the optimal set of task parameters that maximize achieved bitrate in both monkeys [29]. There were no pauses in this task and the subsequent target for the next trial was prompted immediately after a selection was made. If the cursor was outside of the workspace, no selection was made during that time. Monkeys were rewarded with a liquid reward at the end of each correct target acquisition.

2.3. Bitrate

There are many ways to measure the performance of a neurally controlled cursor task. Prior studies have measured values such as success rate and time to target [1, 13, 18, 30, 31], or a measure of channel capacity [5, 32–35]. The issue with metrics such as success rate and time to target is that they are task specific and do not lend themselves well for comparison across studies, making performance difficult to measure. Channel capacity is a task independent measure of performance; it is not a measure of achieved performance, but an upper limit to the throughput of the system. In this study, we sought to evaluate performance in a manner that was task independent to facilitate cross-study comparison as well as represent a meaningful, real-world metric. Achieved bitrate, or information throughput, is a taskindependent measure that accounts for both success rate, selection rate, and adjusts for task complexity. We thus used this metric to evaluate neural prosthetic performance in this study. Achieved bitrate was calculated by measuring the net rate at which correct targets (symbols) were transmitted. For the purposes of a communication neural prosthesis, evaluating bitrate required the transmission of a single symbol out of multiple potential choices, where every symbol had an equal likelihood of being prompted.

Additionally, an important issue regarding calculating bitrate is how to handle error, as communication error heavily penalizes achieved bitrate, but less so for other measures. A theoretic upper-limit known as channel capacity estimates the maximum bitrate of a channel assuming infinitely long channel codes and infinite symbols transmitted. Channel capacity for short selection chains was calculated in prior work [5], and provided an upper bound for the bitrate in that system. This metric is also common with EEG studies [32, 35]. However, despite having units of bits per second, channel capacity is not the same as achieved bitrate. Thus, to avoid overestimating throughput, this study focused solely on reporting achieved bitrate where each transmitted symbol contained independent and unique information. The formula for achieved bitrate, B, under these conditions, is described by equation (1).

$$\boldsymbol{B} = \frac{\log_2(N) \max\left(\left(S_{\rm c} - S_{\rm i}\right), \ 0\right)}{t}.\tag{1}$$

In equation (1), N is the total number of targets on the screen, S_c is the number of correct selections, S_i is the number of incorrect selections, and t is the time elapsed. In this study, t was measured as the total time elapsed between the start of the first trial of a session to the end of the session's last trial (often hours in duration). All inter-trial time was counted as part of the elapsed time. If the quantity $(S_c - S_i) < 0$, then the bitrate was set to a floor of 0 bps since it cannot be negative. Note that trials where a selection was not made (i.e., the cursor was out of the workspace or the trial timed out) were

not errors in symbol transmission and thus they were not counted in S_i since no symbol was transmitted. However, the time spent during such a trial was included in the elapsed time and decreased bitrate appropriately.

An experimental task must meet certain criteria in order to calculate achieved bitrate. However, as achieved bitrate measures actual communicated information, this criteria is not as strict as that for channel capacity [35]. Nevertheless, two of the preconditions are still shared. First, the task must be memoryless and have stable discrete transmission channels. The grid task used here satisfies these requirements by having individual trials that are distinct and separate from all other trials, so that only one symbol may be transmitted at a time. Further, the selection of one symbol does not impact the prompting or selection of the subsequent symbol, as the target prompting sequence is random. The second criteria is that all symbols must be equally likely to be selected. This is satisfied by having all targets identical in size with uniform hold times and again prompted randomly. The third and forth preconditions (i.e., uniform classification accuracy and error) that are necessary for measuring channel capacity do not apply here. This is because the measure of achieved bitrate used in this study penalizes incorrect symbol transmission and does not account for any structure in accuracy or errors. Thus, it does not allocate any information transmission for incorrect transmissions. As a result, equation (1) underestimates the achieved bitrate by fully penalizing for incorrect selections even if the selection was in the vicinity of the prompted target. In fact, equation (1) is the most conservative measure for evaluating bitrate. It sets a floor for the measure of achieved bitrate, and is thus a good estimate of the minimum expected neural prosthetic performance.

Methods for evaluating the structure of transmission errors would elevate bitrate, but these rely on channel codes and measure channel capacity and not achieved bitrate. In the context of a typing task, word completion algorithms are often employed to increase typing speed, but again, this is not a measure of raw performance. For instance, a neural prostheses achieving a 50% success rate that may otherwise appear functional would have an information rate of 0 bps under these metrics. Despite the harsh penalty for errors, this approach more accurately measures the true utility of a neural prosthesis, as transmitting text involves similar challenges and penalties when used in the clinical setting. For reference, on the grid task under arm control, Monkey J achieved 100% success rate with acquire times around 450-600 ms, corresponding to a bitrate in the range of 4.5–5.2 bps. The upper limit on bitrate on this task is 7.1 bps (i.e., $\log_2(25) \cdot \frac{1000}{650} = 7.1$), but this would require near instantaneous movement to each prompted target. The bitrate is not infinite because of the required hold time of 450 ms and the 200 ms post-selection lockout time. Further, as expected, achieved bitrate is zero if the monkey is not participating in the task or the decoder weights are shuffled. A standardized that accurately reflects neural prosthetic measure

performance, such as achieved bitrate measured here, is crucial for tracking progress and the appropriate comparison of neural prostheses across the field.

2.4. Tuning calculation

Channel tuning was evaluated by measuring firing rates on a center out and back eight direction task during neural cursor control using the ReFIT-KF algorithm. This center out task had an 8 cm distance to the target, a 4 cm square acceptance region, a 500 ms hold time, and a 3 s trial timeout. For each direction, a tuning depth was calculated using the average firing rate across the first 200–500 ms after a peripheral target appeared for all trials in that direction on that day. This was repeated for contiguous days using the same baseline decoder built from an earlier experimental day.

2.5. Channel contribution

Channel contribution, which measures the impact a channel has on the decoder on a bin-by-bin basis, was calculated across multiple contiguous experimental days in each monkey. This is mathematically equivalent to taking the average firing rate of the channel over the session and multiplying it by the decoder's assigned velocity weight for the decoder in question. This impact on velocity can then be normalized against all other channels in the session to calculate a fractional contribution to the decoder. These fractional contributions averaged across multiple days are shown in figure 2 and provide insight into a channel's average contribution to the decoder for that monkey. This idea of channel contribution is also extended to the perturbation calculation discussed in detail in the following section.

2.6. Perturbation calculation

Measuring decoder perturbation required analysis of changes across multiple days with a fixed decoder. This was performed with both monkeys and measured as a perturbation in velocity. Velocity perturbation for a single channel was calculated by subtracting the average binned firing rate of that channel on the base experimental day (i.e., the day the fixed decoder was built) from the average binned firing rate on subsequent experimental days. That difference in average binned firing was then multiplied against that channel's neural steady-state weight to velocity (i.e., the weighted linear mapping of channel i to V_x and V_y). This perturbation was measured as a change in velocity m s⁻¹ and represents the change in the contribution to velocity as a result of a channel's change in firing across days. This perturbation is expressed by equation (2).

$$\mathbf{P}_i = \left(\mathbf{f}_{b_i} - \mathbf{f}_{t_i}\right) \cdot \mathbf{V}_i. \tag{2}$$

In equation (2), f_{b_i} is the average firing rate of channel i for the base experimental day, f_{t_i} is the average firing rate of channel i for a subsequent testing experimental day using

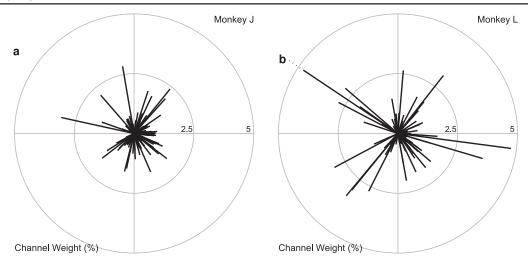


Figure 2. Array status. (a) Weights and directions assigned to channels by the decoder for Monkey J. The magnitude of the black lines represents the percent, on average across multiple consecutive experimental days, that a given channel contributed to the movement of the decoder. Data from four datasets: J120117-J120120. (b) Same plot, but for Monkey L. Channels that exceeded 5% contribution are noted by a dotted line near the tip. Only one such channel exists in Monkey L, which contributed over 9% of the decoder velocity, tuned at approximately $\frac{4\pi}{5}$. Data from five datasets: L120207-L120210 and L120214.

the same decoder, V_i is the decoder's steady state velocity weight, $[v_{x_i} \ v_{y_i}]$, that maps the *i*th channel's firing rate to cursor velocity. The components of P can be positive or negative, depending on the direction of the perturbation in 2D space.

2.7. Signal equipollence

In this study, we also introduce the idea of signal equipollence. It is a measure of the degree of homogeneity in the distribution of contributions to the decoder. These contributions correspond to the weight of the decoder multiplied by the average firing rate of that channel. We have limited this idea largely to a qualitative measure for this study and are introducing it primarily for intuition building and relative comparison between the two monkeys' arrays. If channel contributions were perfectly equal to each other, this would imply that all channels are contributing equally to the decoder. This scenario would be the highest degree of channel signal equipollence possible, because no signal is more important than any other and all are contributing equally to the final decoder estimate. Low channel signal equipollence is characterized by a small handful of channels that have large decoder contribution assigned to them, with the remainder of the channels contributing significantly less. A relative discrepancy in equipollence was seen between monkeys. Monkey J had higher signal equipollence than Monkey L. In an array with low equipollence (e.g., Monkey L's array), a small handful of channels tend to dominate the decoder. It has been previously shown that recording from neurons can be unstable [12], although some studies have shown that some individual neurons can be successfully tracked across time using chronic implants [18, 36]. However, all studies concur that not all neurons recorded from chronic implants are stable across time. Thus, arrays with low equipollence may face stability challenges if heavily contributing neurons are not reliable across time.

3. Results

3.1. Array status

The electrode arrays recorded in this study were 1–4 years old at the time of the experiments, and thus were of varying signal quality. Signal quality can be evaluted in many ways, from measuring peak-to-peak voltage of single units to counting the number of discernible spikes. However, a measure like peak-to-peak voltage only indicates the recording signal quality of the array, and is not correlated with decoder performance [12]. We were most interested in a channel's directional tuning and effect on the decoder, regardless of the voltage of the action potential measured. For the purposes of a neural prosthesis, it is the direction correlated information present in the neural signal (and not the signal trace itself) that determines the utility of the channel. For this reason, we elected to measure the contribution of each channel to the decoder as the indicator of array status and signal quality. This distribution of decoderassigned channel weights for the two monkeys is plotted in figure 2. We found that the older array in Monkey L had lower equipollence (i.e., fewer useful channels to draw from), and thus the decoder placed more weight on those few informative channels. Monkey J, on the other hand, had far more tuned channels to draw from and thus had higher equipollence (i.e., more equally distributed decoder channel contributions), where all but two channels contributed less than 2.5% to the decoder.

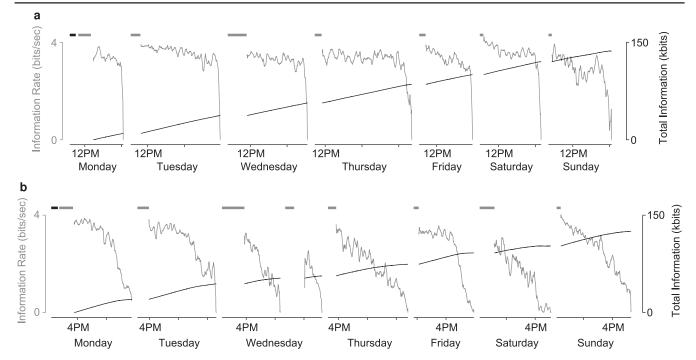


Figure 3. Decoding with retraining sessions. (a) Daily performance plot of seven contiguous days of experiments with Monkey J on the grid task under neural control. The gray line plots the instantaneous, 30 s window smoothed information rate across time. The black line plots the cumulative information transmitted across the experiment. The *x*-axis indicates the time the experiment was run, and each tick mark notes an hour of the day. The dark black bar at the top left represents time spent collecting arm-controlled training data for initializing the decoder. The gray bar at the start of each day represents the time spent for the neural prosthetic control training session. The sharp fall off represents the monkey voluntarily ending the experimental session. (b) Similar performance plot for Monkey L. Plots represent 33,563 trials with Monkey J from datasets J101018-J101024 and 40,210 trials with Monkey L from datasets L101018-L101024.

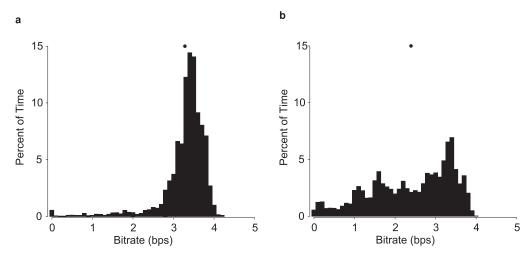


Figure 4. Performance histogram for decoding with retraining sessions. (a) Aggregate bitrate histogram for Monkey J for the seven contiguous days of experiments. This histogram contains every data point used to make the respective time-series plot. The black dot at the top represents the mean bitrate of 3.3 bps. (b) Similar histogram plot for Monkey L, mean bitrate of 2.4 bps. Plots represent the same datasets as in figure 3.

3.2. Stability with daily retraining

In exploring sustained performance, we first evaluated the extent to which performance could be repeatedly sustained with minimal daily retraining. These experiments were attempted only once with each monkey and not restarted or reset at any time. In Monkey J, these experiments were one year after implantation. In Monkey L, they were 2.5 years after implantation. In these week-long experiments, arm-

controlled training sessions were only collected on day 1. Subsequent days used the prior day's decoder to bootstrap decoder training. As the ReFIT-KF algorithm implemented a two-stage training paradigm, stage one (arm-controlled training) was skipped for days 2–7, jumping straight to the second stage of training using the prior day's decoder. Under this protocol, days 2–7 had only one training session: one with the prior day's decoder before starting the

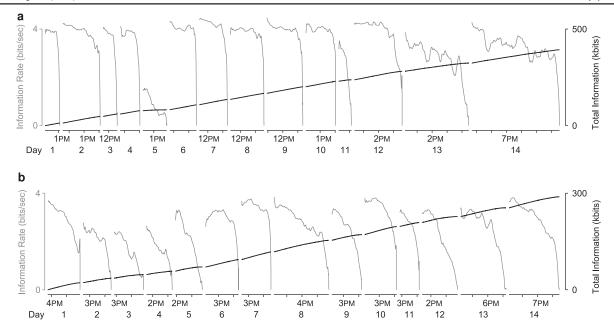


Figure 5. Decoding without retraining sessions. (a) Daily performance plot of 14 contiguous days of experiments with Monkey J. Lines and axes identical to figure 3 except for the lack of training time bars, as there was no training time in this experiment beyond the first day (not shown). (b) Similar performance plot for Monkey L. Plots represent 101,580 trials with Monkey J from datasets J110913-J109026 and 90,188 trials with Monkey L from datasets L110912-L110925.

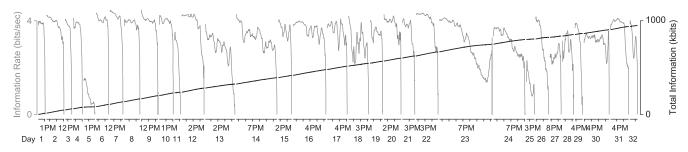


Figure 6. Month long decoding without retraining sessions with Monkey J. This plots the daily performance of the 32 contiguous days of experiments using the decoder built from the last hour of dataset J110913 (day 1 of figure 5(a)). Lines and axes identical to figure 5. The first 14 days of data is identical to that in figure 5(a). Plot represents 252,618 trials from datasets J110913-J111014.

experimental session. In this fashion, as shown in figure 3, high performance of 3-4 bps was repeatedly achieved with both monkeys across a week with daily retraining, never requiring reseeding from arm-controlled training data. The falloff seen in most experiments is behavioral. Monkeys will work for a limited time and then stop working abruptly, not unlike humans on a given task. Further, the more gradual decline in performance seen in Monkey L's performance plots is a reflection of his specific behavior and not that of the decoder's performance. Similar declines in performance are seen with Monkey L under arm controlled trials. Histograms of the aggregate data across seven days for both monkeys is shown in figure 4. The results of these initial experiments suggested that (1) decoders are stable throughout a multi-hour session (2) daily arm-controlled training is not necessary to achieve high performance, and (3) a previously built decoder can serve as the initial decoder for training on the following day.

3.3. Stability without daily retraining

The findings from the prior experiments motivated a subsequent study which minimized training times. These experiments were attempted only once with each monkey and not restarted or reset at any time. In Monkey J, these experiments were two years after implantation. In Monkey L, they were 3.5 years after implantation. In these experiments, a ReFIT-KF decoder was trained from arm-controlled trials and used until no longer functioning. Unlike in the prior experiments where retraining occurred daily, there were no online retraining sessions performed. Instead, the last hour of each day's session served as the training data for a new decoder. This led to a growing pool of decoders available for use. On a subsequent experimental day, if the prior day's decoder did not qualitatively perform well when loaded, we rapidly switched to one of the other decoders from the pool. The decoder selection process never took more than 5 min before a working decoder was found and used for the remainder of

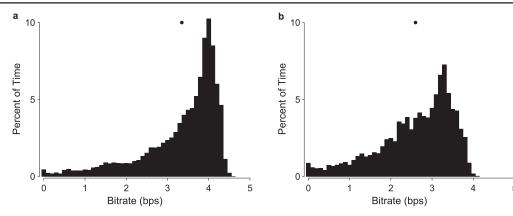


Figure 7. Performance histogram for decoding without retraining sessions. (a) Aggregate bitrate histogram for Monkey J across the 32 contiguous day experiment. This histogram contains every data point used to make the respective time-series plot. The black dot at the top represents the mean bitrate of 3.4 bps. (b) Similar histogram plot for Monkey L, mean bitrate of 2.6 bps. Plots represent the same datasets as in figures 3 and 6.

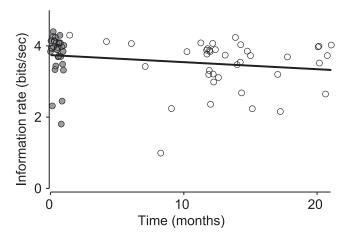


Figure 8. Long-term performance of a single decoder. Performance of Monkey J using a decoder built from the last hour of dataset J110913 (day 1 of figure 5(a)). Circles represent average performance on the day tested (minimum of 200 contiguous trials) across 22 months. The gray-filled circles denote experimental days 1-32 from figure 6). The solid line was regressed against all data points and is the equation: Information rate = $-0.24 \cdot (\text{#years}) + 3.74 \text{bps}$. Data from 70 datasets spanning J110913 to J130606.

that experimental day. Figure 5 shows the results of this daily experiment for both monkeys across two weeks. To achieve this sustained high performance, Monkey L required the use of seven different decoders over the two weeks. Likely due to his array's lower equipollence, a single decoder would not function well for more than a few days before requiring to be switched out to a more recent one. Despite this instability however, no retraining sessions were necessary to maintain performance, as the experimental data from a prior day was sufficient for building a decoder capable of achieving high performance.

This finding is in contrast to what we observed with Monkey J. His decoder was so reliable that only a single decoder was necessary: decoder J100913, the one built with the last hour of neural prosthetic control from day 1. Further,

as shown in figure 6, because of Monkey J's reliable performance, this daily experiment was continued for over a month. On day 5, Monkey J interacted with a new researcher during experiment setup. Monkey J does not handle strangers well and the stress of the encounter led to poor behavior and a low performing experimental day. Histograms of the aggregate data across 14 days with Monkey L and all 32 days with Monkey J, respectively, appear in figure 7. Using the aforementioned experimental protocol, we minimized training time in both arrays with high equipollence, as with Monkey J, and an array with low equipollence, as with Monkey L, achieving reliable high performance over the course of two weeks. This finding suggests that aging arrays even several years old can sustain high performance and should only infrequently require the collection of training data.

3.3.1. Long-term decoder stability. The finding from Monkey J's stable decoder led to a further experiment in which the performance of his decoder, J100913, was periodically rechecked to quantify the rate of performance decline over time. Figure 8 plots the performance of this decoder across the nearly two year period it was evaluated. Over the course of 22 months, the decoder had an average rate of performance decline of $0.24\,\mathrm{bps}\,\mathrm{yr}^{-1}$, which is a 6% decline in performance per year. This time corresponded to 2–4 years post-implantation. Throughout this time, it consistently achieved over 2 bps, often approaching its ceiling of around 4 bps. The decoder was not checked frequently during the 2-10 month period, but was evaluated approximately at least monthly during the 10-22 month period. All days the J100913 decoder was tested during the 22 months are presented in this plot. These evaluations typically occurred when additional experimental time was available at the end of the day for other unrelated experiments. A video demonstrating this decoder's performance across this time appears in the supplementary data (available from stacks.iop. org/JNE/11/066003/mmedia). This finding supports earlier reports [23], and suggests that threshold-based spiking

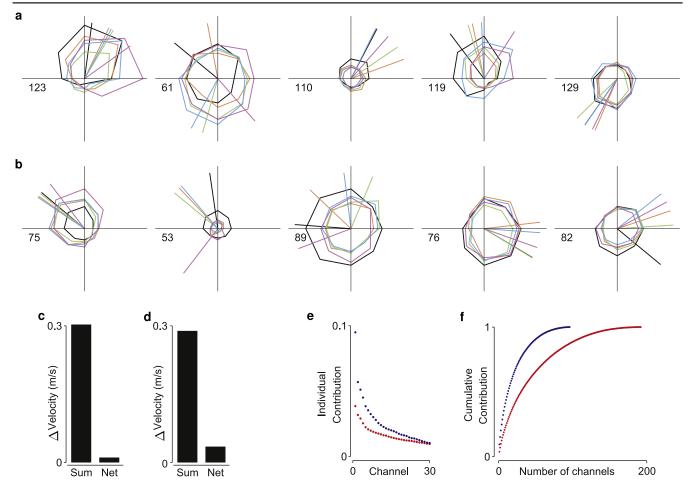


Figure 9. Stability analysis. (a) Tuning curves of top five contributory channels for a six month-old decoder for Monkey J. The number to the left of the axis denotes the channel number. Tuning was computed every $\frac{\pi}{4}$ radians and depicted as the vertices of the polygons. Each colored polygon represents a different consecutive experimental day. The black polygon represents the tuning on the day the decoder was generated (J110722). The straight lines are a channel's cosine fit preferred direction. (b) Same tuning curves as in subpanel a, but for Monkey L. The original decoder was from L110805. (c) Bar graph of perturbation velocity for Monkey J with fixed decoders. The first column plots the unsigned sum of the perturbation while the second column is the true net perturbation in velocity seen by the decoder. Sum: 0.310 m s⁻¹. Net: 0.011 m s⁻¹. (d) Same plot as in subpanel c, but for Monkey L. Sum: 0.289 m s⁻¹. Net: 0.035 m s⁻¹. (e) Plot of individual contribution of top 30 channels for Monkey J (red) and L (blue). (f) Plot of cumulative sorted per channel contribution to decoder. Data for entire figure from datasets J120117-J120120 and L120207-L120210.

activity measured from intracortical arrays can yield stable decoders over years.

3.4. Tuning analysis

We propose that this finding of robust high performance was enabled by several synergistic factors. One factor may be the high performance of the ReFIT-KF decoder that yields cursor control performance 75–85% that of the native arm, potentially facilitating a more consistent behavioral strategy [13]. Another factor may be that unlike studies utilizing individual neuron tuning and action-potential tracking [18], we simply thresholded neural data. This eliminated a large source of signal variation as action potentials were successfully captured by the threshold despite fluctuations within and across days [12]. It should be noted that throughout these experiments performance could not be reasonably attributed to learning. A functional decoder worked immediately at the

start of each day and no adjustment or accommodation time was permitted or required. Additionally, decoder performance did not increase over the course of a multi-hour day, or across days. A decoder that did not work well within the first several seconds of evaluation was halted and another decoder was immediately put in its place. Although LFP could also enable decoder stability [16, 23], we did not utilize this additional signal source in this study.

Decoders based on multielectrode arrays draw upon many channels to estimate the kinematic state. In isolation, a single channel's tuned direction appeared to fluctuate slightly from one day to the next (figures 9(a) and (b)). In analyzing a single channel, this change of tuning would result in a mismatch between the channel's new direction preference and the old direction modeled by the decoder. This mismatch would negatively impact decoder performance since that channel would be firing more strongly for a direction different than

that modeled by the decoder. Thus, the contribution of that channel would be incorrectly assigned.

However, these observed variations in channel tuning were not systematic across channels. In this task, channels projected to a 2D cursor velocity. If channel tuning variations were independent and distributed randomly with zero mean, then in aggregate, there would be no significant net expected perturbation to the decoder's performance, yielding a system that would be consistently controllable over time. The results shown in figures 9(c) and (d) suggest that this hypothesis may be valid. Absolute perturbation, shown in the first columns of figures 9(c) and (d), represents the total cumulative perturbation when the direction of change is ignored, namely, $\sum_i |P_i|$. On the other hand, net perturbation, shown in the second columns figures 9(c) and (d), preserves the direction of change and allows for the mathematical 'cancelling-out' of velocity vectors: $|\sum_i P_i|$.

When viewed in isolation, individual channel tuning changes may appear more significant than they actually are to the decoder. Whereas absolute perturbation emphasizes individual channel tuning changes and may be of more interest scientifically, net perturbation is the more appropriate measure for neural prosthesis analysis because the decoder mathematically respects vector sign and direction when calculating its final velocity. By the triangle inequality, this net perturbation is expected to be smaller than the absolute absolute perturbation. It is eight times smaller for Monkey L, but 30 times smaller for Monkey J. This difference in ratios between subjects may further help explain why a fixed decoder in Monkey J was successfully held constant over time but not in Monkey L.

Monkey J's average net perturbation appeared to be very small, whereas Monkey L's net perturbation, while still many fold smaller than his absolute perturbation, was 3.5 times larger than that of Monkey J. This disparity also suggests that the greater the number of channels, the more stable the decoder will be due to a larger number of available channels with which to 'cancel-out' fluctuations in tuning. This is further supported by figures 9(e) and (f), which demonstrate lower contribution per channel in Monkey J with two arrays compared to Monkey L with one array. These contribution curves in figures 9(e) and (f) were calculated based on the absolute value of the channel contribution to the decoder, as opposed to the net contribution measured in figures 9(c) and (d). Taken together, we concluded that the continued functioning of a fixed decoder over long periods of time is partly enabled by the shifts in tuning cancelling-out after projection to velocity.

Monkey L's array had lower equipollence. The decoder assigned greater weights to the most informative channels: four channels each had over 3% weight. This appeared to make Monkey L's decoder more susceptible to individual channel variation which did not sufficiently cancel out, partially explaining why no single decoder was stable for extended periods of time. This was in contrast to Monkey J, where no channel was assigned greater than 3% weight, and thus individual channel variations had less effect. On average, Monkey L's decoder drew 50% of its information from only a

dozen channels, whereas Monkey J's decoder distributed that same weight across 30 channels. Thus, the availability of two arrays with higher equipollence in Monkey J versus one with lower equipollence in Monkey L partially accounted for the greater fixed decoder stability observed in Monkey J.

4. Discussion

The findings in this paper support two important points. The first is to confirm and extend prior reports of sustainable performance with neural prosthetic decoders over the period of weeks to months [16, 18, 23]. This report demonstrates the highest performing (i.e., over 4 bps) sustained (i.e., 22 months) fixed neural prosthesis in one monkey. Combined with the results from prior studies, these findings suggest that under conditions of high equipollence, fixed decoders can remain functional for months to years. This finding also agrees with recent offline long-term analysis of of signal stability and decoding accuracy [12, 17].

Second, this report highlights the challenge that prosthetic systems face as signal sources degrade. In an array with low equipollence, we found performance to be variable to the point that fixed decoders are not reliable across days. Without a strategy to address this issue, new training data would have to be collected almost every day. In this paper, we addressed this in one monkey with the use of the prior day's experimental data to continually provide an 'up-to-date' decoder for a given day's experiment.

In a clinical setting, this could be accomplished by examining the data when the user is navigating to determinable targets, such as keys on a virtual keyboard or hyperlinks on a web browser. The lack of subsequent activation of a correction symbol (e.g., the backspace key in the virtual keyboard or the back button in the web browser), would imply that the user navigated to the correct target. The system would then flag that portion of movement time as a 'valid trial' (i.e. cursor movement to the known and now presumed intentionally selected target), suitable for inclusion in the training set for building that day's 'up-to-date' decoder.

Other approaches exist, such as actively adapting filters that update the decoder on the fly [37–40], or automatically adjusting the decoder to account for the daily fluctuations in the neural data [41]. One recent online study updated the decoder's tuning every two minutes in a probabilistic manner, using the decoder's output as a proxy for intention and the current decoder model as a prior for the tuning model [42]. Whatever the approach, however, it is important to recognize the added complexity that aging arrays present to sustaining performance and fixed decoders.

We note that the approach taken in this study to utilize all available channels from the array in the decoder under threshold crossings differs from prior reports of neuron stability. Prior studies have relied on careful selection of individual neurons to achieve decoder stability [18]. This hand selection of neurons from chronic arrays can yield neurons that are tracked across time [36]. However, the hand selection of neurons from chronic implants is best suited for scientific

research and may not be ideal for a clinical application [12]. The finding in this study that the gross inclusion of all available channels from a chronic array can yield stable 2D decoder performance provides a clinically-plausible approach to robust decoders. These findings suggest that a reasonable alternative to careful selection of neurons may be the mass aggregation of all channels, thereby minimizing the impact of day to day tuning variations in individual channels.

We also note that signal equipollence provides little information regarding overall decoding performance. An array with high equipollence could perform very poorly if all channels are equally poorly tuned. Similarly, an array with low equipollence could have very high decoding performance if the few channels that have weight assigned are well tuned, as is the case with Monkey L. Additionally, signal equipollence makes no a priori statement regarding decoding stability. An array with high equipollence could be very unstable from one day to the next if there are significant tuning shifts in all of the channels. Similarly, an array with low equipollence could be very stable if those few channels that are assigned significant weight by the decoder undergo little change in tuning across days. Empirically however, we did not find this to be the case in our two monkeys. We found that the monkey with higher equipollence had more stable decoders the one with lower equipollence. We hypothesize that this is because there are some changes in channel tuning from one day to the next. These changes appear to be large enough that an array with low equipollence cannot utilize a fixed decoder for several days at a time, but small enough to enable long-term decoder stability in an array with high equipollence. We note that this classification of stable versus unstable is relative. Decoders build from Monkey L's low equipollence array were nonetheless stable enough to sustain performance for one to three days at a time. They were, however, less stable than Monkey J's decoders. A link between signal equipollence and decoder stability is not immediately intuitive, need not exist, and can only be determined empirically. The findings here do suggest that this link exists, though we cannot claim this definitively with only two monkeys.

Our analysis of channel tuning suggests a hypothesis for the long-term decoder stability seen here and in prior reports. Although significant variability exists from one day to the next, this variability exists in a high-dimensional (100-200 dimensions) neural space. The decoder projects this highdimensional variability down to just two kinematic dimensions, namely x and y velocity. As a result, there is significant 'redundancy' (i.e., overlapping directional tuning) in the mappings of each neural dimension down to kinematics. We hypothesize that the redundancy present in this mapping enables the stability seen in this study. Decoders are most stable when channels have approximately comparable contributions assigned to them, thus explaining the differences in stability seen between Monkeys J and L. This hypothesis would also thus predict that decoder stability becomes more challenging to achieve as the number of kinematic dimensions decoded grows (i.e., beyond 2D control) because there will be less redundancy in the projection from neural to kinematic space. This may be a difficult issue to address as decoders expand to higher dimensions of control, such as the case of three-dimensional robotic arm control [1, 6, 14]. This is an important question for further study that may have significant bearing on the stability and translatability of higher dimensional neural prostheses.

5. Conclusion

In this study, we examined the stability of decoders in the context of differing equipollence. We found that when equipollence is high, fixed decoders appear to be sustained for long periods of time, in agreement with prior reports. We demonstrated the highest-performing, longest-lasting, fixed decoder reported to date. Monkey J's decoder (J100913) sustained communication rates of 4 bps for 22 months in duration, declining at a rate of 0.24 bps yr⁻¹. However, when equipollence was low in the case of Monkey L, a fixed decoder did not sustain performance well, and thus required an alternate strategy. This low equipollence appeared to make the decoder more susceptible to daily fluctuations from variability in direction tuning of these few critical channels. For this case, we developed a decoder protocol that required no additional training time, yet yielded daily high-performance for two weeks. However, such variability appeared to be of limited consequence to a decoder from arrays with high equipollence in Monkey J because the random nature of the changes in tuning tended to cancel one another out when projected to 2D cursor velocity. These findings suggest that even aging intracortical electrode arrays may be viable signal sources for sustainable, high performance neural prostheses that minimize retraining time—an important factor for translation to clinical use.

Acknowledgments

We thank M Mazariegos, J Aguayo, S Kang, W Kalkus, C Sherman, and E Morgan for surgical assistance and veterinary care; S Eisensee, B Davis, and E Castaneda for administrative support; D Haven and B Oskotsky for information technology support.

Funding This work was supported by the Stanford Medical Scholars Program, Howard Hughes Medical Institute Medical Research Fellows Program, Paul and Daisy Soros Fellowship, Stanford Medical Scientist Training Program (PN); National Science Foundation Graduate Research Fellowships (JMF, JCK, and SDS); Stanford Graduate Fellowship (JMF); Christopher and Dana Reeve Paralysis Foundation (SIR and KVS); National Science Foundation IGERT Grant 0734683 (SDS); and the following to KVS: Burroughs Welcome Fund Career Awards in the Biomedical Sciences, Defense Advanced Research Projects Agency Revolutionizing Prosthetics 2009 N66001-06-C-8005 and Reorganization and Plasticity to Accelerate Injury Recovery N66001-10-C-2010, US National Institutes of Health National Institute of Neurological Disorders and Stroke Collaborative Research in Computational Neuroscience Grant R01-NS054283 and Bioengineering Research Grant R01-NS064318 and Transformative Research Award T-R01NS076460, and US National Institutes of Health EUR-EKA Award R01-NS066311 and Director's Pioneer Award 8DP1HD075623.

Author contributions PN was responsible for designing and conducting experiments, data analysis, and manuscript writeup. JMF assisted in designing and conducting experiments and manuscript review. JCK assisted in conducting experiments and manuscript review. SDS assisted in conducting experiments and manuscript review. SIR was responsible for surgical implantation and assisted in manuscript review. KVS was involved in all aspects of experimentation, data review, and manuscript writeup.

References

- Taylor D M, Tillery S I H and Schwartz A B 2002 Direct cortical control of 3D neuroprosthetic devices *Science* 296 1829–32
- [2] Carmena J M, Lebedev M A, Crist R E, O'Doherty J E, Santucci D M, Dimitrov D F, Patil P G, Henriquez C S and Nicolelis M A L 2003 Learning to control a brain–machine interface for reaching and grasping by primates *PLoS Biol.* 1 193–208
- [3] Musallam S, Corneil B D, Greger B, Scherberger H and Andersen R A 2004 Cognitive control signals for neural prosthetics *Science* 305 258–62
- [4] Hochberg L R, Serruya M D, Friehs G M, Mukand J A, Saleh M, Caplan A H, Branner A, Chen D, Penn R D and Donoghue J P 2006 Neuronal ensemble control of prosthetic devices by a human with tetraplegia *Nature* 442 164–71
- [5] Santhanam G, Ryu S I, Yu B M, Afshar A and Shenoy K V 2006 A high-performance brain-computer interface *Nature* 442 195–8
- [6] Velliste M, Perel S, Spalding M C, Whitford A S and Schwartz A B 2008 Cortical control of a prosthetic arm for self-feeding *Nature* 453 1098–101
- [7] Ganguly K, Dimitrov D F, Wallis J D and Carmena J 2011 Reversible large-scale modification of cortical networks during neuroprosthetic control *Nat. Neurosci.* 14 662–7
- [8] O'Doherty J E, Lebedev M A, Ifft P J, Zhuang K Z, Shokur S, Bleuler H and Nicolelis M A L 2011 Active tactile exploration using a brain-machine-brain interface *Nature* 479 228–31
- [9] Ethier C, Oby E R, Bauman M J and Miller L E 2012 Restoration of grasp following paralysis through braincontrolled stimulation of muscles *Nature* 485 368–71
- [10] Hochberg L R et al 2012 Reach and grasp by people with tetraplegia using a neurally controlled robotic arm Nature 485 372–5
- [11] Ifft P J, Shokur S, Li Z, Lebedev M A and Nicolelis M A L 2013 A brain–machine interface enables bimanual arm movements in monkeys Sci. Transl. Med. 5 210ra154
- [12] Chestek C A et al 2011 Long-term stability of neural prosthetic control signals from silicon cortical arrays in rhesus macaque motor cortex J. Neural Eng. 8 045005
- [13] Gilja V, Nuyujukian P, Chestek C A, Cunningham J P, Yu B M, Fan J M, Kao J C, Ryu S I and Shenoy K V 2012 A high-performance neural prosthesis enabled by control algorithm design *Nat. Neurosci.* 15 1752–7
- [14] Collinger J L, Wodlinger B, Downey J E, Wang W, Tyler-Kabara E C, Weber D J, McMorland A J C, Velliste M, Boninger M L and Schwartz A B 2013 High-

- performance neuroprosthetic control by an individual with tetraplegia *Lancet* **381** 557-64
- [15] Green A M and Kalaska J F 2011 Learning to move machines with the mind *Trends Neurosci.* 34 61–75
- [16] Blakely T, Miller K J, Zanos S P, Rao R P N and Ojemann J G 2009 Robust, long-term control of an electrocorticographic brain-computer interface with fixed parameters *Neurosurg*. Focus 27 E13
- [17] Chao Z C, Nagasaka Y and Fujii N 2010 Long-term asynchronous decoding of arm motion using electrocorticographic signals in monkeys *Front. Neuroeng.* 3 3
- [18] Ganguly K and Carmena J M 2009 Emergence of a stable cortical map for neuroprosthetic control *PLoS Biol.* 7 e1000153
- [19] Simeral J D, Kim S P, Black M J, Donoghue J P and Hochberg L R 2011 Neural control of cursor trajectory and click by a human with tetraplegia 1000 days after implant of an intracortical microelectrode array J. Neural Eng. 8 025027
- [20] Flint R D, Lindberg E W, Jordan L R, Miller L E and Slutzky M W 2012 Accurate decoding of reaching movements from field potentials in the absence of spikes J. Neural Eng. 9 046006
- [21] So K, Dangi S, Orsborn A L, Gastpar M C and Carmena J M 2014 Subject-specific modulation of local field potential spectral power during brain-machine interface control in primates J. Neural Eng. 11 026002
- [22] Perge J A, Zhang S, Malik W Q, Homer M L, Cash S, Friehs G, Eskandar E N, Donoghue J P and Hochberg L R 2014 Reliability of directional information in unsorted spikes and local field potentials recorded in human motor cortex J. Neural Eng. 11 046007
- [23] Flint R D, Wright Z A, Scheid M R and Slutzky M W 2013 Long term, stable brain machine interface performance using local field potentials and multiunit spikes J. Neural Eng. 10 056005
- [24] Chestek C A, Batista A P, Santhanam G, Yu B M, Afshar A, Cunningham J P, Gilja V, Ryu S I, Churchland M M and Shenoy K V 2007 Single-neuron stability during repeated reaching in macaque premotor cortex *J. Neurosci.* 27 10742–50
- [25] Fan J M, Nuyujukian P, Kao J C, Chestek C A, Ryu S I and Shenoy K V 2014 Intention estimation in brain–machine interfaces J. Neural Eng. 11 016004
- [26] Nuyujukian P, Fan J M, Gilja V, Kalanithi P S, Chestek C A and Shenoy K V 2011 Monkey models for brain–machine interfaces: the need for maintaining diversity Conf. Proc. IEEE Eng. Med. Biol. Soc. 2011 1301–5
- [27] Davoodi R and Loeb G E 2012 Real-time animation software for customized training to use motor prosthetic systems *IEEE Trans. Neural Syst. Rehabil. Eng.* 20 134–42
- [28] Cunningham J P, Nuyujukian P, Gilja V, Chestek C A, Ryu S I and Shenoy K V 2011 A closed-loop human simulator for investigating the role of feedback control in brain-machine interfaces J. Neurophysiol. 105 1932–49
- [29] Nuyujukian P, Fan J M, Kao J C, Ryu S I and Shenoy K V 2014 A high-performance keyboard neural prosthesis enabled by task optimization *IEEE Trans. Biomed. Eng.* at press doi:10.1109/TMBE.2014.2354697
- [30] Kim S, Simeral J D, Hochberg L R, Donoghue J P and Black M J 2008 Neural control of computer cursor velocity by decoding motor cortical spiking activity in humans with tetraplegia J. Neural Eng. 5 455–76
- [31] Mulliken G H, Musallam S and Andersen R A 2008 Decoding trajectories from posterior parietal cortex ensembles J. Neurosci. 28 12913–26
- [32] Wolpaw J R, Ramoser H, McFarland D J and Pfurtscheller G 1998 Eeg-based communication: improved accuracy by response verification *IEEE Trans. Rehabil. Eng.* 6 326–33

- [33] Bin G, Gao X, Wang Y, Li Y, Hong B and Gao S 2011 A high-speed BCI based on code modulation VEP J. Neural Eng. 8 025015
- [34] Hwang H J, Lim J H, Jung Y J, Choi H, Lee S W and Im C H 2012 Development of an SSVEP-based BCI spelling system adopting a qwerty-style LED keyboard J. Neurosci. Methods 208 59–65
- [35] Yuan P, Gao X, Allison B, Wang Y, Bin G and Gao S 2013 A study of the existing problems of estimating the information transfer rate in online brain–computer interfaces *J. Neural Eng.* **10** 026014
- [36] Fraser G W and Schwartz A B 2012 Recording from the same neurons chronically in motor cortex J. Neurophysiol. 107 1970–8
- [37] Orsborn A L, Dangi S, Moorman H G and Carmena J M 2011 Exploring time-scales of closed-loop decoder adaptation in brain-machine interfaces *Conf. Proc. IEEE Eng. Med. Biol.* Soc. 2011 5436–9
- [38] Orsborn A, Dangi S, Moorman H and Carmena J 2012 Closedloop decoder adaptation on intermediate time-scales

- facilitates rapid bmi performance improvements independent of decoder initialization conditions *IEEE Trans. Neural Syst. Rehabil. Eng.* **20** 468–77
- [39] Dangi S, Gowda S and Carmena J 2013 Likelihood gradient ascent (lga): a closed-loop decoder adaptation algorithm for brain-machine interfaces Engineering in Medicine and Biology Society (EMBC) -35th Ann. Int. Con. of the IEEE 2013) pp 2768-71
- [40] Dangi S, Orsborn A L, Moorman H G and Carmena J M 2013 Design and analysis of closed-loop decoder adaptation algorithms for brain–machine interfaces *Neural Comput.* 25 1693–731
- [41] Bishop W, Chestek C C, Gilja V, Nuyujukian P, Foster J D, Ryu S I, Shenoy K V and Yu B M 2014 Self-recalibrating classifiers for intracortical brain–computer interfaces J. Neural Eng. 11 026001
- [42] Zheng L, O'Doherty J E, Lebedev M A and Nicolelis M A L 2011 Adaptive decoding for brain-machine interfaces through Bayesian parameter updates *Neural Comput.* 23 3162–204