Neural correlates of learning in a linear discriminant analysis brain-computer interface paradigm

Yu Tung Lo[1], Brian Premchand[2], Camilo Libedinsky[3,4], and Rosa Qi Yue So[2]

[1] Department of Neurosurgery, National Neuroscience Institute, 11 Jalan Tan Tock Seng, Singapore 308433
[2] Institute for Infocomm Research (I^2R), A*STAR, 1 Fusionopolis Way, #21-01 Connexis (South Tower), Singapore 138632
[3] Department of Psychology, National University of Singapore, Singapore
[4] Institute of Molecular and Cell Biology, A*STAR, Singapore

E-mail: Yu Tung Lo, jacklo@nus.edu.sg; Rosa So, rosa-so@i2r.a-star.edu.sg

Abstract

Objective: With practice, the control of brain-computer interfaces (BCI) would improve over time; the neural correlate for such learning had not been well studied. We demonstrated here that monkeys controlling a motor BCI using a linear discriminant analysis (LDA) decoder could learn to make the firing patterns of the recorded neurons more distinct over a short period of time for different output classes to improve task performance. *Approach:* Using an LDA decoder, we studied two Macaque monkeys implanted with microelectrode arrays as they controlled the movement of a mobile robotic platform. The LDA decoder mapped high-dimensional neuronal firing patterns linearly onto a lower-dimensional linear discriminant (LD) space, and we studied the changes in the spatial coordinates of these neural signals in the LD space over time, and their correspondence to trial performance. Direction selectivity was quantified with permutation feature importance (FI). *Main results:* We observed that, within individual sessions, there was a tendency for the points in the LD space encoding different directions to diverge, leading to fewer misclassification errors, and, hence, improvement in task accuracy. Accuracy was correlated with the presence of channels with strong directional preference (i.e. high FI), as well as a varied population code (i.e. high variance in FI distribution). *Significance:* We emphasized the importance of studying the short-term / intra-sessional variations in neural representations during the use of BCI. Over the course of individual sessions, both monkeys could modulate their neural activities to create increasingly distinct neural representations.

Keywords: brain-computer interface; neural implant; BCI; BMI; neuroprosthesis; cortical implant

1. Introduction

Brain-computer interface (BCI) technology has seen major advances in the twenty-first century [1–3], particularly with BCI implanted in the motor cortex for movement control. For example, we previously demonstrated that a non-human primate could wirelessly control a mobile robotic platform through a BCI driven by neural recordings [4]. In the brain, precise motor planning utilizes a highly complex and distributed system that involves the participation of multiple brain areas such as the primary and supplementary motor cortices, basal ganglia, cerebellum, and the brainstem [5–7]. BCI systems only sample a very small subpopulation of this complex network of neurons, commonly those in the primary motor cortex (M1) and sometimes the supplementary motor area (SMA) [8,9]. Decoder training also depends on the user and the decoder acquiring a set of neural patterns that are sufficiently mutually distinguishable across the range of output commands, which can be a challenge, especially during the initial phase of BCI learning.

While neural activities conveyed by multi-unit spikes or local field potentials [10–12] and their applications in BCI [8,9,13,14] have been richly researched, the changes in neural firing patterns during BCI learning are less well understood.

Authors have described brain control as skill-learning [15] involving a wide ensemble of brain regions [16]. Functional cortical reorganization also occurs with BCI use, with changes in the contribution of individual neurons to model performance [8].

In BCI paradigms, the measured neural activities are the sole output to the end effector; these neural activities are therefore deterministically and causally linked to the behavior of the end effector. Thus, given a fixed decoder, improvement in decoding accuracy will require the patterns of neural activities to change across trials, guided by real-time feedback from the BCI output. For instance, Fetz demonstrated the adaptability of the neural population in the 1960s by showing that neurons' firing patterns can be cognitively modulated to perform certain tasks [17].

This study comprised a series of experiments performed on two macaque monkeys implanted with microelectrode arrays, employing a linear discriminant analysis (LDA) decoder. BCI users often improve in performance over time with the use of the BCI [5,18]. Despite that, the short-term (i.e. intra-sessional) temporal structure of neural firing patterns and their neural correlates have not been explored in depth by existing studies. Our results provide a more fundamental understanding of the neurons' behavior during BCI control.

We hypothesized that the neural firing patterns do not remain homogenous or stationary within trial blocks. As the neural representations of the recorded neurons comprise the sole input to the decoder, the BCI user would need to gradually calibrate the neural firing pattern to bring about an eventual improvement in decoding accuracy. In these monkey subjects, we investigated the changes in the neural firing patterns as the monkeys improved in their control of the BCI over the course of the individual trial sessions. We also discussed how such short-term learning corresponded to changes in the mapped features in the low-dimensional linear discriminant space as the trials progressed.

2. Materials and Methods

2.1 Subject and hardware

Two adult male macaque monkeys (*Macaca fascicularis*) were trained to control a commercial mobile robotic platform (Adept MobileRobots LLC, USA) using a wireless cortical BCI, which recorded multi-unit activities in terms of the rate of threshold-crossing spikes. Our protocol and setup have been described in a previously published study [1]. Briefly, multiple electrode arrays with a total of 96 and 128 channels were implanted in the upper limb areas of the primary motor cortex for both monkeys (Monkey 1: 2 arrays of 16 electrodes each and 2 arrays of 32 electrodes each; Monkey 2: 4 arrays of 32 electrodes each). The arrays used were made of platinum-iridium leads with 400 μm separation, 1 to 1.5mm length, and 0.5 MΩ impedance; they were arranged in 4x4 or 8x4 grids. For this study, we retrospectively performed offline analyses of the recorded data from three different days spanned over two months (for Monkey 1, the multielectrode array was implanted around 1 year 5 months to 1 year 7 months prior to these sessions; for Monkey 2, the array was implanted around 8 to 10 months prior; see **Supplementary Table 1** for the specific sequences of these blocks). Spikes were detected if they exceeded the threshold defined by the following formula

$$Threshold = 5\sigma; \sigma = median\{|x|/0.6745\}$$

where x is the filtered signal, and σ is an estimate of the standard deviation of the background noise.

It should also be noted that the spikes used in the current study had not been spike-sorted (and in many cases, individual neurons could not be identified). The number of channels used as features was typically around 20 (they differed slightly between trial blocks).

Animal procedures were performed in compliance with the standards of the Agri-Food and Veterinary Authority of Singapore and the Singapore Health Services Institutional Animal Care and Use Committee (Singhealth IACUC #2012/SHS/757) and followed the recommendations in the Guidelines for the Care and Use of Mammals in Neuroscience and Behavioral Research (National Academies Press, 2003). Sample neural recordings of both monkeys for these sessions are available online at https://osf.io/dce96/.

2.1 Experimental set-up

In this paper, we used the term "session" or training "day" to describe the collection of experiments within a single day. Each day was divided into "blocks", which in turn consisted of "trials". The task involved the monkey controlling a movable robotic platform interfaced through a wireless BCI system. Initially for training, monkeys controlled the platform using an attached joystick. Neural spikes recorded during these sessions were used to train an LDA decoder ("Initial Decoder"). Next, this Initial Decoder was used for BCI control of the platform. To assist the monkeys in familiarizing themselves with BCI control, trials were initially performed with 50% or 60% assisted control ("assistance"), where 50% or 60% (respectively) of the issued commands by the decoder were matched to the correct directions, irrespective of the actual decoded commands. The neural activities recorded during these blocks were in turn used to train a second decoder ("Recalibrated Decoder"). Once the monkeys achieved a satisfactory level of performance (>70% success rate), they performed trials with full BCI control using the Recalibrated Decoder (i.e. with no assistance). Only these Recalibrated Decoders were analyzed for the purpose of this paper; furthermore, we focused on the short-term (within trial

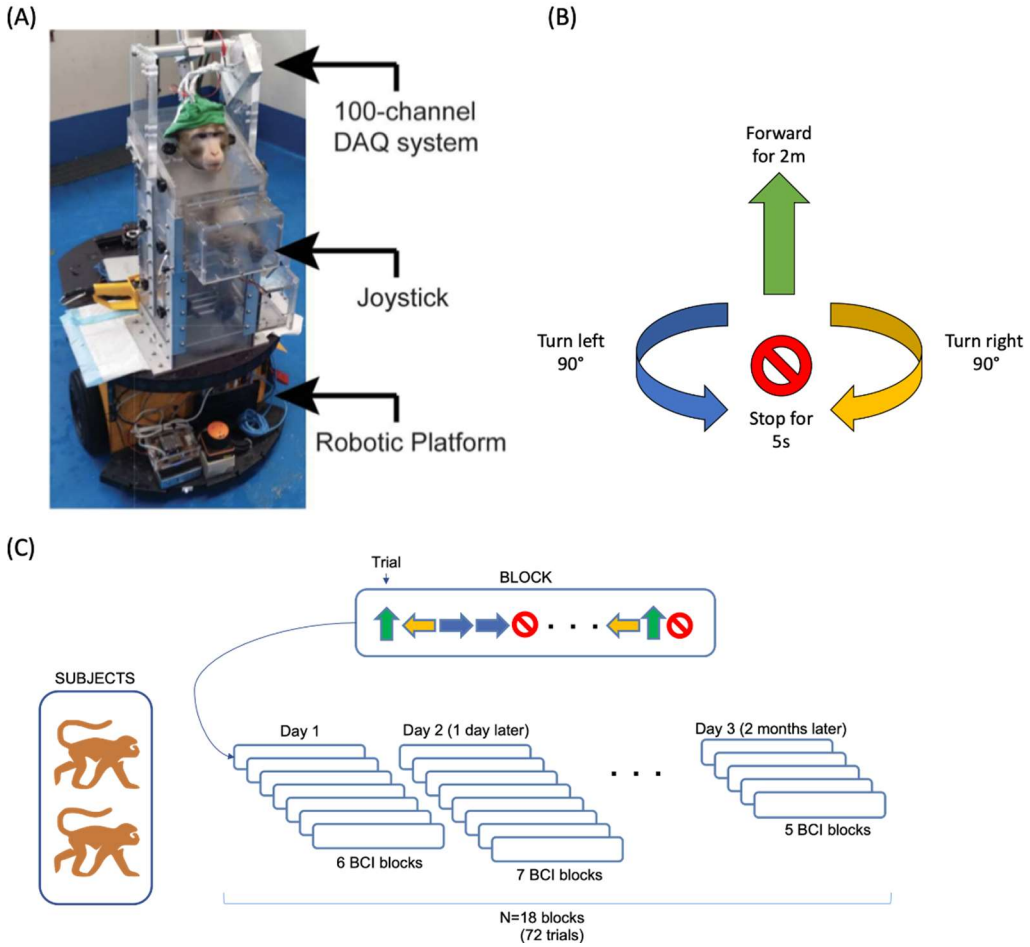

Figure 1. Trial design. (A) A macaque positioned in a mobile robotic platform (image adapted from Libedinsky et al 2016). The two monkeys were respectively implanted 1 year 5 months to 1 year 7 months, and 8 to 10 months prior to these sessions. (B) Four different commands were used to control the platform: turn left 90°, turn right 90°, move forward for 2 m, and stop for 5 s. (C) The sequence of the trial blocks used in the current analysis.

sessions) learning effect. The trial block sequence is described in **Supplementary Table 1**.

For each trial, the monkey was required to produce four different commands ("directions"): move forward for 2 m ("forward"), turn left 90° ("left"), turn right 90° ("right"), or stop for at least 5 seconds ("stop") (**Figure 1**). A trial was considered successful if the monkey reached the target location within 15 s for "forward"/ "left"/ "right" trials, or remained stationary for 5 s for "stop" trials. Food rewards were given to the animal upon successful completion of each task. The monkeys were trained to perform these tasks well using joysticks prior to the BCI sessions. The mean firing rates (MFRs) for each channel were calculated by summing the total number of neural spikes within a moving 500 ms window, sampled every 100 ms.

2.3 Decoding algorithms

The MFRs for each channel were used as input features to the LDA decoder, with the presented direction (or "stop") as the target variables (i.e. the number of features equalled the number of channels with detectable spikes). The MFRs were mapped onto a linear discriminant subspace constituted by two-dimensional axes representing the linear discriminants with the two highest eigenvalues. The centroid (i.e. the vector mean) of the last 100 sample points was recalculated every 100 ms and plotted as time series, for each of the presented directions.

We quantified the increase in cluster separability by taking the mean of the six Euclidean distances between the four cluster centroids (inter-cluster distance) at each instance in time, and trending these means over time. As the values of the linear discriminants were in arbitrary units and differed from one LDA decoder to the next (some trained on different dataset), the dimensions of LD feature space were normalized (each point coordinate mean-subtracted and then divided by the standard deviation of all constituent points) to allow pooling across different decoder feature space. Moreover, as there were variable numbers of trials for different directions, and each trial took place over a variable duration of time, the trial durations had to be standardized to a range of 0 to 1, and

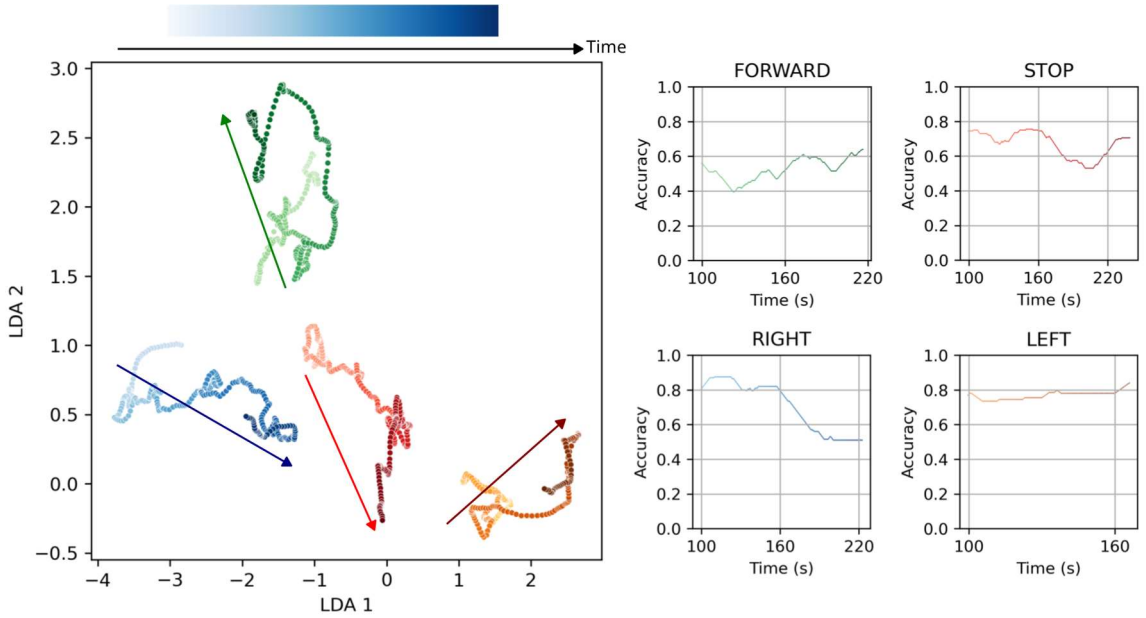

Figure 2. Divergence of points within the linear discriminant (LD) space over the span of one example trial block (Monkey 1) and their associated accuracy for the various directions. (A) Time series of the features in the two-component LD subspace. Each dot represents a centroid of all the points in 100 time-steps. The time component is represented with progressively darker shades (in their respective colors). (B) Decoding accuracy improved over time for the 'forward' direction. This was correlated with increased separation of the cluster over time, as seen in (A). For the 'right' direction, the cluster centroids drifted towards the other clusters, and hence correlated with progressively lower accuracy. The directions are color-coded identically in (A) and (B), with green representing 'forward', red representing 'stop', blue representing 'right' and yellow representing 'left'. The animated version of this plot can be found in *Supplementary Figure 3*.

then discretized into 20 equal-width bins for subsequent pooling and statistical analyses.

We performed further offline analyses using other common representative classification algorithms, namely: logistic regression (LR), support vector classifier (SVC), Naive Bayes classifier (NB), and K-Nearest Neighbor (KNN). Nevertheless, LDA as a dimension reduction algorithm, unlike the others, remained useful for the visualization of the increase in the distinctiveness of neural representation over time.

2.4 Accuracy measure

We defined the decoding accuracy as the number of correctly decoded commands (i.e. those that matched the target direction) as a proportion of the total number of decoded commands. For each block, the moving averages of accuracy were calculated over every consecutive 200 time-steps (each time-step encompassing 500 ms and updated every 100 ms and sent to the robotic platform) and graphed as a smoothed time series.

2.5 Direction preference

We derived feature importance (FI) for each of the neural channels using a permutation importance-based method [19] – the MFRs for each neural channel were randomly shuffled one channel at a time across the entire trial block (and thereby decoupling them from the target direction), keeping the other channels unchanged. Next, the mean accuracy was recalculated from this shuffled dataset. The differences between the originally achieved accuracies and the new accuracies from the shuffled dataset indicated the FIs of the respective channels (i.e. how much they contributed to the decoding accuracy). This process was boot-strapped 100 times (with different shuffles), and the mean differences in FI were used.

2.6 Statistical analysis

The training and live implementation of the decoder were performed in MATLAB R2018b 0 (Mathworks Inc, Massachusetts, USA). *Post hoc* analysis and visualization of the collected data were performed in *Python* version 3.7.1 and implemented in *Jupyter Notebook* version 5.7.4. T-tests were used for statistical comparisons. The rank-based Mann-Kendall test (via the *pyMannKendall* package) was used to test for statistical significance in time trends [20]. For correlation analyses, quadratic curves were fitted and R^2 were used to quantify the proportion of the variation in the dependent variable that was predictable from the independent variable.

3. Results

A total of 18 trial blocks for two monkeys over three days were included for analysis (12 blocks for Monkey 1 and 6 blocks for Monkey 2). The two-component LDA yielded an explained variance of 0.843±0.024 (mean ± SD) across all trial blocks.

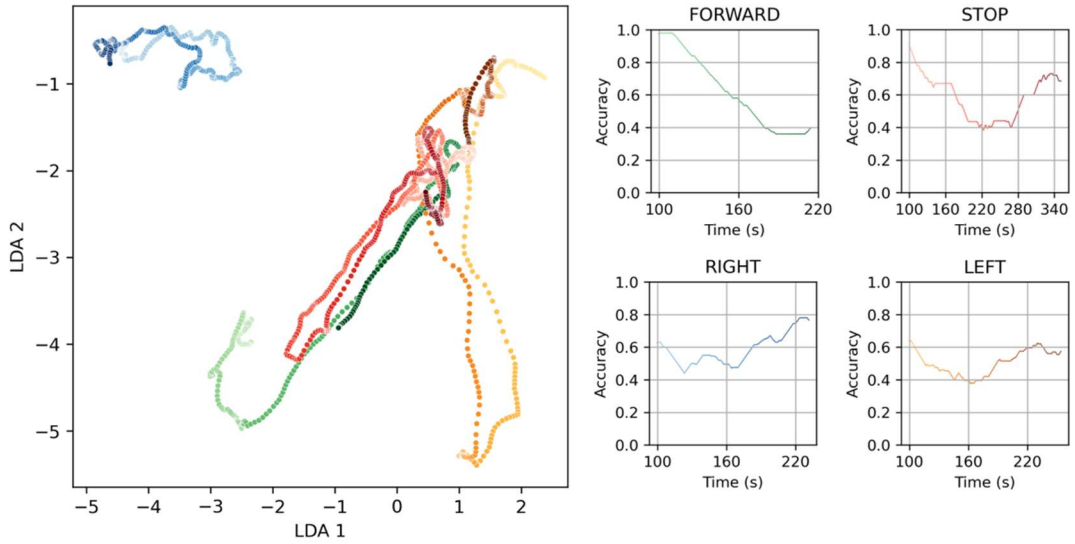

Figure 3. The reduction in decoding accuracy caused by a convergence of feature points within the linear discriminant space. In this example block (Monkey 1), only the right direction (blue) could be accurately decoded, as neural activities for the other directions progressively became more similar and mapped to the same regions in the linear discriminant space. The performance for the forward direction deteriorated the most, although the stop and left directions eventually diverged and the performance recovered. The directions are color-coded identically in (A) and (B), with green representing 'forward', red representing 'stop', blue representing 'right' and yellow representing 'left'. The animated version of this plot can be found in *Supplementary Figure 4*.

3.1 Within-session learning effect

Figure 2 shows the movement of point centroids over time in the LD feature space for a sample block. With the same decoder trained on a previous trial block (Monkey 1 Day-575 Block 5, trained on Block 4 data on the same day, s ee **Supplementary Table 1**), the monkey in this trial block achieved progressive improvement in the forward direction, which can be visualized as the spatial divergence of the feature points within the LD space over time (**Figure 2A**); however, feature points representing the 'right' direction converged in the LD space (particularly with the 'stop' command), with an associated decline in decoder accuracy over time. This suggested that the monkeys were able to make progressive adjustments online using feedback to improve task accuracy. In many other trial blocks, such a pattern of divergence in the linear discriminant space could be observed with correlated improvement in decoding accuracy.

Furthermore, we performed additional analyses using trials where the monkeys performed the same tasks using joysticks with their natural limb ("Joystick trials"), instead of the implanted BCI. Divergence of LD features was not observed (**Supplementary Figure 2**).

3.2 Performance reduction as point convergence in LD space

Conversely, decreases in decoding accuracy were associated with the convergence of points within the LD space. Other than the 'right' direction in **Figure 2**, another example of such a trial is shown in **Figure 3** (Monkey 1 Day-576 Block 11, trained on Block 2 data on the same day; **Supplementary Table 1**); here, only the 'forward' direction remained well separated from the three other directions ('left', 'right', and 'stop'), as the latter displayed similar neural firing patterns and hence were mapped to the same regions in the LD space, with consequently large misclassification errors.

3.3 Well-learned firing patterns

Finally, analyses of later sessions conducted two months later, when the monkeys were more accustomed to using the BCI, revealed stable patterns of firing, with spatially distinct clusters and little movement of the centroids over time within the LD space (**Figure 4**) (Monkey 2 Day-296 Block 15, trained on Block 14 on the same day; **Supplementary Table 1**).

A total of 18 blocks performed over three days, were available for both monkeys (**Figure 1**). The results were pooled and the average trends plotted in **Figure 5**. Over time across all trials within each block, there was an overall increase in average Euclidean distances between clusters and performance accuracy ($p=0.021$ and $p<0.005$ respectively, using the Mann-Kendall test for time series; **Figure 5B and C**).

3.4 Directional preferences of neural units

The direction selectivities of channels were next quantified with permutation feature importance (FI) by shuffling the neural signals one channel at a time, to dissociate the firing

patterns and the target direction. **Figure 6** demonstrates the relationship between decoder accuracy and these FIs.

There were significant variabilities in the FIs (i.e. direction preference) between trial blocks (**Figure 6A**). Within a trial block, if the firing patterns of a specific channel were poorly distinguishable between the four directions, then random permutations (which shuffled and decoupled the MFRs across all these directions) would have minor impacts on the decoding accuracy. Conversely, if a channel preferentially fired in one or more directions, then random permutations of the MFRs of that channel would lead to a large loss in accuracy, resulting in high permutation FI (**Figure 6A**). It should also be noted that random permutations for some channels actually led to an improvement in decoding accuracy, indicating a low signal-to-noise ratio in these channels (i.e. contributing more noise than signals).

Across all 72 trials, higher permutation FI of the most influential channel (i.e. maximal FI) correlated with higher trial accuracy, and over a quarter of the variance in trial accuracy can be explained by the maximal FI (R^2 of 0.28, **Figure 6C**).

The variance in the distribution of FI across the channels was also positively correlated with the decoding accuracy; with an R^2 of 0.18 (**Figure 6D**). In other words, a more varied distribution of neural representations across the channels on the array (as measured by the variance) accounted for 18% of the variability in trial accuracy.

3.5 Comparing other machine learning algorithms

We tested other common classifier algorithms using the same train-test datasets, which included: logistic regression (LR), support vector classifier (SVC), K-nearest neighbor (KNN), and naïve Bayes classifier (NB). All algorithms demonstrated a progressive increase in accuracy as the trials progressed (**Figure 7**).

To further support our findings, we recalculated the FI by setting the respective channel's MFRs to zero, one channel at a time, in lieu of the random permutation method. In addition, the magnitudes of the logistic regression coefficients were also used as an alternative to FI calculated from both the aforementioned methods. The maximal FI and the variance in FI (or, analogously, the logistic regression coefficient magnitudes and their variance) were similarly correlated with trial accuracy, as in our primary analyses (**Supplementary Figure 1**).

4. Discussion

We showed here that during BCI control, the neuronal activities of the recorded neurons would become more mutually distinct to bring about an improvement in decoding accuracy. This divergence in neural representation was not observed when the neurons were recorded but did not contribute to BCI output (i.e. during joystick-controlled trials). Understanding the underlying brain-decoder interactions during BCI learning is important in guiding the future design, implementation, and monitoring of BCI usage in real-world applications. This study characterized the changes in the firing rates over different periods of BCI

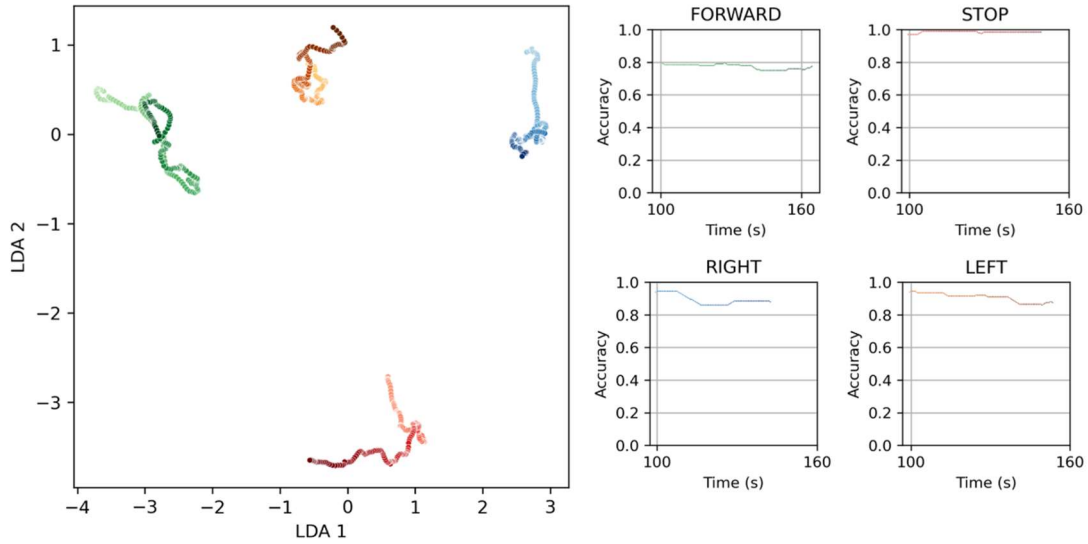

Figure 4. An example of more stable linear discriminant features in a later trial in Monkey 2. The high performance was maintained throughout the entire trial, with very stable clusters with minimal movement within the linear discriminant space, signifying that there was minimal variations in the underlying neural firing patterns across the session. The directions are color-coded identically in (A) and (B), with green representing 'forward', red representing 'stop', blue representing 'right' and yellow representing 'left'. The animated version of this plot can be found in *Supplementary Figure 5.*

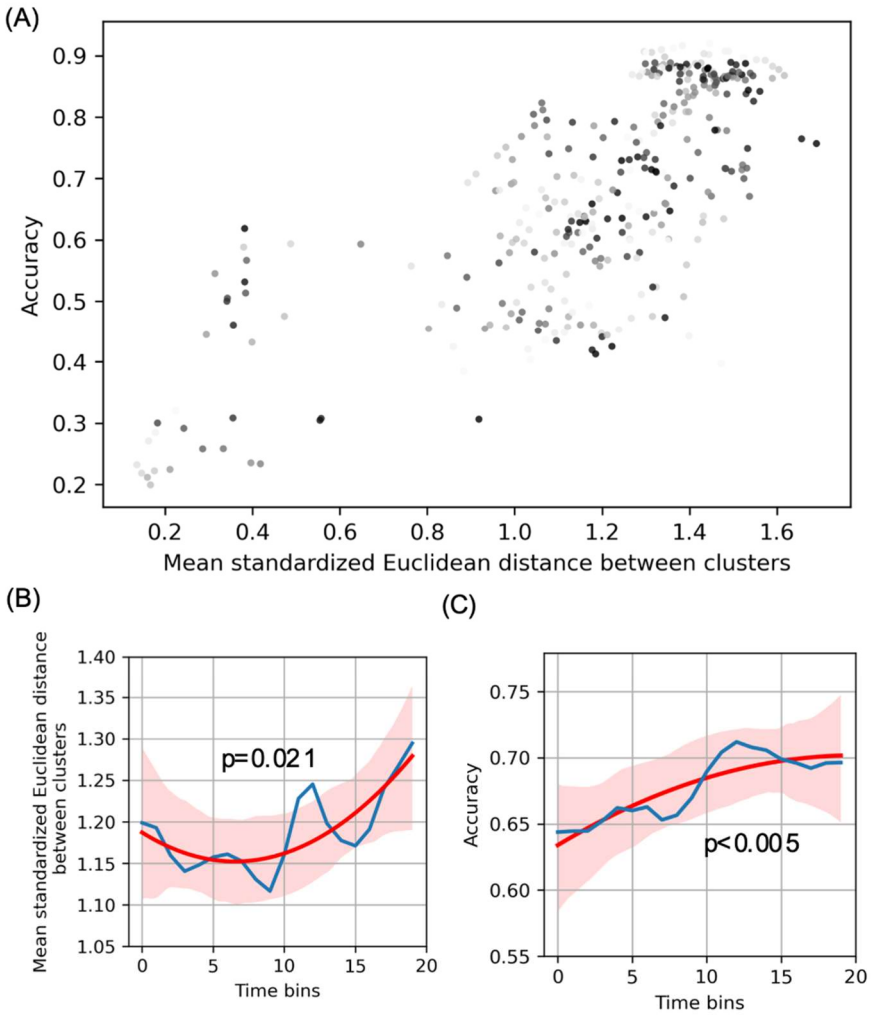

Figure 5. Correlations between trial accuracy, inter-cluster distances, over the course of individual trial blocks. (A) The aggregate plot (for all 18 trials for 2 monkeys, on 3 different days) of performance accuracy against the mean standardized Euclidean distance between the four clusters. The shade of each point indicates time, with lighter shades representing the early timepoints, and darker shades representing the later timepoints (for a total of 20 equal-sized bins). Over time, there was a general increase in performance accuracy as well as the average Euclidean distance between clusters. (B) Plot of the mean standardized Euclidean distance between clusters against time (across 20 bins). (C) Plot of trial accuracy against time (across 20 bins). For both (B) and (C), the blue line plots represent the average of the data points for each of the 20 time bins, and quadratic curves (red curves) were fitted over the data points with 100-fold bootstrapping to generate the 95% confidence interval bands (red bands). Time series trends were tested using the Mann-Kendall test, yielding $p = 0.021$ for increasing trend for the mean standardized Euclidean distance between clusters over time (B), and $p < 0.005$ for accuracy over time (C).

training and across two monkeys by studying the corresponding changes in the LD feature space and the permutation FIs (a measure of direction specificity) of the recorded neural channels. We observed that the monkeys gradually exerted more accurate control over the BCI as the trial block progressed, in the form of greater divergence of the points representing different task directions in the LD space. In addition, while the variations in trial accuracy could be influenced by a plethora of factors, we observed that better decoder accuracy was associated with channels possessing stronger directional selectivity (i.e., higher maximal permutation FI) as well as the more varied directional selectivities across the ensemble, which respectively accounted for 28% and 18% of the variations in trial accuracy (as measured by the R^2).

How neurons behave under BCI control is still poorly understood, as BCI use is an unnatural situation in which cortical neurons bypassed the complex neural mechanisms that modulate movement control to directly bring about a motor output (physically or virtually e.g. cursor control). Many studies to date have focused on the end results of BCI control, and less emphasis on the learning process and its neural substrate. While for non-invasive BCI, some work had been done to study the large changes in brain activities during

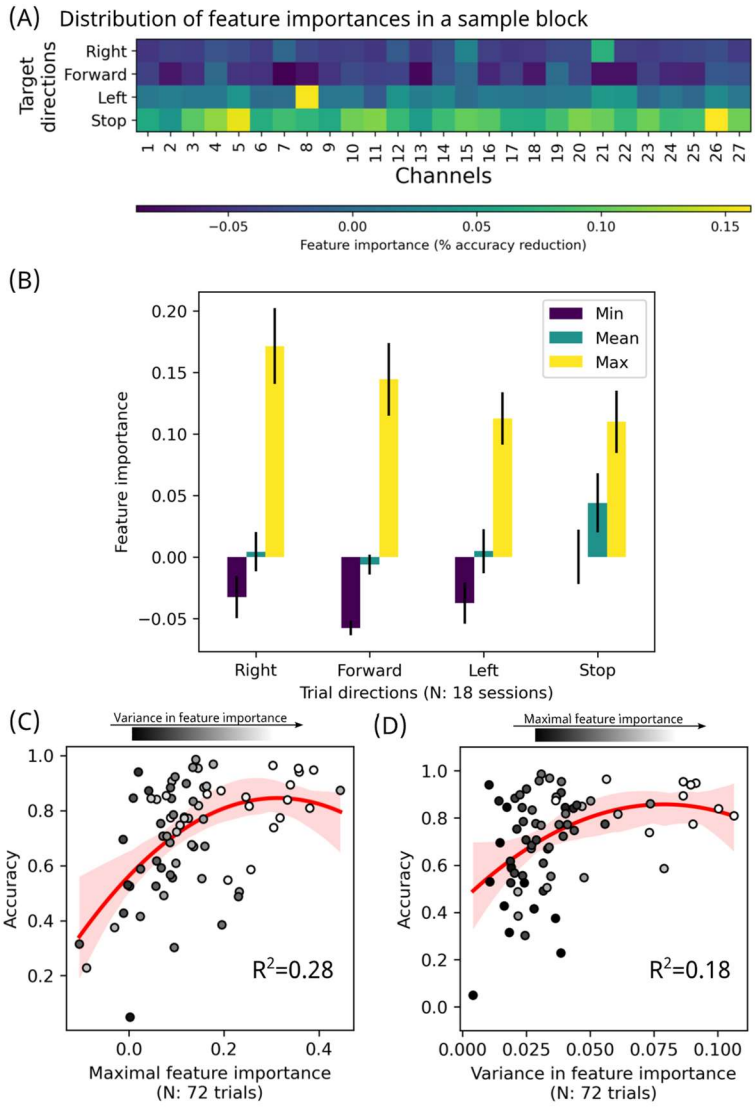

Figure 6. Analyses of permutation feature importance. (A) The permutation feature importances (FI) for all recorded channels in their respective directions for a sample trial block. The channels were randomly shuffled one at a time and the difference in the decoding accuracy from the actual decoding accuracy denoted the feature importance; this random shuffling was repeated 100 times for each channel, the average was taken. (B) Bar graphs summarizing the means and standard errors of means for the minimum, maximum and mean permutation importance across all trials, for each of the four commands. (C) Across all 72 trials, there was a positive correlation between directions with higher maximal FI (i.e., more selective channels) and higher decoding accuracy. The red line represents the best-fit quadratic curve with an R2 of 0.28. The lighter red band represents the 95% confidence intervals for the best-fit curve based on a 100-fold bootstrap. The data points are further shaded by the variance in feature importance across all channels for each trial, with a spectrum from black (low variance) to white (high variance). (D) There was also a positive correlation between trial accuracy and the variance in permutation FI across the channels, for each of the 72 trials. The red line represents the best-fit quadratic curve with an R2 of 0.18. The lighter red band represents the 95% confidence intervals for the best-fit curve based on a 100-fold bootstrap. The data points are further shaded by the maximal FI across all channels for each trial, with a spectrum from black (low maximal FI) to white (high maximal FI). For (C) and (D), the same plots were repeated using FIs calculated by setting the mean firing rates to zero for each channel in turn (instead of random permutation), as well as using logistic regression coefficients to replace the permutation FI; the results are shown in Supplementary Figure 1.

BCI control tasks [21], studies that involved intracortical electrodes so far had not explicitly discussed the learning process in any detail [22–26]. Of note, Colachis *et al*. similarly found that movement with the most distinct neural representation led to the highest accuracy, and vice versa, although the authors stopped short of commenting on whether the neural representations would become more distinct over time [27]. Understanding the neural substrate underlying such learning of BCI control would be especially crucial when planning BCI use that involves higher degree-of-freedom task

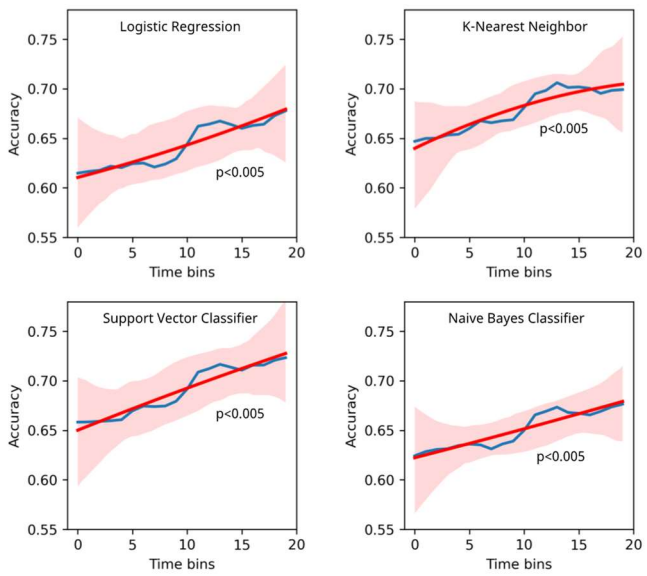

Figure 7. Simulated decoding accuracies based on four other common classifier algorithms: logistic regression, K-nearest neighbor, support vector classifier and naïve Bayes classifier. The blue lines represent the average of the data points for each of the 20 time bins, and the red quadratic curves were fitted over the data points with 100-fold bootstrapping to generate the 95% confidence intervals. Mann-Kendall tests demonstrated significantly increasing time trends ($p < 0.005$) for all classifier algorithms.

kinematics, which had been shown to be more difficult to perform [28].

Previous studies employed simpler signal mapping techniques (such as where the population direction vectors were directly mapped to the intended output) to characterize the neural changes during learning [18]. Here, we used an algorithmically more complex LDA decoder. The neural units adapted in the form of a divergence in the distribution of points encoding each direction within the LD space, suggesting that the monkeys were capable of altering neural activities across the recorded channels to improve performance (through reducing the misclassification errors) even for more complex decoders. Conversely, poor decoder performance was associated with a convergence of points within the LD space; this is to be expected, as neural patterns that were insufficiently distinct would be mapped to the same regions in the LD space [27]. In some trial blocks, we observed that the monkeys were able to produce stable and consistent firing patterns such that their activities mapped to a relatively confined region within the LD space over time, with well-separated clusters. In summary, real-time visualization of such temporal patterns within the LD space during decoding would provide more detailed insights into the neural correlate of the learning process during BCI use.

Secondly, we observed that in some trial blocks the monkeys experienced difficulty in encoding some directions (and conversely some directions were more readily encodable). The exact mechanism of such unequal performance remains unclear, but we suspected that it could be due to uneven sampling of neurons. There might not have been enough neurons in the measured population representing some of the directions to bring about satisfactory performance (although this "disadvantaged direction" could shift between trial sessions). Despite that, initially suboptimal accuracy in certain directions was generally surmountable even without the need to recalibrate the decoder, through such divergence in neural representations. Whether the user should adapt to the decoder, or the decoder should adapt to the user, remained an active area of debate. Furthermore, previous studies have reported fluctuations in neural signals during BCI use, which would introduce further randomness into the decoding process that the subject would need to compensate for [6,13]. Nevertheless, others have described very stable firing patterns especially with regard to directional tuning curves, suggestive of long-term learning effects [5,14,29–31].

4.1 Direction-selectivity of neurons

Here we also characterized the directional preference of the neural unit using permutation FI, a robust, model-agnostic method commonly used in machine learning [19]; repeating the analyses by zeroing the respective channels and using logistic regression coefficients showed similar results. A direction could either be encoded by one (or a small handful) of neural units with strong directional preference (such as the 'right' direction, as seen in **Figure 6B**, with higher maximal FI but the mean FI being near-zero), or by a larger number of units with individually weaker selectivities (such as for the 'stop' direction, as seen in **Figure 6B**, with generally higher mean FIs across the channels, compared to the other three directions). In some trial blocks, certain directions were encoded poorly with few channels selective for them.

Multi-unit or local field potential recordings, like in our experiments, where each channel recorded from multiple neurons with different directional preferences, would result in the respective channel having a composite direction preference for multiple directions. As task performance was related to the separability of point clusters in the LD space, various configurations of neural activities can lead to slightly different directions of divergence from other clusters; this was a convenient combinatorial property as there would be theoretically many different such combinations to reduce signal overlap, and hence ways for the neural ensemble to adapt.

Moreover, as shown in **Figure 6**, performance tended to be higher when more strongly directionally selective channels were present (represented by higher maximal FI) and/or with more heterogeneous neural representations (i.e. higher variance in FI, as opposed to a more homogeneous distribution of channels, i.e. with similar FIs). Another observation was that the "stop" command was generally encoded by a more

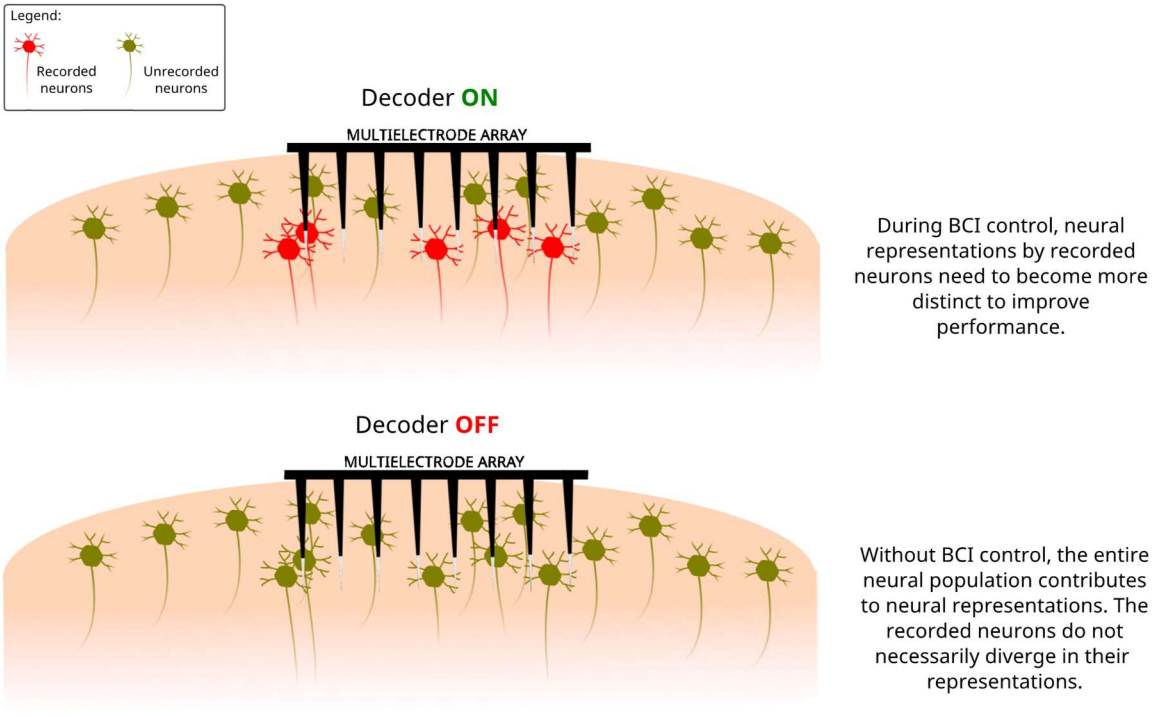

Figure 8. Hypothesis that during BCI use, neurons measured by recording array, and which directly influence BCI output, will be more likely to undergo observable changes in their firing patterns to bring about better signal separation and decoding accuracy. Those not measured by BCI, in contrast, will not experience such changes in their firing patterns (as their activities had no influence on the BCI output). Without BCI control (i.e. joystick control by natural limb), the entire neurons population (largely) contributes equally to movement control, and the neurons that are otherwise measured by the recording array will not have any special influence, and hence their firing patterns will not necessarily become more distinct over time.

distributed set of channels (**Figure 6B**), distinct from the three directions. We postulated that (directionally speaking) there were no "stop"-selective neurons in the primary motor cortex, although many neurons would inhibit motor response during these "stop" trials. Therefore, the encoder had to rely on the synthesis of a broader range of channels to encode the stop command. In previous studies, a small number of recorded cortical neurons (less than 30) were sufficient to control a cursor on a screen [32], or control a three-dimensional neuroprosthesis [33]. For more complicated tasks that required the extraction of more complicated features, a larger neuronal ensemble is recommended [8,33]. How much redundancy should be built into the recording system would be important with these considerations in mind.

We observed that during training, in some trials, the monkeys often had difficulty performing directional BCI tasks in certain directions. For such cases where BCI could not encode or decode a certain direction reliably, we suspected that the poor performance could be contributed by the nature of the measured neural activities. For instance, there might be a deficiency in neurons tuned to a specific direction among the measured population. Such an imbalance in performance could be an even bigger issue for a BCI that involves a wider range of outputs.

4.2 Task-relevant changes in neural code during BCI use

We conjectured that neurons measured by recording array, and hence directly influenced BCI output, will be more likely to undergo observable changes in their firing patterns to make the neural representations more distinct, and therefore increase the decoding accuracy. In contrast, those not measured by BCI were less likely to diverge in their neural representations (as their activities had no direct influence on the BCI output). With natural limb control during the Joystick trials, the movement would be represented by a much larger neural population, and the neurons that would otherwise be recorded by the array would have no special importance and therefore did not diverge in their neural representations (and in fact converged over time, although this was not associated with any accuracy change). This concept is illustrated in **Figure 8**. In addition, BCI users reported that BCI use required concentration [34], so this difference in neural code during task-switching most likely involved some attentional or salience mechanisms, for which little is known. Further studies should be done to further shed light on this.

With the observation that there could be uneven sampling of neurons with imbalanced direction selectivities, future algorithms could for example adaptively recalibrate by presenting directions with less differentiable neural codes

more frequently, allowing the subject to fine-tune the firing patterns to increase the divergence in the LD space; those with already well-differentiated neural firing patterns (and therefore good separation in the LD space) could be presented less. As the number of output commands increases (beyond the four directions in this study) relative to the number of recorded channels, such an adaptive design could be more efficient in giving the user more opportunities to improve the separability of the neural signals.

Another practical use could involve using such a visualization as additional visual feedback for the BCI user. Divergence of the features within the LD space could serve as an intermediate objective in BCI training as a means to fine-tune the accuracy and reduce misclassification errors. This could potentially expedite the training process by providing the BCI user (or the trainer) direct visual feedback on how to maximize inter-cluster distances and reduce misclassification errors.

These considerations would complement the ongoing efforts to design better predictive algorithms by utilizing the ability of the brain to adapt to a decoder, as well as strategies to shorten the long process of decoder calibration and BCI learning through adaptive decoder designs. Future work can be done to test these considerations in the design of BCI training sessions.

4.3 Strengths and limitations

We analyzed a wide range of trials at different stages of training and on two different monkeys with chronically implanted Utah arrays, and discussed specific scenarios with regards to the changes over time in the neural firing patterns and how those corresponded to movements of points within the LD space. Visualization of the temporal structure of the features in the decoder space will contribute to the recent emphasis on explainable artificial intelligence (AI) and help in the troubleshooting and perhaps even promote user adoption of these BCI, as it will help BCI users and external operators better understand the inner mechanisms of the decoders [35–37]. It is also uncertain whether all performance imbalances can eventually be compensated for through neural adaptations alone, and the limits of such adaptive mechanisms should be further evaluated.

Here we described and characterized, as post hoc analyses, the temporal sequences of the neural recording acquired from a prior set of experiments. In these post hoc analyses, the LDA decoders were refitted based on the recorded data to produce the LD graphs, and minor discrepancies existed between the actual decoded commands used for online decoding and the ones produced from the refitted decoders.

5. Conclusion

We have illustrated that for a motor BCI, the learning process can be characterized by studying the temporal and spatial distribution of points within the decoder space (in this case, the linear discriminant space). Temporally stable and spatially well-separated clusters in the decoder space signify that the user has become proficient in the use of the BCI. Before this stage is achieved, however, learning was represented by 'exploratory' movements of points within the decoder space. Analyses of such spatiotemporal changes can lead to better insights into BCI learning and inform future decoder design.

Conflict of interest statement

The authors declare no conflict of interests.

References

[1] 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

[2] Truccolo W, Friehs G M, Donoghue J P and Hochberg L R 2008 Primary motor cortex tuning to intended movement kinematics in humans with tetraplegia *Journal of Neuroscience* **28** 1163–78

[3] Kennedy P R, Bakay R A E, Moore M M, Adams K and Goldwaithe J 2000 Direct control of a computer from the human central nervous system *IEEE Transactions on Rehabilitation Engineering* **8** 198–202

[4] Libedinsky C, So R, Xu Z, Kyar T K, Ho D, Lim C, Chan L, Chua Y, Yao L, Cheong J H, Lee J H, Vishal K V, Guo Y, Chen Z N, Lim L K, Li P, Liu L, Zou X, Ang K K, Gao Y, Ng W H, Han B S, Chng K, Guan C, Je M and Yen S-C 2016 Independent Mobility Achieved through a Wireless Brain-Machine Interface ed G Schalk *PLoS One* **11** e0165773

[5] Ganguly K and Carmena J M 2009 Emergence of a stable cortical map for neuroprosthetic control *PLoS Biol* **7** 1000153

[6] Carmena J M, Lebedev M A, Henriquez C S and Nicolelis M A L 2005 Stable ensemble performance with single-neuron variability during reaching movements in primates *Journal of Neuroscience* **25** 10712–6

[7] Ikeda A, Yazawa S, Kunieda T, Ohara S, Terada K, Mikuni N, Nagamine T, Taki W, Kimura J and Shibasaki H 1999 Cognitive motor control in human pre-supplementary motor area studied by subdural recording of discrimination/selection-related potentials *Brain* **122** 915–31

[8] 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**

[9] Nicolelis M A L, Dimitrov D, Carmena J M, Crist R, Lehew G, Kralik J D and Wise S P 2003 Chronic, multisite, multielectrode recordings in macaque monkeys *Proc Natl Acad Sci U S A* **100** 11041–6

[10] Brincat S, Jia N, Salazar-Gomez A, Panko M, Miller E and Guenther F 2013 Which neural signals are optimal for brain-computer interface control? *Proceedings of the Fifth International Brain-Computer Interface Meeting: Defining the Future* **2**

[11] Georgopoulos A P, Kettner R E and Schwartz A B 1988 Primate motor cortex and free arm movements to visual targets in three-dimensional space. II. Coding of the direction of movement by a neuronal population *Journal of Neuroscience* **8** 2928–37

[12] Georgopoulos A P, Kalaska J F, Caminiti R and Massey J T 1982 On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex *Journal of Neuroscience* **2** 1527–37

[13] Rokni U, Richardson A G, Bizzi E and Seung H S 2007 Motor Learning with Unstable Neural Representations *Neuron* **54** 653–66

[14] Stevenson I H, Cherian A, London B M, Sachs N A, Lindberg E, Reimer J, Slutzky M W, Hatsopoulos N G, Miller L E and Kording K P 2011 Statistical assessment of the stability of neural movement representations *J Neurophysiol* **106** 764–74

[15] Birbaumer N and Chaudhary U 2015 Learning from brain control: clinical application of brain–computer interfaces *eNeuroforum* **21**

[16] Koralek A C, Jin X, Long J D, Costa R M and Carmena J M 2012 Corticostriatal plasticity is necessary for learning intentional neuroprosthetic skills *Nature* **483**

[17] Fetz E E 1969 Operant conditioning of cortical unit activity *Science (1979)* **163** 955–8

[18] Jarosiewicz B, Chase S M, Fraser G W, Velliste M, Kass R E and Schwartz A B 2008 Functional network reorganization during learning in a brain-computer interface paradigm *Proc Natl Acad Sci U S A* **105**

[19] Breiman L 2001 Random Forrests *Mach Learn* **45**

[20] Hussain Md and Mahmud I 2019 pyMannKendall: a python package for non parametric Mann Kendall family of trend tests. *J Open Source Softw* **4**

[21] Stiso J, Corsi M C, Vettel J M, Garcia J, Pasqualetti F, de Vico Fallani F, Lucas T H and Bassett D S 2020 Learning in brain-computer interface control evidenced by joint decomposition of brain and behavior *J Neural Eng* **17**

[22] Hughes C L, Flesher S N, Weiss J M, Downey J E, Boninger M, Collinger J L and Gaunt R A 2021 Neural stimulation and recording performance in human sensorimotor cortex over 1500 days *J Neural Eng* **18**

[23] Downey J E, Schwed N, Chase S M, Schwartz A B and Collinger J L 2018 Intracortical recording stability in human brain-computer interface users *J Neural Eng* **15**

[24] Weiss J M, Flesher S N, Franklin R, Collinger J L and Gaunt R A 2019 Artifact-free recordings in human bidirectional brain-computer interfaces *J Neural Eng* **16**

[25] Downey J E, Brane L, Gaunt R A, Tyler-Kabara E C, Boninger M L and Collinger J L 2017 Motor cortical activity changes during neuroprosthetic-controlled object interaction *Sci Rep* **7**

[26] Handelman D A, Osborn L E, Thomas T M, Badger A R, Thompson M, Nickl R W, Anaya M A, Wormley J M, Cantarero G L, McMullen D, Crone N E, Wester B, Celnik P A, Fifer M S and Tenore F v. 2022 Shared Control of Bimanual Robotic Limbs With a Brain-Machine Interface for Self-Feeding *Front Neurorobot* **0** 140

[27] Colachis S C, Bockbrader M A, Zhang M, Friedenberg D A, Annetta N v., Schwemmer M A, Skomrock N D, Mysiw W J, Rezai A R, Bresler H S and Sharma G 2018 Dexterous control of seven functional hand movements using cortically-controlled transcutaneous muscle stimulation in a person with tetraplegia *Front Neurosci* **12**

[28] Wodlinger B, Downey J E, Tyler-Kabara E C, Schwartz A B, Boninger M L and Collinger J L 2015 Ten-dimensional anthropomorphic arm control in a human brain-machine interface: Difficulties, solutions, and limitations *J Neural Eng* **12**

[29] 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 *Journal of Neuroscience* **27** 10742–50

[30] Dickey A S, Suminski A, Amit Y and Hatsopoulos N G 2009 Single-unit stability using chronically implanted multielectrode arrays *J Neurophysiol* **102** 1331–9

[31] Pels E G M, Aarnoutse E J, Leinders S, Freudenburg Z v., Branco M P, van der Vijgh B H, Snijders T J, Denison T, Vansteensel M J and Ramsey N F 2019 Stability of a chronic implanted brain-computer interface in late-stage amyotrophic lateral sclerosis *Clinical Neurophysiology* **130** 1798–803

[32] Serruya M D, Hatsopoulos N G, Paninski L, Fellows M R and Donoghue J P 2002 Instant neural control of a movement signal *Nature* **416** 141–2

[33] Taylor D M, Tillery S I H and Schwartz A B 2002 Direct cortical control of 3D neuroprosthetic devices *Science (1979)* **296** 1829–32

[34] Kögel J, Jox R J and Friedrich O 2020 What is it like to use a BCI? - Insights from an interview study with brain-computer interface users *BMC Med Ethics* **21**

[35] Shin D 2021 The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI *International Journal of Human Computer Studies* **146**

[36] Jia X, Ren L and Cai J 2020 Clinical implementation of AI technologies will require interpretable AI models *Med Phys* **47**

[37] Samek W, Montavon G, Vedaldi A, Hansen L K and Muller K-R 2019 Explainable AI: Interpreting, Explaining and Visualizing Deep Learning *Lecture Notes in Computer Science (LNCS)* **11700**