Table of Contents

Acknowledgements

I would like to thank Professor John Donoghue for affording me the opportunity to work on this project and providing strong guidance and insight. I would also like to thank Carlos Vargas-Irwin, whose supervision over my work on this project was integral to its completion and fostered many of the ideas presented here.

I would like to thank Corey Triebwasser, Michelle Nevor, John Murphy, and Allan Rydberg for their assistance with animal care and instrumentation design, and Carlos Vargas-Irwin and Lachlan Franquemont who collected the data used in this work and made my analysis possible.

I also gratefully acknowledge the following sources of funding: VA-Rehab R&D; NINDS-Javits (NS25074); DARPA (N66001-10-C-2010), Katie Samson Foundation.

Abstract

Neural encoding of movement-related variables is frequently examined through tasks in which instructions and actions are separated by a temporal delay, and analysis of the relationship between behavior and neural activity typically focuses only on trials where the subject correctly interprets the instructions, choosing an action from a learned cue-behavioral response. However, animals make incorrect responses, and understanding the coding of error trials in neural ensembles can provide insight into the role of particular cortical areas in computing actions and opens the possibility for detecting errors before or while they are occurring. Here, we compared neural activity patterns in ensembles of ventral premotor cortex (PMv) neurons associated with correct and incorrect action selection. We trained three male rhesus macaques to perform a reach-and-grasp task in which an arbitrary color cue instructing one of three known grips is followed by a cue to make that action. Data was recorded from PMv using chronically implanted 96 channel microelectrode arrays and sorted into single units. We projected spiking activity during the delay between grip and go cues into a spike train similarity space (SSIMS) (Vargas-Irwin *et al.*, 2014) to evaluate spiking patterns across time. Recent experiments have shown that correctly executed trials produce a consistent progression of ensemble states that clearly separate clusters associated with specific grip strategies (manuscript under review). Based on clustering of trials immediately before the go cue, we see a mixture of cue-encoding and action-decoding errors with respect to PMv activity. We also observe that, during the instructed delay and movement, error trials diverge from "correct" ensemble states, exhibiting distinct neural activity despite identical actions being performed. During these times, we successfully differentiate between correct and error trials with accuracy significantly above chance. We find that particular small ensembles of approximately ten neurons are sufficient for this classification and that single

unit and small ensemble grip classification performance is not a consistent predictor of error classification performance. These findings suggest the existence of a neural signature of errors present in cortical motor areas which is independent of movement encoding. Early detection of these possible error states may contribute to the design of more accurate neural decoding algorithms for use in brain computer interfaces and provide insight into the role of PMv in computing hand grip types from learned arbitrary cue-action associations.

Introduction

High-density microelectrode arrays have become a powerful tool to extract data from over one hundred neurons at once (Nordhausen *et al*, 1996). This technology has substantially increased investigators' ability to decode signals from motor cortical regions through the analysis of neural activity in non-human primates performing motor tasks, which has driven the development of brain computer interfaces (BCI's) to restore functions to paralyzed individuals such as motor control (Hochberg *et al.*, 2006, Hochberg *et al.*, 2012, Collinger *et al.*, 2013, Wodlinger *et al.*, 2015) and communication (Kim *et al.*, 2007). However, the speed, accuracy and precision of these devices are far from that of a healthy individual performing the same functions.

One potential avenue for improving this accuracy is to better understand the neural encoding of errors in motor tasks – trials in which a subject performs an inappropriate action given the situation. It is unclear at what point in processing these cue-to-action mapping errors arise – whether in encoding a cue into neural activity or in decoding neural activity into an appropriate action – and what spiking patterns differentiate these trials from their correct counterparts. Understanding the error processing properties of motor regions could lead to a

better understanding of motor cortical processing and the design of error detecting features in BCI decoding algorithms.

We focused our analysis on ventral premotor cortex (PMv), a region heavily implicated in movement representation (Ehrsson *et al.*, 2000, Kakei *et al.*, 2001) and goal-oriented action (Buccino *et al.*, 2001). Of particular interest, PMv is critical in transforming the perception of objects into hand actions to manipulate or grasp them. (Rizzolatti *et al.*, 1983, Fogassi *et al.*, 2001), prompting us to study the timing and types of errors seen in PMv during a cued grasping task.

PMv has previously been implicated in motor error processing – in particular, fMRI studies have found PMv active when observing a motor error (Manthey *et al.*, 2003), while single unit recordings have also implicated PMv in error detection during the execution of motor errors (Romo *et al.*, 2004). However, much of the error processing literature has looked at the cingulate cortex, a region which projects to PMv (Pandya *et al.*, 1981, Boussaoud *et al*., 2005, Dancause *et al.*, 2006) and may influence its firing patterns. The anterior cingulate cortex (ACC) was initially implicated in error detection through the discovery of the error-related negativity (ERN) and, later, positivity (ERP) – deflections in electroencephalography (EEG) signals localized in the ACC (Dehaene *et al.*, 1994, Miltner *et al.*, 1997). These error signals have been observed even when the subject is unaware that they have made an error (Nieuwenhuis *et al.*, 2001), and, paradoxically, occasionally during correct trials, though at smaller amplitudes (Vidal *et al.*, 2000). Indeed, more than a rote error detecting, the ACC has since been more generally found to be active in circumstances where errors are likely – that is, when the brain is confronted with conflicting information – leading to theories of its role as a conflict monitor (Carter *et al.*,

1998, Botvinick *et al.*, 1999). This continuous processing suggests the possibility of predicting errors before they occur through recordings of conflict monitoring areas.

To these ends, we trained rhesus macaques to perform a cued reach and grasp task in which an arbitrary color cue instructing a known grip is followed by a cue to make that action. Using chronically implanted 96-microelectrode arrays, we recorded spiking activity from PMv throughout the task. We employ spike train similarity (SSIMS) analysis (Vargas-Irwin *et al.*, 2014), an algorithm which projects ensemble spiking data into a low-dimensional space using a spike train similarity metric and dimensionality reduction, allowing quantitative analysis of the similarity between patterns of neural activity. This methodology has previously been used on data from this task to classify grip and object encoding (submitted manuscript under review) – here, we investigate trials in which the monkey executes the grip which was not cued, thereby making a cue-to-grip-mapping error. We compare the single neuron and ensemble activity patterns associated with correct and erroneous grip planning, showcasing error states before and during movement and distinguishing between encoding and decoding errors in PMv. We demonstrate that errors have statistically distinct activity patterns in PMv which allows for error detection significantly above chance, a feature which could be implemented into decoding algorithms to improve BCI accuracy.

Methods

Behavioral task

Three male rhesus macaque monkeys (designated G, R and S) were trained to perform a cued grasping task with an instructed delay, outlined in detail in Vargas-Irwin *et al.*, 2015 (submitted). This task involves two objects, each of which can be gripped in two ways (Figure

1a) and contain capacitive sensors (two for each grip) allowing identification of which grip was performed. These objects are mounted on sliding rails, allowing them to be lifted 2 cm upward. Both objects included a 10 cm long cylinder 3.5 cm in diameter. This part of the objects was grasped using a power grip. In object A, the cylinder was joined to a horizontal disk with a diameter of 7.5 cm and thickness of 0.5 cm. The disk was grasped using a key grip. In object B, the cylinder was joined to a vertical plate measuring 10 x 7 x 0.5 cm. This part of the object was grasped using a precision grip. For monkey G, the cylinder was the top part of the object. In subsequent experiments with monkeys S and R the cylinder was positioned at the top, as we found that this configuration was easier for the monkeys to visualize and grasp while in head fixation.

The temporal sequence of the task is illustrated in Figure 1b. The head-fixed monkey, sitting in a primate chair in a darkened room, begins the task by holding his hands on touch-sensitive pads for one second. One of the two objects is then illuminated, followed by a light cue indicating one of two grips – key (yellow) or power (red) for object A, and precision (yellow) or power (red) for object B. After a two-second instructed delay (ID), a go cue (green light) is presented, instructing the monkey to remove his hand from the plate and lift the object with the instructed grip for 200 msec. Upon completion with the cued grip, the monkey receives a fruit juice reward.

Errors

An “error” is defined as a trial in which the monkey successfully grasped the object after the go cue but did so using the grip which was not cued. Other ways in which a trial can end unsuccessfully – the monkey removing his hand from the pressure plate prematurely, using

neither of the two standard grips for that object, failing to lift the object within 2 seconds of the go cue, or lifting the object for fewer than 200 ms – are not considered errors for the purpose of this analysis, but are simply aborted trials. In any error or aborted trial, no reward is given. For this analysis, we treat each object separately as its own dataset and only consider datasets in which there are at least five errors but where fewer than 1/3 of all trials were errors, as this provides enough errors for clustering analysis but not so many that the monkey was likely ignoring the cue entirely.

Neural activity recording

Silicon-based microelectrode arrays (Blackrock Microsystems LLC, USA) with 100 electrodes (96 active, 1.0-mm length, 400-μm electrode spacing) were implanted in ventral premotor cortex (PMv). The precise location of arrays is illustrated in Vargas-Irwin *et al.* 2015 (submitted). Data was acquired through a Cerebus multichannel data acquisition system (Blackrock Microsystems LLC, USA). Single units were identified using a template matching algorithm to separate spikes by waveform shape (Vargas-Irwin & Donoghue, 2007). Only units with a signal-to-noise ratio greater than 1.2 were included in this analysis. Data were separated by recording session for analysis.

Spike Train Similarity Analysis

To quantitatively compare spike trains between trials, we use the Spike Train Similarity Space (SSIMS) Toolbox, detailed in Vargas-Irwin *et al.*, 2014. In brief, SSIMS analysis projects trials into a low dimensional space agnostic of trial labels by 1) projecting spiking data into a

high-dimensional space using a spike train similarity metric, and 2) reducing dimensionality. We use spike trains of one-second length.

Spike Train Similarity Metric

The SSIMS toolbox employs the Victor-Purpura (V&P) spike train similarity metric (Victor & Purpura, 1996), which calculates the similarity between two spike trains as the sum of the following costs: 1) a fixed cost of 1 to add or remove a spike, and 2) a variable cost of q to shift a spike in time. The value of q determines the level of temporal precision – we keep q constant at 10 (100 ms precision), though changes to this value have little effect on our results.

In single neuron analysis, the high-dimensional data is a T×T pairwise distance matrix between each trial using the V&P metric, where T is the number of trials. This is treated as a set of T points in a T-dimensional space. For analysis with an ensemble of N neurons, we compute the pairwise distance matrix for each neuron, yielding T points in a T×N-dimensional space. When analysis spans multiple time windows, as when plotting neural trajectories of trials over time, we compute the pairwise distance matrix for each neuron in each time window. Three time windows are used to create the SSIMS for neural trajectories – the object presentation phase, and the first and second halves of grip cue phase.

Dimensionality Reduction with t-SNE

From the high-dimensional space, we use t-SNE (van der Maaten & Hinton, 2008) to produce a low-dimensional SSIMS. In brief, t-SNE seeks to preserve the joint probability that point x would choose point y as its nearest neighbor for all points x and y, and does this by minimizing the Kullback-Leibler divergence of this quantity in the high and low dimensional

spaces. To improve computation time, the high-dimensional space is first reduced to 30 dimensions using principal component analysis (PCA), and then to 10 (for analysis) or 3 (for visualization) with t-SNE.

Classification

We classify trials with a median-based classifier using leave-one-out cross-validation. Let **x** be the low-dimensional vector of the trial to be classified, and define a function f(T) by Eq. 1

$$f(T) = med(|\boldsymbol{x} - \boldsymbol{t}| \ \ \forall \, \boldsymbol{t} \in T) \quad (1)$$

where T is a set of training data from one of four classes (correct or error for either grip), med() is the median function, and |**a**| denotes the norm of vector **a**. **x** is classified by Eq. 2

$$class = argmin(f(T)) \quad (2)$$

where the domain of f is each of the four classes of trials in the training data. In short, this algorithm classifies a trial as that class whose cluster's median distance to the trial is smallest over all points in the cluster. Classification of a trial is considered successful if it is correct and classified as correct (true negative) or an error and classified as an error (true positive), and classification is considered unsuccessful otherwise.

Due to the small number of errors compared to correct trials, we primarily consider "classification accuracy", which we define as the average of the true positive and true negative rates, whereas the term "fraction correctly classified" indicates a ratio of true to false classifications and is thus biased towards true negatives (correct trials) over true positives (error trials). By these definitions, both values have a range from 0 to 1. Chance values and 95% confidence thresholds for classification accuracy are approximated using 1000 Monte Carlo label shuffling simulations.

Greedy Algorithm for Choosing Ensembles

To approximate maximal classification accuracy as a function of ensemble size, we used a greedy algorithm, reducing the high algorithmic complexity of solving the problem exhaustively. We chose "best" subsets as follows: choose the "best 1-neuron subset" as the neuron which maximizes classification accuracy, and choose the "best n-neuron subset" as the subset which maximizes classification accuracy over all n-neuron subsets of neurons containing the best (n–1)-neuron subset.

Results

Summary of Data

Five datasets from three monkeys were analyzed. Two of these datasets regarded Object A and three regarded Object B (two in which Object B was inverted for Monkey G). In total, 287 correct trials and 67 error trials (18.9% of all trials) were recorded. In all datasets, more than 70% of errors occurred for only one of the two grip cues, though which of the two grip cues led to greater numbers of errors was not consistent. A breakdown of the components of each dataset is presented in Table 1.

Characterization of Error Trials

Projection of trials into a SSIMS over the duration of the task (sliding 1-s window with 100-ms steps) reveals that, on average, cue-to-grip mapping errors traverse a neural trajectory qualitatively distinct from correct trials of either the cued or executed grip (Figure 2a). These differences are exemplified by the natural clustering of single trials that occurs when looking at 2-D projections of SSIMS at individual time windows throughout the task evolution (Figure 2b).

Monkeys Perform Both Encoding and Decoding Errors with Respect to PMv Activity

It is unclear at what point in the cue-to-grip mapping process these errors occur. One possible explanation for an incorrect action is an error encoding the cue into neural activity upstream of or in PMv, while another explanation could be an error decoding the neural activity in or downstream of PMv into an action. One would predict that, once the cue has been processed, a trial in which an encoding error occurred would have neural activity similar to trials corresponding to the grip to be executed, while a trial in which a decoding error occurred would be similar to trials corresponding to the grip which was cued. An example of each trial is shown in Figure 2b (top right).

To determine which kinds of errors occurred in our datasets, we attempted to classify each error trial as either a power grip or precision/key grip (depending on the object) based solely on training data from correct trials using a SSIMS calculated during the latter half of the instructed delay (ID) between the grip and go cues. Overall, 70.2% of errors (8/10, 17/21, 5/12, 12/17, and 4/7 for each dataset) clustered best with correct trials corresponding to the uncued grip, suggesting that these incorrect trials are more likely due to encoding errors. Conversely, 29.8% of errors clustered best with correct trials corresponding to the cued grip, suggesting that these incorrect trials are due to decoding errors. These data are summarized in Table 2

Distinct Ensemble States Allow Error Detection During Planning and Movement

To quantitatively determine if and when error trials are distinguishable from correct trials solely from their ensemble activity, we projected the 1-second spike trains into separate SSIMSs every 100 ms and attempted to classify each trial as a correct or error trial at each time (Figure 2c). At the onset of the grip cue, error classification never exceeded 95% confidence (0.702) in

all five datasets (classification accuracy = 0.6574, 0.6395, 0.6722, 0.6528, and 0.6180). After one second of the ID, classification accuracy exceeded 95% confidence in 4 out of 5 datasets (classification accuracy = 0.8241, 0.7449, 0.7778, 0.7050, and 0.6879) At the go cue, classification accuracy exceeded 95% confidence in 3 out of 5 datasets (0.7407, 0.7789, 0.5972, 0.5959, and 0.7873).

We repeated the above analysis aligning trials to contact with the object rather than the grip cue (Figure 3). At contact, classification accuracy exceeded 95% confidence (0.707) in all datasets (classification accuracy = 0.8426, 0.7109, 0.8389, 0.9891, and 0.8307). Classification accuracy peaked after contact, but data after contact is not considered in this analysis as error trials are aborted as soon as an incorrect grip is registered.

This analysis suggests significant grouping of errors into clusters which are separate from the clusters formed by correct trials.in the SSIMS early in the ID (4 of 5 datasets), just before the go cue (3 of 5 datasets), and just before contact (5 out of 5 datasets) Based on this data showing peak ID error classification accuracy early rather than later, error classification in the ID for the remainder of this paper is performed over the first second of the ID (t = 2 in Figure 2c), except where otherwise noted. The data for classification at this time are summarized in Table 2.

Differences in Reaction Times Do Not Account for Error Classification

Error trials are behaviorally similar to correct trials corresponding to the same executed grip, but one major difference is that, on average, monkeys take longer to begin moving after the go cue ($p < .10^{-4}$) when they are about to perform an erroneous grip (Figure 4a). It could thus be argued that the error signal we are decoding is simply a reflection of longer upcoming reaction times. To test this, we redid our classification using all error trials but only correct trials whose

reaction times were higher than the median error reaction time, which left between 1/10 and 1/3 of correct trials for analysis depending on the dataset. We saw no significant difference in classification accuracy between this (classification accuracy = 1.0000, 0.7857, 0.7210, 0.6849, 0.7906) and our classification with all trials ($p = 0.4408$) (Figure 4b), indicating that we are decoding an error signal irrespective to increased reaction times.

Small Ensembles Achieve Maximal Error Classification Accuracy

Intuitively, one might expect larger neural ensembles to represent more information, leading to more accurate classification. To quantify this gathering of information and optimize classification accuracy, we found the approximate maximum classification accuracy as a function of the number of units used for classification using a greedy algorithm (see Methods). Figure 5a shows the approximated maximal classification accuracy as a function of ensemble size and the classification accuracy when the neurons are chosen in the opposite order for two example data sets from different monkeys. In general, classification accuracy began to plateau once the best approximately ten neurons were picked, but began to fall rapidly when the final 5-20 neurons were chosen.

Pooled data (Figure 5b) compares the average classification accuracy of the ensemble to the average classification accuracy of the best 1, 5, and 10-neuron subsets. The best 5-neuron subsets on average outperform the full ensemble but do not achieve significance at the $p < 0.05$ level ($p = 0.0592$), but the best 10-neuron subsets do significantly outperform the ensemble ($p = 0.0465$). This suggests that there exist small subsets of neurons which can more effectively classify correct and error trials than the full ensemble. However, since even our non-exhaustive

algorithm runs with algorithmic complexity $O(n^2)$ where n is the number of neurons, choosing which subsets to use is a nontrivial problem.

A heuristic to choose a subset of neurons for effective error decoding could both improve classification accuracy and reduce computation time. One easily calculated heuristic is a neuron's firing rate, and when we remove neurons which fire sparsely (< 5 Hz) during the instructed delay, we see comparable classification accuracy (Figure 6a) using only approximately 1/3 of the neurons (Figure 6b). This data is summarized in Table 2.

Accurate Grip Decoding is Not a Consistent Predictor of Accurate Error Decoding

Previous results (not yet published) have shown the effectiveness of SSIMS analysis in classifying between motor plans for power and precision/key (depending on the object) grips in correctly-executed trials using PMv ensemble activity. We hypothesized that neurons and ensembles which were effective for this type of grip classification would also be effective for error detection. To test this, we investigated the accuracy of subsets of neurons in classifying both grips and errors. Figure 7a compares the grip and error classification accuracy for single neurons from one dataset from each monkey, and Figure 7b does the same for randomly chosen ensembles of five neurons. Least-squares lines of best fit are drawn for each dataset, but in both the single-neuron and small ensemble case, r^2 values (0.0190, 0.3630, and 0.0015 for single neurons; 0.0030, 0.0353, and 0.0422 for 5-neuron ensembles) suggest no strong correlation between grip and error decoding, except for the single neurons of monkey R for whom a weak positive correlation exists. This suggests that, in general, grip and error decoding are independent processes which can involve both the same and different neurons and ensembles.

Discussion

We have demonstrated the ability to detect cue-to-grip mapping errors during planning and action execution at a rate significantly above chance using spiking activity from ensembles of neurons in PMv. This corroborates previous accounts of error signals in premotor cortex during execution and observation of motor errors (Carter *et al.*, 1999; Manthey *et al.*, 2003, Romo *et al.*, 2004) at higher spatial resolution than previously shown. These results are consistent across the five data sets which contained enough errors for analysis.

We have shown that particular small ensembles of ten or fewer neurons can be sufficient for error decoding, suggesting that errors may appear or be detectable in PMv ensembles by inputs to select classes of neurons. However, this claim would require further investigation of these particular cells. Since determining which small ensemble to classify with is a nontrivial problem, we have demonstrated that using only neurons with a high firing rate (> 5 Hz) during the ID as a reasonable heuristic to determine which neurons to use. This is a mostly arbitrary threshold, chosen to remove many but not all neurons, as higher thresholds tend to leave datasets with very few neurons for analysis. This heuristic is generally robust regardless of the time over which firing rates are sampled during the task – for instance, during movement or after contact rather than during the ID as presented here. Other heuristics were tested, such as using only the neurons whose firing rate distributions were sufficiently not unimodal using Hartigan's dip statistic (Hartigan & Hartigan, 1985), but we found them no better and less selective than the simpler high firing rate heuristic.

A curious pattern consistent across the maximal classification accuracy functions such as those in Figure 4a is the steep drop-off when the final 5-20 neurons are chosen. We hypothesize that this drop-off occurs due to neurons which are irrelevant for error classification drowning out

the information provided by the few best neurons. While using the ensembles chosen by our greedy algorithm to produce the best classification is prone to overfitting, these merely show the limits of what classification rates are theoretically possible and incentivize our attempts to reduce computation time.

We have classified errors as either encoding – errors in mapping a cue to neural activity – or decoding – errors in mapping neural activity to an action – with respect to activity in PMv ensembles and shown that encoding errors are most frequent (70.2% of errors) but decoding errors occur as well (29.8% of errors). This suggests that most errors are made in some process upstream of PMv, though errors could arise in PMv itself and resemble encoding errors if they arise early in PMv processing or decoding errors if they arise late. Analyzing recordings from multiple motor cortical areas at once could allow us to better pinpoint when and where these errors arise.

Our method to classify errors as encoding or decoding comes with two caveats. First, while grip classification with SSIMS is effective, it is imperfect, correctly classifying 88.07% of trials as either power or precision/key grip trials just before the go cue (98.15%, 91.84%, 82.22%, 100.00%, and 71.74% for each dataset). As such, the "true classification" of which grip was being planned right before the go cue in error trials cannot be determined with perfect accuracy, adding some level of uncertainty to our encoding and decoding percentages. Secondly, even just before the go cue, we have shown that 3 of 5 datasets exhibit significant grouping of errors into their own distinct cluster. This information presented in the low-dimensional space is useful for error detection, but may lead to noise when attempting to classify error trials into one of the two grip clusters. As such, this analysis presents only a rough estimate of the rates of encoding versus decoding errors.

Understanding errors and pinpointing when they occur could be valuable for improved control of BCI technology. SSIMS classification could be implemented into BCI decoding algorithms to flag undesired or incorrect actions before or as they are being executed. While this type of classification would require specific timing events for analysis which are not always inherently present in the real world, potential alternatives exist. For instance, it has been possible in certain contexts to detect signals associated with attentional switching with EEG (Escera *et al.*, 1998; Bledowski *et al.*, 2004) or voluntary movement initiation with local field potentials (Masimore *et al.*, 2005). Generalizing these results and potentially improving their accuracy with higher resolution technologies could allow locking onto these natural events instead.

It is difficult to determine precisely what feature of errors we are decoding. While we can rule out potential confounds such as major differences in grip structure (from sensors in the objects) and differences in reaction time (Figure 4), other phenomena such as decreased or misdirected attention likely correlate with the probability of an error, and we cannot determine whether we are decoding that or a signal reflecting a conflict of information. Curiously, we consistently see error classification accuracy above 0.6 even when there is no grip information, while chance levels are 0.5. This is not a significant effect, but it suggests that, at least in some cases, we are decoding ensemble states in which errors are likely rather than the errors themselves. This could be consistent with an attentional or conflict monitoring signal. Nevertheless, whichever quality we are decoding correlates strongly enough with errors to make their detection possible and useful, and additional tests are possible to better narrow down what type of information is represented in the signal we are decoding.

We hypothesize that the signal which distinguishes correct from error trials in PMv may be the downstream result of the ACC recognizing conflict between the action being planned and

the cue. The ACC is just one of several regions upstream of PMv which have been implicated in decision making, such as the prefrontal cortex and basal ganglia (Hoover and Strick, 1993, Carmichael and Price, 1995, Boussaoud *et al.* 2005). However, in decision making literature, these regions have been primarily implicated in action selection (Mink, 1996, Redgrave *et al.*, 2010), evidence accumulation (Ding and Gold 2010) and implementation of control (MacDonald *et al.*, 2000) rather than conflict monitoring, the type of signal which would best explain our ability to predict errors before they occur. As the ACC has been implicated in monitoring conflict between streams of information throughout a task (Carter *et al.*, 1998, Botvinick *et al.*, 1999), we believe the ACC would be the most likely candidate for an upstream source of the PMv error signal we observe. This model makes several testable hypotheses. First, the neurons which are effective for error decoding during one of planning, movement, or conscious realization of errors after a withheld reward should be effective for the other two, since they remain physically connected to some upstream conflict monitor. Second, error detection using similar analysis on ensemble spiking data from the ACC should produce similar results. Thirdly, any error detectable in PMv should also be detectable in the ACC, and, contrapositively, any error not detectable in the ACC should also not be detectable in PMv.

These types of tests would require simultaneous recording from PMv and ACC and a task which is specifically designed for the analysis of errors. This could be done by simply tweaking parameters of our existing instructed delay task. Firstly, error trials are currently aborted as soon as an incorrect grip is detected, prompting the monkey to drop rather than lift the object as he would during a correct trial. These significantly different movements confound our data during this post-action period, but reserving the reward for correct actions until after the monkey has moved his hand back to the starting pad would allow for analysis during this time. Secondly,

turning off the grip cue light partway through the ID would both force the monkey to interpret the cue quickly which would synchronize trials better in time and likely cause more errors, reducing disparities in the number of errors and correct trials. Both of these factors would likely lead to better decoding, and such experiments are currently being pursued.

			# Correct		# Errors		
Dataset	Object	# Neurons	Power	Other	Pow-> Other	Other-> Pow	% Errors
G1	B[1]	60	26	28	9	1	15.63%
G2	B[1]	48	20	29	21	0	30.00%
R1	A	39	48	42	0	12	11.76%
S1	B	206	22	24	1	16	26.98%
S2	A	218	19	29	5	2	12.73%
Total		**571**	**135**	**152**	**36**	**31**	**18.93%**

Table 1: Summary of data. Five datasets were used in this analysis. Each dataset is given a code corresponding to the monkey (designated G, R or S) and a number in chronological order. Each dataset only considers one of object A or B. # Neurons indicates the number of spike sorted units which were recorded from and exceeded a signal-to-noise ratio of 1.2. Power indicates the number of trials for which a power grip was cued, and Other indicates the number of trials for which a precision (for object A) or key (for object B) grip are cued. Pow-> Other Indicates the number of error trials for which a power grip was cued but a precision/key grip was executed, and Other-> Pow indicates the number of error trials for which a precision/key grip was cued but a power grip was executed. % Errors indicates the percentage of all trials which are errors.

		Classification Accuracy				
Dataset	% Encoding	Ensemble	Best 1-N	Best 5-N	Best 10-N	Rate > 5 Hz
G1	80.00%	0.824	0.852	0.907	0.907	0.8611
G2	85.71%	0.745	0.714	0.769	0.748	0.7211
R1	41.67%	0.778	0.703	0.753	0.800	0.7667
S1	70.59%	0.705	0.752	0.924	0.946	0.705
S2	57.14%	0.688	0.803	0.907	0.907	0.6273
Average	**70.21%[2]**	**0.748**	**0.765**	**0.852**	**0.862**	**0.736**

Table 2: Error type and classification accuracies for each dataset. % Encoding gives the percentage of errors which were classified as encoding errors as opposed to decoding errors. The remaining columns give classification accuracy values for each dataset when error classification is done with the entire ensemble (Ensemble), the best single (Best 1-N), 5-neuron (Best 5-N) or 10-neuron (Best 10-N), ensembles, and only neurons with an average firing rate of at least 5 Hz during the ID (Rate > 5 Hz).

[1] For Monkey G, the objects were inverted (see Methods)

[2] The mean % encoding errors is reported as a mean over total number of errors rather than datasets

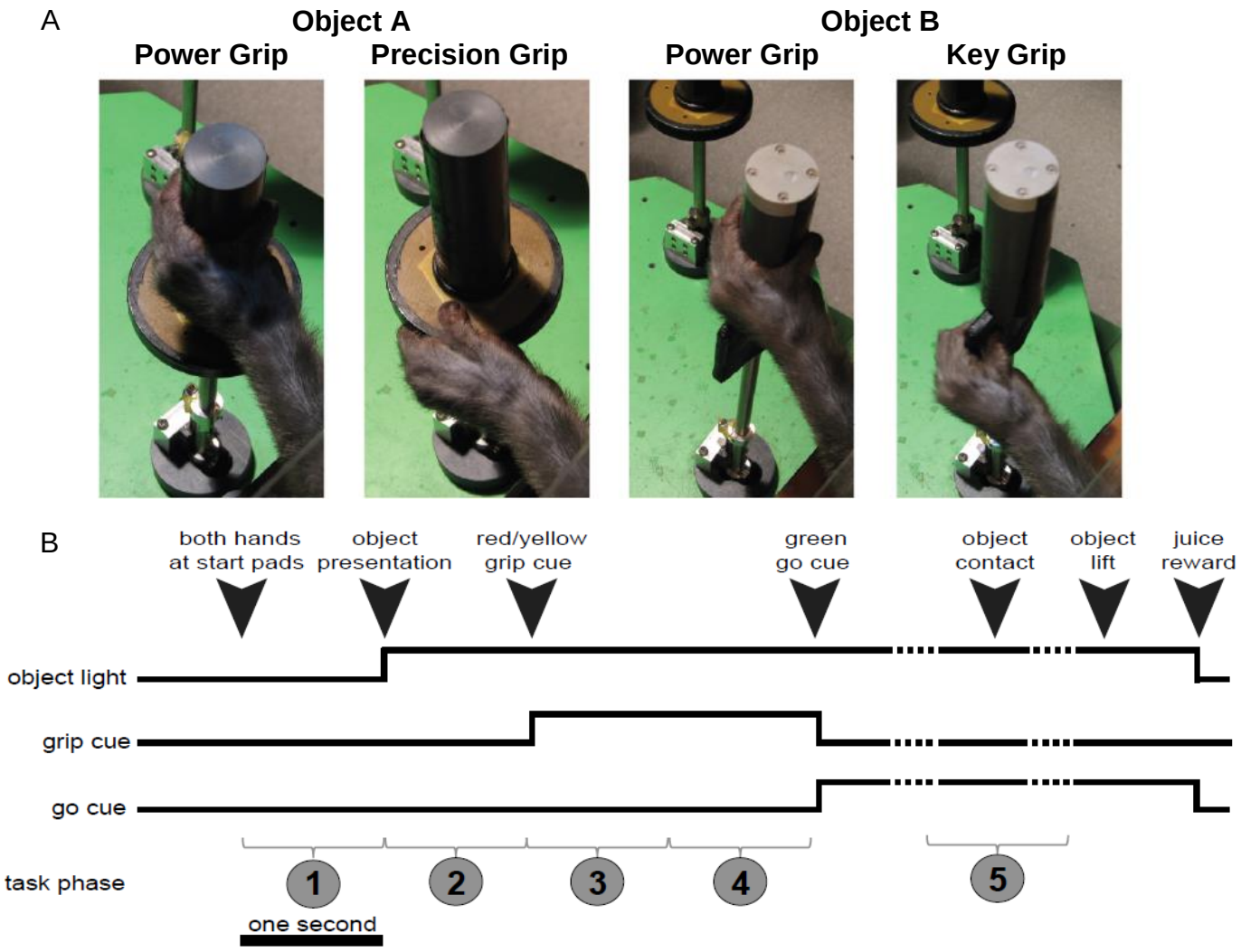

Figure 1: Summary of instructed delay task. **A)** Images of the objects and grips used in the task. Within a recording session, each object is considered its own dataset and only datasets with a sufficient number of errors (see Methods) were used (2 for object A, 3 for object B). Note that in sessions with Monkey G, the objects were inverted (see Methods). **B)** Timeline of task, showing the onset and offset of the object, grip and go cues. Figure adapted from Vargas-Irwin *et al.*, 2015 (submitted)

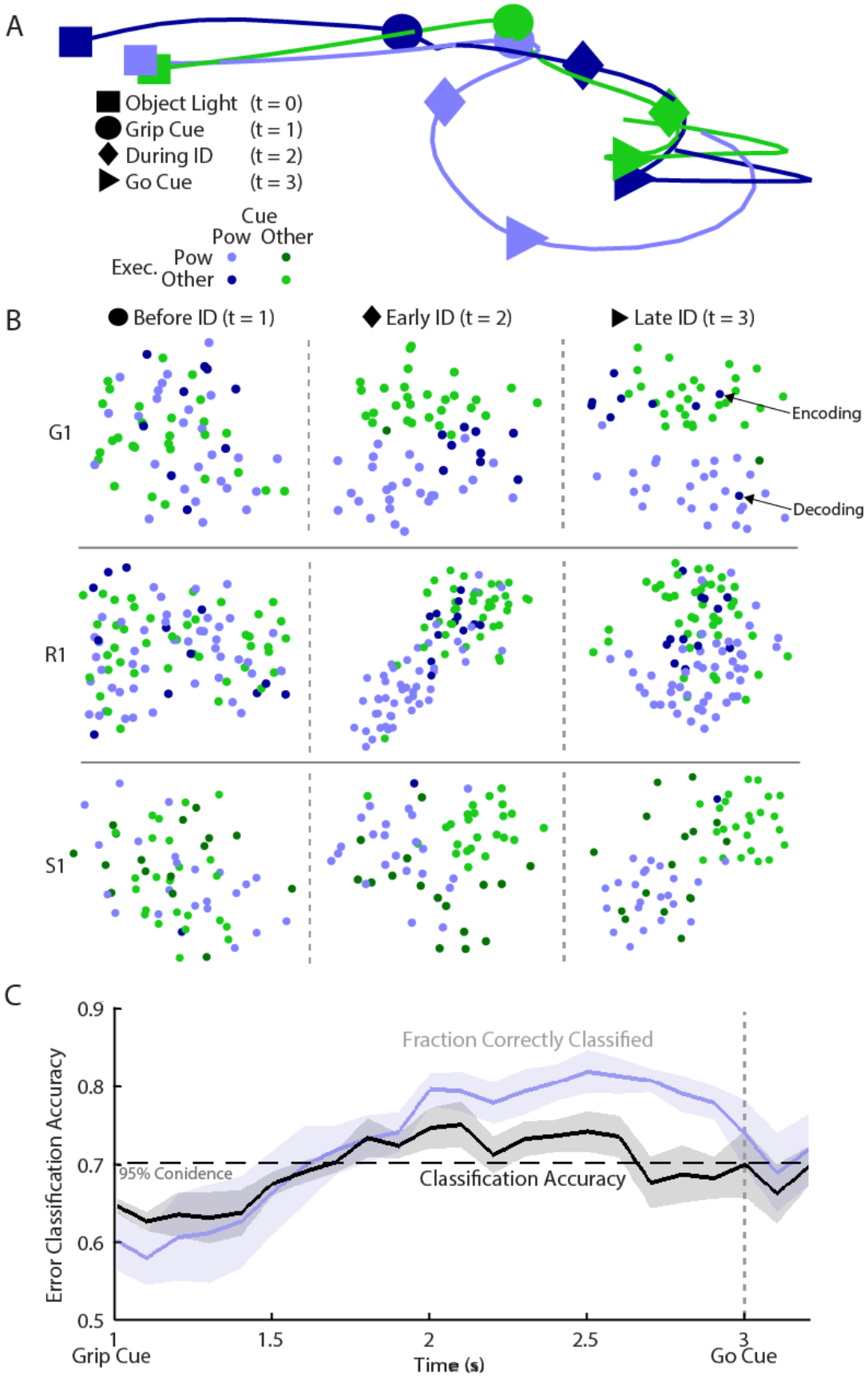
A
Object Light (t = 0)
Grip Cue (t = 1)
During ID (t = 2)
Go Cue (t = 3)
Cue
Pow Other
Exec. Pow
Other
B
Before ID (t = 1)
Early ID (t = 2)
Late ID (t = 3)
G1
R1
S1
Encoding
Decoding
C
Fraction Correctly Classified
95% Conidence
Classification Accuracy
Error Classification Accuracy
0.9
0.8
0.7
0.6
0.5
1
1.5
2
2.5
3
Grip Cue
Time (s)
Go Cue

Figure 2: Error detection is possible significantly above chance. **A)** Example mean neural trajectories for both correct (power and key) and one type of error trial (cued power, executed key) over the course of a task from one data set (G1). Time points of interest – object presentation, grip cue, partway through the instructed delay (the point of classification for the remainder of the paper) and go cue – are labeled for comparison. **B)** Example spike train similarity spaces (one dataset from each monkey) projected into two dimensions calculated before (right), early (middle) and late (left) in the instructed delay illustrating natural clustering of trials using the color scheme from (A). The symbol next to each time window corresponds to the point in the neural trajectory in (A). "Pow" refers to the power grip, "Other" refers to either precision or key grip depending on the object (precision for R, key for G and S). An example encoding and decoding error are noted for G1. **C)** Error classification from pooled data throughout the task locked onto external cues. Classification is reported as classification accuracy (black, average of true positives and true negatives, see Methods) and fraction correctly classified (blue, not accounting for disparities in number of correct and error trials). The 95% confidence interval approximated from Monte Carlo simulations corresponds to the classification accuracy measure. Transparent sheaths show standard error over five data sets.

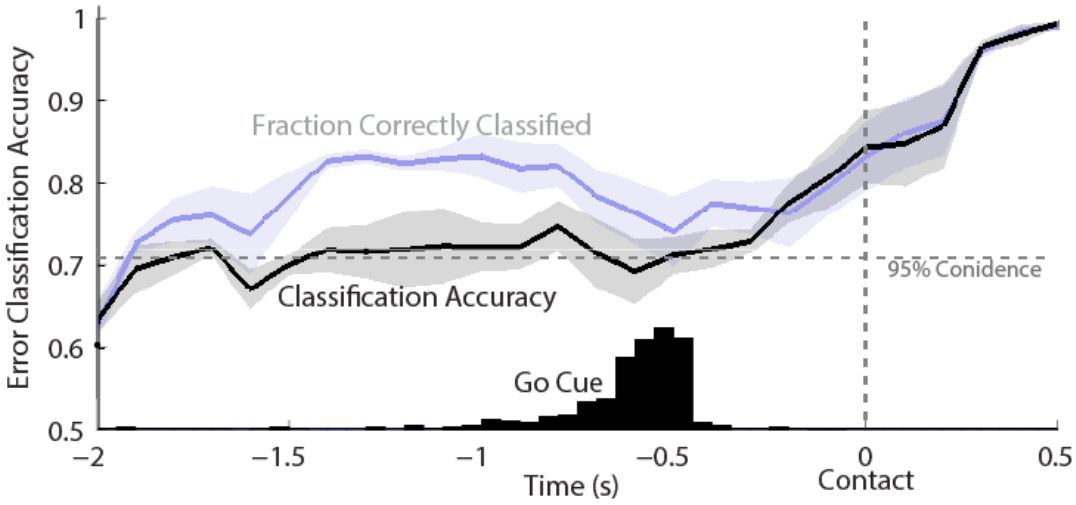

Figure 3: Error detection is possible during movement. Same as Figure 1c with trials locked onto contact with the object. The histogram shows the distribution of go cue times. Classification accuracy peaks after contact, likely due to confounds related to error trials being aborted once an incorrect grip is registered (see Results)

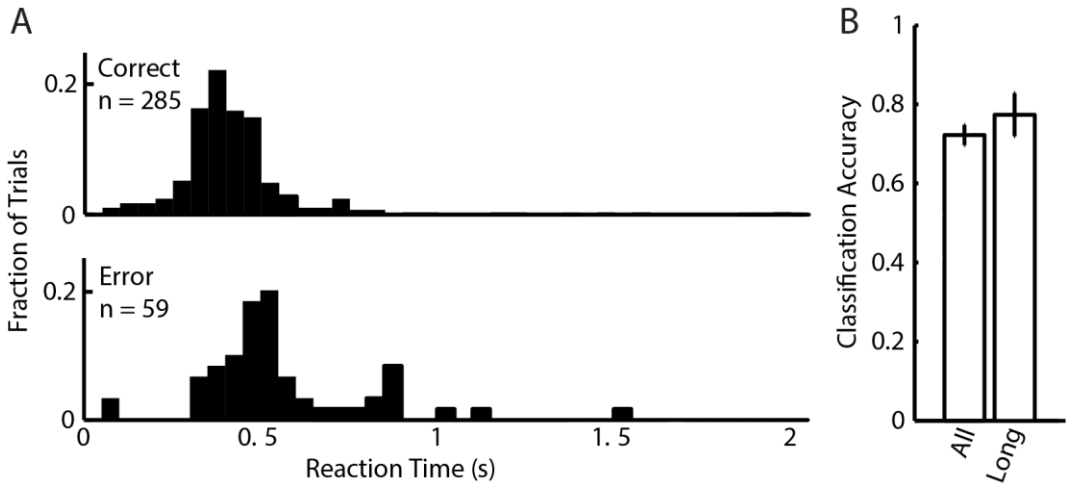

Figure 4: Differences in reaction time do not explain accurate error classification. **A)** Histograms of reaction times (go cue to movement initiation) for correct (top) and error (bottom) trials pooled across all data. Difference in distributions is significant ($p < 10^{-4}$) using a two-sample Student's t-test. **B)** Comparison of SSIMS analysis and classification using all correct and error trials (All) versus using all error trials and only correct trials with reaction times longer than the median error reaction time (Long) shows no significant difference in classification accuracy ($p = 0.4408$) using a two-sample Student's t-test.

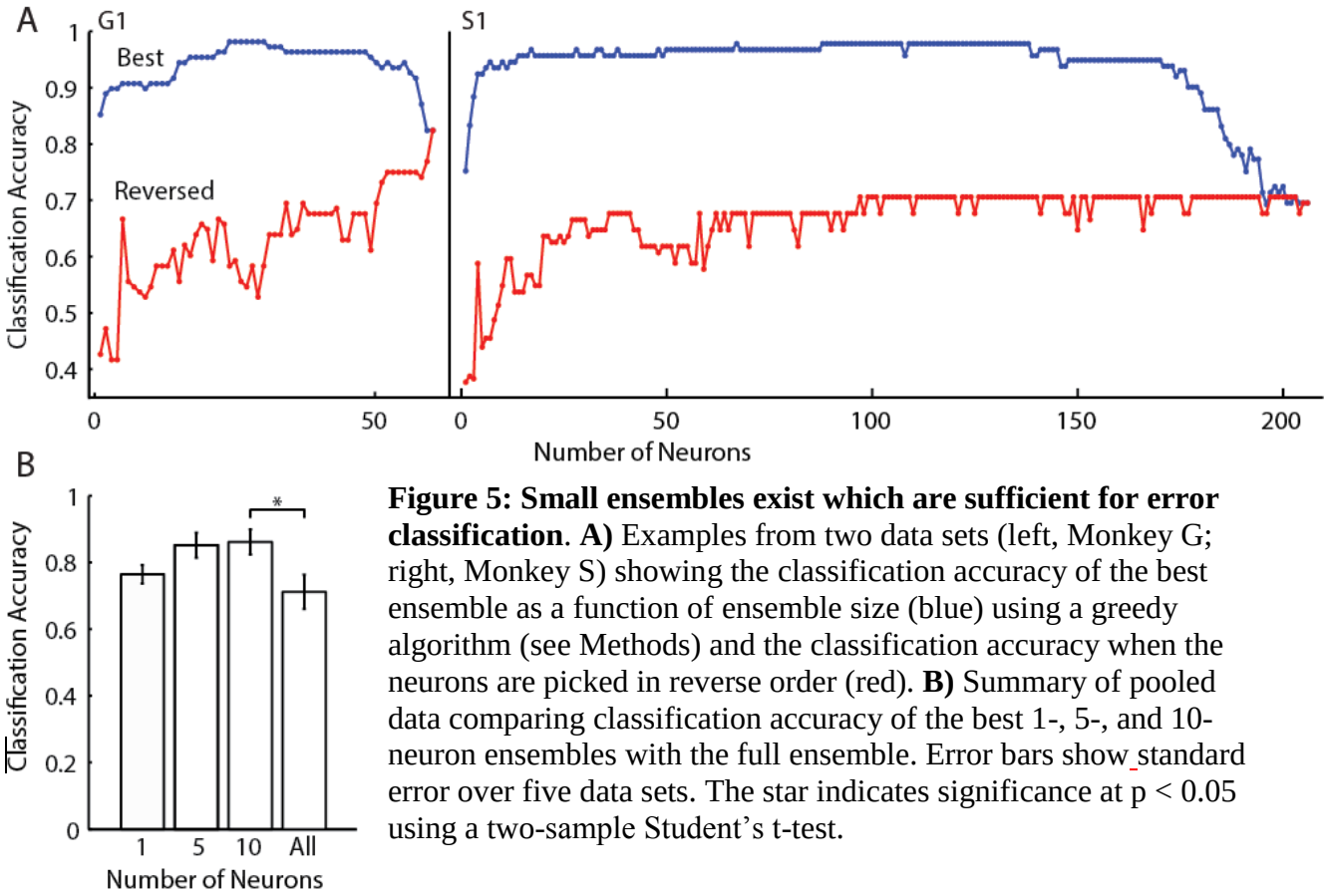

Figure 5: Small ensembles exist which are sufficient for error classification. **A)** Examples from two data sets (left, Monkey G; right, Monkey S) showing the classification accuracy of the best ensemble as a function of ensemble size (blue) using a greedy algorithm (see Methods) and the classification accuracy when the neurons are picked in reverse order (red). **B)** Summary of pooled data comparing classification accuracy of the best 1-, 5-, and 10-neuron ensembles with the full ensemble. Error bars show standard error over five data sets. The star indicates significance at $p < 0.05$ using a two-sample Student's t-test.

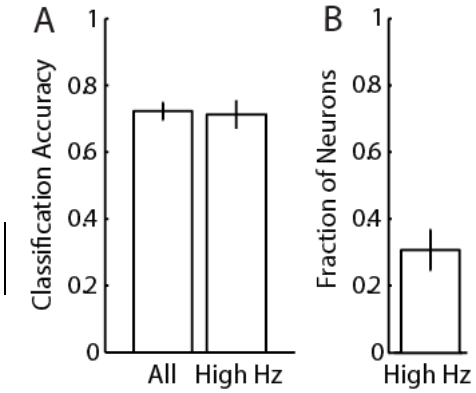

Figure 6: Neurons with high firing rate are sufficient for error decoding. **A)** Classification accuracy from pooled data using all neurons versus using only neurons with an average firing rate greater than 5 Hz during the instructed delay. Error bars show standard error across five data sets. **B)** Fraction of neurons with average firing rate greater than 5 Hz. The error bar shows standard error across 5 data sets.

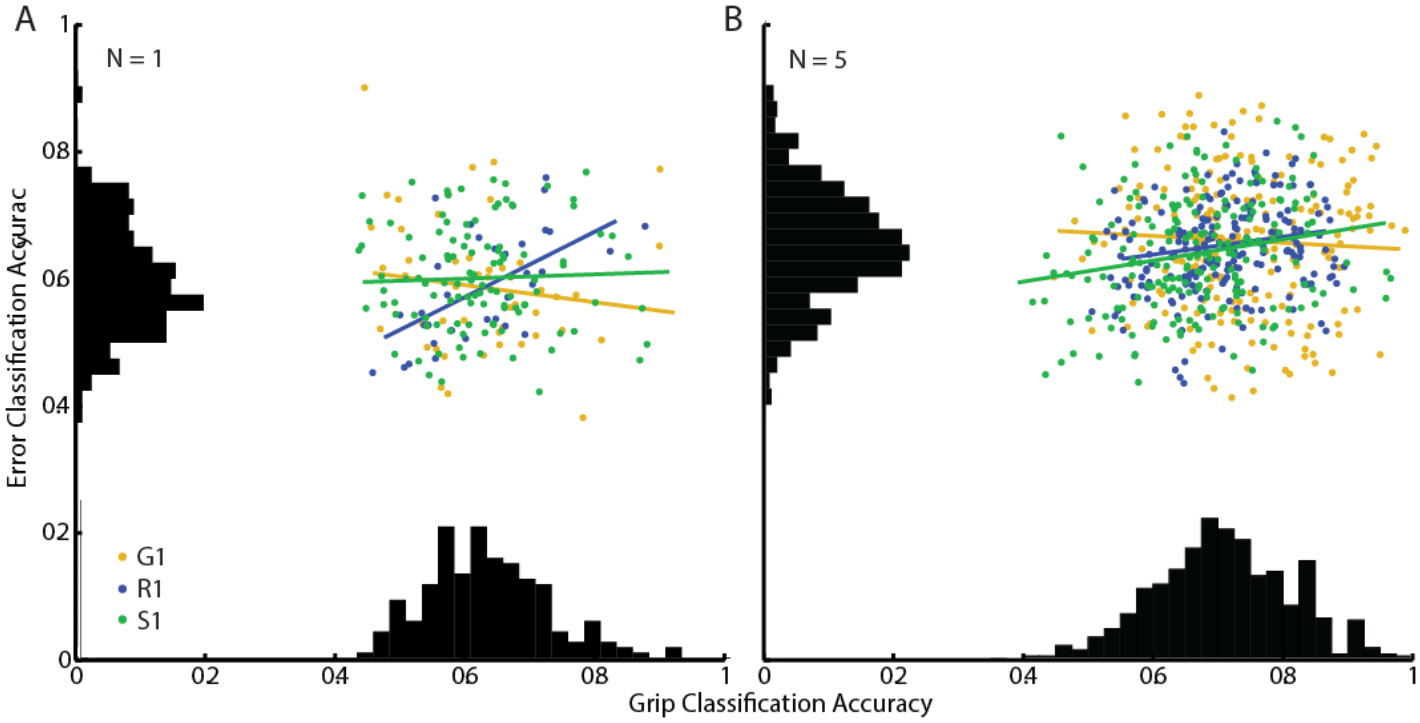

Figure 7: Grip classification accuracy is not a consistent predictor of error classification accuracy. **A)** Scatterplot of single neurons' grip and error classification accuracy with marginal histograms for three data sets (one for each monkey). Least-squares lines of best fit are plotted for each monkey (r^2 = 0.0190 (G1), 0.3630 (R1), and 0.0015 (S1)). **B)** Same as (A) using randomly chosen ensembles of ten neurons (r^2 = 0.0030 (G1), 0.0353 (R1), and 0.0422 (S1))

References

Bledowski, C., Prvulovic, D., Goebel, R., Zanella, F. E., & Linden, D. E. (2004). Attentional systems in target and distractor processing: a combined ERP and fMRI study. *Neuroimage*, *22*(2), 530-540.

Botvinick M, Nystrom LE, Fissell K., Carter, C. S., & Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature, 402*(6758), 179-181.

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., ... & Freund, H. J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European journal of neuroscience*, *13*(2), 400-404.

Carmichael, S. T., & Price, J. L. (1995). Sensory and premotor connections of the orbital and medial prefrontal cortex of macaque monkeys. *Journal of Comparative Neurology*, *363*(4), 642-664.

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, *280*(5364), 747-749.

Carter CS, Botvinick MM, Cohen JD (1999) The contribution of the anterior cingulate cortex to executive processes in cognition. *Reviews in the Neurosciences 10*:49-58.

Collinger, J. L., Wodlinger, B., Downey, J. E., Wang, W., Tyler-Kabara, E. C., Weber, D. J., ... & Schwartz, A. B. (2013). High-performance neuroprosthetic control by an individual with tetraplegia. *The Lancet*, *381*(9866), 557-564.

Dehaene, S., Posner, M. I., & Tucker, D. M. (1994). Localization of a neural system for error detection and compensation. *Psychological Science*, 303-305.

Ding, L., & Gold, J. I. (2010). Caudate encodes multiple computations for perceptual decisions. *The Journal of Neuroscience*, *30*(47), 15747-15759.

Ehrsson, H. H., Fagergren, A., Jonsson, T., Westling, G., Johansson, R. S., & Forssberg, H. (2000). Cortical activity in precision-versus power-grip tasks: an fMRI study. *Journal of neurophysiology*, *83*(1), 528-536.

Escera, C., Alho, K., Winkler, I., & Näätänen, R. (1998). Neural mechanisms of involuntary attention to acoustic novelty and change. *Journal of cognitive neuroscience*, *10*(5), 590-604.

Fogassi, L., Gallese, V., Buccino, G., Craighero, L., Fadiga, L., & Rizzolatti, G. (2001). Cortical mechanism for the visual guidance of hand grasping movements in the monkey A reversible inactivation study. *Brain*, *124*(3), 571-586.

Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. *The Annals of Statistics*, 70-84.

Hochberg, L. R., Serruya, M. D., Friehs, G. M., Mukand, J. A., Saleh, M., Caplan, A. H., ... & Donoghue, J. P. (2006). Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, *442*(7099), 164-171.

Hochberg, M. C., Altman, R. D., April, K. T., Benkhalti, M., Guyatt, G., McGowan, J., ... & Tugwell, P. (2012). American College of Rheumatology 2012 recommendations for the use of nonpharmacologic and pharmacologic therapies in osteoarthritis of the hand, hip, and knee. *Arthritis care & research*, *64*(4), 465-474.

Hoover, J. E., & Strick, P. L. (1993). Multiple output channels in the basal ganglia. *Science*, *259*(5096), 819-821.

Kakei, S., Hoffman, D. S., & Strick, P. L. (2001). Direction of action is represented in the ventral premotor cortex. *Nature neuroscience*, *4*(10), 1020-1025.

Kim, S. P., Simeral, J. D., Hochberg, L. R., Donoghue, J. P., Friehs, G. M., & Black, M. J. (2007, May). Multi-state decoding of point-and-click control signals from motor cortical activity in a human with tetraplegia. In *Neural Engineering, 2007. CNE'07. 3rd International IEEE/EMBS Conference on* (pp. 486-489). IEEE.

MacDonald, A. W., Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science*, *288*(5472), 1835-1838.

Manthey S, Schubotz RI, von Cramon DY (2003) Premotor cortex in observing erroneous action: an fMRI study. *Cognitive Brain Research 15*:296-307.

Masimore, B., Schmitzer-Torbert, N. C., Kakalios, J., & Redish, A. D. (2005). Transient striatal γ local field potentials signal movement initiation in rats. *Neuroreport*, *16*(18), 2021-2024.

Miltner, W. H., Braun, C. H., & Coles, M. G. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a "generic" neural system for error detection. *Journal of cognitive neuroscience*, *9*(6), 788-798.

Mink, J. W. (1996). The basal ganglia: focused selection and inhibition of competing motor programs. *Progress in neurobiology*, *50*(4), 381-425.

Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P., & Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: Evidence from an antisaccade task. *Psychophysiology*, *38*(5), 752-760.

Nordhausen, C. T., Maynard, E. M., & Normann, R. A. (1996). Single unit recording capabilities of a 100 microelectrode array. *Brain research*, *726*(1), 129-140.

Pandya, D. N., Van Hoesen, G. W., & Mesulam, M. M. (1981). Efferent connections of the cingulate gyrus in the rhesus monkey. *Experimental brain research*, *42*(3-4), 319-330.

Redgrave, P., Rodriguez, M., Smith, Y., Rodriguez-Oroz, M. C., Lehericy, S., Bergman, H., ... & Obeso, J. A. (2010). Goal-directed and habitual control in the basal ganglia: implications for Parkinson's disease. *Nature Reviews Neuroscience*, *11*(11), 760-772.

Rizzolatti, G., Matelli, M., & Pavesi, G. (1983). Deficits in attention and movement following the removal of postarcuate (area 6) and prearcuate (area 8) cortex in macaque monkeys. *Brain*, *106*(3), 655-673.

Romo, R., Hernández, A., & Zainos, A. (2004). Neuronal correlates of a perceptual decision in ventral premotor cortex. *Neuron*, *41*(1), 165-173.

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*(2579-2605), 85.

Vargas-Irwin, C., & Donoghue, J. P. (2007). Automated spike sorting using density grid contour clustering and subtractive waveform decomposition. *Journal of neuroscience methods*, *164*(1), 1-18.

Vargas-Irwin, C. E., Brandman, D. M., Zimmermann, J. B., Donoghue, J. P., & Black, M. J. (2014). Spike Train SIMilarity Space (SSIMS): A Framework for Single Neuron and Ensemble Data Analysis.

Victor, J. D., & Purpura, K. P. (1996). Nature and precision of temporal coding in visual cortex: a metric-space analysis. *Journal of neurophysiology*, *76*(2), 1310-1326.

Vidal, F., Hasbroucq, T., Grapperon, J., & Bonnet, M. (2000). Is the 'error negativity'specific to errors?. *Biological psychology*, *51*(2), 109-128.

Wodlinger, B., Downey, J. E., Tyler-Kabara, E. C., Schwartz, A. B., Boninger, M. L., & Collinger, J. L. (2015). Ten-dimensional anthropomorphic arm control in a human brain−machine interface: difficulties, solutions, and limitations. *Journal of neural engineering*, *12*(1), 016011.