Decoding of eye movement trajectory by neuronal ensembles in cortex

Bing Liu and J. Patrick Mayo

Abstract:

The vast majority of BMI research has focused on decoding arm movement signals in primary motor cortex to control external prosthetic devices. However, since the cortical signals used for implementing BMIs and how to translate these signals into movements are still in question, it is useful to investigate the neural responses in the oculomotor system and eye movement behavior, which contain many fewer degrees of freedom than arm movements. Here, we propose a different strategy that decodes eye motion trajectories using neuron ensembles recorded from cortical areas FEF and MT. Our results show that, using different decoding algorithms, the eye motion position could be accurately decoded. The modern neural network methods provide decent decoding performance with 75% (87%) variance explanations (correlation coefficient is 0.87 (0.94), $p<0.001$), with an averaged position error of 1.3deg (0.84deg). The reconstructed eye motion trajectory could successfully finish 61% (84%) of the trials with an error range of 6deg (in behavioral experiment, the range is 3deg). These results demonstrated the feasibility of implementing a generalized gaze BMI, provided the neural substrate of this implementation, and also supplied a powerful framework to investigate neuroscience questions related to eye motion information processing.

The vast majority of BMI (brain-machine interface) research has focused on decoding arm movement signals in primary motor cortex to control external prosthetic devices. Despite recent progress in BMI, questions remain about the recorded cortical signals and how best to translate them into movements. The eye has fewer degrees of freedom so, in principle, it should be easier to develop a BMI for dynamic motor control using neural signals from oculomotor or possibly visual cortex. Here, we investigate the plausibility of an oculomotor BMI by analyzing the relationship between cortical activity and eye movement trajectories in behaving macaques. We recorded ensemble neuronal activity in the frontal eye fields (FEF) in prefrontal cortex and visual area MT via laminar probes while male monkeys tracked moving visual targets. Using multiple decoding algorithms as well as modern neural network methods, we show that eye movements can be accurately decoded with relatively high precision. In an example session with 69 simultaneously recorded neurons (45 FEF; 24 MT), a neural-network decoder accounted for 87% variance of eye position over time (correlation coefficient between decoded trajectories and data is 0.94, $p<0.001$), with an average position error of 0.84 degrees. The reconstructed eye motion trajectories were successfully reproduced on 84% of the trials with an error constraint of 6 degrees (compared to 3 degrees in behavior). Our results demonstrate the feasibility of a gaze BMI, identify an adequate cortical substrate for its implementation, and supply a novel framework for investigating questions in oculomotor control.

Method:

Experiments paradigm:

During the experiment, monkeys perform a pursuit task. Eye motion velocities and positions are recorded through Maestro. The targets used in this experiment have 4 motion directions (90deg apart), 2 contrast (12% and 100%), and 2 speeds (10dps and 20dps). For now, all the data are used to perform the decoding, ignoring different contests in the experiments. During the experiment, first the monkey was required to fix on the target at the beginning of experiment, then pursuit the target motion while target start to move. The constraint of eye positions is inside 1 degree and 3 degrees ranges of target position in horizontal and vertical directions during fixation the pursuit. In the dataset I analyzed, monkey finished 2095 trials. Further analyzing is needed for investigating the impacts of different contexts, number of neurons, and trials to the decoding performance.

Decoding and algorithms:

In general, here we demonstrated how to reconstruct the eye motion trajectories by decoding eye position continuously using population neuron responses recorded from cortical areas of FEF and MT (42 neurons, 22 from FEF and 20 from MT). The monkey performed a typical pursuit task, while a lot of saccades also happened during pursuit. For each trial of behavioral task, we took 500ms and 300ms before stimuli onset and 500ms after stimuli stop. Together with the 800ms target motion time, we have 1600ms (500+800+300=1600) data for each trial. Both the eye motion position and neural data are binned in a non-overlap 50ms time window. During the decoding, we used 6 bins of neural data both before and after the current motion frame (300ms before and after the current 50ms data), to decode the position of current eye motion position.

Multiple algorithms were used to perform the decoding, including Winner Filter (WF), Winner Cascade Filter (WC), Extreme Gradient Boosting (XG Boost), support vector regression (SVR), standard recurrent neural network (Simple RNN), Dense feed forward network (DenseNN), a neural network with gated recurrent units (GRU), Long Short Term Memory networks (LSTM). The decoding tools used are from Kording lab (https://github.com/KordingLab/Neural_Decoding). The Winner Filter methods are normally used in traditional BMI work, while more complicated network methods are more sophisticated and potentially supply better decoding results. Hyperparameter optimizations were performed using BayesianOptimization provided by (https://github.com/fmfn/BayesianOptimization). For now, 70% of data are used to build the decoding model, 15% data are used as testing data set, and 15% data are used as the validation data set.

Quantifying the goodness of decoding

Several methods are used to quantify the goodness of decoding. First, we used a scoring metric of R^2,

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (y_i - \bar{y})^2}$$

where $\hat{y}_i$ are the decoded eye position, y_i are the eye motion position, and $\bar{y}$ are the mean value of y_i. This R^2 indicated the fraction of variances that are accounted for by our decoding models. In our results, the R^2 reported are averaged across decoding horizontal and vertical eye motion positions. We also calculated the Pearson's correlation coefficient between decoded positions and eye positions. These different calculations lead to similar conclusions to our data.

Experiment reconstructions:
We also quantify the goodness of decoding by simulating the behavioral experiment using the reconstructed eye motion trajectory. During this simulation, we used different levels of criteria comparing with that used in behavioral task. Then we calculated the percentage of trials that could be successfully performed by the decoded data. Our results show that using the 2 times of the constraints used in behavioral experiments (2 degrees and 6 degrees during fixation and pursuit), the decoded eye movement could successfully finish 61% of the total trials in the validation dataset.

Neuronal direction presentation:
To understand the neural coding of eye motion in different contests and in different eye motion types, first we begin from the direction tuning. The direction modulation indexes are calculated to quantify the direction tuning of neurons across population in both MT and FEF, and also in different contexts of motion speeds and contrasts.

$$DI = (R_{pref} - R_{null})/(R_{pref} + R_{null})$$

Results show that MT neurons are more modulated by stimuli directions in general, however, there are some neurons in FEF are strongly modulated by stimuli direction (it is interesting to see if these neurons are more involved during decoding). Further more, results show that the stimuli contrasts don't have a strong influence on direction tuning, while stimuli speeds do.

Results:
1/ Eye motion position could be decoded with a decent goodness score using cortical neurons from FEF and MT.

1. Decoding eye position using different decoders. Traditional methods give a worse performance comparing the modern neural network methods (Figure 1A, B).

2. The LSTM performance outperforming other methods, with R^2 of 0.75, and correlation coefficient of 0.87 (Figure 1A, B).

3. The distribution of error residuals is -0.46$\pm$2.5deg and -0.27$\pm$2.2deg for horizontal and veridical positions (Figure 1C).

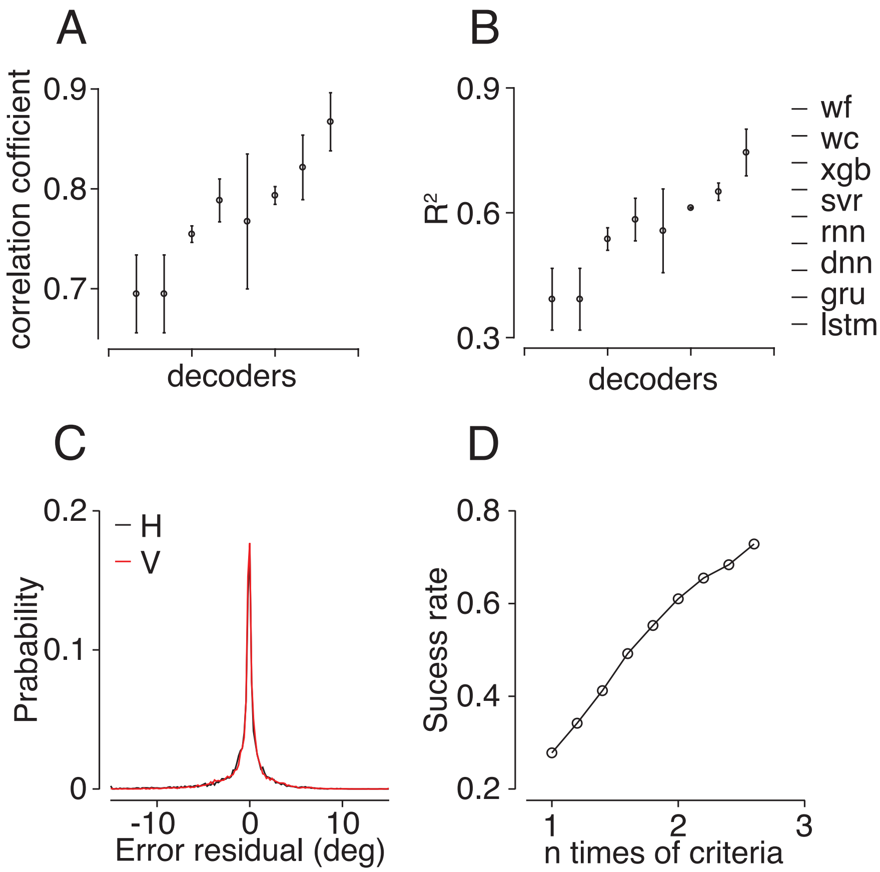


Figure 1. The decoding performance of different algorithms and simulation of reconstructing the experiments. The correlation coefficient (A) and R^2 (B) between decoded position and eye motion position using different algorithms. (C) The distribution of position error of horizontal (black) and vertical (red) positions between the data and decoded results. (D) The rate of successfully performing the behavioral task using the decoded eye motion trajectories with different levels of criteria of task.

4. Further analysis focused on using this LSTM method, we reconstructed the behavioral experiment using decoded data. With criteria of 2 times of behavioral experiments, the decoded data could finish 61% of all the trails (Figure 1D).

5. These results indicate we could decode the eye motion position with a decent goodness scoring using this number of neurons (22 FEF and 20 MT neurons), and trials (2095 trials all together ignoring different contexts). So I think we are good to go further asking more scientific questions.

6. We also show some reconstructed data examples to give a direct visualization of what is going on (Figure 2).

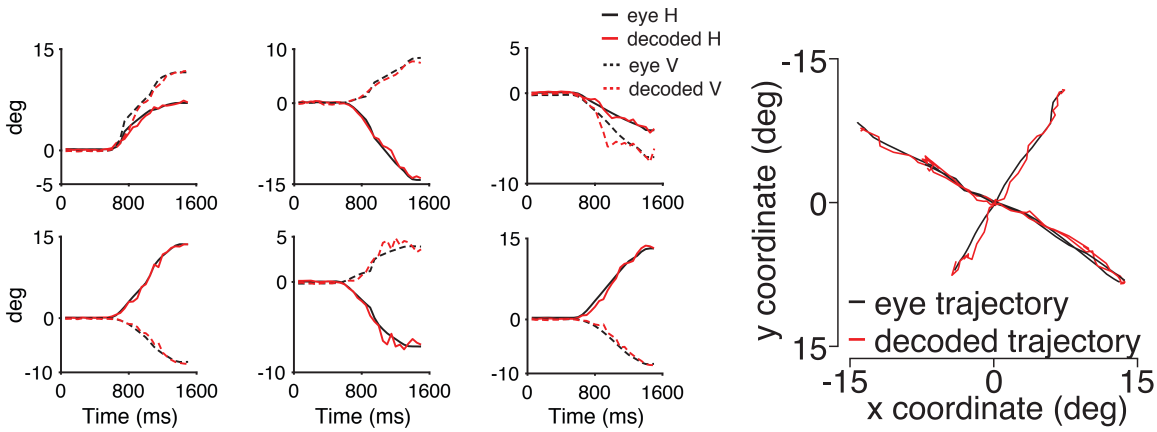


Figure 2. The reconstructed eye motion trajectory using LSTM algorithm. (A) Six example trials of eye motion positions (black) and decoded trajectories (red) in horizontal (solid line) and vertical (dashed line) coordinates across time. (B) Reconstructed eye motion trajectories (red) and their eye motion trajectories (black) in a 2D space. The data are the same as in (A).

Further work:

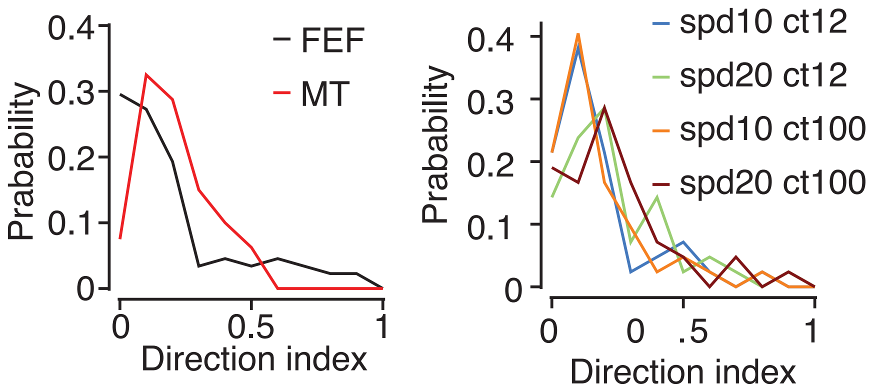
Here are some works I am working on or I plan to work on. I also wish I could give another conclusion before the abstract submission (I am working on the trial/neuron number question and MT/FEF separation question now, question 2 and 3).

1. Improving decoding performance by optimizing the bin window size and the bin numbers used to decode. Especially how would the bin numbers that before and after the current frame influence the decoding will supply some indications on the function of MT and FEF areas (feed forward information or feedback of some efferent copies).

2. Run a simulation experiment to drop neurons and trials, the idea is trying to investigate how many neurons and trials are needed to perform a decent decoding (I am not sure if we can also try to investigate the lateralizing effects, or H/V error-preferred direction correlations).

3. Separate the MT and FEF population, investigating their roles in decoding eye motion positions.

4. How will the visual contexts influence the decoding? From some preliminary analysis of only the neuron data, it seems contrast may only played as a gain function, and the stimuli speed will influence the neuron direction tuning, which potentially will influence the decoding (However, the contrast will also potentially influence the Fano Factor, or the neural correlation level, which also potentially influence the decoding).



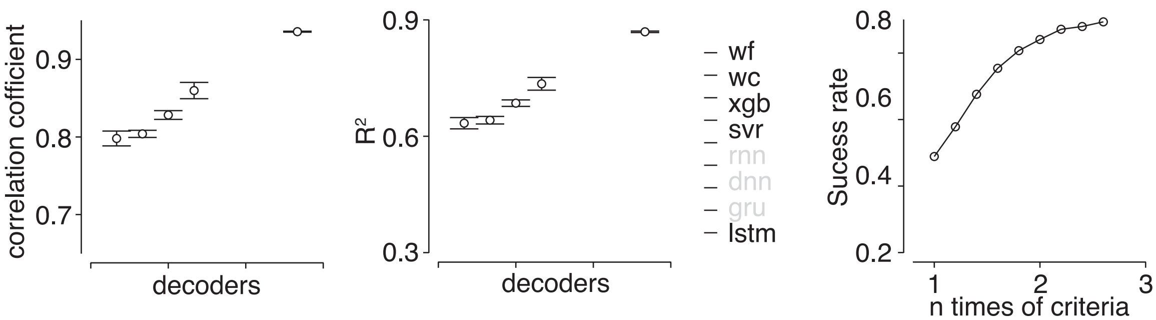
Supplementary Figure 1. The distribution of neural direction tuning indexes for FEF neurons and MT neurons at different contexts conditions. A. The distribution of neural direction tuning indexes for FEF population and MT population. B. The distribution of neural direction tuning indexes at different stimulation conditions.

5. What kind of signal is better to decode the eye movement? I will get LFP with different range of frequencies to test the question. Also the spike and LFP combined signal will be used to run the decoding.

6. For now, the separation of pursuit and saccade is not very important, but it will be good if we can supply some evidences on this. We are decoding the position directly, which is good since we also avoid one problem of “not accurate” recording of eye velocity of saccade using our Maestro system (the filtering range are different, and this is build in at the hardware level, meaning we cannot change it).

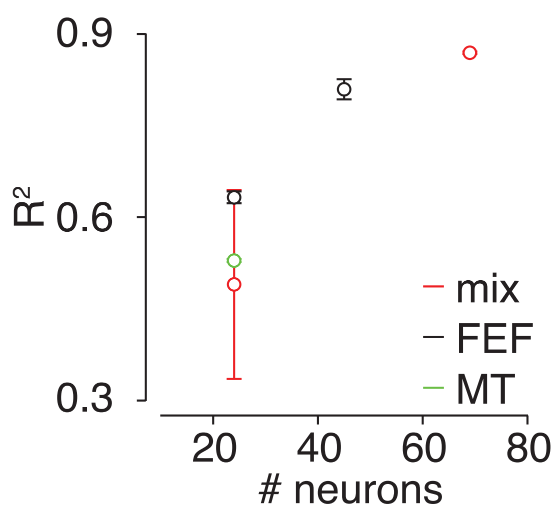
7. It will be great if I can have some free eye moving data to see if the decoder could also work during the free motion.

Results from another data set:



Supplementary Figure 2. The same figures as Figure1, but for another data set. In this data set, we have 2396 trials (which is similar as the previous data set), and 69 neurons with 45 in FEF and 24 in MT (the FEF neurons are almost doubled comparing with previous data set, MT neuron number is similar). During the decoding, I used several methods (but not the rnn, dnn, and gru, which are time consuming to finish, but I tried the algorithm that gave us the best results, LSTM).

Conclusion: In general, these results gave us similar conclusions as we got from previous dataset. However, the decoding results are much better comparing with previous one. It is not very clear why this is the case. But my hypothesis is the number of neurons and trials are both very important for decoding, this improvement in decoding may drive from the increased number of neurons.



Supplementary Figure 3. A simplified comparison contributions of FEF and MT populations to eye motion position decoding.

Methods:
In this dataset, we have 69 neurons (45 FEF, and 24 MT neurons). We run the decoding with LSTM algorithm. To minimizing the effect of neuron number to the decoding, we randomly selected 24 FEF neurons, or 24 mixed FEF/MT neurons from the whole population (I got 10 different random populations), then run the same decoding again.

Preliminary results:
1. This preliminary result indicates FEF contributing more than MT to this eye motion position decoding.
2. The decoding performance with MT neurons dropped comparing with the same number of FEF neurons. However, MT population also supplied some information about the eye motion position.
3. The further detailed analyses are needed.

Several questions related:
1. What kind of neurons provided the most informative responses? More spikes are more important for decoding? Or more directional selective is more important?
2. What parts of the neural responses are important? The Spikes before (intention) the stimuli or the spikes after (feedback or efference copies)?
3. We could also supply the mutual information analysis, which is supposed to provide the similar results.

Patrick's suggestions:
We could start sketching out figures or general topics to cover in figures. 1) MT vs FEF decoder accuracy; 2) improvement in decoder performance when including motion feedback vs no-motion feedback; 3) effect of context on decoder performance; 4) decoder performance as a function of number of neurons/channels; 5) decoder performance for different task conditions (e.g., building a decoder on 100% contrast data and seeing how it performs on the 12% contrast data); 6) LFPs. Those topics would make a very nice and useful paper, I think.

I really like your idea of trying to sketching out figures now. It will clarify our thinking (or maybe we can ask the specific questions we want to understand, before the figures we want to plot).
According to your questions:
1) MT vs FEF decoder accuracy; 2) improvement in decoder performance when including motion feedback vs no-motion feedback;
To me, these are interesting, but complex questions. It includes, 1) decoding motion information (can we do this using MT neurons, and what it really means). 2) decoding sensory information (how to separate sensory and motion information in FEF). 3) which time range to use (are we using the neural encoding or feedback information). 4) how will the combinations improve decoding (but can we use these combinations while in online decoding system). To simplify this, I think we can start from if MT neurons will decode the motion or not (depending on time range), and will this improve the decoding with FEF neurons. I am happy to get more suggestions on these topics.

3) effect of context on decoder performance; 5) decoder performance for different task conditions (e.g., building a decoder on 100% contrast data and seeing how it performs on the 12% contrast data);
I think this is an important point, but does not get enough investigation. In general, neurons will adapt at different contexts. But how would these adaptation would influence the decoding is not clearly understood. I think we should show some detailed influence of the contexts to the decoding. And it will be more interesting if we can try some online experiment in the further.

4) decoder performance as a function of number of neurons/channels; 6) LFPs.
This is the major questions I'd like to answer. I also want to take a look of how good the decoding is with multi-units.

According to these, there are 3 parts of things to investigate: 1) some basic properties of decoding, including the number of neurons, channels, LFPs (I think another question is how would pursuit and saccade be decoded in the same population of neurons). 2) how would the sensory information from other areas, or feedbacks influence decoding. 3) how would the contexts influence the decoding. I will try to get some basic results on these and we can talk about these further while I got some figures.

1. The first question is can we decode the single trial eye motion trajectory? So how do we do that?
The answer is we can try to decode the position, other than the velocity.

2. The second question is what kind of signal can we use to decode this eye position?
Areas: FEF / MT?
Type of signal: single unit / LFP / multi-unit
Time range: what is the best time bin and time range for this decoding?
Only use the signal before the motion, or also use the time after motion (this is the feedback).

3. The third question is if the decoder could be generalized, as the decoder used in 100% can also be used in 12% contrast? This is the decoding in different contexts.

Find the best parameters:
1/ different bin size
2/ with or without the motion feedback signal
3/ the length of data used to decode.

Neuron dropping experiment:
1/ random dropping neurons
2/ only using FEF neurons and dropping.
3/ only using MT neurons and dropping.

BMI in different contexts:
1/ some neuronal presentation
2/ getting decoder using one condition, decode the results in another condition.
3/ which condition will change the decoding, and which won't. The key question is why, and how to leverage this property.

Generalized eye movement prediction:
1/ I need some this kind of data.

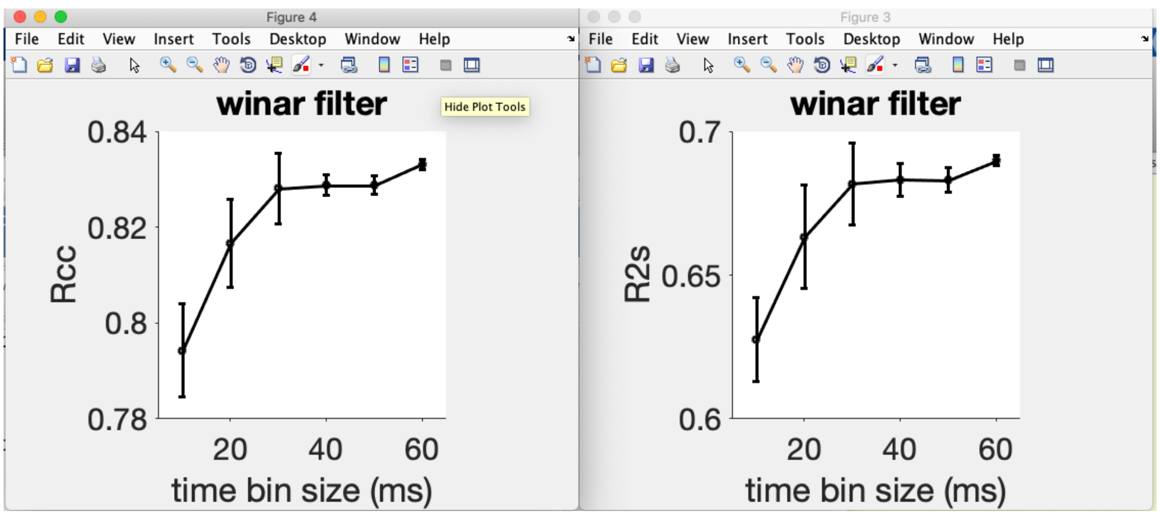
Online experiment:
1/ The decoder may all change.

Some parameter results:

1. For the different decoders, it seems the paramotors are similarly influenced. So fuscous on the simplest decoder (Winner filter, this is faster).

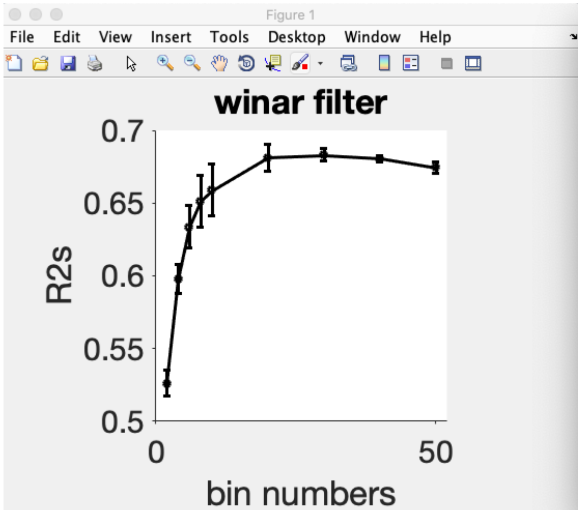
1. bin size:

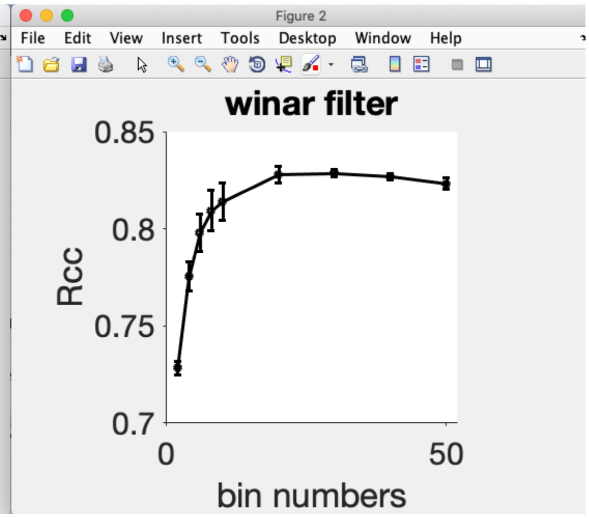
Intuitively, longer bin size should be better for decoding. But this bin size is important for the time scales for online decoding (we prefer real time decoding, which request the shorter bin size). 30ms is also okay, but the variance is higher. 40ms, 50ms is better (we are using 50ms in our data). all data were tested with 30bins. In general, around 1000ms data will get better results.



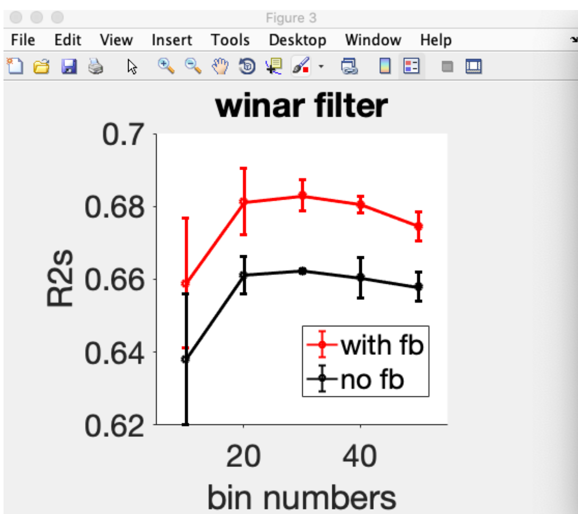
2. bin numbers:

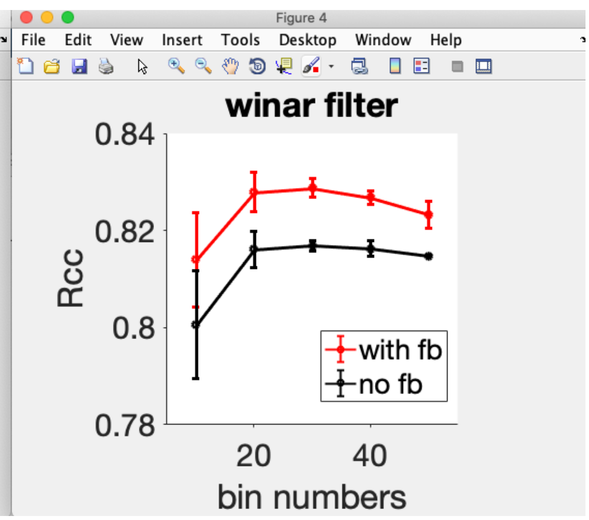
Intuitively, more bin number should be better for decoding. In the results, 20 bins is the best (this is similar comparing with previous results, that, using 1000ms data will give the best decoding results). (we were using 6 bins in our results). All data were tested with 50ms bins. So in general, we should use longer time scale of data to perform the decoding. But we decide to use the 6 bins, which give 50*6=300ms length of data. The reasons are: 1. The decoding results is higher than 95% of the best decoded results (average results through 20 to 50ms). 2. We are trying to investigate the feasibility of future online usage, which will need shorter time of data.





3. The same data as in 2, but without the feedback part (only using the neuron data before the movement). In general, the conclusion is: the decoding results will be better with the motion feedback information (3% higher at the best parameter location, even week, but significant).





4. The effect of the feedback. We used different feedback time bins to calculate the decoding results. 1/ The decoding results are higher if there is feedback. 2. More feedback bins will derive better decoding results. 3. The longer bins are not needed, 5 bins seems enough for the decoding (5*50ms=250ms, but even the 2 bins (100ms) is good enough for the performance saturation).

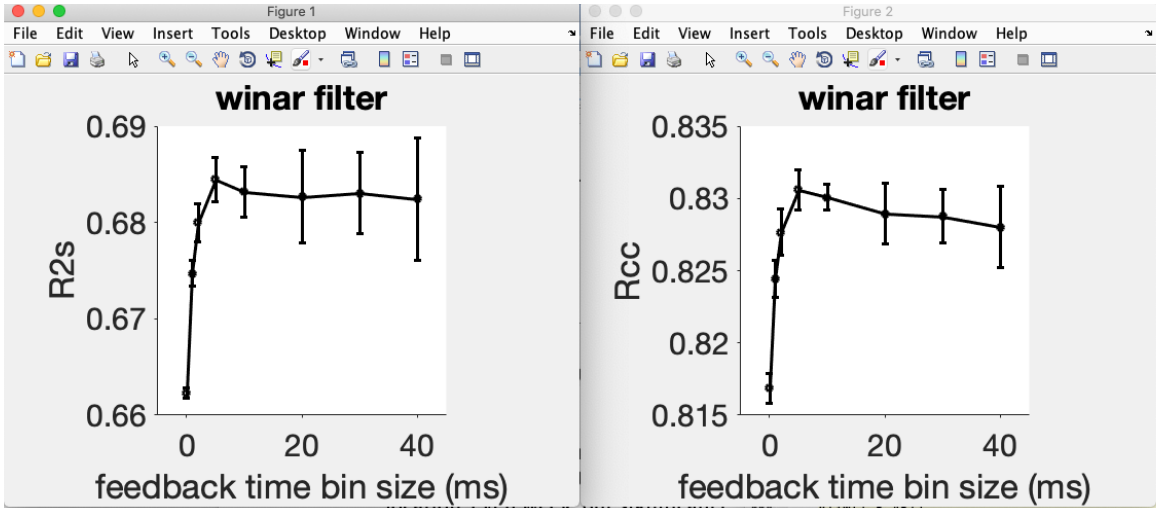


Figure 1
File Edit View Insert Tools Desktop Window Help
winar filter
R2s
0.69
0.68
0.67
0.66
0
20
40
feedback time bin size (ms)
Figure 2
File Edit View Insert Tools Desktop Window Help
winar filter
Rcc
0.835
0.83
0.825
0.82
0.815
0
20
40
feedback time bin size (ms)