1- "Significance:...difficulty of the task": We will add the following to sec.4: The DRNN relies on neural data inputs–not just the kinematic feedback or target information–based on the following evidence. First, target information is not explicitly provided to the DRNN. Any target information available to the DRNN is learned from the neural data and/or feedback components. Second, DRNN outputs change substantially based on different feature engineering approaches (Figures 5,6) and over different trials (with the same features) (Figure 4, Rebuttal Figure 1a). Finally, predictions fail when the DRNN uses only feedback (Feedback-Only), feedback with noise substituted for neural data (Feedback-Noise), or feedback with the neural data provided only at the beginning of the trials (Short-Neural) (Rebuttal Figure 1b).

2- "Clarity...additional description of the motivation": See sec.2 and lines 239-246. Our DRNN is unique since it learns a matrix that establishes a relationship between the past and present derivative states unlike the conventional DRNN. Also our DRNN, which learns all the hyper-parameters of the model by using back propagation through time (BPTT), is distinct from F-DRNN that only learns the output weight by using RLS algorithm. Moreover, its application differs from existing decoders that have been applied to motor cortex data of a non-human primate. We present the first demonstration of applying feedback and scheduled sampling to a DRNN and comparing different learning based decoders operating on different features to predict kinematics from PPC data of a human subject in the BMI setting.

3- "...how real the hardware problem is...": See sec.5. We will add the following: BMIs are intended to operate as wireless, implantable systems that require low-power circuits, small physical size, wireless power delivery, and low temperature deltas ($\leq 1^{\circ}$C) (Dewhirst 2003). By choosing efficient algorithms that map well to CMOS technologies, ASIC implementations could offer substantial power and mobility benefits.

4- "...show the wavelet features were best for a range of decoders...": We will add results for all the decoders to supplementary material. Wavelet features show superior results for all the decoders (Example: Rebuttal Figure 1c).

5- "Clarity...why does feeding back the decoders' output recurrently result in better performance/more robust decodes?": We will add the following to sec.2: Feeding the output back to the input recurrently in addition to the input neural data provides more information to the DRNN to make predictions, which results in a smaller network with less history. Analogous to the gain term of the Kalman filter, the DRNN learns the relative importance of the neural data and feedback. Integrating both state and neural information in this way leads to smoother predictions (Figure 4a).

6- "...compare with standard approaches": We did compare (Figures 2, 4, 7, 8) our DRNN with 12 standard decoders that are addressed in (Glaser 2017) including Kalman filter, SVR, XGBoost, NN, RNN, GRU, and LSTM.

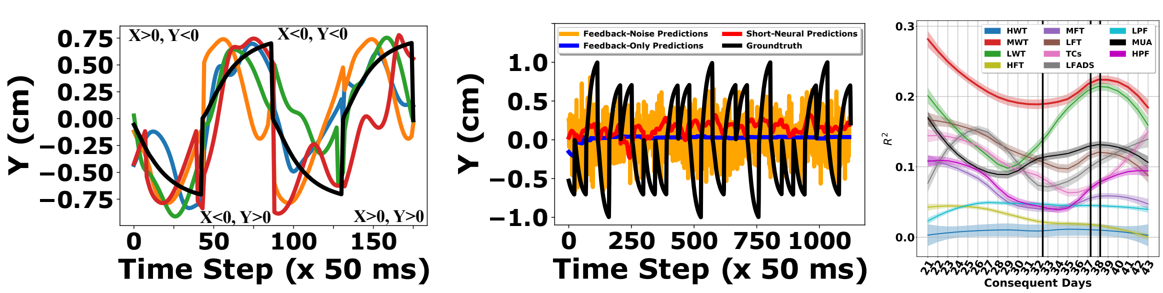
7- "...how the decoding performance differed from traditional BMI brain regions": While a full discussion is outside the scope of this paper, we will add the following after line 48: PPC processes a rich set of high-level aspects of movement including sensory integration, planning, and execution (Aflalo 2015) and may encode this information differently (Zhang 2017). These characteristics of PPC differentiate it from other brain areas and, while providing a large amount of information to the decoder, also require new paradigms, such as those discussed here, to extract useful information.

8- Question about trial definition: We will add this definition to sec.4: A trial is one trajectory of the cursor from the center of the screen to one of the eight targets on a unit circle (Figure 3).

9- "...disconnect between Figure 7 and ... Figure 8": There is no disconnect since Figure 7 reports the R^2 for single days, whereas Figure 8 shows the R^2 averaged over 23 test days by excluding 3 worst days.

10- "...why the more complex model is doing worse?": More layers do not always result in superior performance of neural networks (lines 66-71). See supplementary table 1 for optimum parameters of the DRNN.

11- "Clarity...lines 192-194...": Since training and testing are done on single days in cross-day analysis, these lines do belong under the single-day analysis section.



(a) DRNN predictions for sample targets (b) DRNN predictions - no/short neural data (c) LSTM predictions

Rebuttal Figure 1. (a) and (b) Comment 1: true target motion (black) and reconstructions (colored), (c) comment 4.