nature biomedical engineering

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41551-020-0542-9

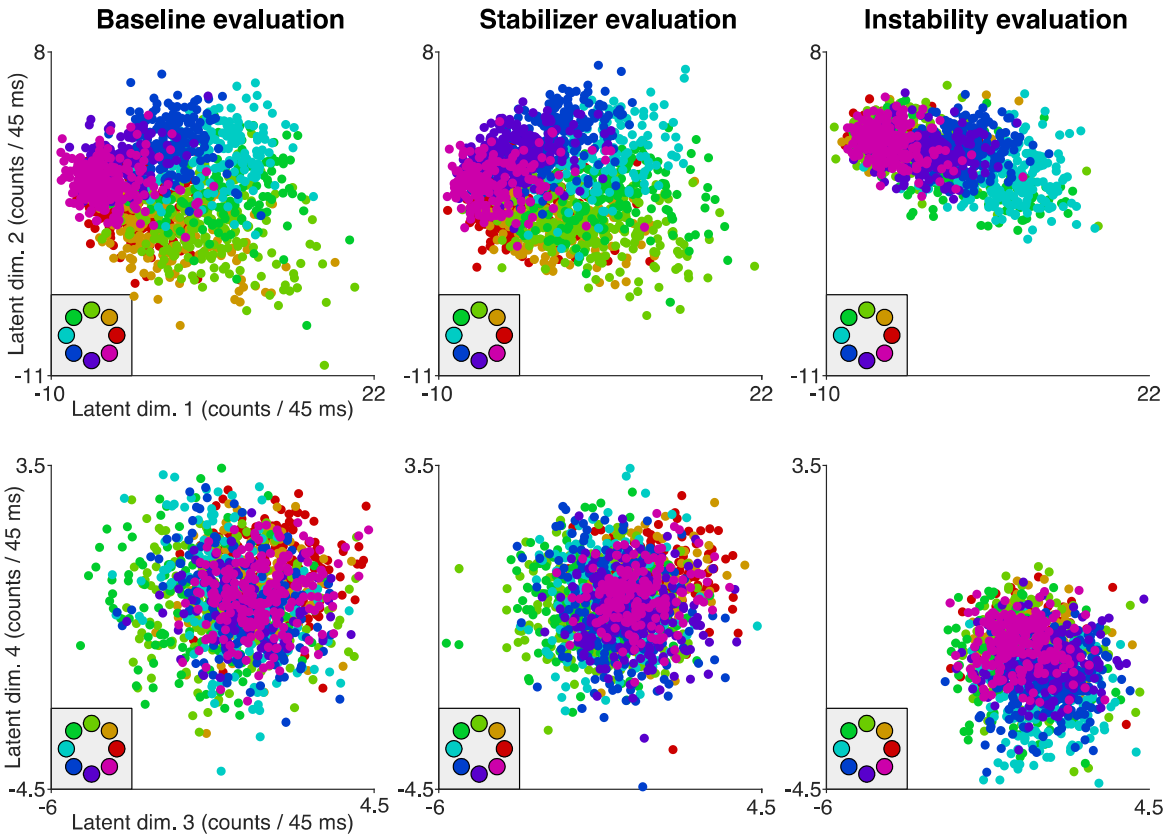
In the format provided by the authors and unedited.

Stabilization of a brain–computer interface via the alignment of low-dimensional spaces of neural activity

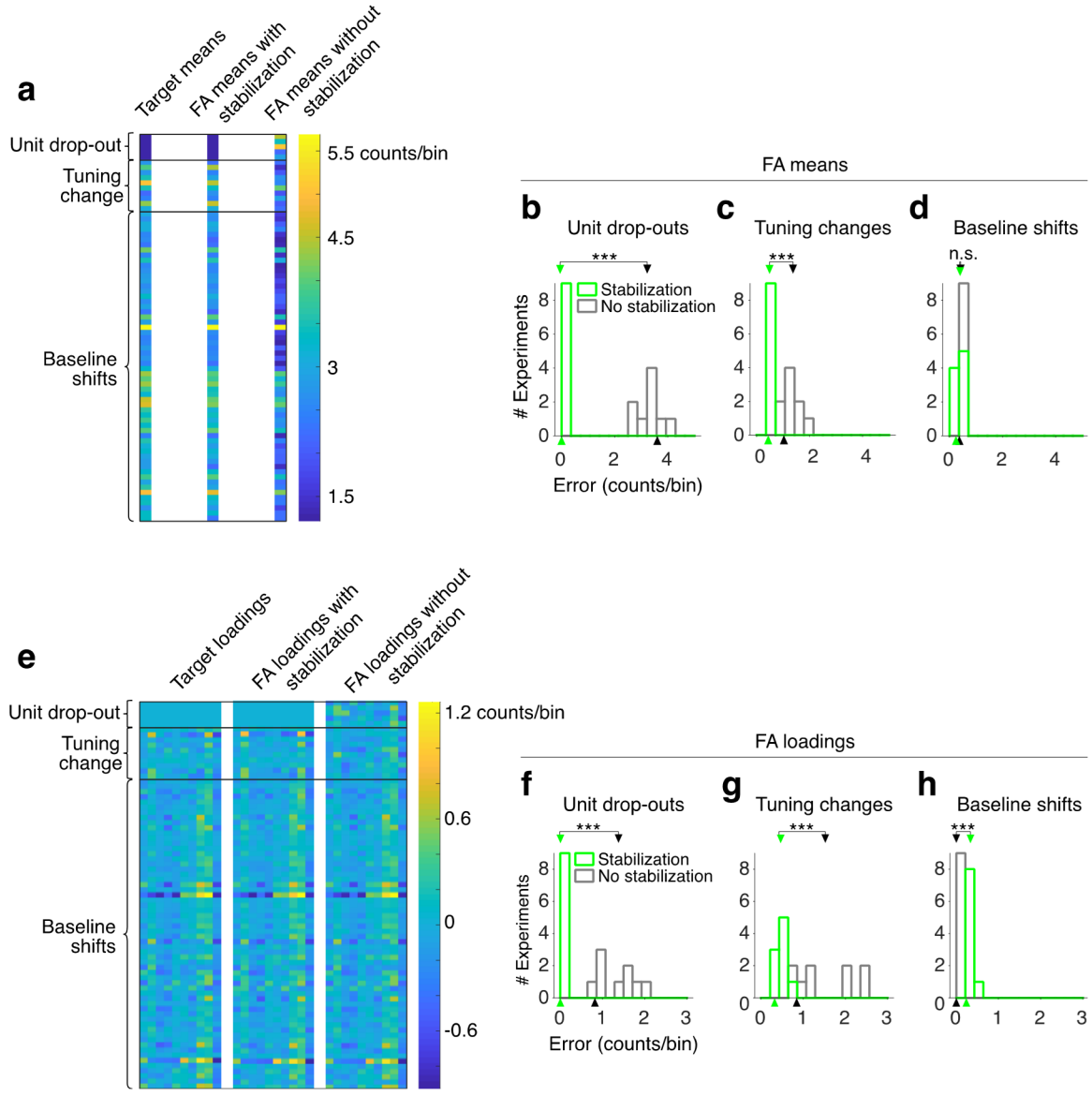
Alan D. Degenhart[1,2,3,4,5,14], William E. Bishop[3,6,7,14], Emily R. Oby[2,3,4,5,8], Elizabeth C. Tyler-Kabara[2,9,10,11], Steven M. Chase[3,12,13,15], Aaron P. Batista[2,3,4,5,15] and Byron M. Yu[1,3,12,13,15] ✉

[1]Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA. [2]Department of Bioengineering, University of Pittsburgh, Pittsburgh, PA, USA. [3]Center for the Neural Basis of Cognition, Pittsburgh, PA, USA. [4]Brain Institute, University of Pittsburgh, Pittsburgh, PA, USA. [5]Systems Neuroscience Center, University of Pittsburgh, Pittsburgh, PA, USA. [6]Department of Machine Learning, Carnegie Mellon University, Pittsburgh, PA, USA. [7]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA. [8]Department of Neurobiology, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA. [9]Department of Neurological Surgery, University of Pittsburgh, Pittsburgh, PA, USA. [10]Department of Physical Medicine and Rehabilitation, University of Pittsburgh, Pittsburgh, PA, USA. [11]McGowan Institute for Regenerative Medicine, University of Pittsburgh, Pittsburgh, PA, USA. [12]Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA, USA. [13]Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA, USA. [14]These authors contributed equally: Alan D. Degenhart, William E. Bishop. [15]These authors jointly supervised this work: Steven M. Chase, Aaron P. Batista, Byron M. Yu. ✉e-mail: byronyu@cmu.edu

Table of contents

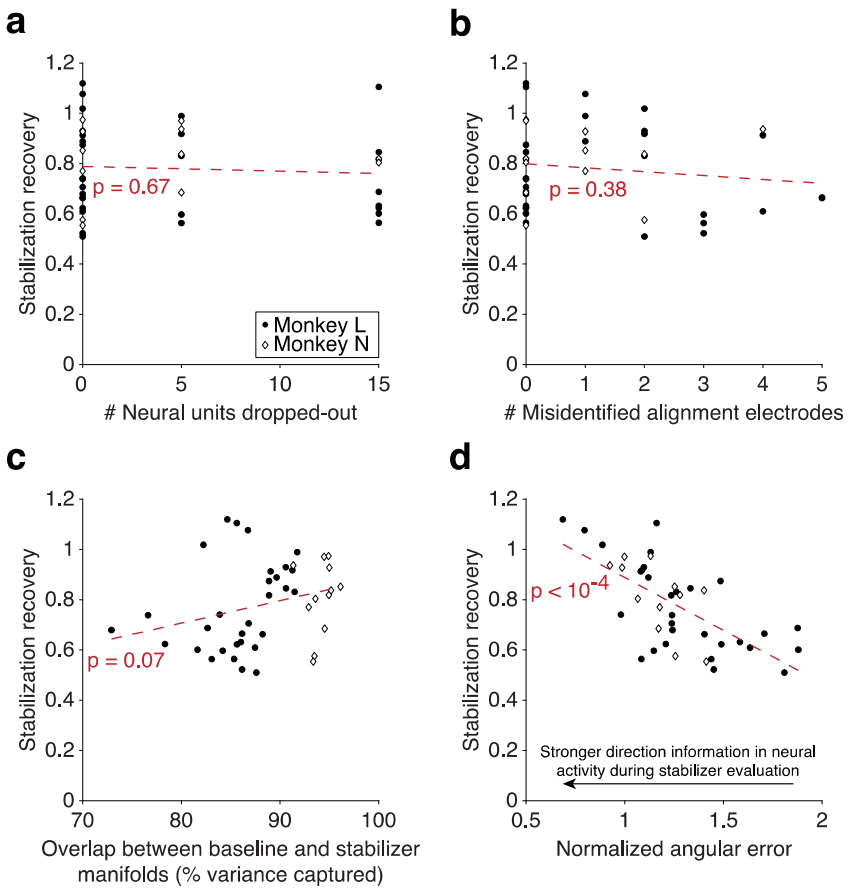


Supplementary Fig. 1 | Visualization of stabilized latent states. Shown are two projections of the latent states for the single-day experiment depicted in Fig. 4 (L20160325). The left column shows neural activity during the baseline evaluation block, before the introduction of a neural recording instability. Following the introduction of the instability and application of stabilization, the structure of the neural activity during the stabilizer evaluation block (middle) closely matches that of the baseline evaluation block, allowing BCI performance to be restored. The neural activity during the instability evaluation block (right) is shifted relative to that during the baseline evaluation block, which results in a decrease in BCI performance when using a non-stabilized decoder. Each row corresponds to a different two-dimensional orthogonal projection of the same 10-dimensional latent space. Neural activity is colored by target position (see panel inset). The projections shown are those capturing the most shared variance of the trials used to calibrate the base decoder and stabilizer parameters, which correspond to the top four orthonormalized dimensions of the base factor analysis model. Axes represent orthonormalized dimensions of the base factor analysis model (e.g., latent dimension 1 is the first orthonormalized dimension). Latent dimensions 1, 2, 3, and 4 capture 56.4%, 14.1%, 8.1% and 5.3% of the shared variance of the calibration trials for this experiment, respectively. Spike counts are mean-centered during factor analysis, which allows spike count values to be negative.



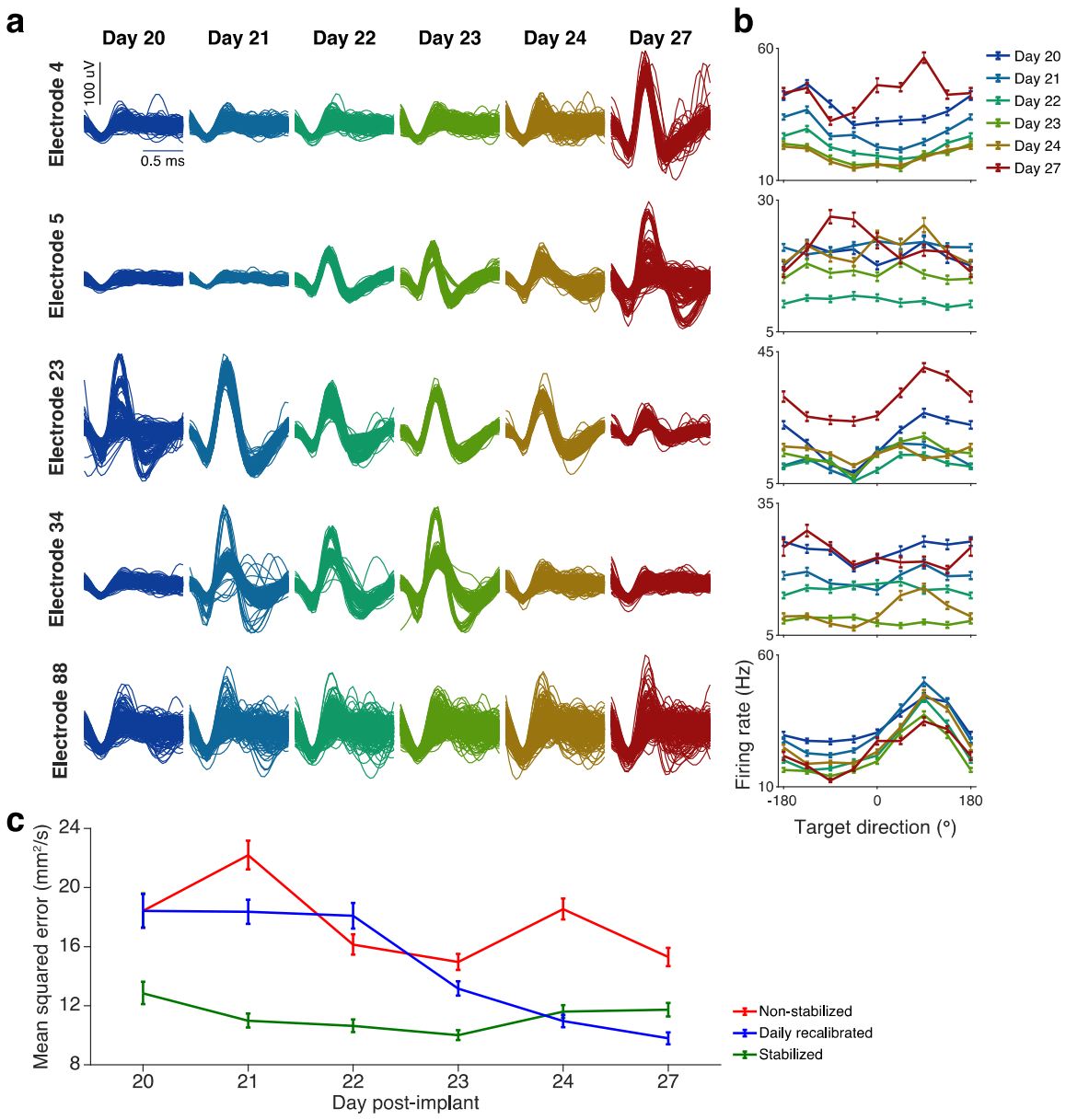
Supplementary Fig. 2 | Stabilization accurately recovers factor analysis parameters to account for instabilities. To better understand how stabilization is able to correct for neural recording instabilities, we compared the parameters of the factor analysis (FA) model (describing the relationship of the recorded neural activity to the manifold) with and without stabilization. Specifically, we compared the mean vector and loading matrix of the FA model with and without stabilization to the "target" parameters representing perfect recovery by stabilization. Target mean and loading matrix parameters were formed by using knowledge of the applied instabilities to modify the baseline FA parameters. **a**, Target mean parameters for each electrode and mean parameters of the FA model with and without stabilization for an example combination-instability experiment (L20160324). Rows of the mean vectors have been sorted so that rows corresponding to electrodes with unit-drops, tuning changes and baseline shifts are grouped together. **b-d**, Histograms of average errors between target and FA model mean parameters across all combination-instability experiments (9 experiments total) for each of the different types of applied instabilities. Average error across experiments is significantly reduced with stabilization for electrodes with unit drop-outs and tuning changes ($p < 0.001$ in both cases, permutation test). Average error for tuning changes with stabilization is non-zero because tuning change instabilities also included a baseline shift. Average error across experiments was not significantly different ($p = 0.38$, permutation test) when considering electrodes with baseline shifts. This is likely due to the small baseline shifts which were applied (offsets were drawn from a $\mathcal{N}(0.375 \text{ cnts/bin}, (0.25 \text{ cnts/bin})^2)$ distribution). For each experiment, the average error between target and FA model means was calculated for each group of electrodes with different types of instabilities. The average error for a single experiment was the average absolute value of the difference

between the FA mean parameters and the target mean parameters across electrodes. Histograms show the number of experiments with error values in the indicated binned ranges with (*green*) and without (*gray*) stabilization. Arrows near the top of each plot denote the average error across experiments. Arrows in the bottom indicate the measured average errors for the sample experiment shown in (**a**). We note that stabilization was not given knowledge of which electrodes corresponded to unit drop-out or tuning change instabilities, meaning that the ability of stabilization to recover the appropriate parameters in these cases is not trivial. **e**, The target loading matrix and the loading matrix of the FA model with and without stabilization for the same experiment as in (**a**). Rows are sorted as in (**a**). **f-h**, Same as in (**b-d**), but for the errors between target loadings and loadings with and without stabilization. Error for an electrode is defined as the L^2 distance between its row in the target loading matrix and its row in an FA model's loading matrix. Across experiments, stabilization significantly reduced the average error of loadings for subsets of electrodes with unit drop-outs and tuning changes ($p < 0.001$, permutation test). For baseline shifts, error was higher with stabilization ($p < 0.001$, permutation test). However, the target loadings for electrodes with baseline shifts was defined to be equal to their loadings in the baseline FA model, so the measured error between the target loadings and those without stabilization for baseline shift electrodes was by definition 0. All statistical tests in (**b-d**) and (**f-h**) are two-sided.



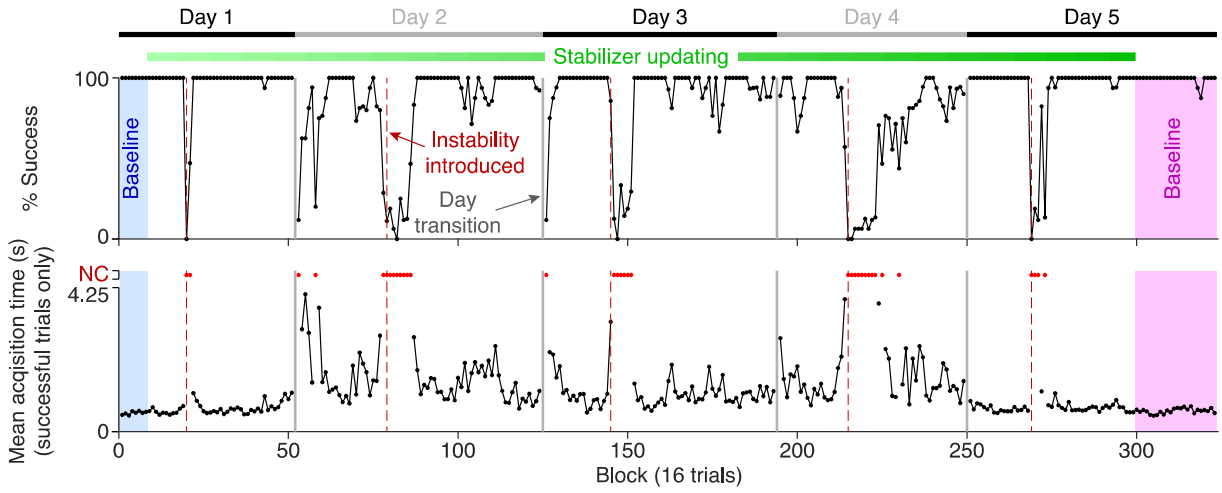
Supplementary Fig. 3 | Factors influencing stabilization performance. To investigate why stabilization performance varied across single-day experiments, we measured the correlation between stabilizer performance recovery and a set of factors hypothesized to affect stabilization. Stabilization recovery was defined as the mean target acquisition rate (TAR) during the stabilization evaluation block divided by the mean TAR during the baseline evaluation block. A stabilization recovery value of 1 indicates stabilization performance equal to that observed during the baseline evaluation block, whereas stabilization recovery values less than 1 indicate that stabilization did not fully recover the performance observed during the baseline evaluation block. *p*-values indicating significance of Pearson's correlation, assessed using a two-sided permutation test (10^4 permutations), are reported in each panel. **a**, Effect of the number of dropped-out neural units on stabilization recovery. The drop-out of neural units may lead to a reduction in cursor velocity information in the neural activity, possibly resulting in lower stabilization performance. We found a small but statistically insignificant decrease in stabilizer performance for larger numbers of dropped-out electrodes. This suggests there was sufficiently redundant cursor velocity information in the neural activity recorded from the remaining electrodes to sustain decoding performance. **b**, Effect of the number of unstable electrodes used for alignment on stabilization performance. Our ability to properly align manifolds could decrease if our method for identifying stable electrodes mistakenly identifies unstable electrodes as stable, possibly resulting in poorer stabilizer performance. We found a small but statistically insignificant decrease in performance as more unstable electrodes are used for alignment. This suggests that stabilization performance is robust to a small number of unstable electrodes incorrectly used for manifold alignment. **c**, Effect of manifold recovery on stabilizer performance. Variability in our ability to recover the underlying manifold could conceivably affect stabilizer performance. The amount of manifold overlap, quantified by the percent variance captured (see Methods), was calculated between the baseline calibration manifold and final stabilizer manifold. The percentage of variance captured computes how much of the variance of the latent state described by one manifold is also captured by another. If the manifold identified by the stabilizer matches perfectly with the manifold identified during baseline calibration, the percent of variance captured would be 100%. We observed a slight but statistically insignificant increase in stabilizer performance with higher degrees of manifold overlap. However, for all single day experiments the stabilizer manifold captured over 70% of the variance of the baseline calibration manifold. This indicates that even though the degree of manifold overlap varied across experiments, it was still high in all cases. **d**, Effect of the

strength of cursor direction signals during stabilizer evaluation trials on stabilization recovery. Even if stabilization is able to perfectly correct for recording instabilities, stabilization performance could vary if the amount of directional information in the neural activity changes over time. We used a decoding analysis to measure the strength of directional information in the neural activity before and after stabilization. We calibrated decoders to predict intended cursor velocity during the baseline and stabilizer evaluation blocks of each experiment. Separate decoders were calibrated for each evaluation block. These decoders were used only for assessing the amount of directional information in the neural activity and were separate from those used during closed-loop control. Calibration was performed using neural activity from the final two bins (90ms) of the initial 300ms "freeze" period at the beginning of each BCI trial. During this period, the animal could see the BCI target and begin to modulate their neural activity, but was not yet provided control of the cursor. We chose to decode neural activity from the open-loop period in order to minimize the influence of corrective movements which can occur during closed-loop control. We found a significant relationship between the amount of cursor-related information in the neural activity (as quantified using normalized angular error) and stabilization performance. Decoders were calibrated and angular error was calculated in the same manner as described in the "Comparing stabilization to supervised retraining" section of the Methods. Average angular error of decoded cursor velocities during the stabilizer evaluation block was normalized by dividing by the error of the baseline evaluation block. Even though stabilization can outperform supervised recalibration because it is less sensitive to the amount of directional information present in the neural activity than supervised recalibration (cf. Fig. 8), decoding performance with stabilization is still sensitive to the amount of direction information present in the neural activity. The amount of directional information present in the neural activity can vary possibly due to the animals' motivation in volitionally modulating its neural activity or because the animals sometimes attempted to compensate for the instability in a way that leads the neural activity to have less directional information. Note that while the we decode neural activity from the same trials used to measure stabilization recovery, this analysis is not circular because we fit new decoders during the cross-validation procedure. Furthermore, these analyses use data from non-overlapping parts of the trials; stabilization performance is measured during the closed-loop part of the trial, whereas the offline decoding is based on the open-loop part of the trial. Finally, it is important to note that in contrast to Fig. 8d, which shows how stabilizer performance varies with the amount of directional information in the data used for stabilization, this analysis shows how directional information in the stabilizer evaluation block (when the parameters of the stabilizer were fixed) is related to stabilizer performance recovery. Results in (**a-d**) are based on 42 independent experiments.

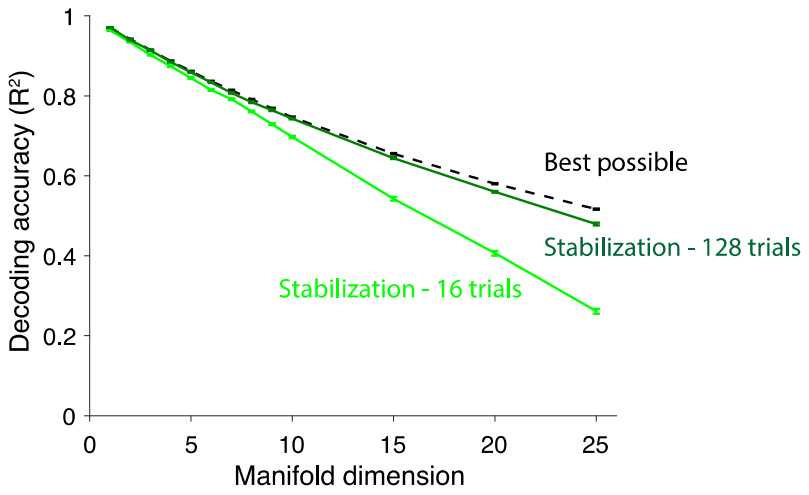


Supplementary Fig. 4 | Stabilization overcomes natural neural recording instabilities. a, Waveforms from 5 selected electrodes over the first 6 recording days following electrode implantation for Monkey N. Data were recorded during a center-out reaching task where the animal was cued to reach to one of 8 possible targets. Day numbers indicate the number of days after electrode implantation. For some electrodes (e.g., electrodes 4, 5, 23, and 34) the waveform shape changes across days, indicating the presence of neural recording instabilities. For other electrodes (e.g., electrode 88), the waveforms are consistent across days, suggesting the existence of a stable set of electrodes to be used in stabilization. **b**, Tuning curves for selected electrodes for the same days as shown in panel (**a**). Average firing rate, estimated over a 500ms window aligned to the onset of the reaching movement, is plotted as a function of target direction. The tuning curves for the selected electrodes with variable waveform shapes (electrodes 4, 5, 23, and 34) exhibit changes over the course of the 6 recording sessions, indicative of neural recording instabilities. The color of each tuning curve corresponds to the recording day, as indicated in panel (**a**). Tuning curves were calculated based on 500 trials per day, randomly distributed across target direction (range: 51-80 trials/target direction). **c**, Hand velocity decoding performance using fixed (*red*), daily-recalibrated (*blue*), and stabilized (*green*) decoders. Stabilization is able to sustain decoding performance in the presence of neural recording instabilities, significantly outperforming the non-stabilized decoder for each of the 6 days tested ($p < 10^{-3}$ each day, two-sided Wilcoxon signed-rank test). Furthermore, stabilization outperforms the daily-recalibrated decoder for the first 4 days (days 20 to 23, $p < 10^{-3}$, two-sided Wilcoxon signed-rank test), suggesting that the stabilized decoder is able to compensate for

natural recording instabilities that occur during this time period. In all three cases, a base decoder was calibrated on 128 target-balanced trials collected at the beginning of the first recording day (Day 20). We then compared performance of the three methods over the course of 6 days. For the non-stabilized decoder, the decoder calibrated on the first day was used for all subsequent days. For the daily-recalibrated decoder, a new decoder was calibrated each day using 128 target-balanced trials collected at the beginning of the day. For the stabilized decoder, stabilization parameters were updated every 16 trials using a 128-trial sliding window. Decoding performance is quantified by the mean-squared error between the actual and decoded hand velocities using each of the three methods. The same set of evaluation trials was used for all three methods. Note that performance of the daily-recalibrated decoder improves over the course of the 6 days due to the strengthening of neural signals recorded by the electrode array post-implant. As the signals on the array improve (e.g., through an increase in the number of isolatable single units), the information about user intent is increases, leading to improvements in the performance of the daily-recalibrated decoder. Error bars represent the 95% confidence interval of the mean. For this analysis, we used all available electrodes (96) in the decoder, set number of alignment electrodes (B) to 75, and used an alignment threshold (T) of 0.01 counts/bin (see Methods, “Identifying stable electrodes”).



Supplementary Fig. 5 | Multi-day experiment for Monkey L. Success rate (*top*) and target acquisition time (*bottom*) are shown over the course of a 5-day experiment for Monkey L. Same conventions as in Fig. 7.



Supplementary Fig. 6 | Stabilization is robust to increases in manifold dimensionality. We investigated the performance of stabilization as the dimensionality of the underlying manifold increases, as could be the case for tasks with increasing task complexity (e.g., controlling a robotic limb with many degrees of freedom). To do this, we assessed how stabilization was able to recover simulated *m*-dimensional control signals embedded in an *m*-dimensional manifold in the presence of applied recording instabilities. Conceptually, the control signals can be thought of as representing the independent task parameters that a BCI user seeks to control (e.g., velocity of different joints of a robotic limb). Stabilization performance was quantified as the R^2 between ground-truth control signals and those estimated with stabilization after the introduction of instabilities (mean $\pm$ SE). Stabilization performance was compared to an upper bound on performance ("best possible"), which was calculated by estimating control signals with a ground-truth stabilizer which perfectly captured the instabilities applied in each simulation. Upper-bound performance decreases with increasing dimensionality due to the need to estimate an increasing number of control signals from a fixed number of electrodes. We found that stabilization performance decreases relative to that of the upper bound with increasing manifold dimensionality. This is due to the need to estimate more parameters defining a manifold with a fixed amount of data. We find that increasing the amount of data used for stabilization (e.g., going from 16 to 128 trials) recovers performance, suggesting that stabilization should continue to work with increasingly complex BCI tasks as long as sufficient data is used for stabilization.

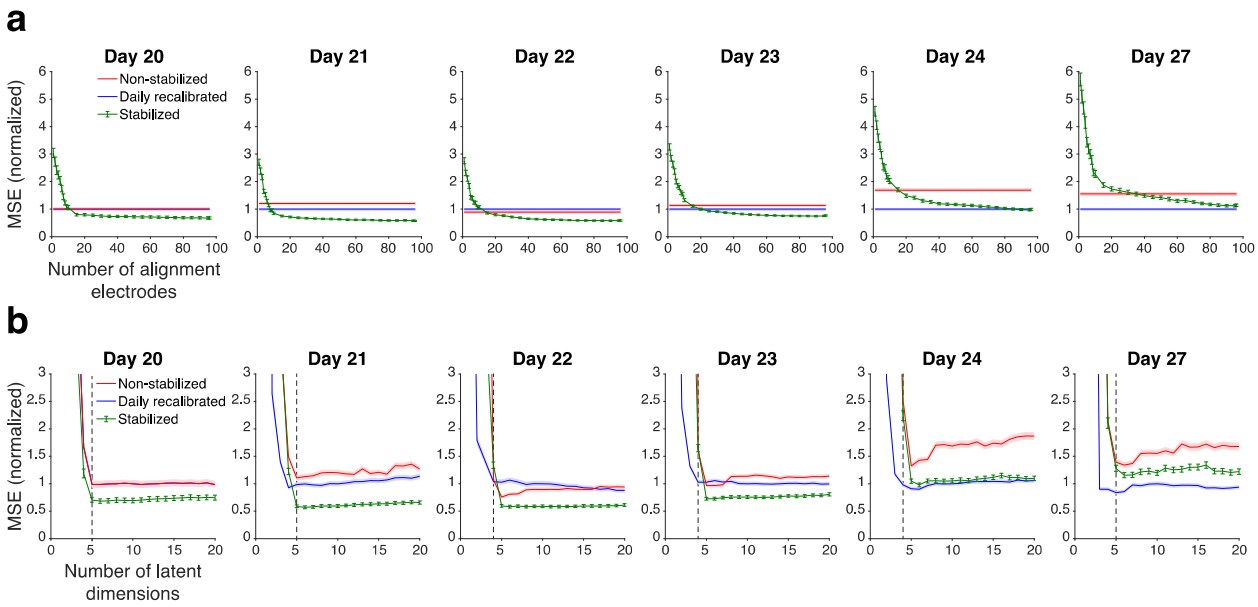
We simulated neural activity from 85 electrodes in the following way. Control signals were sampled from $z_t \sim \mathcal{N}(0, I)$, where *I* represents an $m \times m$ matrix. Neural activity was generated according to $u_t = \Lambda z_t + \mu + \epsilon_t$ for a matrix $\Lambda \in R^{85 \times m}$ and $\epsilon_t \in R^{85}$ where $\epsilon_t \sim \mathcal{N}(0, \Psi)$ and $\Psi \in R^{85 \times 85}$ is a diagonal covariance matrix. Parameters for this model were randomly generated in a manner so that the statistics of simulated neural activity qualitatively matched that of the spike counts recorded during the single-day experiments. Entries of Λ were drawn i.i.d. from a $\mathcal{N}(0.02, 0.27^2)$ distribution and entries of μ were drawn i.i.d. from a $\mathcal{N}(2.1, 0.83^2)$ distribution, where means and standard deviations of these distributions were selected to match those of the entries of Λ and μ estimated during calibration of the single-day baseline stabilizers. Diagonal entries of Ψ were drawn from a $\mathrm{Uniform}(1,2)$ distribution and then all entries of Ψ were scaled so that the percent shared variance of the neural population, defined as $100 * \mathrm{trace}(\Lambda\Lambda^T)/\mathrm{trace}(\Lambda\Lambda^T + \Psi)$, was equal to 32%, which was the average percent shared variance measured across the single-day baseline stabilizer models.

Each simulation consisted of three blocks of trials: (1) a calibration block in which a baseline stabilizer was fit to data from 128 trials, (2) a stabilization block in which the stabilizer was updated using either 16 or 128 trials, and (3) a stabilization evaluation block in which stabilization performance was evaluated using 16 trials. Between the calibration and stabilization blocks, an instability was applied to the simulated neural activity. We chose to use either 16 or 128 trials for stabilization updates because 16 trials was the minimum number of trials used for the first stabilization updates in the closed loop experiments, and 128 trials was the number of trials used for stabilization updates when the sliding buffer of our implemented stabilizer was full.

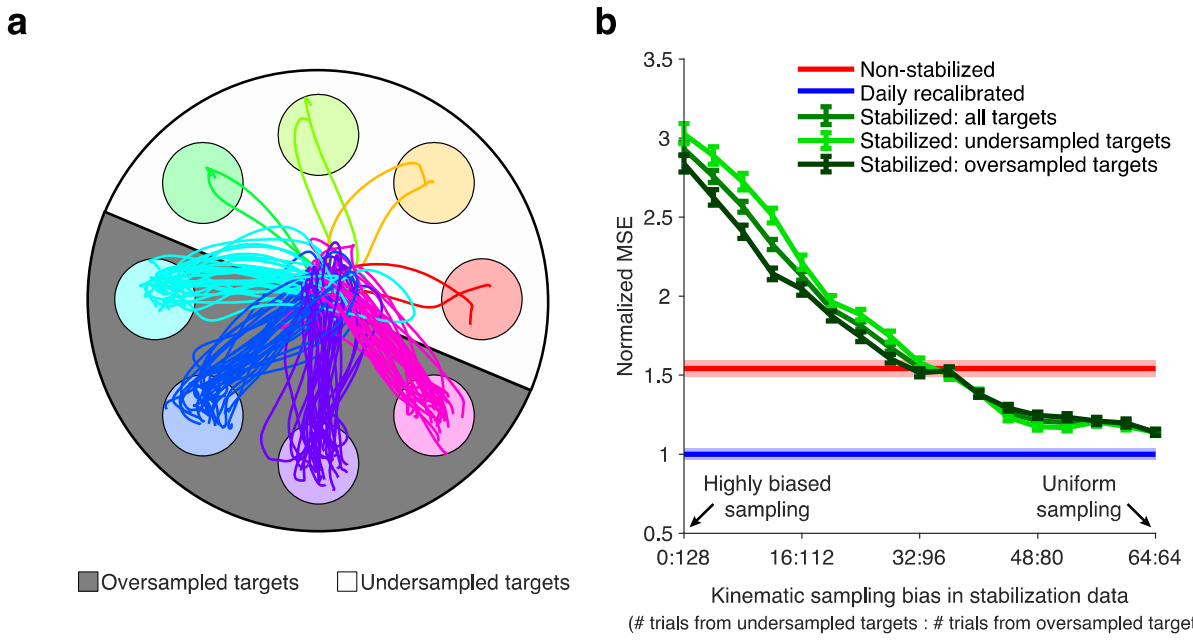
During normal BCI use, a decoder would estimate control signals from the latent state returned by the stabilizer. Here, the latent state returned by the stabilizer, up to a rotation, is a direct representation of the control signals. Therefore, in place of

a decoder we found a rotation on the baseline data to align the control signals estimated from the baseline stabilizer to the ground truth control signals. This rotation was found and held fixed before updating the stabilizer to instabilities, just as the decoder in closed-loop experiments was trained on baseline data and then held fixed.

During supervised calibration, a baseline stabilizer was fit to simulated neurons 1-75, leaving 10 electrodes to be used to create tuning change instabilities. Between the supervised calibration and stabilization blocks, instabilities were introduced into the model for generating neural activity. This was accomplished by randomly selecting 10 of the 75 neurons used in the supervised calibration block and swapping them with the 10 held-out electrodes. Unit drop-out instabilities were applied to an additional 5 electrodes. Baseline shifts were drawn i.i.d from a $\mathcal{N}(0.375, 0.25^2)$ distribution and applied to the entries of μ for the remaining neurons. With these instabilities applied, additional trials of data were generated for fitting and evaluating the stabilizer. Stabilization was performed in the same manner as during the closed-loop experiments, including requiring the stabilizer to identify stable electrodes for alignment. Simulations were repeated 32 times for each manifold dimensionality.



Supplementary Fig. 7 | Effect of number of alignment electrodes and assumed latent dimensionality on stabilization performance. To determine how stabilization performance varies with the number of alignment electrodes and the assumed dimensionality of the manifold, we used the stabilized BCI to decode hand velocity from neural activity recorded during a center-out reaching task (see Supplementary Fig. 4). **a**, Stabilized BCI performance for different numbers of alignment electrodes. Stabilization performance improves with the number of alignment electrodes used. When enough alignment electrodes are used, stabilization (*green*) outperforms the non-stabilized decoder (*red*), and is better than or comparable to the daily-recalibrated decoder (*blue*). While one might expect that using all electrodes for alignment would degrade performance, as some unstable electrodes would then be used to determine the alignment of the manifolds, in practice we did not observe this to be the case. While this might at first be surprising, we note that for these data the majority of electrodes exhibited baseline shifts, a minority exhibited changes in tuning, and no electrodes exhibited drop-out. The alignment process can accurately identify firing rate offsets and unit drop-out, leaving tuning change as the primary source of alignment error. Electrodes with different tuning changes lead to different alignment errors and they can effectively cancel each other out, as long as the number of stable electrodes included is sufficiently high. Performance is quantified using mean-squared error (MSE), normalized each day to the error of the daily-recalibrated decoder. Error bars represent the 95% confidence interval of the mean. For this analysis, we set the manifold dimensionality to 10 and the alignment threshold to 0.01 counts/bin, the same values as used in the BCI experiments reported in the main text. **b**, Stabilized BCI performance for different assumed dimensionalities of the manifold. Same conventions as in (**a**), but for the number of latent dimensions used in factor analysis. Error is normalized with respect to that the daily-recalibrated decoder using 10 dimensions. The estimated dimensionality for these days, indicated by dashed black lines, was between 4 and 5. Dimensionality was determined by first selecting the number of dimensions maximizing the cross-validated likelihood of 128 trials collected at the beginning of each day. We then found the number of dimensions need to explain 95% of the shared variance of the neural activity[1]. For this analysis, we set number of alignment electrodes to 75 and the alignment threshold to 0.01 counts/bin. Using an assumed dimensionality smaller than that of the recorded neural activity results in a dramatic increase in error. Conversely, using an assumed dimensionality greater than the true dimensionality of the data results in a minimal impact in stabilization performance; stabilizer mean-squared error is equal to or lower than that of the daily-recalibrated decoder for 5 of the 6 days when using the maximum number of dimensions (20). Although one might have expected error to increase if the number of latent dimensions used for stabilization is greater than the number of dimensions that can be identified in the neural activity, these additional dimensions are unlikely to contain task-relevant information, and thus would not be weighted highly in the BCI decoder. As long as the behaviorally-relevant dimensions of neural activity are appropriately aligned by the stabilizer, inclusion of extra dimensions should not detrimentally affect performance. Mean-squared error values were calculated based on 576, 1120, 1184, 1728, 1408, and 1232 trials for Days 20, 21, 22, 23, 24, and 27, respectively.



Supplementary Fig. 8 | Effect of non-uniform kinematic sampling on stabilizer performance. To assess how stabilization performance varies with non-uniform sampling of the kinematic space, we re-ran stabilization using data in which the number of trials used per target was biased towards one half of the animal's workspace. For this analysis, we used neural activity recorded during a center-out reaching task (see Supplementary Fig. 4). Stabilizer performance was compared to that of fixed and daily-recalibrated decoders. Stabilization performance improves as the sampling becomes more uniform, performing better than the non-stabilized decoder when the kinematic sampling was more uniform than 40:88 trials (40 trials from the under-sampled targets, 88 trials from the over-sampled targets). This suggests that stabilization is tolerant of moderate non-uniformities in kinematic sampling, and can still yield performance improvements when approximately twice the amount of data is used from one half of the workspace as from the other half. Furthermore, for non-uniformities in kinematic sampling stabilization performs better for the oversampled targets (*dark green*) compared to the undersampled targets (*light green*). This highlights how compromises must be made when the stabilizer is updated, and that the parameters of the stabilizer are more appropriate for the targets for which there is more data. **a**, Example of non-uniform sampling of trials. Sampling bias was varied from 0:128 to 64:64 (uniform sampling of all targets). The example shown corresponds to a 8:120 trial split, where 8 trials (2 per target) come from half of the workspace and the remaining 120 trials (30 per target) come from the other half. **b**, Performance of the stabilizer (*green*), daily-recalibrated (*blue*) and non-stabilized (*red*) decoders. Performance of the daily-recalibrated decoder provides a measure of the "best-case" scenario, where the decoding parameters are recalibrated in an optimal manner at the beginning of the session to account for any instabilities. In contrast to the analyses shown in Supplementary Figs. 4 and 7, where the stabilizer was updated continuously over the course of each experimental session, in this analysis stabilization was performed using 128 trials collected only at the beginning of the session and evaluated on the remaining trials. This was done because non-uniform sampling required sub-selecting trials to use for stabilization; this meant that it was not possible to both run stabilization continuously and evaluate performance on the same set of trials for each sampling bias condition. Performance was quantified using mean-squared error normalized with respect to that of the daily-recalibrated decoder. Error bars represent the 95% confidence interval of the mean. Stabilizer performance was averaged over the 8 possible oversampled/undersampled workspace splits. Parameters of the non-stabilized decoder and those of the initial stabilizer were found using data from the first recording day (Day 20). Parameters of the daily-recalibrated decoder were found using trials collected at the beginning of Day 27. In all cases, decoder parameters were initialized using uniformly-sampled data. Decoding performance was then evaluated on trials from Day 27, with parameters of the stabilizer subsequently updated using data with non-uniform sampling.

References

1. Williamson, R. C. et al. Scaling Properties of Dimensionality Reduction for Neural Populations and Network Models. PLoS computational biology **12**, e1005141 (2016).