Supplementary Methods

Experimental methods

Two rhesus monkeys (*Macaca mulatta*) were trained to perform center-out arm movements. Briefly, the monkey grasped two manipulanda (one with each hand) that were movable in the horizontal plane and controlled two cursors on a vertical screen in front of the monkey. In response to a visual cue, the monkey had to move either the left or the right arm to one out of eight directions (regularly arranged on a circle, 45° apart). We analyzed LFPs, SUA and MUA that were simultaneously recorded from four glass-coated tungsten electrodes (arranged at relative distances of 350 to 700 μm) in each hemisphere. The experimental data from the two monkeys were pooled. All experimental procedures were in accordance with the Hebrew University and NIH regulations for the Care and Use of Laboratory Animals. For details on the experimental setup see refs. 1,2.

Decoding of movement target and movement arm

LFPs and single-unit/multi-unit spike trains were preprocessed by smoothing the signals with a Gaussian kernel of 125 ms standard width. To decode neuronal activity with respect to the movement target or arm in a given trial (the test trial) we trained a classifier (see below) using a set of 15 trials for each target with the test trial excluded. This procedure was repeated using each trial as test trial exactly once. The decoding power was then computed as the number of correct classifications divided by the total number of test trials. As classifier we used the following methods:

1. Population vector (PV): The population vector approach[3] was applied on a singletrial basis for SUA/MUA by using the spike counts in a window ranging from 150 ms before movement onset to movement end. To decode the LFP we took the single-trial peak-to-peak amplitude between the negative peak within $\pm$250 ms around movement onset and the first succeeding positive peak of the LFP. Cosine tuning curves were fitted to the trial-averaged activity measures for the 8 movement targets by least-squares regression on the training data. For decoding, single-trial population vectors were constructed from the sum of vectors for each single channel, oriented along the LFP/SUA/MUA's preferred direction, with a length proportional to the appropriately scaled single-trial amplitude of the LFP/SUA/MUA channel. We then chose the target whose direction was closest to the direction the population vector pointed to.

2. Bayesian decoding based on a multivariate Gaussian model[4] : For each target, we modeled the spike counts of N SUAs/MUAs or the peak-to-peak amplitudes of N LFPs, yielding in both cases an N-dimensional signal vector s, as samples from a

multivariate Gaussian distribution:

$$p\left(s|\text{target}\right) = \frac{1}{\sqrt{(2\pi)^N |C|}} e^{-\frac{1}{2}(s-\mu)^T C^{-1}(s-\mu)}$$

The mean vector μ and the covariance matrix C were computed from the training data by maximum likelihood estimation. We used two different estimates of the covariance matrix: either the full covariance matrix with N * (N + 1) / 2 parameters, or the covariance matrix C was constrained to be diagonal with only N free parameters. The latter can result in advantageous estimates in the case of uncorrelated signals. Applying Bayes' rule allowed us to compute the posterior probability distribution

$$p\left(\text{target}|s\right) = \frac{p\left(\text{target}\right) p\left(s|\text{target}\right)}{\sum_{\text{target'}} p\left(\text{target'}\right) p\left(s|\text{target'}\right)}$$

where s is the LFP/SUA/MUA signal vector of the single-trial values of the test trial. We chose the target yielding the highest probability.

3. Support vector machine[5] (SVM): For the SVM we did not reduce the activity of each LFP/MUA/SUA channel to a single value as was done to apply the population vector approach and the multivariate Gaussian model. Instead, the activity between -150 ms before movement onset and movement end was binned at 50 ms and taken as a multidimensional vector of dimension D for each LFP/SUA/MUA channel. In case of N channels, the corresponding measurements were concatenated, yielding a D * N-dimensional vector. These vectors were used as input to a SVM with either a linear or a radial basis-function kernel.

4. Penalized linear discriminant analysis[6,7] (PLDA): We used the same D * N-dimensional signal vector as for the SVM. The PLDA projects this vector linearly onto a seven-dimensional space (one dimension less than the number of targets) where signal vectors belonging to different targets are maximally separated. This projection was found by penalized optimal scoring[6,7]. To classify the projected data vectors (PDV) we used two different approaches, which yielded approximately the same results: In the first approach the PDVs were classified by a linear SVM. In the second approach the distribution of the PDVs were modeled for each target separately as a multivariate Gaussian distribution with a certain mean and diagonal covariance matrix. We then computed the posterior probability distribution using Bayes' rule and finally chose the target, which yielded the highest probability. When we combined data from different recording days (only done in **Fig. 3a**), we applied only the first approach and the linear projection was performed separately for the data from each day.

For the LFP, SUA and MUA, the PLDA and SVM approach led, on average, to the highest decoding power (**Supplementary Fig. 2a**). The PLDA as well as the SVM yielded a considerably higher decoding power than the multivariate Gaussian model and the population vector approach. For all three types of signals the impact of the width of the smoothing kernel on the decoding power was relatively weak (**Supplementary Fig. 2b**) and the chosen width of 125 ms was more or less optimal. To compute the decoding power as a function of the number of recording electrodes (as done in **Fig. 3a**), we successively combined data from different recording sessions, each session containing the neuronal activity recorded simultaneously with eight electrodes.

We also examined the temporal evolution of the decoding power relative to movement onset (**Fig. 3b**). For that purpose, we took the signals recorded simultaneously from 8 electrodes and provided our classifier with data segments from a sliding window of 50 ms width. The window was moved in steps of 12.5 ms from 175 ms before to 350 ms after the onset of movement. Here, we truncated the Gaussian smoothing kernel at its maximum to avoid any acausal preview into the future.

Prediction of movement trajectories

To predict movement trajectories from neuronal activity, we modeled the hand velocity vector $v(t)$ at time t as a function of the neuronal signal $n(t)$ strictly before time t

$$v(t) = f\left[n(\tilde{t}); \tilde{t} \leq t\right]$$

The neuronal data were binned at 50 ms by averaging over 50 ms windows after smoothing with a Gaussian kernel of 30 ms standard width. We restricted f to depend only on neuronal activity up to 500 ms prior to the movement. We tested a linear and a non-linear model for the function f. The linear model is fully specified by a $(T+1)$-dimensional parameter vector β, which relates the neuronal signal to the instantaneous velocity vector as follows

$$f_{x,y}(\{n_t, n_{t-1}, \ldots, n_{t-T+1}\}) = (n_t, n_{t-1}, \ldots, n_{t-T+1}, 1) \cdot \beta_{x,y}$$

For the non-linear model we used ν-SVM regression[5,8] with a radial basis-function kernel. To compute the prediction of movement from the neuronal signals, we splitted the complete data (17-50 trials per target) into two mutually exclusive sets: a training set comprising all trials except one for each target, and a test set containing the remaining eight trials. For both models we learned f from the training set. The learning of the linear model, i.e. the estimation of the parameter vector β, was performed by least-squares regression. Training a support vector machine using an algorithm presented in ref. 9 and implemented in the software package "libsvm"[10] performed the learning of the non-linear model.

For each test trial we applied the fitted regression functions f to predict the 2-dimensional velocity vector of the hand in steps of 50 ms, using only neuronal activity recorded before the present point in time. Attaching these vectors tip-to-tail lead to an approximate reconstruction of the full 2-dimensional movement trajectories. To quantify the accuracy of the trajectory prediction we calculated, separately for the x- and y-components of the time course of position, the correlation coefficient (cc) between the predicted and the real trajectories. These two correlation coefficients were then averaged. As a second measure we computed the correlation coefficient between the real and the predicted time course of the speed (length of velocity vector).

For each recording session and each movement type (left- or right-handed), we performed the above learning and testing procedure for ten training and testing sets, which were randomly chosen from the complete data of that session. The associated cc of one session and one handedness then was the average over these ten training and testing sets. **Fig. 3f** shows the distribution of twenty correlation coefficients from ten different recording sessions including left- and right-handed movements.

Supplementary Figure 2c shows the prediction accuracy of the non-linear SVM regression in comparison to the prediction accuracy of the linear filter (LF). For all three signals the SVM regression outperformed the LF. The difference was most pronounced for the LFPs, where the non-linearity of the SVM regression turned out to be crucial to achieve a prediction accuracy substantially above zero.

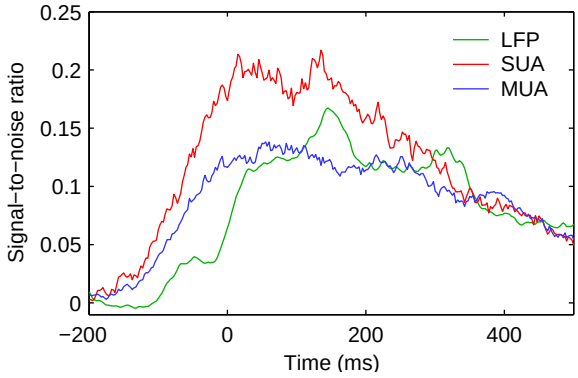
Test of significance for a decoding power above chance level

The probability to predict K times the correct target in N trials by chance is given by the binomial distribution

$$P(K) = \frac{N!}{(N-K)!\,K!} \left[\frac{1}{T}\right]^K \left[\frac{T-1}{T}\right]^{N-K},$$

where T refers to the number of targets.

For the neuronal signals, we computed the time-resolved decoding power of multiple LFPs, SUAs and MUAs from 10 different recording sessions with left- and right-handed movements and a total of 120 trials each (**Fig. 3c**). We counted the number of correct target predictions for all N = 10 * 2 * 120 = 2400 trials. Using the binomial distribution we were able to establish the statistical significance of the observed values for the decoding power.



Supplementary Figure 1: Temporal evolution of tuning strength. Colors depict the different types of signals: LFPs (green), SUAs which were checked for stability (red) and MUAs (blue). Tuning strength was calculated as the signal-to-noise ratio of the tuning curves in windows of 50 ms width from the unsmoothed signals.

Results

Tuning properties of local field potentials and single-unit activity

Most averaged LFPs exhibited a clear time-dependent modulation of activity with a negative peak around movement onset and a successive positive peak (for details, see ref. 11). To determine the tuning function of an LFP, we computed the peak-to-peak amplitude as a function of the target direction. For many LFPs, the tuning was unimodal and smooth, similar to the example shown in **Fig. 1b**. As we have demonstrated in ref. 12, the tuning curves of LFPs and SUAs recorded from the same electrode exhibited a weak, but significant similarity.

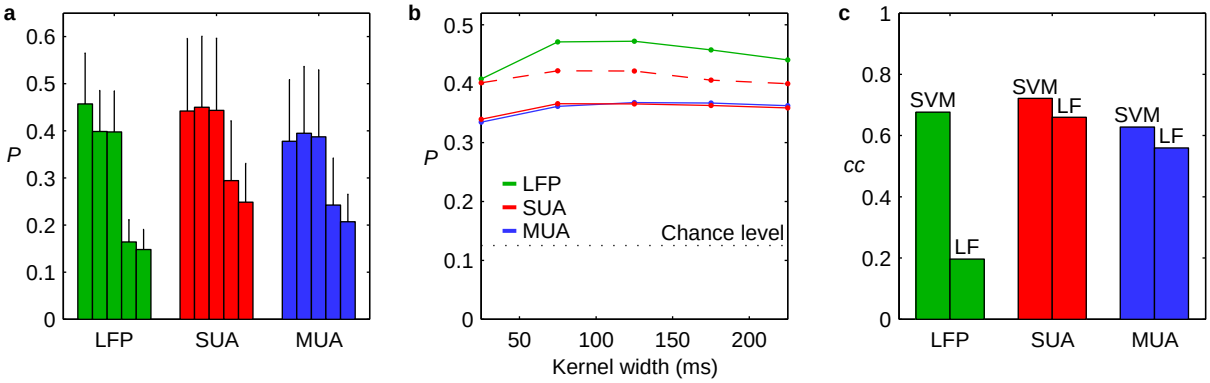
Temporal evolution of tuning strength

To establish that LFPs and spike signals exhibit tuning already before the onset of movement, we computed the tuning strength as a function of time. As a measure of tuning strength we used the signal-to-noise ratio (SNR), which is here defined as

$$SNR = \frac{\sigma_s^2 - \sigma_e^2}{\sigma_n^2}$$

where σ_s^2 is the variance of the signal, i.e. the tuning curve, σ_n^2 is the variance of the trial-by-trial fluctuations and σ_e^2 is the variance of the signal that would be expected by chance for flat tuning curves. Due to limited sampling, σ_e^2 is not equal to zero and can be computed as follows

$$\sigma_e^2 = \frac{\frac{1}{8}\sum_{target=1}^{8} \sigma_{target}^2}{N_T}$$



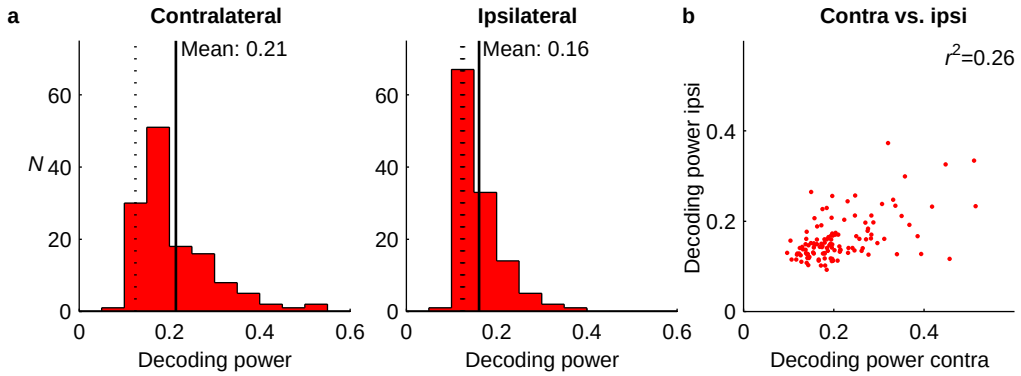
Supplementary Figure 2: Decoding power and accuracy of trajectory prediction for different algorithms. We used the neuronal signals (LFP, SUA, MUA) that were recorded simultaneously by eight electrodes. The graphs depict averages over 10 recording sessions and both, left- and right-handed movements. In (**a**), (**b**) and (**c**) results are shown for the whole recording set of single-units. Subfigure (**b**) includes results for SUAs additionally checked for stability across trials. **(a)** Decoding power for different classification algorithms: PLDA, SVM with radial basis function kernel, SVM with linear kernel, multivariate Gaussian model, population vector approach (from left to right). Error bars depict the standard deviation. **(b)** Decoding power as a function of the width of the smoothing kernel. Colors and line styles as in **Fig. 3a**. **(c)** Accuracy of 2D trajectory prediction (average correlation coefficients across sessions and handedness) using either SVM regression (SVM) or the linear filter (LF).

where σ^2_{target} denotes the variance of trial-by-trial fluctuations for a specific target direction and N_T is the number of trials per target.

We calculated the time-resolved signal-to-noise in sliding windows of 50 ms width (**Supplementary Fig. 1**). The SNR of LFPs started to rise above zero about 100 ms before movement onset. This is in good agreement with **Fig. 3c**, which showed that the decoding power rose significantly above chance level around the same time. The SNR of SUAs and MUAs started to rise roughly 50 ms earlier than the SNR of the LFPs.

Decoding power of individual single-units taken from the whole recording set regardless of stability

We performed our SUA analysis twice: once for SUAs obtained by spike sorting only (spike detection and sorting was done online by MSD(R), Alpha-Omega, Nazareth, Israel), and once for SUA additionally checked for stable activity across trials. **Supplementary Figure 3** shows the decoding power of individual single-units that belong to the first group. As can be seen when comparing **Figure 2** (top row) to **Supplementary Figure 3**, SUA checked additionally for stability across trials yield a higher decoding power on average than SUA obtained without this criterion.



Supplementary Figure 3: Decoding power of individual single-units taken from the whole recording set regardless of stability. **(a)** Distribution of decoding power for individual SUA, shown for contralateral (left) and ipsilateral (right) movements. Dotted lines depict the chance level (0.125). **(b)** Contra- and ipsilateral decoding power was largely uncorrelated.

References

1. Cardoso de Oliveira, S., Gribova, A., Donchin, O., Bergman, H. & Vaadia, E. *Eur. J. Neurosci.* **14**, 1881-1896 (2001).

2. Donchin, O., Gribova, A., Steinberg, O., Bergman, H. & Vaadia, E. *Nature* **395**, 274-278 (1998).

3. Georgopoulos A.P., Schwartz A.B. & Kettner R.E. *Science* **233**, 1416-1419 (1986).

4. Maynard, E.M. *et al. J.Neurosci.* **19**, 8083-8093 (1999).

5. Schölkopf, B. & Smola, A. J. Learning with kernels. (The MIT Press, Cambridge, Massachusetts, 2002).

6. Hastie, T.J., Buja A. & Tibshirani, R. *Ann. Statist.* **23**, 73-102 (1995).

7. Roth, V. & Steinhage, V. In: S. A. Solla, T. K. Leen, K.-R. Müller (Eds.) Advances on Neural Information Processing Systems, NIPS (1999).

8. Schölkopf, B., Smola, A.J. & Williamson, R. *Neural Comput.* **12**, 1207-1245 (2000).

10. Chang, C.C. & Lin, C.J. *Neural Comput.* **14**, 1959-1977 (2002).

11. http://www.csie.ntu.edu.tw/~cjlin/libsvm.

12. Donchin, O. *et al. Exp. Brain. Res.* **140**, 45-55 (2001).

13. Mehring, C. *et al. Proceedings of the 1st International IEEE EMBS Conference on Neural Engineering* ISBN 0-7803-7579-3, 28-31 (2003).