

Patterns

Representing the dynamics of high-dimensional data with non-redundant wavelets

Highlights

- WCMI can extract meaningful information from high-dimensional data
- Extracted features from neural signals are non-redundant
- Simple decoders can read out these features with superb accuracy

Authors

Shanshan Jia, Xingyi Li, Tiejun Huang,
Jian K. Liu, Zhaofei Yu

Correspondence

j.liu9@leeds.ac.uk (J.K.L.),
yuzf12@pku.edu.cn (Z.Y.)

In brief

A crucial question in data science is to extract meaningful information embedded in high-dimensional data. Such information is often formed into a low-dimensional space with a set of features that can represent the original data at different levels. A method combining wavelet analysis with conditional mutual information can extract non-redundant features from various types of neuroscience data. Simple decoders can read out meaningful information with high accuracy using only a small set of these features.



Article

Representing the dynamics of high-dimensional data with non-redundant wavelets

Shanshan Jia,^{1,2} Xingyi Li,³ Tiejun Huang,^{1,2} Jian K. Liu,^{4,*} and Zhaofei Yu^{1,2,5,*}

¹Institute for Artificial Intelligence, Peking University, Beijing 100871, China

²Department of Computer Science and Technology, Peking University, Beijing 100871, China

³Center for Neurointelligence, School of Medicine, Chongqing University, Chongqing 400030, China

⁴School of Computing, University of Leeds, Leeds LS2 9JT, UK

⁵Lead contact

*Correspondence: j.liu9@leeds.ac.uk (J.K.L.), yuzf12@pku.edu.cn (Z.Y.)

<https://doi.org/10.1016/j.patter.2021.100424>

THE BIGGER PICTURE One of the essential questions in data science is to extract meaningful information from high-dimensional data. A useful approach is to represent data using a few features that maintain the crucial information. The leading property of spatiotemporal data is foremost ever-changing dynamics in time. Wavelet analysis, as a classical method for disentangling time series, can capture temporal dynamics with detail. Here, we leveraged conditional mutual information between wavelets to select a small subset of non-redundant features. We demonstrated the efficiency and effectiveness of features using various types of neuroscience data with different sampling frequencies at the level of the single cell, cell population, and coarse-scale brain activity. Our results shed new insights into representing the dynamics of spatiotemporal data using a few fundamental features extracted by wavelet analysis, which may have wide implications to other types of data with rich temporal dynamics.



Proof-of-Concept: Data science output has been formulated, implemented, and tested for one domain/problem

SUMMARY

A crucial question in data science is to extract meaningful information embedded in high-dimensional data into a low-dimensional set of features that can represent the original data at different levels. Wavelet analysis is a pervasive method for decomposing time-series signals into a few levels with detailed temporal resolution. However, obtained wavelets are intertwined and over-represented across levels for each sample and across different samples within one population. Here, using neuroscience data of simulated spikes, experimental spikes, calcium imaging signals, and human electrocorticography signals, we leveraged conditional mutual information between wavelets for feature selection. The meaningfulness of selected features was verified to decode stimulus or condition with high accuracy yet using only a small set of features. These results provide a new way of wavelet analysis for extracting essential features of the dynamics of spatiotemporal neural data, which then enables to support novel model design of machine learning with representative features.

INTRODUCTION

The ubiquitous property of neural systems in the brain is functioned as an information processor to compute incoming stimuli, extract relevant contents, and make meaningful decisions thereon. Understanding these behaviors relies on investigating recorded neural signals of the brain. Depending on the experimental techniques used for recording, there are different formats of neural signals across multiple scales of time and space, either

discrete 0 or 1 binary neural spikes,^{1–4} continuous analog signals, such as slow calcium imaging data,^{5–8} coarse-scale brain activity of electrocorticography (ECoG) signals,⁹ and functional magnetic resonance imaging (fMRI).¹⁰ Nevertheless, a peculiar feature of neuronal computation in a neuron is using binary spikes to represent information, where spike timing and rate are both important.^{11–13} The question is then how to extract meaningful information in recorded neural signals. It is useful not only for understanding the mechanism of neural computation



for neural function^{13,14} but also for designing decoding algorithms to control the devices of brain-machine interface.^{15–17}

Neural activities are represented as spatiotemporal patterns, in which the dimensionalities of both time and space are high.^{18–20} One strategy is to reduce the dimension by selecting a subset of features, while keeping information as much as possible.²¹ Wavelet analysis is a unique method that captures information embedded in signals through hierarchical analysis of multiple scales of time and space.^{22,23} Recent studies show that a method, named wavelet information (WI),²⁴ using wavelet decomposition together with feature selection using mutual information, can achieve superb results to extract information in neural spikes and electroencephalography signals with a small subset of wavelet coefficients.²⁵ This method makes full use of wavelet analysis to decompose spatiotemporal patterns into multiple scales manifested by a set of wavelet coefficients. However, it leaves out high-order dependency between features without taking into account potential redundancy in the set of features, as mutual information of each wavelet is independently ranked. In addition, for the neural population, it is well known that neural coding utilizes a collaborative structure for encoding information.²⁶ The question is then how to employ wavelets to evaluate the information of synergy in a neural code.²⁷

To address this issue, we considered a new method combining wavelet analysis with conditional mutual information, named wavelet conditional mutual information (WCMI), to extract non-redundant features of spatiotemporal data. In contrast to WI, WCMI takes a step further implementing feature selection leveraging conditional mutual information to justify the importance of wavelet coefficients beyond the individual level. For this purpose, we developed the computational approach of WCMI combining WI²⁴ with the feature selection method CMI-COT²⁸ to extract meaningful wavelets using conditional mutual information. We demonstrated that WCMI takes full account of the dependency and redundancy between features extracted in neural signals, and achieves a better performance of decoding using simulated spike patterns and recorded neural spikes and calcium imaging data in different conditions. The efficiency and effectiveness of WCMI were shown by decoding of stimuli or conditions using both fast signals of spikes at 1,000 Hz and slow signals of calcium data at 10 Hz at the level of a signal trial of individual neurons and neural populations, as well as human ECoG signals. Particularly, WCMI can capture the high-order dependency of neural population in response to natural images and suggest the notion of sparse coding incorporating the statistics of natural images. Taken together, our results suggest that wavelet analysis unitizing conditional mutual information can provide a new way of extracting meaningful features embedded in spatiotemporal data.

RESULTS

Non-redundant features extracted by WCMI

We first used simulated cells of spike patterns to demonstrate the utility of WCMI. Four simulated cells in response to four stimuli, previously used to test the applicability of WI,²⁴ have distinguished temporal patterns of spikes with different timings, firing rates, and noise levels, as shown in Figure 1A. Each cell contains spike responses to four different stimuli. Wavelet analysis with

four-scale Haar wavelets was applied to every trial of spike sequence to obtain a set of coefficients: detail coefficients at the levels of D1, D2, D3, and D4, and approximate coefficients A4, according to wavelet decomposition (see [experimental procedures](#)). Example coefficients of four trials are shown as wavelet scalograms (Figures 1B and S1) representing the distribution of coefficients with their magnitudes in the parameter space, scale, and translation of wavelets. With the full set of coefficients across stimuli and trials, feature extraction approaches of WI and WCMI were implemented. WI selects features according to the rank of the amount of mutual information with stimulus, where each coefficient is computed independently. In contrast, WCMI ranks each coefficient according to its conditional mutual information demonstrated by other coefficients through a *t*-way interaction (see [experimental procedures](#)), where *t* is a parameter indicating the degree of high-order dependency between coefficients. For illustration, 25 coefficients extracted by WI and 6-way WCMI are visualized in Figure 1C. The distributions of these features are dramatically different between WI and WCMI. The WI coefficients align closely with neural activity triggered by stimulus, such that more spikes give a localization of coefficients at particular time points. This is due to the fact that WI selects each feature according to its information quantity. Therefore, all the selected features have the highest information by ranking, yet they are highly redundant as a set of features. In contrast, the coefficients selected by WCMI are more distributed, and are independent to each other as computed by conditional mutual information, rather than absolute mutual information. As a result, WCMI is able to select those coefficients with high-order dependency without redundancy between features.

To quantify the meaningfulness of the features, we used the linear discriminant analysis (LDA) decoder to classify four given stimuli for each trial of spikes. WI has been shown to have a good decoding ability with this dataset.²⁴ Similar to WI, WCMI also shows a great capability of features selection for decoding. The resulted confusion matrices of classification obtained by both methods are shown in Figure 1D, with the overall accuracy averaged over four cells for WI as $98.12\% \pm 3.75\%$ and WCMI as $98.59\% \pm 2.81\%$ (mean \pm SD [standard deviation]). The accuracy of WI and WCMI on these cells was tested by *t* test ($p = 0.39$) and Wilcoxon rank-sum test ($p = 1$), indicating that there was no significant difference between WI and WCMI on these simple simulation datasets. Given simulated data represent simple scenarios, we demonstrate the benefits of WCMI through a series of complex data in the rest of this work.

Efficient decoding of WCMI on neural spikes in response to natural images

To demonstrate the advantage of WCMI, we applied it to spike responses of a population of cells under the stimulation of natural images. This dataset contains neural spikes of 80 retinal ganglion cells in response to 300 different natural images,²⁹ where each image was presented for 200 ms then removed with a gray background of 800 ms to account for delayed responses before the next image. Since each cell has a unique response pattern for each individual stimulus image, we decoded different image identities from the whole population of neural spikes.

Figure 2A illustrates three sample images overlaid with the receptive fields of the whole population of 80 cells, together

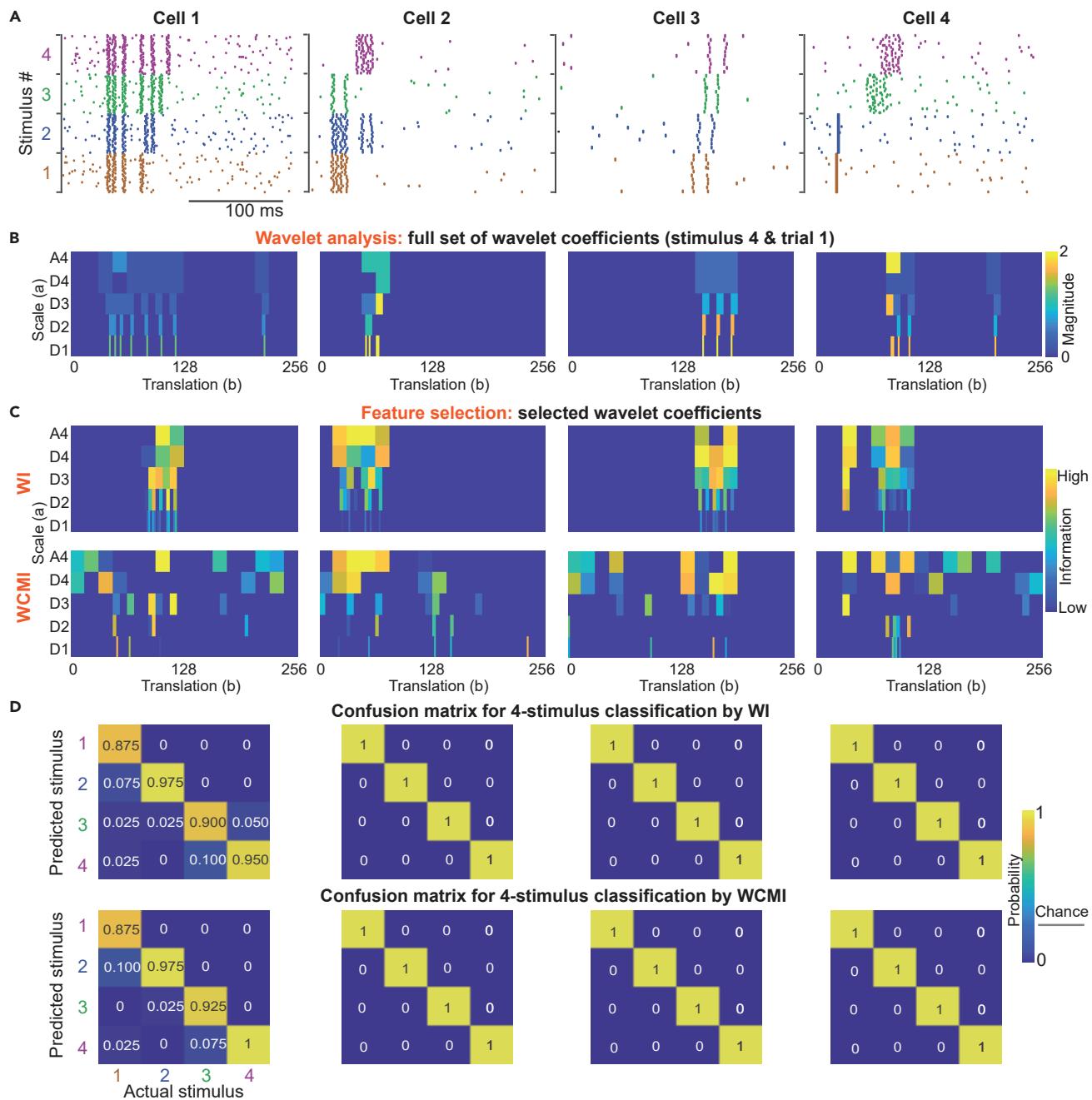


Figure 1. WCMI utilizes high-order dependency for feature selection

(A) Simulated cells of spike patterns in response to four stimuli.

(B) Wavelet analysis. Example scalograms of the whole set of wavelet coefficients after decomposing one trial of spikes. The distribution of 256 coefficients with different magnitudes from trial 1 spikes for stimulus 4, visualized in the parameter space of wavelet translation (b) and scale (a).

(C) Feature selection. Similar to (B) but for a subset of 25 coefficients selected by WI and WCMI.

(D) Decoding stimulus using features. Confusion matrices of four-stimulus classification were obtained by features of WI and WCMI. Each value indicates the accuracy probability with the chance level as 0.25.

with their response spike patterns. For each image, there are distinguished population neural patterns. These population patterns were decoded to classify 300 images with different decoders using a wide range of the number of features selected by WI and WCMI (Figure 2B). Both WI and WCMI show good per-

formance. However, WCMI outperforms WI systematically. The results are independent of the specific decoder used, either the default LDA, or support vector machine (SVM) and naive Bayes (NB) classifier. To quantify the difference between the two methods, the performance curves were fitted with a Hill

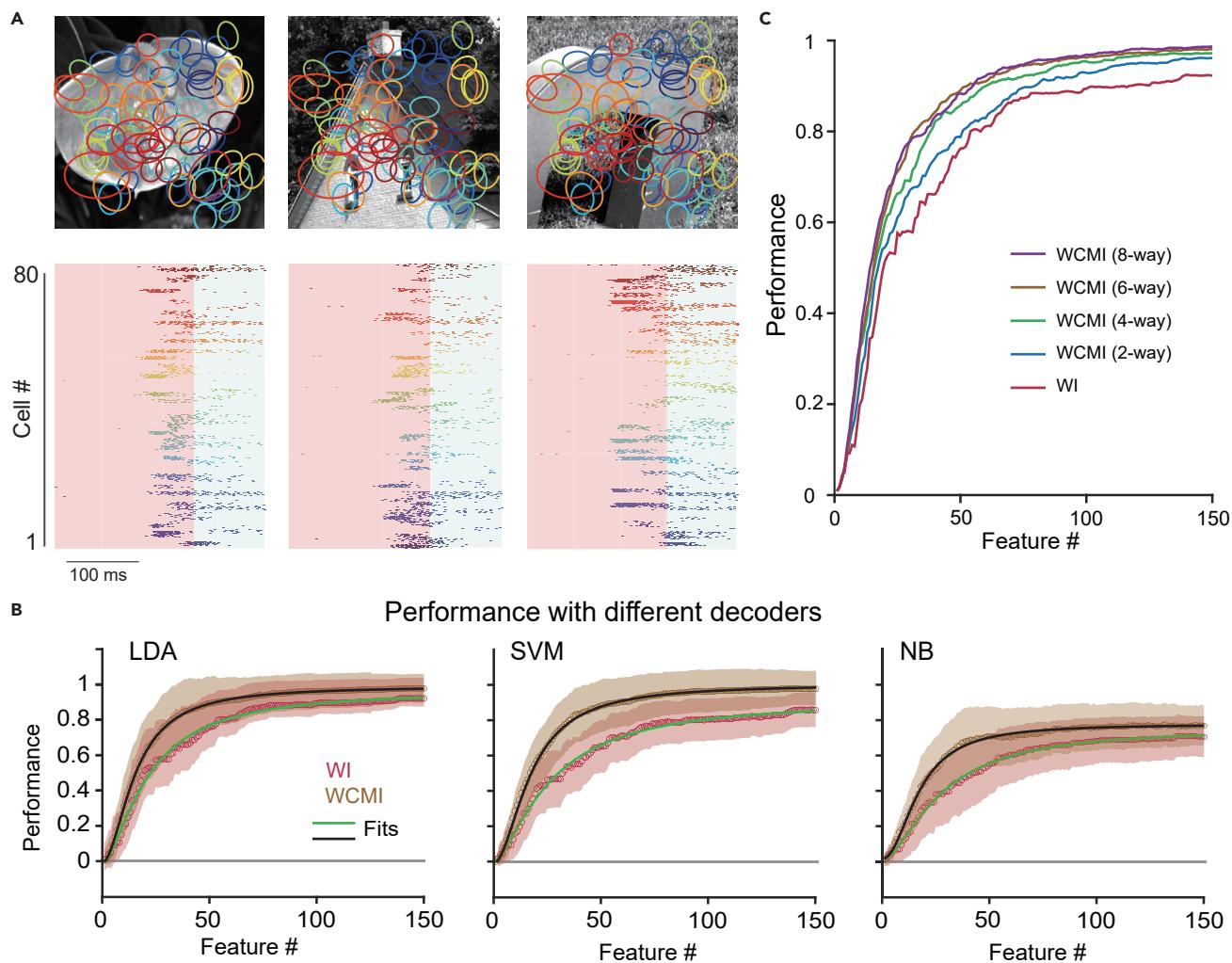


Figure 2. Decoding neural patterns in response to natural images

(A) Example images and neural responses. (Top) Three example images overlaid with the receptive fields (colored outlines) of a population of 80 cells. (Bottom) Corresponding population spike patterns. Image presentation 200 ms (red) followed by 100 ms (blue) of no image.

(B) The performance of WI and WCMI with more features using different decoders. Gray lines are the chance level. Solid lines are fits to a Hill function. Circles (mean); shades (SD).

(C) The profile of mean performance with different values of t -way dependency in WCMI. WI is equivalent to $t = 1$.

function that gives the maximal performance and value of the feature number at which the performance reaches the half maximum (Table S1). The p values of t test and Wilcoxon rank-sum test are close to zero ($p < 10^{-10}$), indicating that the difference in performance between WI and WCMI is significant. With the LDA decoder, WCMI uses 15 features reaching a performance of 50%, while WI needs 22 features. These fitted values indicate that WCMI is able to use fewer features while reaching higher accuracy.

To further explore the effect of high-order dependency between features on decoding, we varied the degree of t -way interaction in WCMI using different values of t (Figure 2C). WI is a special case of WCMI with $t = 1$, where there is no high-order interaction considered. With larger t as 2, 4, 6, and 8 in WCMI, higher performance while using fewer features was obtained and quantified by fitting with a Hill function (Figure S2; Table

S2). Not surprisingly, both WI and WCMI outperform on decoding using the feature of spike count (Figure S3). We also considered a different scenario by decoding cell identity from the population of neural responses and found similar results (Figure S4). These results indicate that WCMI has a good capability of using a few features to capture the information embedded in high-dimensional spatiotemporal spike patterns. To represent the information encoded in the data, WCMI is more efficient with fewer features but more effective with higher performance. Throughout this work, we used the six-way WCMI for feature extraction and the LDA decoder for decoding, except where mentioned in specific cases.

Sparse coding of natural images captured by WCMI

Compared with WI, WCMI utilizes high-order dependency between features to avoid over-representation of similar features.

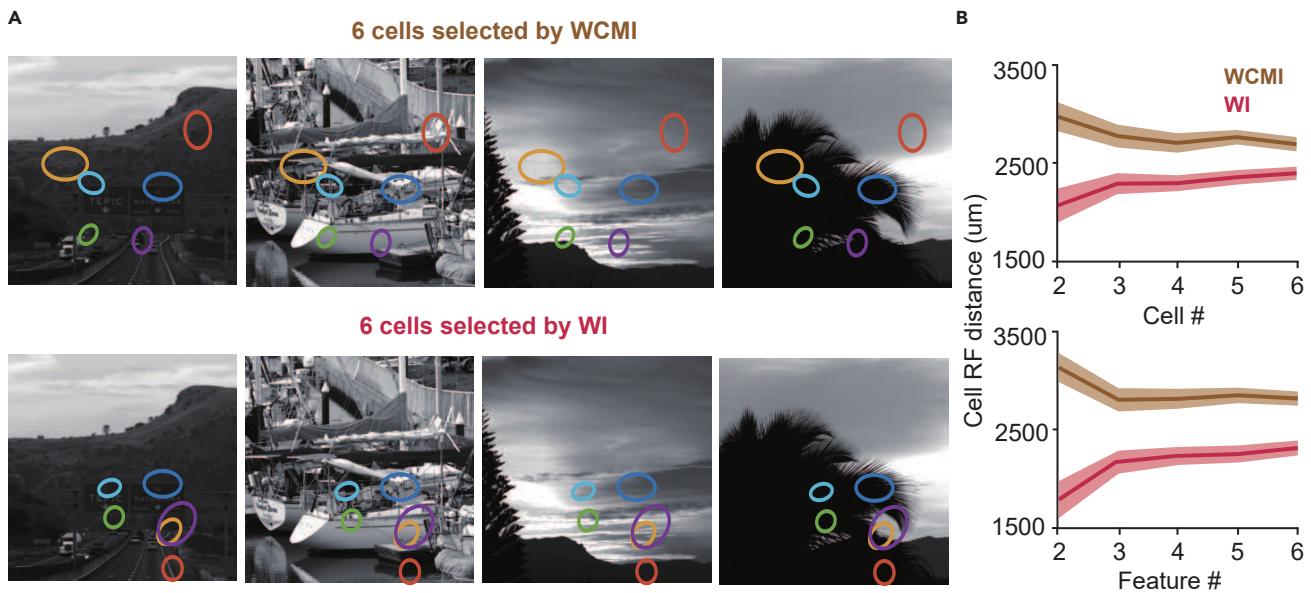


Figure 3. WCMI suggests neural sparse coding of natural images

(A) The layout of receptive fields of six cells selected by WCMI and WI on four example natural images.

(B) The change of the distance between cells with the increasing number of cells (top) or features (wavelet coefficients) ranked by WCMI and WI. The means (solid lines) and standard errors (shadowed areas) were computed over 50 random groups, where each group has four images and all the responses of 80 cells.

Removing redundancy means that WCMI is able to capture a low-dimensional structure of the population signal, presumably relating to the notion of sparse coding. To investigate this question, we utilized the same dataset of a population of retinal ganglion cells recorded in response to a set of natural images. It is suggested that sensory neurons, including retinal cells, show sparse coding, particularly for natural images with rich statistical structures.³⁰ Using the features extracted by WI and WCMI, we examined how these features are distributed in the space of natural images.

At the population level, the dependency between cells can be naturally removed by WCMI, but not by WI. Indeed, we found that WCMI takes into account the high-order response of neural population as represented in the space of natural images. After wavelet decomposition, a set of wavelet coefficients can be obtained for each cell, which was further selected by WCMI. We then extracted the neuronal identity that is corresponding to the selected wavelet coefficients. As a result, for a subset of selected features, there is a subset of cells aligning with them. An overlaid plot of the receptive fields of these cells selected by WI and WCMI is shown in Figure 3A, where four sample images with six cells were visualized. To quantify the distribution of cells in the space of images, we computed the distance between the cells, e.g., the distance between the centers of the cell receptive fields (RFs) distance in terms of physical distance measured by electrodes (see [experimental procedures](#)). The change of the RF distance over the number of cells, corresponding to the number of features or wavelet coefficients, shows that WCMI has a larger separation distance among cells in the images, e.g., the cells are more distributed over images, compared with WI (Figure 3B). Interestingly, the distance is larger with fewer cells selected, which suggests the sparse coding of images using fewer cells in a population. These results suggest that the

neural population structure in response to natural images revealed by WCMI is more dispersed. In other words, WCMI is able to capture high-order dependency between cells in a population in response to natural images with rich statistics. This indicates that the feature selection principle carried out by WCMI relates to the sparse coding of natural images in neural population.

Decoding motion direction from neural spikes and calcium signals

To confirm the feature extraction capability of WCMI to different types of data, we applied it to two different datasets, fast spikes at 1,000 Hz and slow two-photon calcium imaging data at 10 Hz, under the same experimental stimulus of direction selection.

For experimental data of different spike patterns, we used the spiking responses of 80 retinal ganglion cells to the stimulus of moving grating images recorded in salamanders.³¹ With grating images of eight different directions, each cell shows different spiking patterns (Figure 4A). Traditionally, spike count is used to compute the index of directional selectivity visualized in the polar plot. Using spike count, cell 1 has no preference for directions as it has equal spikes for all directions. Cell 2 is a classic direction-selective cell preferring the left direction, as there are more spikes to the left than to the right. However, the neural response is expressed as a temporal sequence of spike trains, which cannot be captured by spike count. From spike trains of cell 1, we find that these spike patterns are distinguished in response to different directions, whereas spike counts are similar across all directions. Thus, it is difficult to decode stimulus using spike count for cell 1. The challenge is to read out the difference in timings of spike patterns, which is the test cornerstone for the algorithm of wavelet analysis.

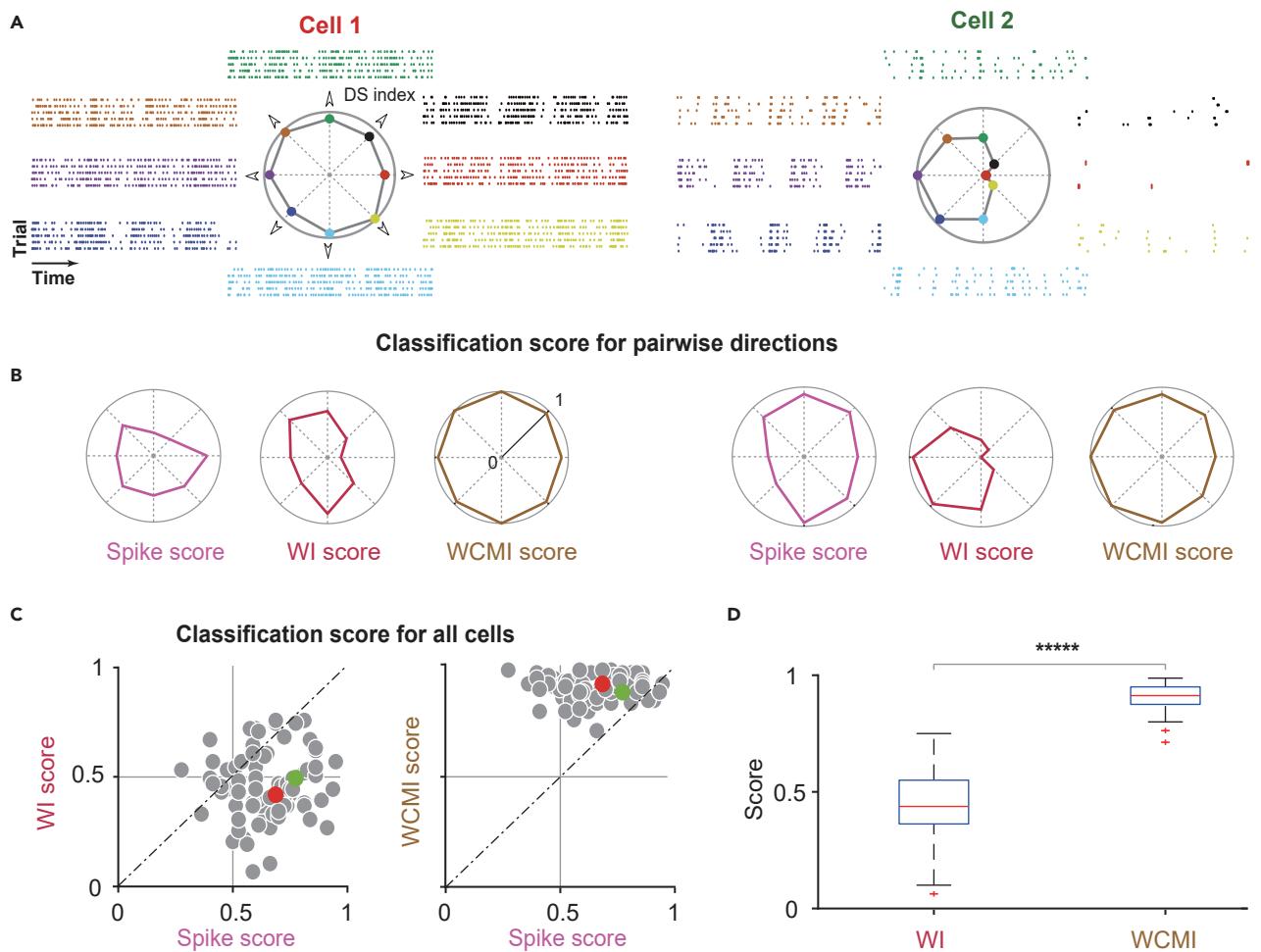


Figure 4. Decoding direction using spikes

(A) Two example retinal cells with spike response in eight directions. (Center) The direction-selectivity (DS) index computed using spike count. Each cell has five trials for the same direction.

(B) Polar plots of classification scores for pairwise adjacent directions using spike count (spike score), wavelet coefficients selected by WI (WI score), and WCMI (WCMI score): 1 indicates perfect decoding, while 0 indicates failed decoding.

(C) Decoding scores for all cells. The points marked by red and green are the example cells in (A). Gray lines are the chance level of classification.

(D) Boxplots of the performance score of WI and WCMI shown in (C). The p values of t test and Wilcoxon rank-sum test of both methods are significant ($****p < 10^{-10}$).

Based on the response of each cell, we used three ways of extracting features, spike count, and wavelets by WI and WCMI, to decode a pair of two adjacent directions. Classification scores for pairwise directions are represented as a polar plot (Figure 4B). Each data point in the plot is a classification score of two adjacent directions, with 1 indicating perfect classification and 0 indicating failure. Across three types of features, we found that the spike score using spike count is good for cell 2 but failed for cell 1 since there are equal number of spikes in all directions. The WI score also failed to explain the difference shown in the spike patterns. In contrast, WCMI discriminates the difference in both spike timings as in cell 1 and spike counts as in cell 2. We averaged all the classification scores for pairwise directions of each cell to obtain the overall score of a population (Figure 4C). It shows that WCMI has the best performance ($90.81\% \pm 5.59\%$), while WI ($45.41\% \pm 15.03\%$) is not as good as spike

count ($65.72\% \pm 14.52\%$). Figure 4D shows that WCMI is significantly better than WI. As both WCMI and WI are based on the framework of wavelet analysis, the difference in performance lies in the way of selecting a subset of wavelet coefficients. These results imply that the improvement of classification is mainly from the effectiveness of WCMI leveraging a novel way of extracting features from data.

We further examined the applicability of WCMI to the signal with low sampling frequency. Compared with spikes, neural calcium imaging data are continuously changing much more slowly. Although the relationship between calcium signals and spikes has not been proved to be strictly linear or nonlinear, it is believed that a larger calcium signal peak is induced by more spikes.³² For comparison, we used two datasets of neural responses of the primary visual cortex under the similar protocol of direction selectivity recorded by two-photon calcium imaging microscopy

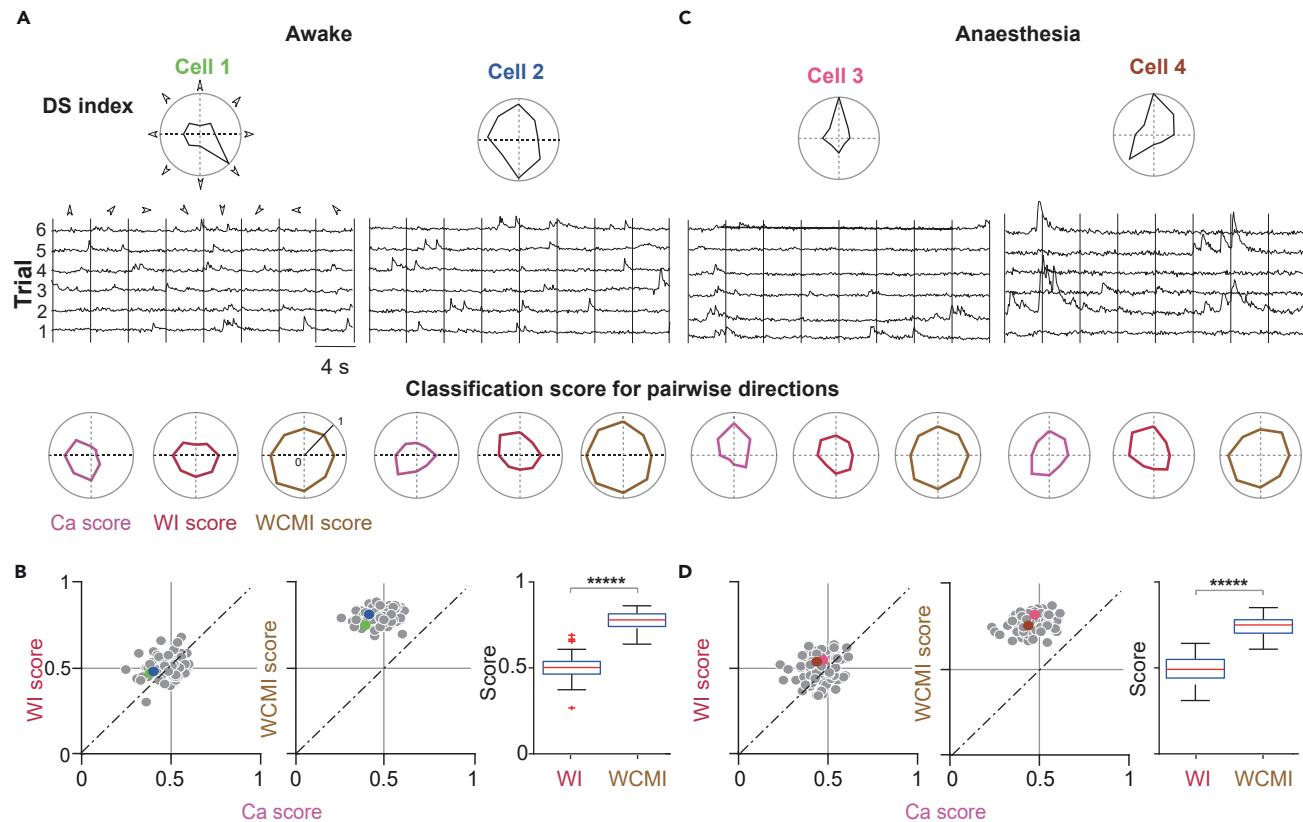


Figure 5. Decoding direction using calcium imaging signals

(A) The calcium signal traces of two example cells from awake mice. (Top) Direction-selectivity (DS) index computed with the amplitude of calcium (Ca) signal trace. (Middle) Ca signal traces of neural responses for eight directions. (Bottom) Polar plots of classification scores for pairwise adjacent directions using amplitude (Ca score), features selected by WI (WI score) and WCMI (WCMI score): 1 indicates perfect decoding, while 0 indicates failed decoding. (B) Decoding scores for all cells. The colored points are the example cells in (A). Gray lines are the chance level of classification. The difference between WI and WCMI is significant: t test and Wilcoxon rank-sum test ($p < 10^{-10}$). (C and D) Similar to (A and B) but for calcium data from anesthesia mice. The difference between WI and WCMI is significant with t test and Wilcoxon rank-sum test.

sampling at 10 Hz.³³ Firstly we used data recorded in awake mice. Figure 5A shows two example cells with their calcium responses and direction-selectivity indices computed by the amplitude of calcium signals (Ca score). Similar to the approach of retinal spike data of direction selectivity, both WI and WCMI were applied to calcium signal traces. We found that WCMI has the best classification score, compared with the Ca score and WI score. The population plot (Figure 5B) of averaged performance of pairwise directions for all cells shows that WCMI ($79.34\% \pm 4.42\%$) has significantly better performance than WI ($49.93\% \pm 6.32\%$); whereas WI is comparable with the direct analysis of the amplitude of calcium signals ($47.42\% \pm 7.80\%$), and both are close to the chance level of decoding. We then validated our results employing another dataset with 95 cells recorded from anesthetized mice under the same stimulation, and found similar results as those in awake mice with the performance of WCMI as $76.72\% \pm 4.48\%$, WI $48.71\% \pm 6.79\%$, and Ca $44.99\% \pm 7.55\%$, and WCMI significantly outperforms WI (Figures 5C and 5D).

Taken together, these results indicate that, despite the property of low frequency and higher noise level of calcium imaging

signals, WCMI allows us to extract useful features to quantify the information coded by neurons. Thus, WCMI is applicable to both fast spikes and slow calcium signals within a wide range of sampling frequency.

Decoding spontaneous brain states using calcium imaging signals

The above analysis was conducted for neural signals in response to external stimulation, where clear patterns of responses were seen. We then examined a more general condition where there is no stimulus applied. We utilized a calcium imaging dataset without any external stimulation, where spontaneous neural activities from three different brain areas, the primary visual cortex (V1), the primary motor cortex (M1), and the hippocampal CA1 region (CA1), were recorded simultaneously in mice.⁶ There are three similar populations of neurons from each brain area recorded firstly in the anesthesia state and later in the awake state of the same mouse. The advantage of these data is that the same neurons have been recorded in different brain states. Figure 6A shows neural activity of sample cells from each brain area, V1, M1, and CA1, in both states.

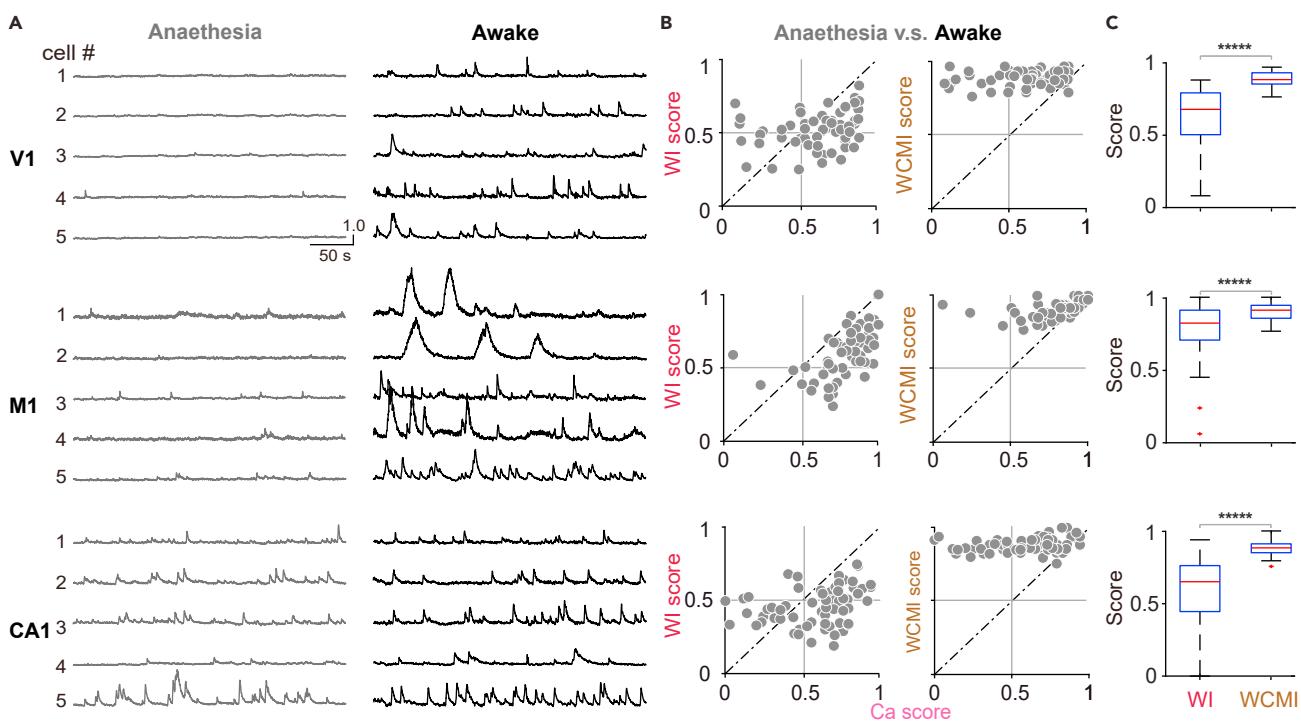


Figure 6. Decoding spontaneous neural activity of anesthesia and awake state from three brain areas

(A) Example calcium traces of neural activity from three brain areas (V1, M1, and CA1) in anesthesia and awake state, respectively.
(B) Decoding scores of anesthesia versus awake state for each cell of three brain areas using calcium amplitude (Ca), WI, and WCMI.
(C) WCMI significantly outperforms WI with t test and Wilcoxon rank-sum test for all three brain areas.

We decoded the anesthesia versus awake state using calcium traces of the same cell. The overall view of all the cells shows that WCMI is able to tell these two states with higher accuracy, compared with CA and WI score, in all three brain areas (Figure 6B, with the overall score of all three areas as WCMI $89.19\% \pm 5.10\%$, WI $53.11\% \pm 13.52\%$, and Ca $66.80\% \pm 20.82\%$). WCMI significantly outperforms WI for all three brain areas (Figure 6C). These results indicate that WCMI is applicable to general neural signals without stimulation from various parts of brain areas.

Decoding speech syllables from human ECoG signals

To further demonstrate the applicability of the method, we employed a dataset of ECoG signals recorded in humans when subjects read 57 different syllables.⁹ Thirteen active channels recorded from the ventral sensorimotor cortex (vSMC) of humans show significant cortical activity (Figure 7A), which was used for our analysis. Features selected by WI and WCMI were used to decode all 57 syllables. When using various *t*-way dependencies in WCMI, selected features are more distributed across 13 channels, while WI only chooses features from a few restricted channels (Figures 7B and S5). The decoding performance with a wide range of features shows that WCMI is able to use much fewer features while achieving higher performance (Figures 7C and S6). In this case, the efficiency of WCMI is even more striking: WCMI can obtain a performance close to 90% with 700 features (45% with 55 features) computed from fitted Hill functions, while WI has to use much more features

yet reach lower performance (Table S3). The *p* values of *t* test and Wilcoxon rank-sum test were close to 0 ($p < 10^{-10}$), indicating that the difference in performance between WI and WCMI is significant. Compared with previous decoding results (around 60% accuracy) with the same data using the feature of frequency bands and the decoder of deep learning networks,³⁴ our results yield higher performance while using fewer features. These results suggest that an efficient and effective way of feature selection is more important for decoding, and WCMI is highly efficient to analyze coarse-scale brain activity.

DISCUSSION

In this study, we propose a novel method, WCMI, that enables us to extract non-redundant, meaningful information leveraging wavelet analysis, information theory, and feature selection. WCMI selects features of spatiotemporal neural signals represented by wavelet coefficients taking into account the interaction and dependency between features to avoid redundancy and over-representation. The capability of WCMI was demonstrated on various types of neural data, simulated spikes, experimental neural spikes and calcium signals, and ECoG signals from humans. The features selected by WCMI can be feasibly used for decoding without needing as large a set of features as WI. Moreover, WCMI is able to capture the distributed and sparse coding property of neural populations. These results provide new insights on utilizing wavelet analysis to extract reliable information embedded in spatiotemporal data.

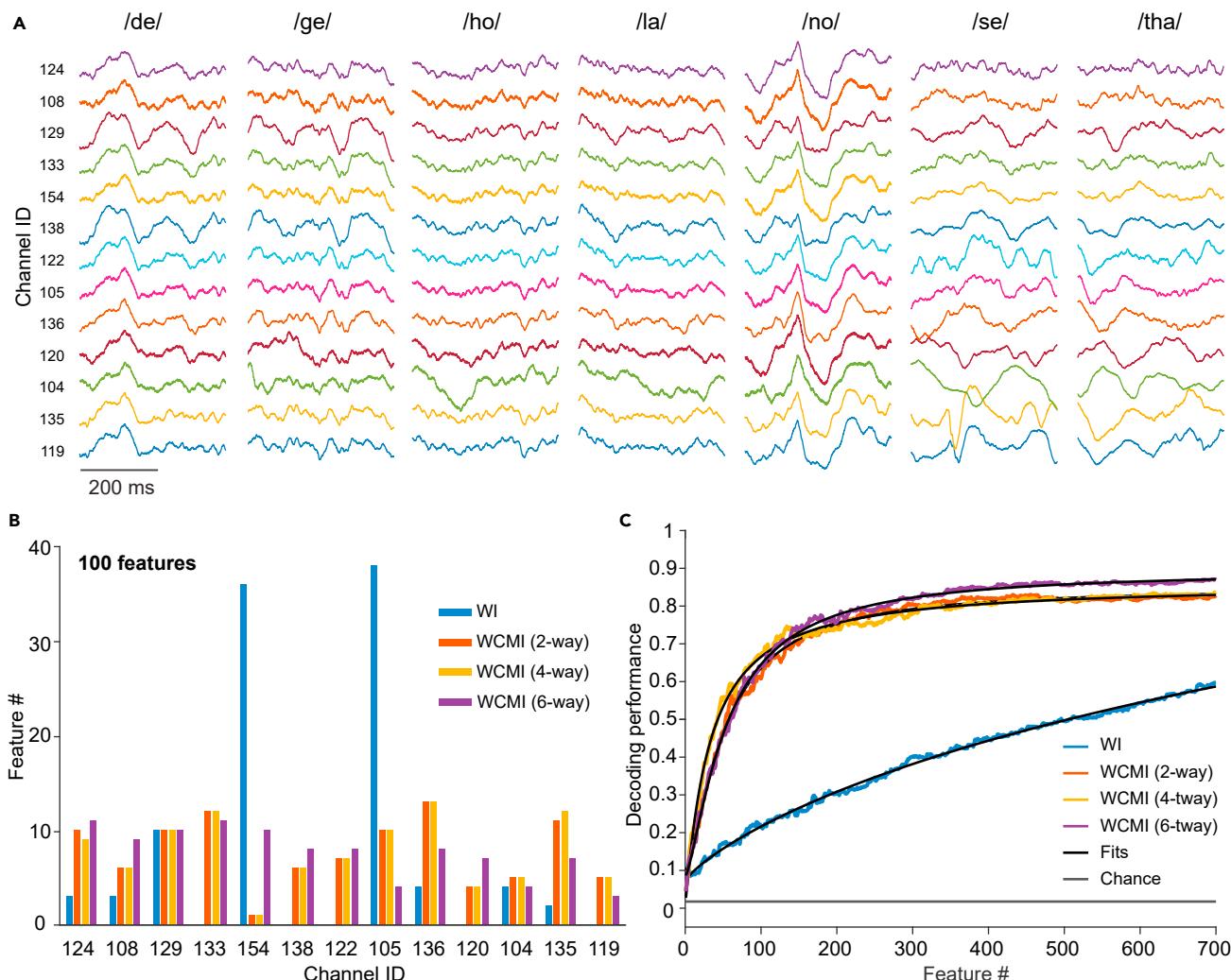


Figure 7. Decoding speech syllables from human ECoG signals

(A) Example ECoG signals from 13 channels in vSMC during syllable articulation of 7 syllables.

(B) The distribution of all 100 selected features by WI and WCMI.

(C) The decoding performance using features selected by WI and WCMI. Different values of t-way dependency were used for WCMI. Gray lines are the chance level. Solid black lines are fits to a Hill function.

Extracting information from neural signals

Although the question of rate-coding versus temporal-coding in neural coding is still under debate, it is acknowledged that both these coding strategies are utilized by neurons in various parts of the neural system.^{13,35} The reflection on this point can also impact how to extract information coded by neural signals sampling at a low rate, such as two-photon calcium imaging data. For that, it remains unclear how to tackle information coded calcium data compared with spikes,^{8,36} despite the fact that a number of techniques have been developed to infer spikes from calcium signals.³⁷

Here, we applied WCMI to two-photon calcium imaging data recorded with 10 Hz. We found that features selected by WI from calcium signals are in general not good enough for decoding. However, WCMI is able to obtain superb decoding performance. This disparity is due to the way of selecting the features

between WCMI and WI, as WI ignores the dependency between features. As a result, feature redundancy causes serious degradation of decoding for WI. Further application to ECoG signals shows that WCMI is more dramatic to extract useful features across channels for better decoding. In contrast, using the advanced decoder, such as deep learning networks but with different types of features (frequency bands), can only obtain a fairly good decoding performance.³⁴ However, the final decoding performance relies on both features and decoders, which needs to be explored systematically in detail in future work. Our current results suggest that WCMI is able to extract meaningful information from calcium signal and coarse-scale brain activity, and could potentially be used for analyzing other types of neural activity, such as fMRI³⁸ and wide-field calcium imaging signals.³⁹

Information extraction by dimensionality reduction

Many studies have shown that the patterns of neural activity carry the meaningful information of stimulus and behavior. Recent advances in experimental techniques allow us to record neurons simultaneously in large scales of space (multiple brain areas) and time (multiple days), which raises a question on analysis known as the curse of dimensionality.⁴⁰

Although many methods have been proposed to reduce dimensionality, either linearly or nonlinearly, to a small set, the question about the meanings of reduced or selected dimensions is still open.²¹ For instance, the classic principal-component analysis (PCA) can provide a useful picture on the data structure,⁴¹ while the dimensions uncovered by PCA do not match the real biological underpinnings.³¹ The dimensions revealed by nonlinear methods are more meaningful for computation,^{31,42} however, they are difficult to be interpreted by biological correspondences, except for a few cases.^{31,43–45} A practical way of justifying these dimensions is to use a decoder for computation.¹³ The WCMI proposed here can extract features well for decoding. In addition, these features correspond to the statistics in neuronal responses in the retina and high-density recordings of human ECoG signals.

Features selected by wavelet analysis

Wavelet analysis, as a classic technique for signal processing,²³ has been widely used for describing temporal signals.⁴⁶ Features obtained by wavelet decomposition were shown as a useful way of dealing with neural data in multiple scales.^{24,25} Taking advantage of the powerful framework of information theory,¹³ computing mutual information carried out by features has been widely used for characterizing the importance of features.^{47,48} However, the dependency and redundancy between features are prevailing in neural coding.^{30,31,49} The WCMI proposed here utilizes conditional mutual information to handle high redundancy between features.

The features, represented by wavelet coefficients selected by two methods, are dramatically different. Features from WI closely align with neural activity over time, e.g., concentrating with a high frequency of firing rate at those time locations. In contrast, features from WCMI are widely distributed over all time locations, indicating a distributed coding strategy and leaving out redundancy between features. Consistent with other studies suggesting that decorrelation and efficient coding can be implemented by neurons in early visual systems, including retinal ganglion cells⁵⁰ and lateral geniculate nucleus,⁵¹ the RF of those cells selected by WCMI are more distributed, while those by WI are more compact, over the image space. Our results indicate that features selected by WCMI relate to the sparse coding of neural populations within the context of the statistics of natural images. Our results presented here are potentially useful for understanding the internal state of neural activity and meaningful features contributing to the decoding of external stimuli⁵² and downstream behavior.⁵³ Our results provide a new way of representing features of spatiotemporal data by utilizing a small set of non-redundant wavelets. Such an efficient approach of feature representation could be potentially applied to construct a low-dimensional feature space for facilitating the design of novel machine learning algorithms.

Advantages and limitations

Based on information theory and wavelet analysis, WCMI is a straightforward extension of WI by taking into account the dependency between wavelet coefficients yielding non-redundant features. Here, we demonstrate the benefit of this extension on extracting useful features from time series of neural signals. Such an extension is nontrivial as WCMI can obtain more effective features archiving higher performance on decoding while using fewer features. With the same ECoG dataset, the previous study obtained a classification accuracy of around 60% using the feature of different frequency bands of signals decoded with deep learning networks.³⁴ In contrast, both WI and WCMI using wavelets are able to achieve better performance. However, WCMI only needs a few features, achieving even higher accuracy close to 90%. Taken together, our results suggest that WCMI provides a more efficient way of selecting features from spatiotemporal data, and these informative features can be used for decoding without requiring sophisticated decoders.

Wavelet analysis, as a popular method for time series, can obtain a set of wavelet coefficients with rich information. The scalogram of the whole set of coefficients, represented as an image, has been used for decoding using various types of deep learning neural network decoders.^{54,55} Here, we only selected a set of prominent coefficients as a feature set for decoding, and showed that a small set selected by WCMI is able to decode a significant portion of information. This feature extraction step can capture essential information embedded in the whole set of coefficients. Thus, further work is needed to use the feature set selected by WCMI as a preprocessing step of data to facilitate the training of large-scale deep learning networks for decoding.

Real-life data are always formatted as a population of time series where both space and time dimensions are prevalent. Neuroscience, as a highly interdisciplinary field of data science, generates different types of time-series data, which are testbeds for various computational methods.⁵⁶ Here, we used a population of neural responses at different sampling rates. Single-cell resolution guarantees that the time series of each individual cell is unique, so that wavelet analysis can extract meaningful information across different timescales from one sequence for each cell, which does not contain the dimensionality of space. Whereas, a population of cells simultaneously recorded enables us to extract information in both time and space while taking into account the interaction between cells and across timescales. This is in contrast to other methods of dimensionality reduction, such as PCA,⁵⁷ non-negative matrix factorization (NMF),³¹ Uniform Manifold Approximation and Projection (UMAP),⁵⁸ and *t*-distributed stochastic neighbor embedding (*t*-SNE).⁵⁹ These methods are mostly used for reducing the dimensionality of space or cell population into a few components while keeping temporal dimension unchanged, which are useful for analyzing low-dimensional dynamical systems over time. Our approach, instead, extracts a set of features from the spatiotemporal patterns of signals taking away both space and time dimensions. Thus, it is more suitable for decoding purposes. However, it is possible to combine our approach with PCA, NMF, UMAP, *t*-SNE, and other dimensionality reduction methods. One can use these methods to obtain a few reduced components first,

then apply WCMI to these principal or significant dimensions for feature extraction. Further work is needed to systematically investigate the mutual benefit between wavelet analysis and dimensionality reduction methods.

EXPERIMENTAL PROCEDURES

Resource availability

Lead contact

The lead contact for this study is Zhaofei Yu (yuzf12@pku.edu.cn).

Materials availability

This study did not generate unique materials.

Data and code availability

The data and code are available at <https://doi.org/10.5281/zenodo.5722875>.

Wavelet analysis

Wavelet analysis, as local transformation in temporal and frequency domains, can perform multi-scale analysis on a signal through operation functions, such as expansion and translation. The wavelet bases used in wavelet transform are finite and decaying wavelets, which slide on the signal to obtain the frequency and locate the time. Given a time-series signal $x(t)$, the wavelet coefficient for a wavelet transform is

$$w_{\Psi}(a, b) = \langle x, \Psi_{a,b} \rangle,$$

where $\Psi_{a,b} = |a|^{-1/2}\Psi(t-b/a)$ is a wavelet function. Basic wavelets have different frequencies and time shifts, controlled by the parameters a and b , which are the scale and the translation, respectively. The scale parameter controls the expansion and contraction of the wavelet function, whereas the translation parameter changes its temporal location. The wavelet transform value, i.e., the coefficient of a wavelet function used to fit the original signal, can be obtained by calculating the inner product between the original signal and each wavelet function.

To implement wavelet analysis, we used a similar approach as in previous works.²⁴ In brief, Haar wavelets were used to implement a four-scale dyadic wavelet decomposition. The wavelet transform was computed by the efficient multiresolution decomposition algorithm,²⁵ which decomposes the signal into a series of detail scales and final approximate values corresponding to time-localized activities in different frequency bands. For the scale of a , we used four scales to obtain four detailed coefficients, D1, D2, D3, and D4, and one approximate coefficient A4. The translation b is in line with the temporal dimension of the time series.

Feature selection using wavelet coefficients

After the neural signal was decomposed by Haar wavelet, a set of wavelet coefficients was obtained. Selecting important coefficients from the whole set can be regarded as a process of feature selection. Here, we compared two strategies for selecting wavelet coefficients.

WI: in the previous study,²⁴ mutual information was used to select wavelet coefficients based on surrogate testing of the significance of information values named WI. In brief, mutual information with a stimulus was calculated first to obtain the information content of each coefficient corresponding to the stimulus. For a set of stimuli or conditions of interest S and a set of values $w_{a,b}$ of a wavelet coefficient at the scale of a and translation of b , the mutual information was computed as

$$I(S; w_{a,b}) = \sum_{S, w_{a,b}} P(S, w_{a,b}) \log_2 \frac{P(S, w_{a,b})}{P(S)P(w_{a,b})},$$

where $P(S)$ and $w_{a,b}$ are the probabilities of having condition S and coefficient $w_{a,b}$, respectively. $P(S, w_{a,b})$ is the joint probability. With the information computed from each coefficient, the significance of the information was obtained by surrogate testing with shuffling trials, so that unbiased information was obtained by comparing the original information values with the threshold of 95% percentiles of the shuffled distribution. Finally, the coefficients were ranked according to the unbiased information, where the number of features was taken in order of rank.

WCMI: WI does not consider the potential redundancy between features embedded in the data, which makes the rank of the information a sub-optimal approach without taking into account interaction or complementarity between features. Instead, we considered an approach, named WCMI, leveraging a new way of feature selection based on conditional mutual information and the sequential forward selection strategy, such that it is able to account for high-order dependency between features and removes their redundancy.

Instead of computing the mutual information $I(S; w_{a,b})$ for each coefficient individually, we considered a t -way feature interaction in a set of coefficients. We started with a feature set K initializing with the maximum mutual information of $I(S; w_{a,b})$, e.g., $K := \{\text{argmax}_{\{w_{a,b}\}} I(S; w_{a,b})\}$, which is the top 1 in the rank of the WI method. Then for a candidate coefficient $w_{a,b}$, the conditional mutual information of S and $w_{a,b}$ given the feature set K was computed as

$$I(S; w_{a,b}|K) = I(S; w_{a,b}, K) - I(S; K),$$

which measures the amount of additional information about S carried by $w_{a,b}$ compared with the existing feature set K . This was done iteratively with the sequential forward selection strategy for remaining coefficients $w_{a,b}$ to find $\max_{\{w_{a,b}, w_{a,b} \notin K\}} I(S; w_{a,b}|K)$ by an algorithm CMICOT.²⁸ Then the newly selected coefficient was added into the set K . When the size of K reaches t , an additional optimal complementary strategy can be used to add more features to form the final optimal feature K set with any desired number of features, while any element in K carries non-redundant information about S . The CMICOT is a fast algorithm with low computational complexity and is able to accommodate a higher value of the parameter t , depending on the size of the data and computer hardware. For a typical value of $t = 6$, the running cost for our data presented here is reasonable. We also tested different values of the parameter t , the outcomes are similar with higher t good for better decoding performance. The benefit of WCMI is that it needs significantly fewer features than WI, yet achieves higher performance for decoding.

To visualize the wavelet coefficients, we used the MATLAB function `wscalogram` and plotted them as a scalogram with the full set of coefficients obtained in one trial of data. The subset of features selected by WI and WCMI were visualized with a scalogram where each feature/coefficient was assigned to rank-order values according to their mutual information of WI or conditional mutual information of WCMI.

Decoding using selected features

Once features were selected, we qualified the efficiency and effectiveness of these features using a decoder to classify a set of stimuli or conditions. The main purpose of this decoding step is to test the meaningfulness of features and compare WI and WCMI, rather than demonstrate the quality of the decoder itself. We used several decoders to avoid the bias of choosing a specific decoder. (1) LDA: the LDA was implemented with the MATLAB function `classify`. (2) SVM: the SVM was implemented using the LIBSVM toolbox.⁶⁰ (3) NB classifier: the NB was implemented using the MATLAB function `fitcnb`. All these decoders were tested in Figure 2, and the results are similar in terms of the behaviors of WI and WCMI, in which WCMI has higher performance with fewer features. Thus, we used LDA as the default decoder in all of our analyses. The wavelet coefficients corresponding to each trial were taken as the input for the decoder, and the output of the decoder is the stimulus/condition category. For populating analysis, we concatenated all the cells or channels for one trial as one input. The training of the decoders follows a scheme of 13-/40-fold or leave-one-out cross-validation, and the additional test data were used to obtain the decoding performance shown in this work. The accuracy of test trials was used to obtain a confusion matrix quantifying the performance of decoding results. The overall accuracy was computed as the mean, while the SD of the accuracy was obtained across the whole confusion matrix. The numbers of categories N to be classified in all the datasets used in this work are balanced. Therefore, for each dataset, the theoretical chance level is $1/N$. To verify the performance difference between WI and WCMI, we used both t test and Wilcoxon rank-sum test for the hypothesis test, where the p value is used to measure whether the difference is significant.

To quantify the change of the performance P over a number of features F from WI and WCMI, we used the following Hill function to fit the curves: $P = P_{\max}/(1 + (F_{50}/F)^n) + P_0$, where n is the exponent factor, P_0 the offset,

and P_{\max} the maximum performance. F_{50} is the value of F at which F reaches half maximum performance. The fitted parameters were reported in supplemental information.

Dataset 1: Simulated spiking data with different patterns

We used the same set of simulated spike data with various types of specific spike patterns as previously.²⁴ The data are to simulate spike responses of a single cell to different stimuli with background noise included. Four simulated cells have various spike patterns of the same or different firing rates with different delays and timescales, which can be used to test and compare different methods for classifying spike patterns. There are four stimuli simulated, and each stimulus triggered 40, 20, 10, and 20 trials of spike responses for cells 1, 2, 3, and 4, respectively. Decoding was done for 4-class discrimination using single trial data, so that the chance level is 25%. The decoders were trained by the leave-one-out cross-validation, and train-test ratios were 159:1, 79:1, 39:1, and 79:1, respectively, for cells 1–4. The time series of one trial was 256 ms long with a sampling rate of 1,000 Hz. Wavelet analysis was implemented on each trial to get a set of 256 coefficients, from which features were extracted by WI and WCMI. Each feature can be mapped to a pair of wavelet parameters (a, b) with $a = \{D1, D2, D3, D4, A4\}$ and $b = [1, 256]$, as shown in Figure 1C. The same procedure was done for all other datasets below.

Dataset 2: Neural spikes in response to natural images

The data were recorded from ganglion cells of salamander retinas using multi-electrode arrays, as described previously.⁴⁸ This was to test neural encoding of natural images, as described previously.^{29,31,61} The stimulus set contains 300 grayscale natural images using a McGill calibrated color image database.⁶² Each image was displayed for 200 ms following an 800 ms interval with a gray background as one trial. There are 80 cells and each cell has 13 trials for one image. The time series of one trial is 1,000 ms long with a sampling rate of 1,000 Hz. We used the first 300 ms spiking data for our analysis as this time window was enough to capture the spiking activity of image stimulation.^{29,31} Then each trial per image per cell is a time series of 300 points. Wavelet analysis was applied to each trial to get a set of 300 coefficients. With a population of 80 cells, a set of 24,000 coefficients was obtained, from which features were extracted by WI and WCMI. Note that the final features were selected across all the cells, e.g., there is no bias of which cell should be used or not. As a result, in the final set selected, there could be more features from some cells and fewer features from the other cells. In total, 3,900 samples with 13 trials of 300 natural images were used for decoding of a 300-class task with the chance level as 0.3%. We used the leave-one-out cross-validation with the training set of 3,899 samples and test with 1 sample. We also did the 13-fold cross-validation with the training set of 3,600 samples and test with 300 samples. Both scenarios of cross-validation have similar outcomes. After decoding, a confusion matrix of 300×300 as the decoding performance was obtained. In the data, the RFs of cells were given as a 2D Gaussian function, where the centers and covariances were given in the unit of physical distance, μm .²⁹ To compute the distance between cells, the Euclidean distance between the centers of Gaussian RFs was computed for pairwise cells. The process was repeated for 50 groups of images, where each group has 4 random images out of all the 300 images.

Dataset 3: Direction selectivity of neural spikes

Experimental spiking data recorded from ganglion cells of the salamander retinas was to test direction selectivity, as described previously.³¹ In brief, to determine the direction preference of each cell, moving grating images were used to stimulate cells. The gratings were arranged in eight equally spaced directions of motion. The conventional way of describing direction selection is using spike counts by calculating the number of spikes, e.g., spike count, of each cell for each direction. A polar plot with spike response in all eight directions was used to represent the direction selectivity of a cell. There are 80 cells and each cell has 5 trials for one direction. The time series of one trial is 6,670 ms long with a sampling rate of 1,000 Hz. Wavelet analysis was applied to each trial to get a set of 6,670 coefficients. To generate a polar plot of decoding score, pairwise decoding was conducted for two adjacent directions with the classification chance level as 50%, then, for each direction, scores were averaged over its neighbor directions. Specifically, if the score for direction

0 and 45 is $S_{(0,45)}$, and the score for 0 and -45 is $S_{(0,-45)}$, then the score for the direction 0 is $(S_{(0,45)}+S_{(0,-45)})/2$. Averaging over all pairs can obtain the polar plot of the score for each cell, as shown in Figure 4B for two sample cells. We used the leave-one-out cross-validation with the train set of nine samples and test of one sample.

Dataset 4: Direction selectivity of calcium imaging signals

The data of calcium imaging neural responses of 10 Hz were recorded from mice stimulated by moving gratings for direction selectivity.³³ Two-photon calcium imaging was used to record neural responses in the primary visual cortex of mice. Evoked responses of layer 2/3 neurons were recorded in awake and anesthetized mice. The grating images of eight directions were displayed sequentially, where each direction was fixed for 2 s and then moved for 2 s. In total, 95 neurons and 6 trials were recorded in anesthetized mice, and 73 neurons and 6 trials were recorded in awake mice. The time series of one trial is 4-s long with a sampling rate of 10 Hz, yielding 40 data points per trial. Wavelet analysis was applied to each trial to get a set of 40 coefficients. Similar to dataset 3, the polar plot of decoding scores of pairwise directions with the chance level of 50% for each cell was obtained, as shown at the bottom of Figures 5A and 5C for four sample cells. We used the leave-one-out cross-validation with the train set of 11 samples and test of 1 sample.

Dataset 5: Spontaneous neural activity of calcium imaging signals

The data are calcium imaging signals recorded by multiarea two-photon real-time *in vivo* explorer, as described previously.⁶ In brief, three brain areas were recorded simultaneously using two-photon calcium imaging with single-cell resolution. In these data, spontaneous neural activities were recorded in the V1, M1, and CA1 of mice in both anesthetized and awake states. Neural responses of one mouse were recorded in both anesthetized and awake states, so that all the cells identified have activity in both states. In total, 71 cells from V1, 74 from M1, and 79 from CA1 were recorded in the anesthetized state first. After the mice woke up, neural responses of the same set of cells were recorded again. The time series of one cell is 300-s long with a sampling rate of 10 Hz, yielding 3,000 data points. Wavelet analysis was applied to each cell to get a set of 3,000 coefficients. The two-class decoding of either anesthetized and awake states was conducted for each cell with the chance level of 50%. As there is no breakdown trial in this dataset, we split a 300-s long dataset into a set of 50-s long each as a sample. We used the leave-one-out cross-validation with the train set of 11 samples and test of 1 sample.

Dataset 6: Syllable articulation of human ECoG signals

The data are ECoG signals recorded in human patients when they read different syllables, as described previously.^{9,34} The data were recorded from the vSMC that is closely related to the cortical control of pronunciation. A high-density, subdural 256-channel ECoG array was implanted in human patients as part of their clinical treatment of epilepsy. The ECoG signals were collected when subjects read out 57 different syllables (consonants and vowels) in random order from a list. A total of 3 participants were recruited. A time series of each syllable has 1,200 data points with a sampling rate of 3,052 Hz. Out of 256 channels, 13 active channels identified showing significant responses⁹ were concatenated into one trial of 15,600 data points. In total, there are 57 classes of syllables with 40 trials in each class. The 57-class decoding of each syllable was conducted with the chance level of 1.75%. We used the 40-fold cross-validation with the train set of 2,223 samples and test of 57 samples.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.patter.2021.100424>.

ACKNOWLEDGMENTS

We would like to thank Hongbao Jia, Rodrigo Quian Quiroga, and Stefano Panzeri for helpful discussion. This work was supported by the National Natural Science Foundation of China (62176003, 62088102, and 61961130392), the Beijing Major Science and Technology Project (Z191100010618003), the

Chongqing Postdoctoral Science Special Foundation (2019LY18), and the Royal Society Newton Advanced Fellowship, UK (NAF-R1-191082).

AUTHOR CONTRIBUTIONS

Conceptualization, J.K.L. and Z.Y.; data curation, S.J., X.L., and J.K.L.; formal analysis, S.J.; investigation, S.J.; methodology, S.J., J.K.L., and Y.Z.; writing, S.J., J.K.L., and Z.Y.; funding acquisition, X.L., T.H., J.K.L., and Z.Y.; supervision, T.H., J.K.L., and Z.Y.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 26, 2021

Revised: September 22, 2021

Accepted: December 9, 2021

Published: February 11, 2022

REFERENCES

- Victor, J.D., and Purpura, K.P. (1996). Nature and precision of temporal coding in visual cortex: a metric-space analysis. *J. Neurophysiol.* **76**, 1310–1326.
- Strong, S.P., Koberle, R., van Steveninck, R.R.D., and Bialek, W. (1998). Entropy and information in neural spike trains. *Phys. Rev. Lett.* **80**, 197–200.
- Rieke, F., Warland, D., de Ruyter van Steveninck, R., and Bialek, W. (1999). *Spikes: Exploring the Neural Code (Computational Neuroscience)* (The MIT Press), ISBN 9780262181747.
- Panzeri, S., Petersen, R.S., Schultz, S.R., Lebedev, M., and Diamond, M.E. (2001). The role of spike timing in the coding of stimulus location in rat somatosensory cortex. *Neuron* **29**, 769–777.
- Grienberger, C., and Konnerth, A. (2012). Imaging calcium in neurons. *Neuron* **73**, 862–885.
- Yang, M., Zhou, Z., Zhang, J., Jia, S., Li, T., Guan, J., Liao, X., Leng, B., Lyu, J., Zhang, K., et al. (2019). MATRIEX imaging: multiarea two-photon real-time *in vivo* explorer. *Light Sci. Appl.* **8**, 1–11.
- Ledochowitsch, P., Huang, L., Knoblich, U., Oliver, M., Lecoq, J., Reid, C., Li, L., Zeng, H., Koch, C., Waters, J., et al. (2019). On the correspondence of electrical and optical physiology in *in vivo* population-scale two-photon calcium imaging. *BioRxiv*. <https://doi.org/10.1101/800102>.
- Wei, Z., Lin, B.J., Chen, T.W., Daie, K., Svoboda, K., and Druckmann, S. (2020). A comparison of neuronal population dynamics measured with calcium imaging and electrophysiology. *PLoS Comput. Biol.* **16**, e1008198.
- Bouchard, K.E., Mesgarani, N., Johnson, K., and Chang, E.F. (2013). Functional organization of human sensorimotor cortex for speech articulation. *Nature* **495**, 327–332.
- Logothetis, N.K. (2002). The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* **357**, 1003–1037.
- de Ruyter van Steveninck, R.R., Lewen, G.D., Strong, S.P., Koberle, R., and Bialek, W. (1997). Reproducibility and variability in neural spike trains. *Science* **275**, 1805–1808.
- Shadlen, M.N., and Newsome, W.T. (1994). Noise, neural codes and cortical organization. *Curr. Opin. Neurobiol.* **4**, 569–579.
- Quiroga, R., and Panzeri, S. (2009). Extracting information from neuronal populations: information theory and decoding approaches. *Nat. Rev. Neurosci.* **10**, 173–185.
- Harris, K.D., Quiroga, R.Q., Freeman, J., and Smith, S.L. (2016). Improving data quality in neuronal population recordings. *Nat. Neurosci.* **19**, 1165–1174.
- Bensmaia, S.J., and Miller, L.E. (2014). Restoring sensorimotor function through intracortical interfaces: progress and looming challenges. *Nat. Rev. Neurosci.* **15**, 313–325.
- Yu, Z., Liu, J.K., Jia, S., Zhang, Y., Zheng, Y., Tian, Y., and Huang, T. (2020). Toward the next generation of retinal neuroprosthesis: visual computation with spikes. *Engineering* **6**, 449–461.
- Shah, N.P., and Chichilnisky, E.J. (2020). Computational challenges and opportunities for a bi-directional artificial retina. *J. Neural Eng.* **17**, 055002.
- Sussillo, D., and Barak, O. (2013). Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks. *Neural Comput.* **25**, 626–649.
- Bassett, D.S., and Sporns, O. (2017). Network neuroscience. *Nat. Neurosci.* **20**, 353–364.
- Zheng, Y., Jia, S., Yu, Z., Liu, J.K., and Huang, T. (2021). Unraveling neural coding of dynamic natural visual scenes via convolutional recurrent neural networks. *Patterns* **2**, 100350.
- Cunningham, J.P., and Yu, B.M. (2014). Dimensionality reduction for large-scale neural recordings. *Nat. Neurosci.* **17**, 1500–1509.
- Strang, G., and Nguyen, T. (1996). *Wavelets and Filter Banks (SIAM)*.
- Mallat, S. (2008). *A Wavelet Tour of Signal Processing* (Elsevier LTD), ISBN 0123743702.
- dos Santos, V.L., Panzeri, S., Kayser, C., Diamond, M.E., and Quiroga, R.Q. (2015). Extracting information in spike time patterns with wavelets and information theory. *J. Neurophysiol.* **113**, 1015–1033.
- dos Santos, V.L., Rey, H.G., Navajas, J., and Quiroga, R.Q. (2018). Extracting information from the shape and spatial distribution of evoked potentials. *J. Neurosci. Methods* **296**, 12–22.
- Harris, K.D. (2005). Neural signatures of cell assembly organization. *Nat. Rev. Neurosci.* **6**, 399–407.
- Brenner, N., Strong, S.P., Koberle, R., Bialek, W., and de Ruyter van Steveninck, R.R. (2000). Synergy in a neural code. *Neural Comput.* **12**, 1531–1552.
- Shishkin, A., Bezzubtseva, A., Drutsa, A., Shishkov, I., Gladikh, E., Gusev, G., and Serdyukov, P. (2016). Efficient high-order interaction-aware feature selection based on conditional mutual information. *Adv. Neural Inf. Process. Syst.* **4637**–4645.
- Liu, J.K., and Gollisch, T. (2021). Simple model for encoding natural images by retinal ganglion cells with nonlinear spatial integration. *BioRxiv*. <https://doi.org/10.1101/2021.08.29.458067>.
- Beyeler, M., Rounds, E.L., Carlson, K.D., Dutt, N., and Krichmar, J.L. (2019). Neural correlates of sparse coding and dimensionality reduction. *PLoS Comput. Biol.* **15**, e1006908.
- Liu, J.K., Schreyer, H.M., Onken, A., Rozenblit, F., Khani, M.H., Krishnamoorthy, V., Panzeri, S., and Gollisch, T. (2017). Inference of neuronal functional circuitry with spike-triggered non-negative matrix factorization. *Nat. Commun.* **8**, 149.
- Huang, L., Ledochowitsch, P., Knoblich, U., Lecoq, J., Murphy, G.J., Reid, R.C., de Vries, S.E., Koch, C., Zeng, H., Buice, M.A., et al. (2021). Relationship between simultaneously recorded spiking activity and fluorescence signal in GCaMP6 transgenic mice. *eLife* **10**, e51675.
- Jia, H., Rochefort, N.L., Chen, X., and Konnerth, A. (2010). Dendritic organization of sensory input to cortical neurons *in vivo*. *Nature* **464**, 1307–1312.
- Livezey, J.A., Bouchard, K.E., and Chang, E.F. (2019). Deep learning as a tool for neural data analysis: speech classification and cross-frequency coupling in human sensorimotor cortex. *PLoS Comput. Biol.* **15**, e1007091.
- Brette, R. (2015). Philosophy of the spike: rate-based vs. spike-based theories of the brain. *Front. Syst. Neurosci.* **9**, 151.
- Wang, M., Liao, X., Li, R., Liang, S., Ding, R., Li, J., Zhang, J., He, W., Liu, K., Pan, J., et al. (2020). Single-neuron representation of learned complex sounds in the auditory cortex. *Nat. Commun.* **11**, 1–14.
- Berens, P., Freeman, J., Deneux, T., Chenkov, N., McColgan, T., Speiser, A., Macke, J.H., Turaga, S.C., Mineault, P., Rupprecht, P., et al. (2018). Community-based benchmarking improves spike rate inference from two-photon calcium imaging data. *PLoS Comput. Biol.* **14**, e1006157.

38. Logothetis, N.K. (2008). What we can do and what we cannot do with fMRI. *Nature* **453**, 869–878.
39. Stirman, J.N., Smith, I.T., Kudnerov, M.W., and Smith, S.L. (2016). Wide field-of-view, multi-region, two-photon imaging of neuronal activity in the mammalian brain. *Nat. Biotechnol.* **34**, 857–862.
40. Kayser, C., Montemurro, M., Logothetis, N., and Panzeri, S. (2009). Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. *Neuron*, 597–608.
41. Optican, L.M., and Richmond, B.J. (1987). Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. Information theoretic analysis. *J. Neurophysiol.* **57**, 162–178.
42. Gallego, J.A., Perich, M.G., Miller, L.E., and Solla, S.A. (2017). Neural manifolds for the control of movement. *Neuron* **94**, 978–984.
43. Shenoy, K.V., Sahani, M., and Churchland, M.M. (2013). Cortical control of arm movements: a dynamical systems perspective. *Annu. Rev. Neurosci.* **36**, 337–359.
44. Jia, S., Yu, Z., Onken, A., Tian, Y., Huang, T., and Liu, J.K. (2021). Neural system identification with spike-triggered non-negative matrix factorization. *IEEE Trans. Cybernetics*, 1–12.
45. Jia, S., Xing, D., Yu, Z., and Liu, J.K. (2021). Dissecting cascade computational components in spiking neural networks. *PLoS Comput. Biol.* **17**, e1009640.
46. Quiroga, R.Q., Nadasdy, Z., and Ben-Shaul, Y. (2004). Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Comput.* **16**, 1661–1687.
47. Brunel, N., and Nadal, J.P. (1998). Mutual information, Fisher information, and population coding. *Neural Comput.* **10**, 1731–1757.
48. Liu, J.K., and Gollisch, T. (2015). Spike-triggered covariance analysis reveals phenomenological diversity of contrast adaptation in the retina. *PLoS Comput. Biol.* **11**, e1004425.
49. Latham, P.E. (2005). Synergy, redundancy, and independence in population codes, revisited. *J. Neurosci.* **25**, 5195–5206.
50. Pitkow, X., and Meister, M. (2012). Decorrelation and efficient coding by retinal ganglion cells. *Nat. Neurosci.* **15**, 628–635.
51. Dan, Y., Atick, J.J., and Reid, R.C. (1996). Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. *J. Neurophysiol.* **16**, 3351–3362.
52. Zhang, Y., Jia, S., Zheng, Y., Yu, Z., Tian, Y., Ma, S., Huang, T., and Liu, J.K. (2020). Reconstruction of natural visual scenes from neural spikes with deep neural networks. *Neural Networks* **125**, 19–30.
53. Harris, K.D., and Thiele, A. (2011). Cortical state and attention. *Nat. Rev. Neurosci.* **12**, 509–523.
54. Byeon, Y.H., Pan, S.B., and Kwak, K.C. (2019). Intelligent deep models based on scalograms of electrocardiogram signals for biometrics. *Sensors* **19**, 935.
55. Halberstadt, A.L. (2020). Automated detection of the head-twitch response using wavelet scalograms and a deep convolutional neural network. *Sci. Rep.* **10**, 1–12.
56. Akil, H., Martone, M.E., and Essen, D.C.V. (2011). Challenges and opportunities in mining neuroscience data. *Science* **331**, 708–712.
57. Kobak, D., Brendel, W., Constantinidis, C., Feierstein, C.E., Kepecs, A., Mainen, Z.F., Qi, X.L., Romo, R., Uchida, N., and Machens, C.K. (2016). Demixed principal component analysis of neural population data. *eLife* **5**, e10989.
58. McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for dimension reduction. *arXiv*, 1802.03426v3.
59. van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605.
60. Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* **20**, 273–297.
61. Onken, A., Liu, J.K., Karunasekara, P.C.R., Delis, I., Gollisch, T., and Panzeri, S. (2016). Using matrix and tensor factorizations for the single-trial analysis of population spike trains. *PLoS Comput. Biol.* **12**, e1005189.
62. Olmos, A., and Kingdom, F.A.A. (2004). A biologically inspired algorithm for the recovery of shading and reflectance images. *Perception* **33**, 1463–1473.