Research Article

Single-Trial Classification of Bistable Perception by Integrating Empirical Mode Decomposition, Clustering, and Support Vector Machine

Zhisong Wang,[1] Alexander Maier,[2] Nikos K. Logothetis,[3] and Hualou Liang[1]

[1] School of Health Information Sciences, University of Texas Health Science Center at Houston, 7000 Fannin,
Suite 600, Houston, TX 77030, USA

[2] Unit on Cognitive Neurophysiology and Imaging, National Institute of Health, Building 49, Room B2J-45, MSC-4400,
49 Convent Dr., Bethesda, MD 20892, USA

[3] Max Planck Institut für biologische Kybernetik, Spemannstraße 38, 72076 Tübingen, Germany

Correspondence should be addressed to Hualou Liang, hualou.liang@uth.tmc.edu

We propose an empirical mode decomposition (EMD-) based method to extract features from the multichannel recordings of local field potential (LFP), collected from the middle temporal (MT) visual cortex in a macaque monkey, for decoding its bistable structure-from-motion (SFM) perception. The feature extraction approach consists of three stages. First, we employ EMD to decompose nonstationary single-trial time series into narrowband components called intrinsic mode functions (IMFs) with time scales dependent on the data. Second, we adopt unsupervised K-means clustering to group the IMFs and residues into several clusters across all trials and channels. Third, we use the supervised common spatial patterns (CSP) approach to design spatial filters for the clustered spatiotemporal signals. We exploit the support vector machine (SVM) classifier on the extracted features to decode the reported perception on a single-trial basis. We demonstrate that the CSP feature of the cluster in the gamma frequency band outperforms the features in other frequency bands and leads to the best decoding performance. We also show that the EMD-based feature extraction can be useful for evoked potential estimation. Our proposed feature extraction approach may have potential for many applications involving nonstationary multivariable time series such as brain-computer interfaces (BCI).

1. INTRODUCTION

Spiking activity has been extensively studied in brain research to determine its relationship with perceptual reports or behavioral choices during ambiguous visual stimulation in the middle temporal (MT) visual area of macaque monkeys [1, 2]. However, spiking data as collected with standard neurophysiological techniques only provide information about the outputs of a small number of neurons within a given brain area. The local field potential (LFP) has recently received increasing attention in the analysis of the neuronal population activity [3, 4]. LFP is thought to largely arise from the dendritic activity of local populations of neurons and is dominated by the excitatory synaptic inputs to a cortical area as well as intra-areal local processing. The investigation of correlations between perceptual reports and LFP oscilla-tions during physically identical but perceptually different conditions may provide new insights on the mechanism of neural information processing and the neural basis of visual perception and perceptual decision making.

It is of great interest to study the temporally correlated LFP voltage fluctuations to understand the cognitive and perceptual processes. To reveal the temporal structure of LFP, the LFP spectrum at a certain time and frequency is often analyzed. Perhaps the most commonly used method for spectral analysis is the Fourier transform, which provides a general method for estimating the global power-frequency distribution of a given random process, assuming that it is stationary. As for neurobiological time series, however, Fourier analysis may be insufficient because the underlying processes are mostly nonstationary. The short-time Fourier transform (STFT) provides a means of joint time-frequency

analysis by applying moving windows to a signal and by Fourier-transforming it within each window [5]. The shortcoming of the short-time Fourier transform, however, is that a single fixed analysis window is used. Consequently, signals with different spectral components are treated with the same frequency resolution. In comparison with Fourier transform-based techniques, wavelet transform- (WT-) [6] based methods offer new capacity and advantages. For example, the basis functions in WT are not limited to sinusoidal waves, but rather are chosen by the user to address various problems of time-frequency resolution. In principle, WT uses short windows at high frequencies and long windows at low frequencies, which renders WT more suitable for dealing with nonstationary time series. Nonetheless, wavelet analysis is also limited by the fundamental uncertainty principle, in which both time and frequency cannot simultaneously be resolved with the same precision. Moreover, the results of WT analysis depend on the choice of the mother wavelet, which is arbitrary and may not be optimal for the time series under scrutiny. In contrast to the STFT and WT approaches, the empirical mode decomposition (EMD) method [7] adaptively decomposes nonstationary time series into narrow-band components, namely, intrinsic mode functions (IMFs), by empirically identifying the physical time scales intrinsic to the data without assuming any basis functions. These IMF components allow the calculation of a meaningful multicomponent instantaneous frequency by virtue of the Hilbert transform. Thus, one can potentially localize events in both time and frequency, even in nonstationary time series. Because of its versatility and generality, the EMD approach has found wide use in a variety of applications ranging from geophysics, biomedicine, neuroscience, financial engineering, and meteorology to seismology [8–12].

The availability of multichannel recordings offers opportunities to study how populations of neurons interact to produce a certain perceptual outcome. On one hand, different channels of LFP may reflect different aspects of brain activity correlated with the percept, and it is of interest to combine the LFP signals from various channels to exploit the different but complementary information embedded in the data simultaneously recorded from multiple channels. On the other hand, LFP signals from multiple channels may contain irrelevant and redundant information. By extracting the most discriminative features, we not only can reduce the data dimension but also can attain faster computation and more accurate classification. The common spatial patterns (CSPs) method has become a popular feature extraction approach in EEG-based brain-computer interface (BCI) applications [13–15]. CSP essentially finds spatial filters that maximize the variance for one class and simultaneously minimize the variance for the other class. A preprocessing step of CSP is the use of bandpass filtering on the EEG signals in a given frequency band. However, the most predictive frequency content for CSP may be subject specific and unknown a priori in some tasks. To deal with this problem, we first employ EMD on nonstationary single trials from each channel to extract the signal energy associated with various intrinsic time scales. Since different trials of LFP

may be decomposed by EMD into IMFs with different numbers, we use K-means clustering [16] to group the IMFs and residues into a number of clusters across all trials and channels. We then apply CSP to filter the clustered spatiotemporal signals and extract features from the CSP-filtered signals for each cluster. Based on the extracted features, we decode the monkey's perception for each trial using the support vector machine (SVM) classifier. In doing so, we have discovered that the cluster in the gamma frequency band carries the most discriminative information about the percept and hence it is the best feature for decoding perception under these circumstances. By taking advantage of both the temporal and spatial correlations of single-trial time series, we have proposed a general feature extraction approach that may have potential for a plethora of applications involving nonstationary multivariable time series.

In this paper, we examine the use of EMD to study neuronal activity in visual cortical area MT of a macaque monkey performing a bistable structure-from-motion (SFM) task [17]. The rest of the paper is organized as follows. In Section 2, we first present the materials and feature extraction approach consisting of empirical mode decomposition of single-trial LFP time series, K-means clustering, and common spatial patterns-based spatial filtering. We then briefly introduce the SVM classifier. In Section 3, we explore the application of the EMD-based feature extraction approach for decoding the bistable SFM perception and estimating evoked potentials. Finally, Section 4 contains the conclusions.

2. MATERIALS AND METHODS

2.1. Subjects and neurophysiological recordings

Electrophysiological recordings were performed in a healthy adult male rhesus monkey. After behavioral training was complete, an occipital recording chamber was implanted and a craniotomy was made. Intracortical recordings were conducted with a multielectrode array while the monkey was viewing SFM stimuli, which consisted of an orthographic projection of a transparent sphere that was covered with randomly distributed dots on its entire surface. Stimuli rotated for the entire period of presentation, giving the appearance of three-dimensional structure. The monkey was well trained and required to indicate the choice of rotation direction (clockwise or counterclockwise) by pushing one of two levers. Correct responses for disparity-defined stimuli were acknowledged with application of a fluid reward. In the case of fully ambiguous (bistable) stimuli, where the stimuli can be perceived in one of two possible ways and no correct response can be externally defined, the monkey was rewarded by chance. The bistable stimuli effectively dissociate percepts from the visual inputs and such uncoupling between stimuli and percepts serves as an entry point for research into the neural correlate of perception [18]. Only the trials of data corresponding to bistable stimuli are analyzed in the paper. The recording site was the MT area of the monkey's visual

cortex, which is commonly associated with visual motion processing.

2.2. Feature extraction

The feature extraction approach consists of three stages. First, we employ the EMD method to decompose nonstationary single-trial time series into narrow-band components called IMFs with time scales intrinsic to the data. Second, we adopt unsupervised K-means clustering to group the IMFs and residues into a number of clusters across all trials and channels. Third, we use the CSP approach to design spatial filters for the clustered spatiotemporal signals such that the resulting filtered signals carry the most discriminative information about the percept. Using the extracted features of MT LFP responses, the perceptual reports can be determined with high accuracy on a single-trial basis. In what follows, we provide more details for each stage of feature extraction.

2.2.1. Empirical mode decomposition

The central idea of the EMD time series analysis is to decompose a time series into a set of IMFs by a sifting process in order to empirically identify the physical time scales intrinsic to the time series [7]. A time series must satisfy two criteria to be an IMF: (1) the number of extrema and the number of zero crossings are either equal or differ at most by one; and (2) the mean of its upper and lower envelopes equals zero. The first criterion is similar to a narrow-band requirement. The second criterion is necessary to ensure that the instantaneous frequency will not have unwanted fluctuations as induced by asymmetric waveforms.

The sifting process for extracting IMFs from a given time-series $x(t)$ is described as follows. First, two smooth splines are constructed connecting all the maxima and minima of $x(t)$ to obtain its upper envelope, $x_u(t)$, and its lower envelope, $x_l(t)$. Once the extrema are identified, all the maxima are connected by a cubic spline line as the upper envelope. The procedure is repeated for the local minima to produce the lower envelope. Second, the mean of the two envelopes is subtracted from the data to obtain their difference $d(t) = x(t) - (x_u(t) + x_l(t))/2$. Third, the process is repeated for $d(t)$ until the resulting signal $c_1(t)$, the first IMF, satisfies the criteria of an intrinsic mode function. The residue $r_1(t) = x(t) - c_1(t)$ is then treated as a new time series subject to the sifting process as described above, yielding the second IMF from $r_1(t)$. The procedure continues until either the recovered IMF or the residual time series gets too small, or the residual time series has no turning points. Once all of the IMFs have been extracted, the final residual component represents the overall trend of the time-series. At the end of this process, the time series $x(t)$ can be expressed as follows:

$$x(t) = \sum_{j=1}^{n} c_j(t) + r_n(t), \tag{1}$$

where n is the number of IMFs, and $r_n(t)$ denotes the final residue. By the nature of the decomposition procedure, the

technique decomposes a time series into n fundamental components, each with a distinct time scale. More specifically, the first component has the smallest time scale which corresponds to the fastest time variation of data. As the decomposition process proceeds, the time scale increases, and hence, the mean frequency of the mode decreases. Since the decomposition is based on the local characteristic time scale of the time series to yield adaptive basis, it is applicable to nonlinear and nonstationary data analysis.

In practice, the resulting time-series after a certain number of iterations in the sifting process does not carry significant physical information and a pure frequency-modulated signal of constant amplitude can result from oversifting. To maintain the physical sense of the IMF components in terms of amplitude and frequency modulations, it is typical to stop the sifting process by limiting the standard deviation from two consecutive sifting results between 0.2 and 0.3 [7]. In addition, the cubic spline fitting in the sifting process may suffer from the end effect problem: large swings near the ends of the time series may occur and propagate inward. To overcome this problem, characteristic waves with two consecutive extrema for both the frequency and amplitude can be added at the ends [7].

2.2.2. K-means clustering

The goal of clustering is to partition a dataset into clusters such that the data within each cluster are similar but the data across distinct clusters are different. The K-means clustering algorithm is a popular algorithm for partitioning a dataset into K clusters with each cluster represented by its mean [16]. Initially, the K-means clustering algorithm generates K random points as cluster means. Then it iterates two steps, namely, the assignment step and update step until convergence. In the assignment step, each data point is assigned to the cluster so that the distance from the data point to the mean of the cluster is smaller than that from the data point to the means of other clusters. In the update step, the means of all clusters are recomputed and updated based on the data points assigned to them. The convergence criterion can be that the cluster assignment does not change. The K-means clustering algorithm is simple and fast but the clustering results depend on the initial random assignments. To overcome this problem we can take the best clustering from multiple random starts.

To determine the number of clusters, we use the silhouette value that measures how similar a data point is to points in its own cluster compared to points in other clusters [19]. It is defined as follows:

$$s(i) = \frac{\min_l b(i,l) - a(i)}{\max (a(i), \min_l b(i,l))}, \tag{2}$$

where $a(i)$ is the average distance from the ith data point to the other points in its cluster, and $b(i,l)$ is the average distance from the ith point to points in another cluster l. The silhouette value ranges from -1 to $+1$ with 1 meaning that data are separable and correctly clustered, 0 denoting poor clustering, and -1 meaning that the data are wrongly clustered.

2.2.3. Common spatial patterns

The CSP technique has become a popular feature extraction approach in EEG-based BCI applications [13–15]. The CSP algorithm essentially finds spatial filters that maximize the variance for one class and simultaneously minimize the variance for the other class.

Assume that we have M channels of neuronal signals and each trial has T time samples. Let $\mathbf{X}_j(i)$ denote the $M \times T$ data matrix for the ith trial from class j, where $j \in \{-1, +1\}$ denotes a perceptual report for the rotation direction with $j = -1$ for the clockwise direction and $j = +1$ for the counterclockwise direction. Let N_j denote the number of single trials for class j. Then the mean vector for each class is

$$\overline{\mathbf{X}}_j = \frac{1}{N_j} \sum_{i=1}^{N_j} \mathbf{X}_j(i). \tag{3}$$

Let

$$\mathbf{V}_j = \mathbf{X}_j(i) - \overline{\mathbf{X}}_j,$$

$$\mathbf{R}_j = \frac{1}{N_j} \sum_{i=1}^{N_j} \frac{\mathbf{V}_j \mathbf{V}_j'}{\text{tr}(\mathbf{V}_j \mathbf{V}_j')}, \tag{4}$$

where $\mathbf{R}_j$ is the average normalized covariance matrix for class j, $\text{tr}(\cdot)$ represents the sum of the diagonal elements, and $(\cdot)'$ means transpose.

Let the eigendecomposition of $\mathbf{R}_{-1} + \mathbf{R}_{+1}$ be

$$\mathbf{R}_{-1} + \mathbf{R}_{+1} = \mathbf{U}\Sigma\mathbf{U}', \tag{5}$$

where $\mathbf{U}$ is a unitary matrix and Σ is a diagonal matrix. Let

$$\mathbf{S}_{-1} = \mathbf{P}\mathbf{R}_{-1}\mathbf{P}', \qquad \mathbf{S}_{+1} = \mathbf{P}\mathbf{R}_{+1}\mathbf{P}', \tag{6}$$

where $\mathbf{P} = \Sigma^{-1/2}\mathbf{U}'$ is the whitening transformation matrix for $\mathbf{R}_{-1} + \mathbf{R}_{+1}$. Let the eigendecomposition of $\mathbf{S}_{-1}$ be

$$\mathbf{S}_{-1} = \mathbf{B}\Sigma_{-1}\mathbf{B}'. \tag{7}$$

Then it can be shown that

$$\mathbf{S}_{+1} = \mathbf{B}\Sigma_{+1}\mathbf{B}', \qquad \Sigma_{-1} + \Sigma_{+1} = \mathbf{I}, \tag{8}$$

where $\mathbf{I}$ is the identity matrix. Let

$$\mathbf{W} = \mathbf{B}'\mathbf{P}. \tag{9}$$

It can be shown that

$$\mathbf{R}_{-1} = \mathbf{W}'\Sigma_{-1}\mathbf{W}, \qquad \mathbf{R}_{+1} = \mathbf{W}'\Sigma_{+1}\mathbf{W}. \tag{10}$$

Hence $\mathbf{W}$ simultaneously diagonalizes $\mathbf{R}_{-1}$ and $\mathbf{R}_{+1}$, and each row of $\mathbf{W}$ is an eigenvector. In general, only the eigenvectors (rows of $\mathbf{W}$) corresponding to the largest and smallest few eigenvectors are used as spatial filters for CSP since they are most suitable for the discrimination purpose. Note that the spatial filters of CSP can also be determined by solving a generalized eigenvalue problem:

$$\mathbf{R}_{-1}\mathbf{v} = \lambda\mathbf{R}_{+1}\mathbf{v}. \tag{11}$$

2.3. Support vector machine classifier

SVM is a promising classifier that minimizes the empirical classification error and at the same time maximizes the margin by determining a separating hyperplane to distinguish different classes of data [20, 21]. Robust to outliers and having a good generalization capability, SVM has been successfully used in many applications.

Assume that $\mathbf{x}_k$, $k = 1, \ldots, K$ are the K training feature vectors for decoding and the class labels are $y_k \in \{-1, +1\}$. SVM solves the following optimization problem:

$$\min \|\mathbf{w}\|^2 + C\sum_{k=1}^{K} \xi_k \tag{12}$$
$$\text{subject to} \quad y_k(\mathbf{w}'\mathbf{x}_k + b) \geq 1 - \xi_k, \quad \xi_k \geq 0,$$

where $\mathbf{w}$ is the weight vector, $C > 0$ is the penalty parameter of the error term chosen by cross-validation, ξ_k is the slack variable, and b is the bias term. It turns out that the margin of the two classes is inversely proportional to $\|\mathbf{w}\|^2$. Therefore, the first term in the objective function of SVM is used to maximize the margin. The second term in the objective function is the regularization term that allows for training errors for the inseparable case.

The optimal solution of $\mathbf{w}$ and b to the above optimization problem can be found by using the Lagrange multiplier method. Assume that $\mathbf{t}$ is the testing feature vector. Then testing is done simply by determining on which side of the separating hyperplane $\mathbf{t}$ lies, that is, if $\mathbf{w}'\mathbf{t} + b \geq 0$, the label of $\mathbf{t}$ is classified as $+1$, otherwise, the label is classified as -1.

SVM belongs to a class of methods, known as kernel methods, which use a kernel function to map data into a high-dimensional feature space in order to increase the expressive power [20, 22, 23]. The kernel trick can also be applied to other algorithms such as Fisher's linear discriminant analysis (LDA), principal components analysis (PCA), and so on [24]. Nevertheless, the linear SVM has been widely used due to its simplicity, robustness, and interpretability, see for example, [25]. In general, linear classifiers are less prone to overfitting than the nonlinear ones for the limited training data. In addition, the relative importance of individual features can be assessed by examining the weights in the linear classifier.

3. EXPERIMENTAL RESULTS

In this section, we provide experimental examples to demonstrate the performance of the proposed feature extraction approach for predicting perceptual decisions from the neuronal data. Simultaneously collected nine-channel LFP data were used for demonstration. For K-means clustering, we use 50 random starts to find the best clustering and adopt the correlation between the power spectra of IMFs and residues from all trials and channels as the distance metric. For CSP, we use the pair of eigenvectors corresponding to the largest and smallest eigenvectors as spatial filters. In this paper, we employ the linear SVM classifier from the LIBSVM package [26]. We use decoding accuracy as a performance measure and calculate it via leave-one-out cross-validation (LOOCV).
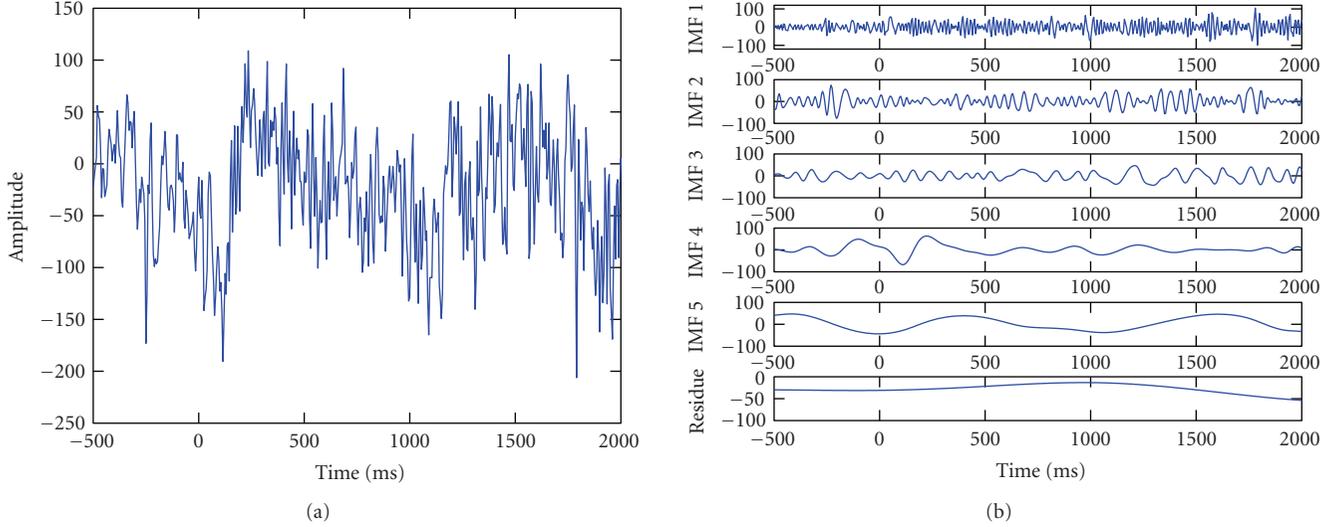
Figure 1: (a) A typical trial of LFP time series and (b) its five intrinsic mode functions (IMFs) and residue. Time 0 indicates the stimulus onset.

In particular, for a dataset with N trials, we choose $N-1$ trials for feature extraction and training and use the remaining 1 trial for testing. This is repeated for N times with each trial used for testing once. The decoding accuracy is obtained as the ratio of the number of correctly decoded trials over N.

Figures 1(a), 1(b) show a typical trial of LFP time series and its five IMFs and residue, respectively. Similar results were observed from other trials though there was variation from trial to trial in the number of IMF components produced by EMD. Figure 2 shows the mean and standard deviation of the average silhouette values obtained by clustering the IMFs and residues across all trials and channels using the K-means algorithm with different random starts as a function of the number of clusters. Note that the average silhouette values significantly increase with the number of clusters until the number of clusters is equal to five (comparing the average silhouette values obtained when the number of clusters is five with those obtained when the number of clusters is six via t-test, P-value > .05). Hence we choose the number of clusters to be five. Figures 3(a), 3(b) show the time series and power spectrum, respectively, of cluster 1–5 as the grand average across all trials and channels. It can be seen that the time series of different clusters differ in the smoothness and physical time scales. These clusters fall into five spectral bands: the first cluster in the gamma band (30–70 Hz), the second cluster in the alpha and beta bands (9–30 Hz), the third cluster in the theta band (4–8 Hz), the fourth cluster in the delta band (1–3 Hz), and the fifth cluster is mainly the DC component.

Next, we compare the SVM decoding accuracy based on different features and present the results in Tables 1, 2, focusing on a nonoverlapping moving window of 200 milliseconds in length. Note that for each feature, we report only the decoding accuracy of the moving window that yields the best decoding performance. In Table 1, we compare the decoding accuracy based on the features obtained from the
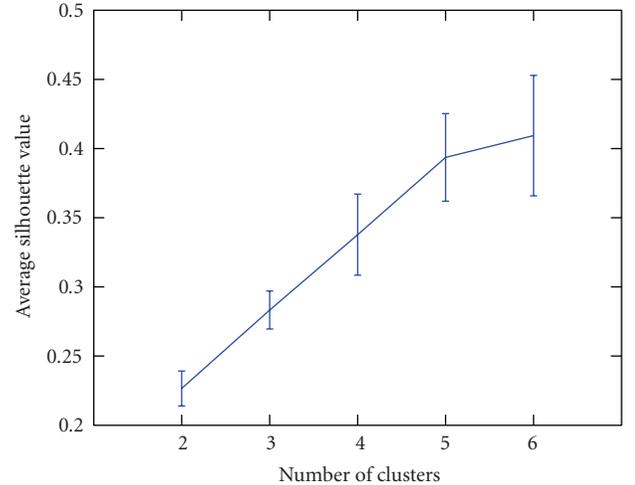
Figure 2: The mean and standard deviation of the average silhouette values obtained by clustering the IMFs and residues across all trials and channels using the K-means algorithm with different random starts as a function of the number of clusters. Note that the average silhouette values significantly increase with the number of clusters until the number of clusters is equal to five (comparing the average silhouette values obtained when the number of clusters is five with those obtained when the number of clusters is six via t-test, P-value > .05).

raw signal and clusters 1–5 (denoted as c1–c5, resp.) and their CSP counterparts. It is clear from Table 1 that the feature obtained by using CSP on cluster 1 yields the best decoding accuracy of 0.76. Hence it is the most discriminative feature of the LFP data of all the above features. In Table 2, we compare the decoding accuracy based on the features obtained from the bandpass filtered signal and cluster 1 (denoted as c1) and their CSP counterparts. Here the bandpass-filtered
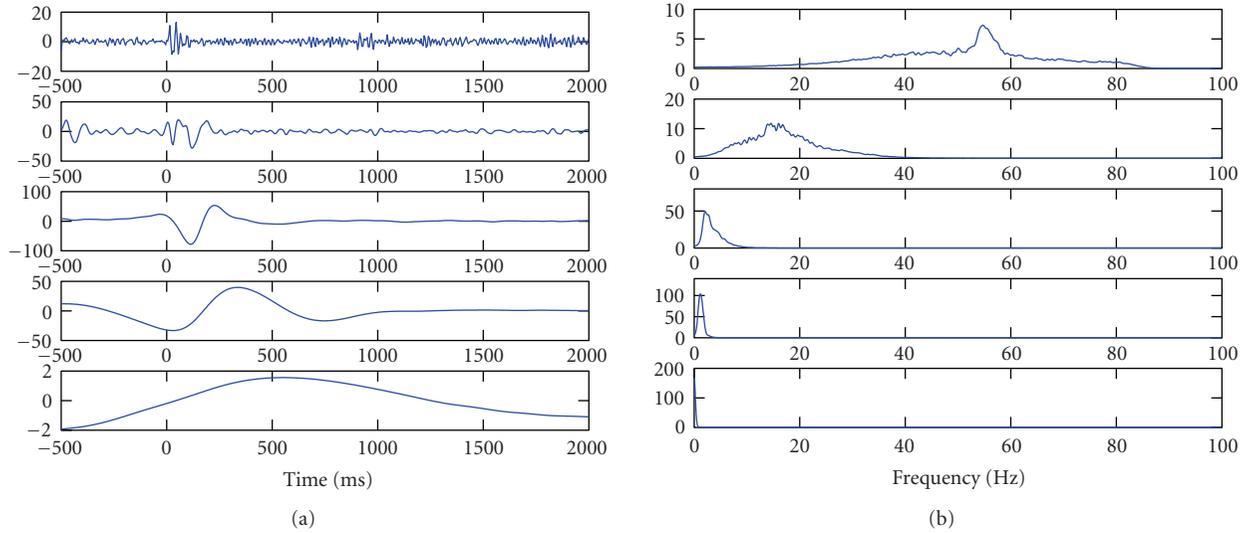
FIGURE 3: The grand average across all trials and channels of (a) time series and (b) power spectrum corresponding for clusters 1–5. The subfigures from top to bottom correspond to clusters 1–5, respectively. Note that the clusters fall into distinct spectral bands: the first cluster in the gamma band (30–70 Hz), the second cluster in the alpha and beta bands (9–30 Hz), the third cluster in the theta band (4–8 Hz), the fourth cluster in the delta band (1–3 Hz), and the fifth cluster is mainly the DC component.
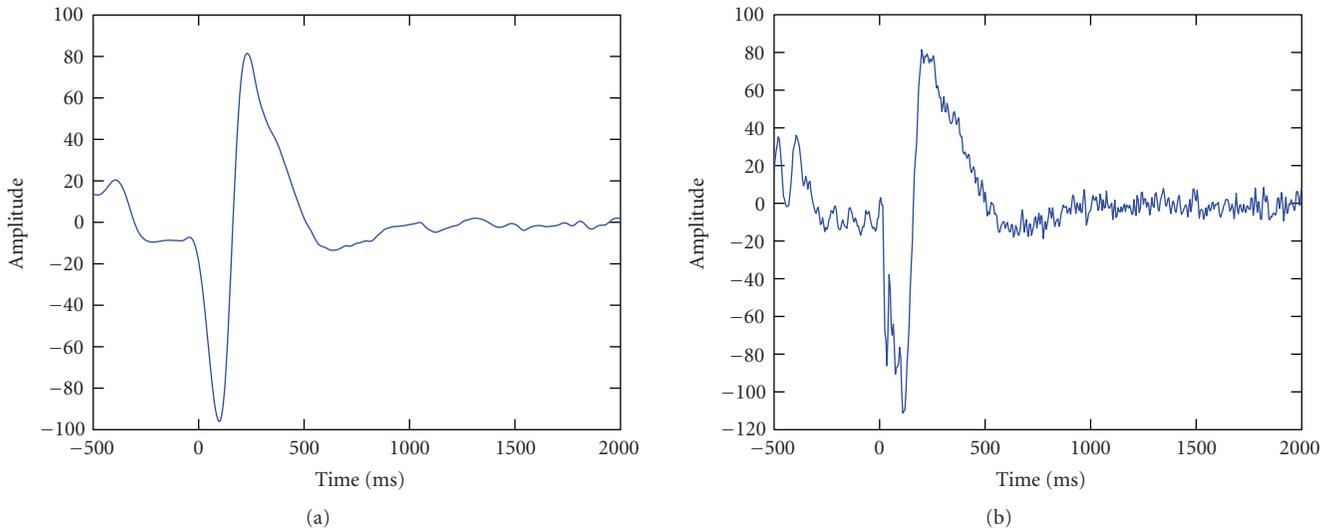
FIGURE 4: The average evoked potential from a single channel obtained as (a) the combination of clusters 3–5 and (b) ensemble averaging. Note that there is a close match between the average evoked potential obtained via these two approaches.

TABLE 1: Comparison of the decoding accuracy based on the features obtained from the raw signal and clusters 1–5 (denoted as c1–c5, resp.) and their CSP counterparts.

Approach	Raw signal	c1	c2	c3	c4	c5
Decoding accuracy	0.70	0.71	0.70	0.61	0.61	0.61
Approach	Raw signal + CSP	c1 + CSP	c2 + CSP	c3 + CSP	c4 + CSP	c5 + CSP
Decoding accuracy	0.69	**0.76**	0.66	0.66	0.66	0.64

TABLE 2: Comparison of the decoding accuracy based on the features obtained from the bandpass filtered signal and cluster 1 (denoted as c1) and their CSP counterparts.

Approach	Bandpass filtered signal	Bandpass filtered signal + CSP	c1	c1 + CSP
Decoding accuracy	0.65	0.71	0.71	**0.76**

signal represents the signal obtained by directly bandpass-filtering the raw single-trial signal in the gamma band (30–70 Hz). We can observe that although the first cluster also falls into the gamma band, its decoding accuracy significantly outperforms that based on the bandpass-filtered signal in the same band. The different performance underscores the fundamental difference between bandpass filtering and EMD in that EMD adaptively identifies the physical time scales intrinsic to the nonstationary LFP time series.

Interestingly, the EMD-based feature extraction is also useful for evoked potential estimation. From Figure 3, we see that clusters 3–5 correspond to the low-frequency components and hence contribute to the average evoked potential. Figure 4(a) shows the average evoked potential from a single channel as the combination of clusters 3–5. Note that the stimulus appears at $t = 0$ second and remains until two seconds. As can be seen, the average evoked potential captures the event-related neural activity. Furthermore, the resulting average evoked potential is quite smooth even when the number of trials is not very large. Figure 4(b) shows the average evoked potential from a single channel obtained by the commonly used ensemble averaging, which averages across all trials the potentials measured over repeated presentations of a stimulus. It is actually the same as the combination of clusters 1–5. Obviously, the average evoked potential obtained by ensemble averaging is poor and probably contains background activity and noise, which is due to the fact that ensemble averaging generally requires a large number of trials to achieve satisfactory performance.

4. CONCLUSIONS

In this paper, we have shown how to apply an EMD-based method to extract features from the LFP in monkey visual cortex for decoding its bistable SFM perception. We have employed EMD to decompose nonstationary single-trial time series into narrow-band components called IMFs with time scales dependent on the data. We have adopted unsupervised K-means clustering to group the IMFs and residues into a number of clusters across all trials and channels. We have used the supervised CSPs approach to design spatial filters for the clustered spatiotemporal signals such that the resulting filtered signals carry the most discriminative information about the percept. We have exploited the SVM classifier on the extracted features of the LFP response to decode the reported perception on a single-trial basis. We have applied the feature extraction approach to the multichannel intracortical LFP data collected from the MT visual area in a macaque monkey performing an SFM task. Using these techniques, we have demonstrated that the CSP feature of the cluster in the gamma frequency band outperforms the features in other bands and leads to the best decoding performance. We have also shown that the EMD-based feature extraction can be useful for evoked potential estimation.

The advantages of our proposed feature extraction approach can be summarized as follows. First of all, the approach is data-driven and can identify the physical time scales intrinsic to the data. Hence it does not demand prior knowledge of the discriminative frequency bands and is robust against subject variability. In addition, it does not require stationarity of the time series or any functional basis. Because of these advantages, the EMD-based feature extraction approach may have potential for a large variety of applications involving nonstationary multivariable time series such as brain-computer interfaces (BCI).

REFERENCES

[1] K. H. Britten, W. T. Newsome, M. N. Shadlen, S. Celebrini, and J. A. Movshon, "A relationship between behavioral choice and the visual responses of neurons in macaque MT," *Visual Neuroscience*, vol. 13, no. 1, pp. 87–100, 1996.

[2] J. V. Dodd, K. Krug, B. G. Cumming, and A. J. Parker, "Perceptually bistable three-dimensional figures evoke high choice probabilities in cortical area MT," *Journal of Neuroscience*, vol. 21, no. 13, pp. 4809–4821, 2001.

[3] B. Pesaran, J. S. Pezaris, M. Sahani, P. P. Mitra, and R. A. Andersen, "Temporal structure in neuronal activity during working memory in macaque parietal cortex," *Nature Neuroscience*, vol. 5, no. 8, pp. 805–811, 2002.

[4] G. Kreiman, C. P. Hung, A. Kraskov, R. Q. Quiroga, T. Poggio, and J. J. DiCarlo, "Object selectivity of local field potentials and spikes in the macaque inferior temporal cortex," *Neuron*, vol. 49, no. 3, pp. 433–445, 2006.

[5] L. Cohen, *Time Frequency Analysis*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1995.

[6] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, Calif, USA, 1999.

[7] N. E. Huang, Z. Shen, S. R. Long, et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time-series analysis," *Proceedings of the Royal Society of London A*, vol. 454, no. 1971, pp. 903–995, 1998.

[8] W. Huang, Z. Shen, N. E. Huang, and Y. C. Fung, "Engineering analysis of biological variables: an example of blood pressure over 1 day," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 9, pp. 4816–4821, 1998.

[9] H. Liang, S. L. Bressler, E. A. Buffalo, R. Desimone, and P. Fries, "Empirical mode decomposition of field potentials from macaque V4 in visual spatial attention," *Biological Cybernetics*, vol. 92, no. 6, pp. 380–392, 2005.

[10] H. Liang, Q.-H. Lin, and J. D. Z. Chen, "Application of the empirical mode decomposition to the analysis of esophageal manometric data in gastroesophageal reflux disease," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 10, pp. 1692–1701, 2005.

[11] N. E. Huang and N. O. Attoh-Okine, Eds., *The Hilbert-Huang Transform in Engineering*, CRC, Boca Raton, Fla, USA, 2005.

[12] C.M. Sweeney-Reed and S. J. Nasuto, "A novel approach to the detection of synchronisation in EEG based on empirical mode decomposition," *Journal of Computational Neuroscience*, vol. 23, no. 1, pp. 79–111, 2007.

[13] Z. J. Koles, "The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG,"

Electroencephalography and Clinical Neurophysiology, vol. 79, no. 6, pp. 440–447, 1991.

[14] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clinical Neurophysiology*, vol. 110, no. 5, pp. 787–798, 1999.

[15] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, pp. 441–446, 2000.

[16] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 281–297, Berkeley, Calif, USA, June-July 1967.

[17] Z. Wang, A. Maier, D. A. Leopold, N. K. Logothetis, and H. Liang, "Single-trial evoked potential estimation using wavelets," *Computers in Biology and Medicine*, vol. 37, no. 4, pp. 463–473, 2007.

[18] R. Blake and N. K. Logothetis, "Visual competition," *Nature Reviews Neuroscience*, vol. 3, no. 1, pp. 13–21, 2002.

[19] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, John Wiley & Sons, New York, NY, USA, 1990.

[20] V. N. Vapnik, *The Nature of Statisical Learning Theory*, Springer, New York, NY, USA, 1995.

[21] C. Cortes and V. N. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[22] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the 5th Annual Workshop on Computational Learning Theory*, pp. 144–152, Pittsburgh, Pa, USA, July 1992.

[23] B. Schölkopf, C. J. C. Burges, and A. J. Smola, Eds., *Advances in Kernel Methods: Support Vector Learning*, MIT Press, Cambridge, Mass, USA, 1999.

[24] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, Cambridge, UK, 2004.

[25] C. P. Hung, G. Kreiman, T. Poggio, and J. J. DiCarlo, "Fast readout of object identity from macaque inferior temporal cortex," *Science*, vol. 310, no. 5749, pp. 863–866, 2005.

[26] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," 2001, http://www.csie.ntu.edu.tw/~cjlin/libsvm.