Classification of Motor Imagery BCI Using Multivariate Empirical Mode Decomposition

Cheolsoo Park, David Looney, Naveed ur Rehman, Alireza Ahrabian, and Danilo P. Mandic

***Abstract*—Brain electrical activity recorded via electroencephalogram (EEG) is the most convenient means for brain–computer interface (BCI), and is notoriously noisy. The information of interest is located in well defined frequency bands, and a number of standard frequency estimation algorithms have been used for feature extraction. To deal with data nonstationarity, low signal-to-noise ratio, and closely spaced frequency bands of interest, we investigate the effectiveness of recently introduced multivariate extensions of empirical mode decomposition (MEMD) in motor imagery BCI. We show that direct multichannel processing via MEMD allows for enhanced localization of the frequency information in EEG, and, in particular, its noise-assisted mode of operation (NA-MEMD) provides a highly localized time-frequency representation. Comparative analysis with other state of the art methods on both synthetic benchmark examples and a well established BCI motor imagery dataset support the analysis.**

***Index Terms*—Brain–computer interface (BCI), electroencephalogram (EEG), empirical mode decomposition, motor imagery paradigm, noise assisted multivariate extensions of empirical mode decomposition (NA-MEMD).**

I. Introduction

BRAIN–COMPUTER interface (BCI) is an emerging technology dealing with computer-aided control using exclusively brain activity, and has found application across bioengineering fields and in neuroprosthetics. Most common BCI are based on the electroencephalogram (EEG) owing to its noninvasive nature and affordable recording equipment which facilitates real-time operation [1]. In particular, motor imagery BCI, that is, the imagination of a motor action without any actual movement of limbs, has clear practical significance [2]. However, motor imagery BCI is notoriously difficult to analyze, requiring long training times and exhibiting limited BCI channel capacity.

The neurophysiological basis for motor imagery BCI are the so-called mu (8–12 Hz) and $beta$ rhythms (18–25 Hz) in EEG [3], which have been observed in the central region of the brain when subjects plan and execute hand or finger movements [4], [5]. Results by Nikouline *et al.* [6] demonstrated that somatosensory stimuli suppressed mu rhythms at both the contralateral and the ipsilateral somatosensory cortex (SI), while Pfurtscheller *et al.* [4] described changes of EEG activity in the low-frequency bands, including mu and $beta$ rhythms, caused by voluntary movements. It has been suggested by Yuan *et al.* [5] that such changes to the mu and $beta$ rhythms are due to the reflection of phase coherence in thalamocortical circuits.[1] The blocking (ERD) of the mu rhythm over the contralateral scalp, and the enhancement (ERS) over the ipsilateral area during motor imagery were also demonstrated in [5].

The so observed changes to mu and $beta$ rhythms are utilized by several existing BCI systems, however, studies so far have employed mostly standard signal processing techniques, mainly those based on Fourier analysis [5], [6], [9]. These methods are based on a projection onto a predefined set of basis functions and thus inherit the well-known problem associated with standard spectrum estimations: poor time-frequency localization [10]. In addition, fixed linear orthogonal basis functions used in standard spectrum estimation are not suitable for processing real-world data such as EEG, which are almost invariably nonlinear and nonstationary [11]. It is well established that complex biological systems, like the brain, do not produce tones with fixed frequencies, and thus the rhythms in brain electrical responses will drift within different EEG bands, making the estimation based on sine-cosine bases inadequate. Finally, the analysis of multivariate responses (multichannel EEG) on a channel-by-channel basis prevents consideration of a key feature of multivariate data—the cross-channel interdependence.

To account for the nonstationarity, multichannel natures and the inherent drifts in the frequency estimation of real-world signals, and at the same time to bypass the problems associated with techniques which employ fixed basis functions, in this paper we set out to analyze motor imagery responses using recent multivariate extensions of a fully data-driven time-frequency analysis technique, the empirical mode decomposition (EMD) [10]. Properties of the EMD-based decomposition ensure that frequency and amplitude information can be analyzed locally, enabling robust analysis of signal dynamics across time and frequency. There are no prior assumptions on the data and, as such, EMD is suitable for the analysis of nonlinear and nonstationary phenomena, such as intracortical signals and EEG [12]–[15].

We here make use of the physical insight into the EEG recording and propagation, and combine the locally orthogonal and narrowband IMF bases with a tool for discriminating between different classes of EEG activities based on their spatial configuration across electrodes, the common spatial patterns (CSP) algorithm. It has already been shown that the combina-

Manuscript received May 10, 2012; revised September 13, 2012; accepted October 07, 2012. Date of publication November 27, 2012; date of current version January 04, 2013.

C. Park is with the Department of Bioengineering, University of California-San Diego, La Jolla, CA 92093 USA (e-mail: charles586@gmail.com).

D. Looney, A. Ahrabian, and D. Mandic are with the Electrical and Electronic Engineering Department, Imperial College London, SW7 2BT London, U.K. (e-mail: david.looney06@imperial.ac.uk; alireza.ahrabian06@imperial.ac.uk; d.mandic@imperial.ac.uk).

N. ur Rehman is with COMSATS Institute of Information Technology, Islamabad, Pakistan (e-mail: nrehman.comsats@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNSRE.2012.2229296

[1] A spectral power decrease is called event-related desynchronization (ERD) [7] while an increase is termed event-related synchronization (ERS) [8].

tion of the EMD and CSP algorithms is a potentially powerful and unified approach to feature extraction across space, time, and frequency for nonlinear and nonstationary data, however, this was achieved without explicitly taking into account a key feature of EEG—its multichannel nature. For instance, Wang *et al.* applied the CSP algorithm to an IMF obtained using standard EMD for the classification of bistable perception [12]. Despite the potential of EMD-CSP, there are several obstacles which limit its usefulness when processing multichannel data, most significantly the problem of uniqueness associated with single channel EMD. The problem of uniqueness refers to the phenomenon whereby IMFs for each data channel typically do not correspond in number and/or frequency [16]. In this way, it is difficult to establish a consistent pattern of multi-channel IMF relevance without sacrificing accuracy, reducing the performance of the CSP analysis. In our previous work [11], [16], [17] we addressed direct bivariate (2-channel) data analysis and were able to show that complex extensions of EMD can be used to circumvent the problem of uniqueness and yield a more accurate estimation of the IMFs when the narrowband oscillations of interest are common to each channel, giving enhanced spectrum estimates. We here extend the analysis to multivariate extensions of EMD and propose an MEMD-based CSP approach to motor imagery classification, fully benefiting from its enhanced localization properties, the use of cross-channel information, ability to identify signal-bearing components across the data channels, and increased robustness to noise and recording artifacts.

II. Empirical Mode Decomposition Algorithm

Empirical mode decomposition is a fully data-driven method for obtaining highly localized time-frequency estimation for nonlinear and nonstationary signals [10], whereby the signal in hand is decomposed into a finite set of AM/FM components (IMFs). The two conditions required for a signal to be considered an IMF are: 1) the number of extrema and the number of zero crossings differ at most by one, 2) the mean of the envelopes connecting respectively the local maxima and local minima is approximately zero. Every IMF can thus be regarded as a narrow-band signal, reflecting a different temporal scale intrinsic to the data—a key property giving EMD an advantage over Fourier techniques [10]. The EMD operation is outlined in Algorithm 1.

Algorithm 1 The Standard EMD Algorithm

1) Let $\tilde{v}(t) = v(t)$ ($v(t)$ is original signal)
2) Identify all local maxima and minima of $\tilde{v}(t)$
3) Find a lower "envelope," $e_l(t)$ that interpolates all local minima
4) Find an upper "envelope," $e_u(t)$ that interpolates all local maxima
5) Calculate the local mean, $\bar{m}(t) = (e_l(t) + e_u(t))/2$
6) Subtract the local mean from $\tilde{v}(t)$, $c_i(t) = \tilde{v}(t) - \bar{m}(t)$ (i is an order of IMF)
7) Let $\tilde{v}(t) = c_i(t)$ and go to step 2); repeat until $c_i(t)$ becomes an IMF

The first IMF is subtracted from the original data, $r(t) = v(t) - c_1(t)$, and the procedure is applied iteratively to the residue, $r(t)$, until it becomes constant or contains no more oscillations; this so called sifting process is controlled by a suitably defined stopping criterion [18]. The signal $v(t)$ is then

$$v(t) = \sum_{i=1}^{M} c_i(t) + r(t) \tag{1}$$

where $c_i(t), i = 1, \ldots, M$, are the IMFs and $r(t)$ the remaining residue. The narrowband nature of the IMFs satisfies the conditions under which the Hilbert transform

$$H(c_i(t)) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{c_i(t')}{t - t'} dt' \tag{2}$$

can be applied to obtain a localized time-frequency spectrogram, where symbol P indicates the Cauchy principal value, and the analytic signal is then obtained as

$$V(t) = \sum_{i=1}^{M} (c_i(t) + jH(c_i(t))) = \sum_{i=1}^{M} a_i(t) e^{j\theta_i(t)} \tag{3}$$

and is described by its amplitude and phase functions, $a_i(t)$ and $\theta_i(t)$. The phase function, $\theta_i(t)$, is differentiated to produce the instantaneous frequency, $w_i(t) = d\theta_i(t)/dt$[19]. A plot of the amplitude $a_i(t)$ versus time t and instantaneous frequency $w_i(t)$, that is, amplitude contours on the time-frequency plane, is called the Hilbert–Huang spectrogram (HHS), $\mathbf{H}(w, t)$, and represents a time-frequency spectrogram of a nonlinear and nonstationary signal.

A. Obstacles to Multichannel Data Analysis by EMD

Owing to the random natures of EEG signals and noise, the uniqueness problem refers to the fact that the IMFs obtained for different EEG channels can be different in number and properties (frequency), heavily compromising any analysis or fusion of multicomponent signals obtained in a channel-by-channel basis. This is reflected by the different decompositions obtained for signals with similar statistics, and the phenomenon of mode-mixing, whereby similar frequencies appear across different IMFs. To address this problem, Wu *et al.* proposed a noise-assisted data analysis method, the ensemble EMD (EEMD), which defines the IMF components as the mean of an ensemble of IMFs obtained by applying standard EMD on the signal corrupted by added white noise of finite amplitude [20], [21]. However, despite being a significant step forward, EEMD does not fully resolve the uniqueness problem and is further limited by its computational and its univariate nature.

In our previous work [11], [16], [17], we used complex extensions of EMD to solve the uniqueness problem for problems pertaining to 2-channel data sources.[2] This was achieved by applying bivariate EMD to decompose the different channels "simultaneously" and then separating the real and imaginary parts of the decompositions, giving two sets of IMFs with the following desired properties.

- The bivariate IMFs are matched in number and frequency; even if mode-mixing is present, it occurs simultaneously in both the real and imaginary components and thus an IMF-by-IMF comparison make sense [16], [17].

[2]There are three different ways to extend the real-valued EMD to the bivariate or complex domain, "rotation invariant empirical mode decomposition (RIEMD)" [22], "complex empirical mode decomposition (CEMD)" [23] and "bivariate empirical mode decomposition (BEMD)" [24].

- Any shared activity, e.g., common oscillations of a given frequency, between the data channels are identified though bivariate IMFs that have the same oscillatory properties at every level and enhance robustness to noise.

However, the bivariate EMD can cater only for 2-channel data, or a multidimensional data where the channels are analyzed pair-wise. We here extend our bivariate analysis to multichannel data sets using the recently developed multivariate extensions of EMD algorithm (MEMD) [25] and show that it enables a matched-scale decomposition across multichannel data (two or more), thus allowing for multi-channel pattern estimation at the intrinsic scales of the signal—the IMF level. There are many advantages in simultaneously analyzing the intrinsic modes from multichannel data, especially for EEG where the background noise is broadband and the useful information narrowband, both exhibiting various degrees of nonstationarity and spatial and temporal dependence.

III. Multivariate Empirical Mode Decomposition

The multivariate EMD, recently introduced by Rehman and Mandic [25], is a natural and generic extension of the standard EMD and BEMD. Standard EMD computes the local mean using the average of upper and lower envelopes. However, the local mean of n-dimensional signals cannot be defined directly,[3] and thus the multiple n-dimensional envelopes are generated by projecting the signal along different directions in n-variate spaces, those projections are then averaged to obtain the local mean. For a uniform set of direction vectors used to project the input multivariate signal, low discrepancy Hammersley sequences are used to obtain quasi-uniform points on high dimensional spheres [26]. The details of MEMD are outlined[4] in Algorithm 2 [25].

Algorithm 2 The Multivariate EMD Algorithm (MEMD)

1) Choose a suitable point set for sampling on an $(n-1)$ sphere.
2) Calculate a projection, denoted by $\{p^{\theta_k}(t)\}_{t=1}^{T}$, of the input signal $\{\mathbf{v}(t)\}_{t=1}^{T}$ along the direction vector $\mathbf{x}^{\theta_k}$, for all k (the whole set of direction vectors), giving $\{p^{\theta_k}(t)\}_{k=1}^{K}$ as the set of projections.
3) Find the time instants $t_j^{\theta_k}$ corresponding to the maxima of the set of projected signals $\{p^{\theta_k}(t)\}_{k=1}^{K}$.
4) Interpolate $[t_j^{\theta_k}, \mathbf{v}(t_j^{\theta_k})]$ to obtain multivariate envelope curves $\{\mathbf{e}^{\theta_k}(t)\}_{k=1}^{K}$
5) For a set of K direction vectors, the mean $\mathbf{m}(t)$ of the envelope curves is calculated as $\mathbf{m}(t) = 1/K \sum_{k=1}^{K} \mathbf{e}^{\theta_k}(t)$.
6) Extract the "detail" $\mathbf{c}_i(t)$ using $\mathbf{c}_i(t) = \mathbf{v}(t) - \mathbf{m}(t)$ (i is an order of IMF). If the "detail" $\mathbf{c}_i(t)$ fulfills the stoppage criterion for a multivariate IMF, apply the above procedure to $\mathbf{v}(t) - \mathbf{c}_i(t)$, otherwise apply it to $\mathbf{c}_i(t)$.

[3]For example, it is impossible to find maxima of the complex signals as the ordering operators are not defined.

[4]The MATLAB code of MEMD is available from http://www.commsp.ee.ic.ac.uk/~mandic/research/emd.htm.

The sifting process for a multivariate IMF can be stopped when all K projected signals fulfill any stoppage criterion adopted in standard EMD; we employed a combination of EMD stoppage criteria, given in [18] and [27], for MEMD sifting. The MEMD algorithm acts as a dyadic filter bank on each channel (variate) when applied to multidimensional white Gaussian noise (WGN), exhibiting greatly enhanced alignment of the corresponding IMFs from different channels across the same frequency range compared to EMD. Using this property of MEMD, Rehman and Mandic [28] introduced a noise-assisted MEMD (NA-MEMD) to further alleviate the mode mixing problem. This was achieved by increasing the dimensionality, including a subspace containing multivariate independent white noise, and processing the resulting composite signal using MEMD. Notice that in this way the noise is never mixed with the useful data channels, as it resides in a different subspace, and is used to enforce a filterbank structure, and thus alleviate the problem of mode mixing and provide much better definition of frequency bands inherent to the data. A set of IMFs corresponding to only the original input signal is kept by discarding the IMF subspace associated with the noise. Owing to the noise subspace, the alignment of IMFs obeys the dyadic filter bank structure and also ensures that IMFs associated with the original input signal are aligned, having the same information at the same level of decomposition, and hence providing an intuitive and rigorous tool for the analysis of narrowband but nonstationary rhythms from biomedical data. The details of the NA-MEMD method are described in Algorithm 3.

Algorithm 3 Noise-Assisted MEMD (NA-MEMD)

1) Create an uncorrelated white Gaussian noise time-series (q-channel) of the same length as that of the input.
2) Add the noise channels (q-channel) created in Step 1 to the input multivariate (n-channel) signal, obtaining an $(n+q)$-channel signal.
3) Process the resulting $(n+q)$-channel multivariate signal using the MEMD algorithm listed in Algorithm 2, to obtain multivariate IMFs.
4) From the resulting $(n+q)$-variate IMFs, discard the q channels corresponding to the noise, giving a set of n-channel IMFs corresponding to the original signal.

A. Common Oscillatory Modes of Multivariate IMFs

We shall illustrate the operation of MEMD in multichannel data decomposition, where common oscillatory modes are present [25], and a 3-channel synthetic signal $[x(t),\ y(t),\ z(t)]$, shown in the top row of Fig. 1, where

$$x(t) = \begin{cases} \sin(2\pi f_1 t) + q_1(t), & t = 1, \ldots, 2048 \\ \sin(2\pi f_2 t) + q_2(t), & t = 2049, \ldots, 4096 \end{cases}$$

$$y(t) = \sin(2\pi f_1 t) + \sin(2\pi f_2 t) + q_3(t), \quad t = 1, \ldots, 4096$$

$$z(t) = \sin(2\pi f_2 t) + \sin(2\pi f_3 t) + q_4(t), \quad t = 1, \ldots, 4096$$

$$f_s = 2048, \quad f_1 = \frac{5}{f_s}, \quad f_2 = \frac{11}{f_s}, \quad f_3 = \frac{23}{f_s}$$

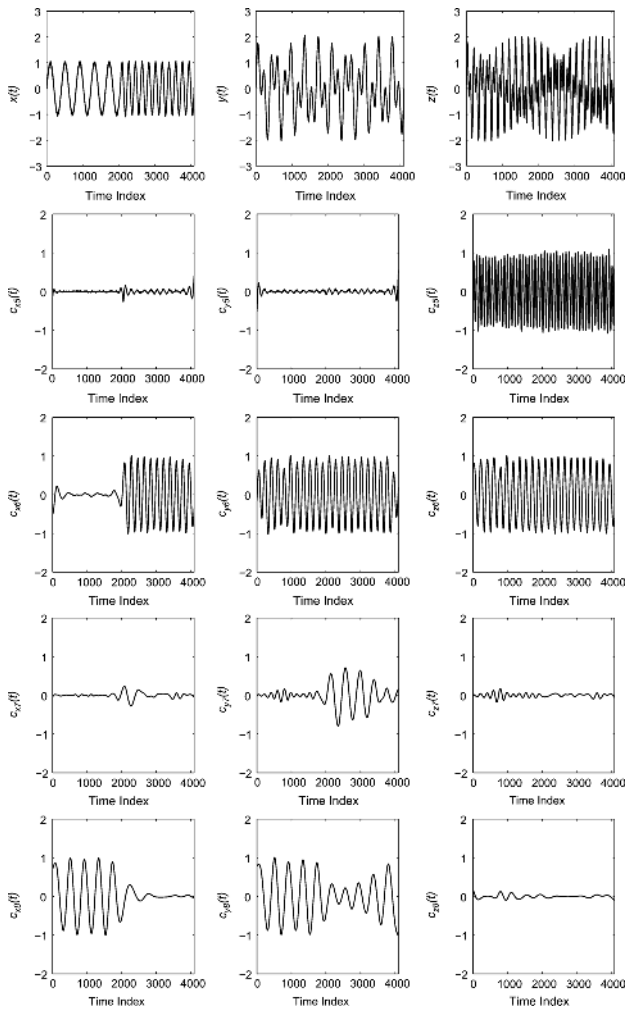


Fig. 1. MEMD-based decomposition of a synthetic multivariate signal $[x(t), y(t), z(t)]$, which contains multiple frequency components. The common oscillatory modes are aligned at the same IMF level, however, the problem of mode mixing is present in the seventh and eighth IMFs, $c_7(t)$ and $c_8(t)$.

and $q_1(t)$, $q_2(t)$, $q_3(t)$ and $q_4(t)$ denote realizations of 30 dB WGN. All the data channels contained continuous sinusoid components, and MEMD was applied to this multivariate synthetic data and the resulting three-variate IMFs are shown in Fig. 1. Observe the common oscillatory modes are aligned at the same IMF level, where the 11 Hz sinusoid, common to the three channels, is present at the sixth IMF level. However, the mode mixing problem is evident in the seventh and eighth IMF as shown in Fig. 1. When the NA-MEMD method was applied to the same data with one additional noise channel, $n(t)$, unlike MEMD, IMFs obtained using NA-MEMD in Fig. 2 fully alleviated the problem of mode mixing and the two different frequency components of $x(t)$ were extracted correctly. This illustrates the ability of MEMD to identify and align information at similar scales in different data channels [29], and that NA-MEMD can furthermore help alleviate the mode mixing problem, indicating that an MEMD-based approach enables a robust scale-by-scale comparison between the data channels—a key requirement for a posterior analysis at the IMF level.

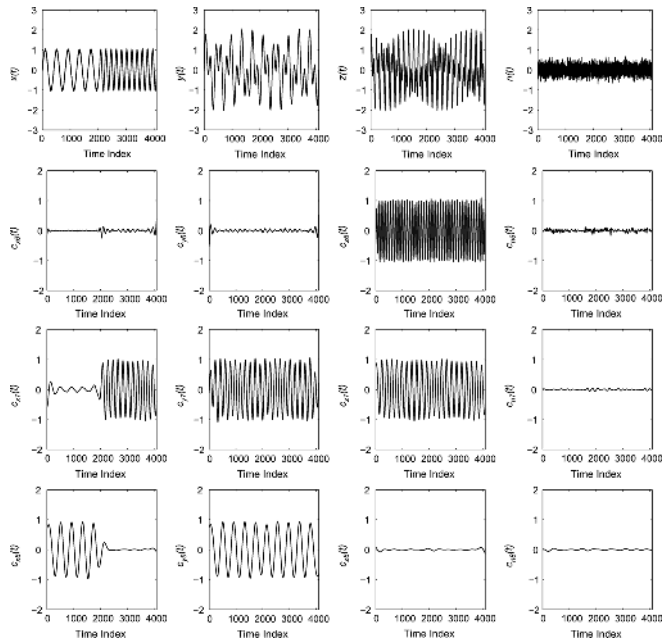


Fig. 2. Decomposition of a synthetic multivariate signal $[x(t), y(t), z(t), n(t)]$ using the noise assisted MEMD (NA-MEMD), where $n(t)$ is an additional white noise channel. Note that each IMF contains a single narrow-band component with no mode mixing present.

B. Component Estimation Using Multivariate IMFs

In this section, we examine the ability of MEMD to achieve a more robust and localized estimate of components at the IMF level compared to the single channel EMD algorithm. Particular emphasis, following our previous work in [11], [16], is on illuminating that MEMD can extract IMFs that have physical meaning and are better aligned with the components of interest, thus enhancing robustness to noise.

We decomposed the signal $x = s_f + q$ using EMD and EEMD, where s_f was a sinusoid of frequency f and q a realization of WGN, and subsequently applied the Wiener filter[5] to the IMFs to obtain estimates of the sinusoid, $\hat{s}_{\mathrm{EMD}}$ and $\hat{s}_{\mathrm{EEMD}}$, denoting the estimates using EMD and EEMD respectively. In the EEMD decomposition,[6] noise with a ratio of 0.4 between the standard deviation of the noise and data was added and the number of ensembles for each case was 100. The more accurate the estimate of s_f, the more accurately the IMFs represent the original oscillating components of the input. Additionally, MEMD was performed on a multi-channel sinusoid data set, where all channels contained the same sinusoid, s_f, corrupted with different realizations of WGN, that is

$$x_1 = s_f + q_1, x_2 = s_f + q_2, \ldots, x_n = s_f + q_n$$

where q_n is a realization of WGN in the nth channel. The performance of MEMD was examined by increasing the number of channels from 4 to 12. This analysis was performed for several frequencies (5, 11, 23 Hz), and over three signal-to-noise

[5]For more details, we refer to [30].

[6]For MATLAB code, see http://rcada.ncu.edu.tw/research1_clip_program.htm.

TABLE I
SINUSOID RECONSTRUCTION FOR DIFFERENT FREQUENCIES AND INITIAL NOISE LEVELS ($\mathrm{SNR}_{\mathrm{INIT}}$). $\mathrm{MEMD}_{(n)}$ DENOTES AN N-CHANNEL MEMD OPERATION. NOTE THE IMPROVED PERFORMANCE WHEN INCREASING THE NUMBER OF CHANNELS

Frequency	Algorithm / $\mathrm{SNR}_{\mathrm{INIT}}$	EMD	EEMD	$\mathrm{MEMD}_{(4)}$	$\mathrm{MEMD}_{(6)}$	$\mathrm{MEMD}_{(8)}$	$\mathrm{MEMD}_{(10)}$	$\mathrm{MEMD}_{(12)}$
5 Hz	5dB	11.6	13.7	15.2	15.9	16.2	16.3	16.1
	0dB	7.9	9.9	11.4	12.0	12.2	12.3	12.2
	-5dB	4.8	7.4	7.8	8.3	8.2	8.3	8.6
11 Hz	5dB	9.1	10.9	11.5	11.9	11.8	12.0	12.1
	0dB	5.8	7.5	9.0	9.4	9.4	9.5	9.8
	-5dB	3.3	4.9	5.5	5.9	6.2	5.9	6.1
23 Hz	5dB	6.8	8.2	9.7	10.0	10.2	10.3	10.4
	0dB	4.1	5.5	6.0	6.1	6.4	6.5	6.7
	-5dB	2.1	2.9	2.9	3.0	3.2	3.3	3.2

ratios ($\mathrm{SNR}_{\mathrm{INIT}}$) of 5 dB, 0 dB, and -5 dB. The sampling frequency was 256 Hz and the signal length 5 s. The average SNRs of the reconstructed sinusoids over 50 simulations using the EMD, EEMD, and MEMD algorithms are given in Table I. In all the scenarios, MEMD outperformed the single channel decomposition algorithms, EMD and EEMD. Observe, in the case of MEMD, that an increase in the number of channels resulted in an increase in performance. These results illustrate that when common activity (oscillating components) exists between several data channels, MEMD has the ability to generate a more accurate estimate of the signal envelope (see the envelope estimation stage as described in Algorithm 2) and thus identify the common activity between multiple data channels more robustly.

IV. MULTIVARIATE ANALYSIS OF MOTOR IMAGERY DATA

This section evaluates the MEMD performance using two motor imagery datasets.

A. Materials: Motor Imagery EEG Datasets

1) BCI Competition IV Dataset I: The data[7] was recorded from four healthy subjects using the BrainAmp MR plus EEG amplifier with 59 electrodes sampled at 1000 Hz [31]. Each subject selected two motor imagery tasks among three: $left\ hand$, $right\ hand$ and $foot$ (both feet). Specifically, subject **a** chose $left\ hand$ and $foot$, subject **b** chose $left\ hand$ and $right\ hand$, subject **f** chose $left\ hand$ and $foot$, and subject **g** chose $left\ hand$ and $right\ hand$. Subjects performed a total of 200 trials. In each trial, the subject imagined one of the two possible tasks (one task per trial) for a duration of 4 s. The selected task was randomised between trials so that the subject imagined each task 100 times in total. Before each task, baseline recordings were made and the subject was presented with a series of pre-task cues so that the onset and termination times were well defined. Based on neurophysiological insight, out of the 59 EEG channels, 11 were selected for analysis, "FC3," "FC4," "Cz," "C3," "C4," "C5," "C6," "T7," "T8," "CCP3," and "CCP4" according to the 10–20 system [32], since the motor imagery response is primarily associated with the central area of the brain [4], [5]. The data was down sampled to a 100 Hz sampling rate.

[7]The BCI Competition IV Dataset I was recorded for both human and artificially-generated motor imagery data and only the human-generated motor imagery data was considered in our simulations. There were two sessions of collected data, calibration and evaluation sessions, and only the calibration session was used. The data is available from http://www.bbci.de/competition/iv/.

2) Physiobank Motor/mental Imagery Database: We used the Physiobank Motor/Mental Imagery (MMI) database recorded using the BCI2000 system [33], available through Physionet[8][34]. Subjects performed different motor imagery tasks while 64-channel EEG were recorded according to 10–10 system, sampled at 160 Hz. We chose the blocks where the subjects imagined movement of $left\ hand$ and $right\ hand$. Subjects performed a total of 45 trials and imagined one of the two tasks for a duration of 4 s. Out of 64 EEG channels, 11 were selected for analysis, "FC3," "FC4," "Cz," "C3," "C4," "C5," "C6," "T7," "T8," "CP3," and "CP4."

B. Time-Frequency Analysis Using MEMD

A comprehensive comparative study was performed to illustrate the ability of MEMD to produce more accurate spectrogram estimates over the short-time Fourier transform (STFT), continuous wavelet transform (CWT), and synchrosqueezed wavelet transform (SST). Time-frequency spectrograms for the $left\ hand$ motor imagery datasets of subject **g** ("BCI Competition IV Dataset I") from electrode C3 and C4 were estimated using the STFT, CWT, SST, and NA-MEMD. A 0.3 s sliding Hamming window with 29 data points overlap was applied to create the STFT representation. The CWT time-frequency representation was calculated using the commonly used Morlet wavelet.[9] The SST [35], [36] is a recent extension of the CWT which reduces redundancy in the wavelet spectrogram and increases its localization [37], see Appendix A for the mathematical formulation. The NA-MEMD decomposed all data channels with the aid of two additional noise channels and the individual HHS were calculated from the 14-variate IMFs. The spectra of C3 (left hemisphere) and C4 (right hemisphere) were produced to examine ERS of mu and $beta$ rhythms over the ipsilateral hemisphere and ERD over the contralateral hemisphere corresponding to the $left\ hand$ motor imagery task.

Fig. 3 illustrates the percentage changes of power in logarithmic scale, relative to the power in a baseline interval from -0.5 to 0 s prior to the stimulus (Event-related spectral

[8]http://www.physionet.org/pn4/eegmmidb/#experimental-protocol

[9]The continuous wavelet transform is given by

$$\bar{W}(\dot{a},\dot{b};V,\psi) = |\dot{a}|^{-1/2}\int_{-\infty}^{\infty} V(t)\psi^*\left(\frac{t-\dot{b}}{\dot{a}}\right)dt \tag{4}$$

where ψ^* is the mother wavelet, $\dot{a}$ the dilation factor and $\dot{b}$ the translation of the origin.

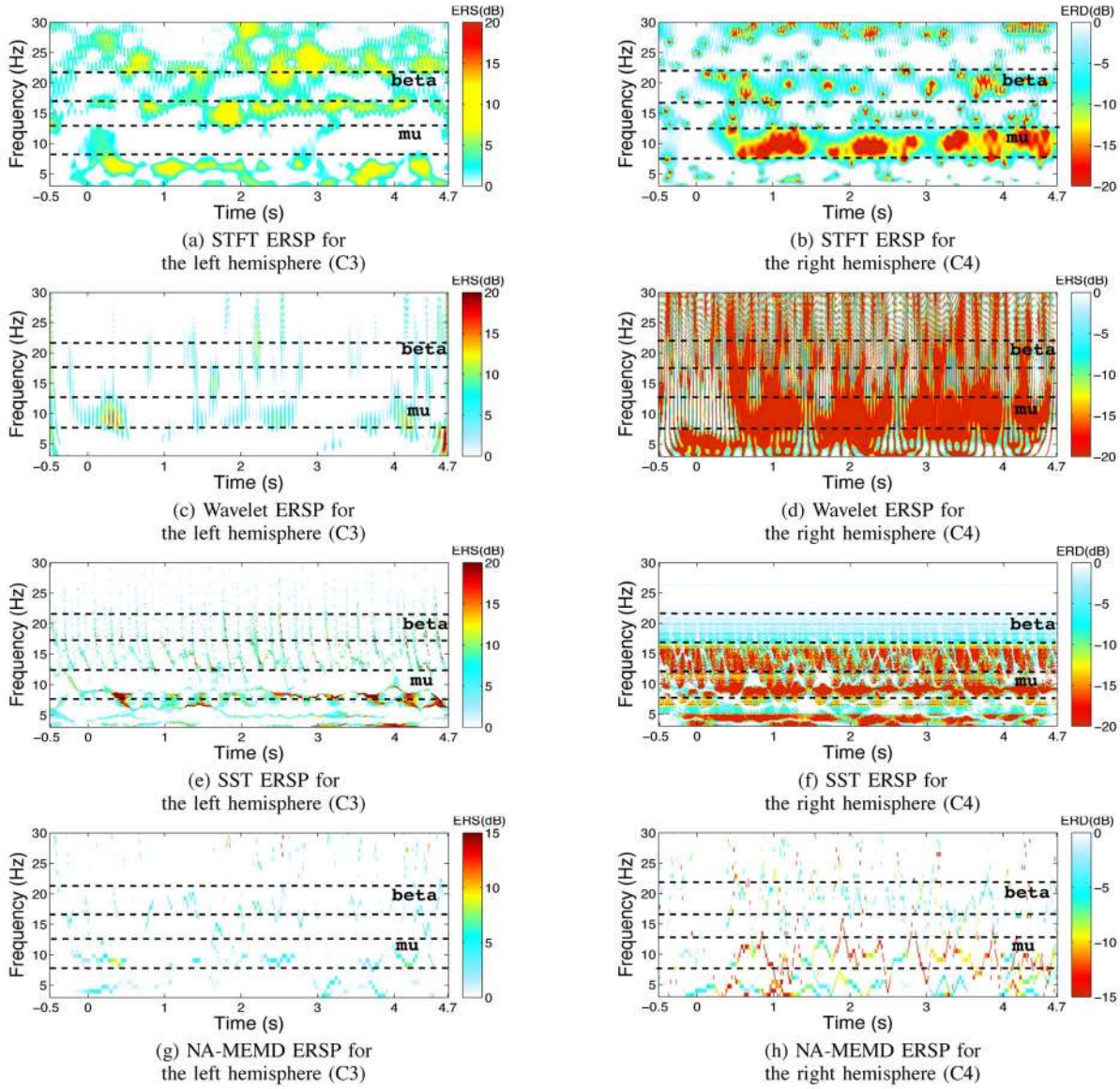


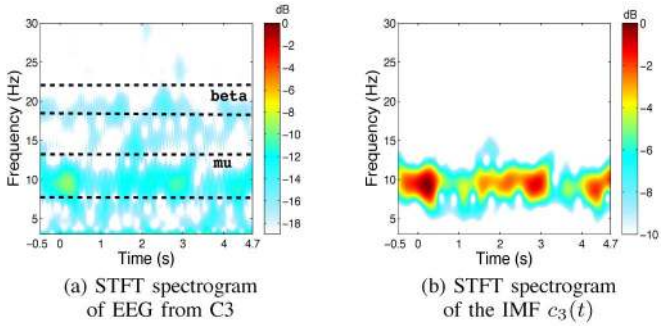
Fig. 3. Event-related spectral perturbation (ERSP) obtained using the short-time Fourier transform (STFT), wavelet (Morlet), synchrosqueezed wavelet (SST) and NA-MEMD HHS for $left\ hand$ motor imagery tasks. NA-MEMD estimated more localized and salient power increases (ERS in C3) and decreases (ERD in C4) of the mu and $beta$ rhythms after 0 s.

perturbation (ERSP) [38]), where the motor imagery task was performed within the time period 0–4 s. Subfigures in the left column show the spectra over the ipsilateral hemisphere while subfigures in the right column display the spectra over the contralateral hemisphere. The ERSPs obtained using all four methods illustrate obvious contralateral decrease (ERD) in the alpha (8–12 Hz) and beta (13–30 Hz) bands on the sensorymotor cortex for the duration of the motor imagery task [5], while ipsilateral increase (ERS) is not shown using STFT. Observe that the STFT and the CWT obtained less localized ERD than the SST and NA-MEMD. However, the SST lost $beta$ rhythm information over approximately 20 Hz.

To compare NA-MEMD and standard methods within the same time-frequency domain, Fig. 4 shows the STFT of the complete EEG signal from C3 and the IMF $c_3(t)$ (mu rhythm) obtained for the EEG using NA-MEMD. Observe that the NA-MEMD-based approach gave highly localized analysis, where the time-frequency component of $c_3(t)$ was very close to the alpha band component in (a).

The extracted mu and $beta$ rhythm time-series from -2 to 6 s obtained using the fifth-order Butterworth filter (BF), CWT, SST, and NA-MEMD are shown in Fig. 5. The time series display the amplitude changes of the signal relative to the mean amplitude in a baseline interval, which were normalized by the standard deviation of baseline waveform. All the methods detected an ipsilateral (C3) increase (ERS) and contralateral (C4) decrease (ERD) in the mu and $beta$ rhythms after the onset time (0 s). In particular, the ERS and ERD within the mu rhythms are more prominent than those observed within the $beta$ rhythms. Note the power changes of the time series relative to the baseline power in Fig. 6; the power was estimated by the squared envelope of the time series.

In order to compare the component estimation performance, we generated noisy realizations of C3 and C4 by adding white



(a) STFT spectrogram of EEG from C3 (b) STFT spectrogram of the IMF $c_3(t)$

Fig. 4. STFT spectra of EEG and its IMF $c_3(t)$. Note that alpha activity has been better estimated by performing the Fourier analysis at the IMF level.

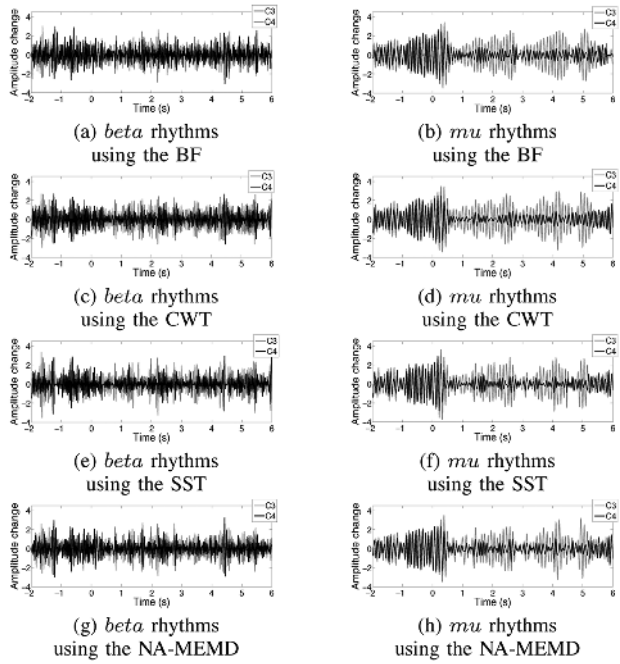
noise to the extracted mu and $beta$ rhythms from the previous set of simulations, defined as: $(t = -2\ s, \ldots, 6\ s)$

$$
\begin{aligned}
u_{C3}(t) &= \mu^B_{C3}(t) + \mu^W_{C3}(t) + \mu^S_{C3}(t) + \mu^M_{C3}(t) \\
\zeta_{C3}(t) &= \beta^B_{C3}(t) + \beta^W_{C3}(t) + \beta^S_{C3}(t) + \beta^M_{C3}(t) \\
u_{C4}(t) &= \mu^B_{C4}(t) + \mu^W_{C4}(t) + \mu^S_{C4}(t) + \mu^M_{C4}(t) \\
\zeta_{C4}(t) &= \beta^B_{C4}(t) + \beta^W_{C4}(t) + \beta^S_{C4}(t) + \beta^M_{C4}(t) \\
x_{C3}(t) &= u_{C3}(t) + \zeta_{C3}(t) \\
x_{C4}(t) &= u_{C4}(t) + \zeta_{C4}(t)
\end{aligned}
$$

where $x_{C3}(t)$ is composed of mu and $beta$ rhythms from channel C3 ($\mu^B_{C3}(t)$, $\mu^W_{C3}(t)$, $\mu^S_{C3}(t)$, $\mu^M_{C3}(t)$, $\beta^B_{C3}(t)$, $\beta^W_{C3}(t)$, $\beta^S_{C3}(t)$, $\beta^M_{C3}(t)$, where B denotes BF, W wavelet, S SST and M NA-MEMD) and $x_{C4}(t)$ is composed of those from channel C4 ($\mu^B_{C4}(t)$, $\mu^W_{C4}(t)$, $\mu^S_{C4}(t)$, $\mu^M_{C4}(t)$, $\beta^B_{C4}(t)$, $\beta^W_{C4}(t)$, $\beta^S_{C4}(t)$, $\beta^M_{C4}(t)$). The mu and $beta$ rhythms were estimated from $x_{C3}(t) + q_1(t)$ and $x_{C4}(t) + q_2(t)$ ($q_1(t)$ and $q_2(t)$ denote realizations of 0 dB WGN) using the four methods, and the mean squared errors (MSEs) were calculated by comparing the estimated mu with original mu ($u_{C3}(t)$ and $u_{C4}(t)$) and $beta$ ($\zeta_{C3}(t)$ and $\zeta_{C4}(t)$) rhythms. Using 100 different realizations of $q_1(t)$ and $q_2(t)$, the averaged MSEs for each channel and frequency band are shown in Table II. Note that NA-MEMD always outperformed the other considered methods.

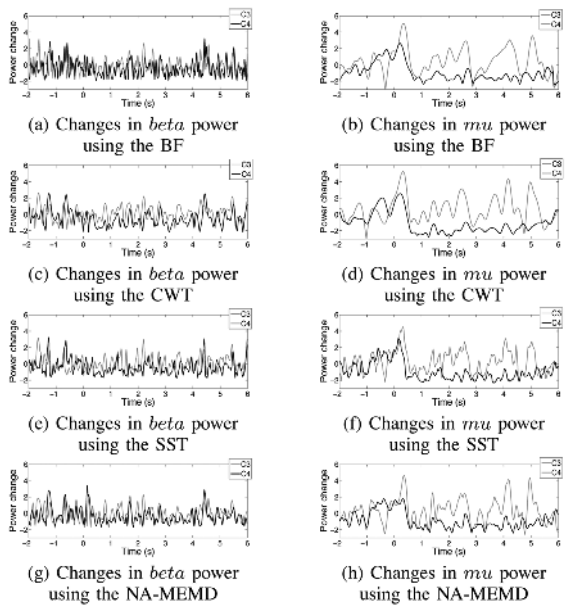
C. MEMD-Based CSP Feature Estimation

1) Preprocessing: The motor imagery data was band-pass filtered to occupy the frequency band 8–30 Hz [9], [39], using a fifth-order Butterworth filter, and subsequently the CWT, the SST, EMD, EEMD, MEMD, and NA-MEMD. In the case of the CWT and the SST, the scales were reconstructed to obtain the band-pass filtered signals. As before, the standard single channel EMD algorithms (EMD, EEMD) were applied to each channel separately while MEMD was applied to the 11 channels simultaneously. The NA-MEMD algorithm was used to decompose the 11-channel EEG signals with two additional noise channels (SNR 20 dB). The IMFs obtained using the EMD approaches were retained or omitted based on inspection, with the most relevant ones added together to obtain the band-pass filtered signals. Relevance of the IMFs was estimated by calculating the combination which gave the best classification performance with respect to the common spatial patterns algorithm.



(a) $beta$ rhythms using the BF (b) mu rhythms using the BF
(c) $beta$ rhythms using the CWT (d) mu rhythms using the CWT
(e) $beta$ rhythms using the SST (f) mu rhythms using the SST
(g) $beta$ rhythms using the NA-MEMD (h) mu rhythms using the NA-MEMD

Fig. 5. Amplitude changes in $beta$ and mu rhythms based on the mean and standard deviation of baseline (from −2 to 0 s) waveform, which were estimated using the BF, the CWT, the SST and the NA-MEMD from the EEG responses at C3 and C4. After the onset time (0 s), an ipsilateral (C3) increase (ERS) and contralateral (C4) decrease (ERD) in the $beta$ and mu rhythms can be observed.



(a) Changes in $beta$ power using the BF (b) Changes in mu power using the BF
(c) Changes in $beta$ power using the CWT (d) Changes in mu power using the CWT
(e) Changes in $beta$ power using the SST (f) Changes in mu power using the SST
(g) Changes in $beta$ power using the NA-MEMD (h) Changes in mu power using the NA-MEMD

Fig. 6. Power changes in $beta$ and mu rhythms based on the mean of baseline (from −2 s to 0 s) power, which were calculated using the waveforms in Fig. 5. After the onset time (0 s), ipsilateral (C3) increase (ERS) and contralateral (C4) decrease (ERD) in the $beta$ and mu rhythms can be noted.

2) Common Spatial Patterns: Features relevant to motor imagery responses were extracted using the CSP algorithm, widely used in BCI applications [9], [39]. It determines spatial filters that maximize the variance of signals in one class and simultaneously minimize the variance of signals in the other class. In this way, CSP filters can discriminate between ERD/ERS caused by changing mental states as their operation is sensitive to subtle

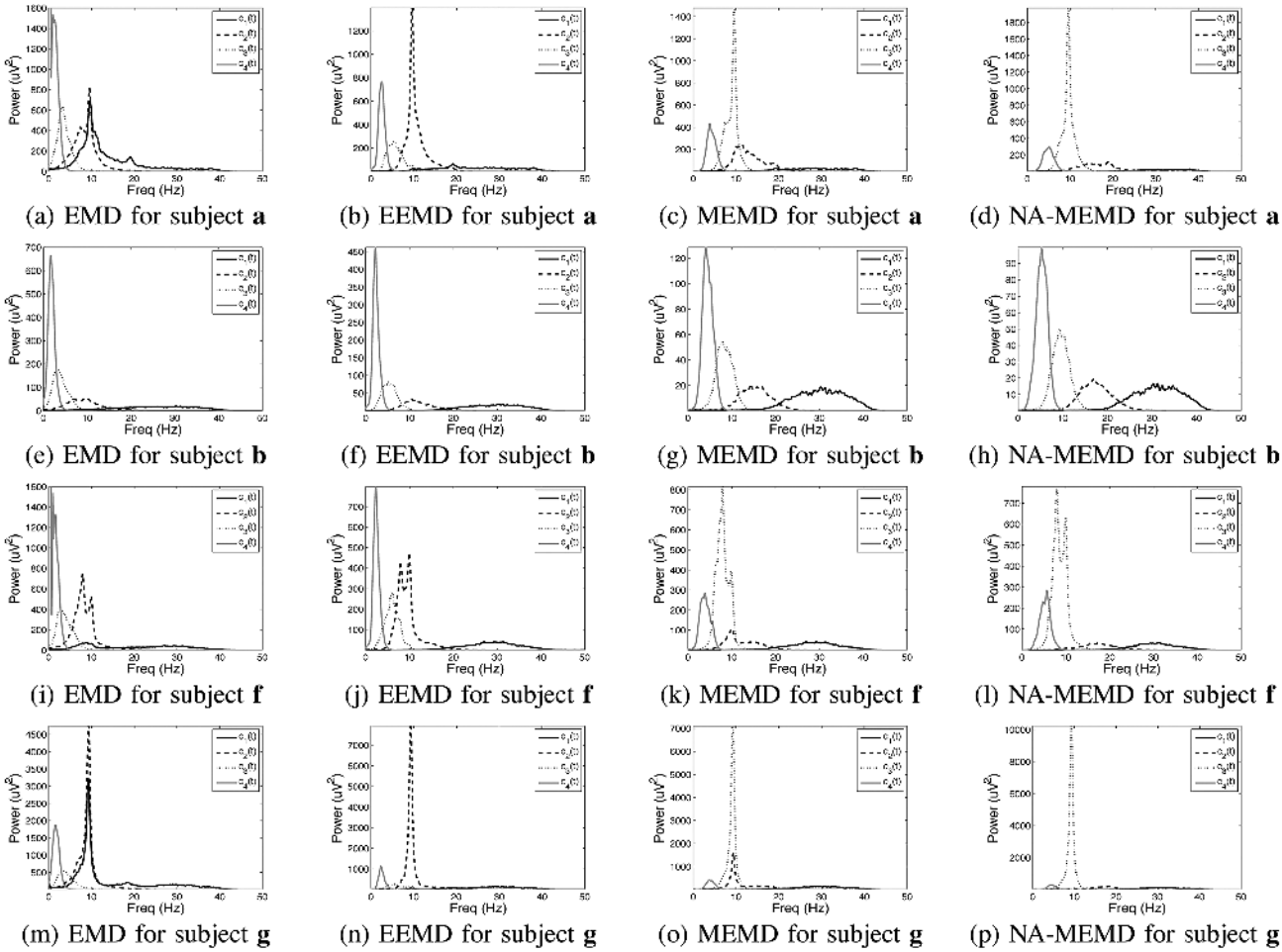


Fig. 7. Average power spectra of $c_1(t) - c_4(t)$ for each subject, obtained using EMD, EEMD, MEMD and NA-MEMD. Note that the $c_2(t)$ and $c_3(t)$ of MEMD and NA-MEMD correspond to the $beta$ and mu rhythms respectively, contrary to the instability of the standard EMD algorithm, caused by mode mixing—components of interest are spread across different IMFs.

TABLE II
MEAN SQUARED ERRORS BETWEEN SYNTHETIC mu AND $beta$ RHYTHMS AND THE ESTIMATED mu AND $beta$ RHYTHMS USING BF, CWT, SST AND NA-MEMD. NOTE THAT NA-MEMD OUTPERFORMS THE OTHER METHODS

	$C3$		$C4$	
Algorithm	mu	$beta$	mu	$beta$
BF	0.102	0.340	0.110	0.396
CWT	0.047	0.085	0.047	0.100
SST	0.045	0.100	0.050	0.133
NA-MEMD	**0.039**	**0.081**	**0.042**	**0.090**

changes in the power of band-pass filtered data. For a detailed description of the CSP algorithm, see Appendix B.

3) Classification: The feature vectors $\mathbf{f} = [f_1, f_{11}]$ and $\mathbf{f} = [f_1, f_2, f_{10}, f_{11}]$, obtained by (13) in Appendix B for $m = 1$ and $m = 2$, respectively, (m defines the number of spatial filters), were classified using a support vector machine (SVM) [40] with a Gaussian kernel.[10] The 200 trial data for each subject in "BCI Competition IV Dataset I" was divided into 140 training and 60 testing sets and the 45 trial data in "Physiobank MMI database" was divided into 32 subsets for training and 13 for test sets.[11] The upper limit of confidence intervals between two classes corresponding to the number of trials was 56.9% for 200 trials and 64.0% for 45 trials (cf. [42]). For cross-validation, classification was repeated 100 times while mixing the sample order, and the average of these outcomes was the final classification rate.

4) Results: Fig. 7 illustrates the average power spectra of the first four IMFs $c_1(t) - c_4(t)$ across the EEG channels for all 200 trials of "BCI Competition IV Dataset I." Note that, compared to standard EMD, the average frequency distributions at each IMF level using EEMD, MEMD, and NA-MEMD were more localized in frequency and that greater separation was obtained between the IMF frequency distributions. This indicates that there is greater consistency in the pattern of extracted EEG components between trials and that these components have been better separated. In particular, the $beta$ and mu rhythms have been clearly separated using MEMD and NA-MEMD, contained in the IMFs $c_2(t)$ and $c_3(t)$. These two different rhythms were extracted erroneously within a single IMF component, $c_2(t)$, using the univariate EEMD, illustrating poorer separability compared to the multichannel algorithms. Observe that the different frequency ranges corresponding to the IMFs of NA-MEMD were more consistently distributed across the subjects than those of MEMD, illustrating the benefits of imposing a data-driven filterbank structure.

Table III shows the classification performances for the four subjects of "BCI Competition IV Dataset I" using the BF, CWT, SST, EMD, EEMD, MEMD, and NA-MEMD, where the IMFs

[10]The MATLAB code can be downloaded from [41].

[11]The CSP filter parameters were defined using only the training data set.

TABLE III
CLASSIFICATION RESULTS (IN % AND ASSOCIATED VARIANCE) FOR ALL THE ALGORITHMS CONSIDERED, $m = 1$ AND $m = 2$ [SEE (13)] AND FOUR SUBJECTS OF "BCI COMPETITION IV DATASET I." SYMBOL "BF" DENOTES THE BUTTERWORTH FILTER, "CWT" MORLET WAVELET TRANSFORM AND "SST" SYNCHROSQUEEZED WAVELET TRANSFORM. THE ENSEMBLE SIZE FOR EEMD WAS 100. TWO NOISE CHANNELS WERE ADDED TO OBTAIN THE NA-MEMD DECOMPOSITION AND m DEFINES THE NUMBER OF SPATIAL FILTERS. NOTE THAT THE MEMD-CSP AND NA-MEMD-CSP APPROACHES, USING $c_2(t)$ AND $c_3(t)$, GAVE THE BEST RESULTS FOR ALL SUBJECTS

	Algorithm	IMFs	$m = 1$	$m = 2$
a	BF		62.0 ± 11.2	82.3 ± 4.8
	CWT		66.2 ± 10.6	84.5 ± 5.5
	SST		67.1 ± 11.6	84.0 ± 4.5
	EMD	$c_1(t) - c_3(t)$	57.0 ± 6.6	62.6 ± 6.0
	EEMD	$c_2(t)$	63.0 ± 8.2	78.4 ± 5.1
	MEMD	$c_2(t) - c_3(t)$	**70.5** ± 11.2	85.7 ± 4.0
	NA-MEMD	$c_2(t) - c_3(t)$	69.8 ± 10.6	**85.9 ± 3.9**
b	BF		57.6 ± 7.5	58.6 ± 6.1
	CWT		71.4 ± 6.5	71.0 ± 5.7
	SST		68.4 ± 7.6	70.5 ± 5.7
	EMD	$c_1(t) - c_4(t)$	52.1 ± 5.7	57.3 ± 5.5
	EEMD	$c_1(t) - c_2(t)$	67.9 ± 8.0	69.8 ± 5.5
	MEMD	$c_2(t) - c_3(t)$	75.6 ± 5.2	73.9 ± 5.8
	NA-MEMD	$c_2(t) - c_3(t)$	**78.7 ± 3.7**	**77.6 ± 4.8**
f	BF		52.6 ± 6.9	60.2 ± 6.8
	CWT		52.9 ± 5.7	58.5 ± 7.3
	SST		54.4 ± 7.1	72.2 ± 5.6
	EMD	$c_1(t) - c_4(t)$	52.2 ± 5.9	57.2 ± 6.0
	EEMD	$c_2(t) - c_3(t)$	53.5 ± 11.2	69.7 ± 7.2
	MEMD	$c_2(t) - c_3(t)$	**57.5 ± 13.4**	77.8 ± 4.3
	NA-MEMD	$c_2(t) - c_3(t)$	57.3 ± 14.2	**78.8 ± 4.4**
g	BF		86.9 ± 7.4	85.6 ± 4.6
	CWT		78.8 ± 9.4	88.1 ± 4.6
	SST		91.4 ± 2.9	90.5 ± 3.5
	EMD	$c_1(t) - c_3(t)$	65.5 ± 10.8	72.3 ± 7.5
	EEMD	$c_2(t)$	89.4 ± 3.9	88.6 ± 3.7
	MEMD	$c_2(t) - c_3(t)$	**91.9 ± 3.0**	**91.5 ± 3.5**
	NA-MEMD	$c_2(t) - c_3(t)$	91.0 ± 3.3	90.9 ± 3.5

of EMD, EEMD, MEMD, and NA-MEMD were selected based on the IMF power spectra in Fig. 7 and the optimal classification performance. The classification performances were calculated for $m = 1$ and for $m = 2$ [see (13)], where the best result among the seven different methods is indicated in bold. Observe that MEMD-CSP and NA-MEMD-CSP features always gave classification rates above the upper limit of confidence interval of 56.9%, and outperformed the other methods. On average (when $m = 2$), the NA-MEMD-CSP approach gave the best classification performance of 83.3%, a 11.6% improvement over the BF, a 7.8% improvement over the CWT, a 4% improvement over the SST, a 20.9% improvement over EMD, a 6.7% improvement over EEMD and a 1.1% improvement over MEMD. These results compare favourably with other methods applied to the same BCI dataset using a similar number of EEG channels[12].

The classification performances for the second dataset, "Physiobank MMI database," are shown in Table IV, and were obtained using the BF, CWT, SST, and NA-MEMD with CSP ($m = 2$, which mostly gave higher or equal classification rates compared to the case of $m = 1$ for "BCI Competition IV Dataset I"). Among the EMD-based algorithms, we chose NA-MEMD since it obtained the highest classification rates for "BCI Competition IV Dataset I." The first 10 subjects from the

[12]Within the BCI competition dataset considered, submission 11 (16 channels), submission 13 (three channels), submission 14 (13 channels), and submission 17 (10 channels) gave average MSE values of 0.557, 0.679, 0.692, and 0.915, while our average MSE value is 0.668 (see http://www.bbci.de/competition/iv/results/#dataset1avsr).

TABLE IV
CLASSIFICATION RESULTS (IN %) FOR "PHYSIOBANK MMI DATABASE" OBTAINED USING BF, CWT, SST, AND NA-MEMD. THE NA-MEMD APPROACH ENABLED THE BEST OVERALL CLASSIFICATION PERFORMANCE, AND THE BEST PERFORMANCE FOR SEVEN OUT OF TEN OF THE SUBJECTS

Subject	BF	CWT	SST	NA-MEMD
1	74.2 ± 11.7	73.2 ± 10.5	65.1 ± 12.7	**77.2 ± 9.7**
2	**86.1 ± 8.0**	86.1 ± 7.7	73.0 ± 12.8	86.0 ± 9.1
4	61.2 ± 11.6	50.6 ± 11.7	54.6 ± 13.4	**66.4 ± 10.6**
7	96.0 ± 5.86	96.5 ± 4.9	78.1 ± 12.4	**97.1 ± 4.9**
12	62.7 ± 13.5	62.9 ± 13.0	54.6 ± 12.9	**64.0 ± 13.1**
13	**65.6 ± 12.4**	64.3 ± 11.3	59.1 ± 13.8	65.5 ± 11.3
15	69.5 ± 14.2	70.9 ± 13.2	64.2 ± 13.4	**71.2 ± 12.1**
25	61.9 ± 12.3	51.4 ± 12.7	60.1 ± 14.7	**77.6 ± 10.9**
26	**74.3 ± 12.6**	70.8 ± 13.3	57.9 ± 14.4	74.2 ± 12.0
29	96.6 ± 4.8	95.9 ± 4.7	81.9 ± 12.8	**97.4 ± 4.0**
mean	74.8	72.3	64.9	**77.7**

"Physiobank MMI database" were chosen, and the classification results were above the upper limit of the confidence interval of 64%. In the simulations, $c_3(t)$ and $c_4(t)$ were decomposed using NA-MEMD, containing $beta$ and mu rhythms. Among 10 subjects, NA-MEMD gave the best classification rates for seven; in the case of the remaining three subjects (2, 13, and 26), the results obtained with other methods only marginally outperformed NA-MEMD by 0.1%. On average, NA-MEMD gave a 2.9% improvement over BF, 5.4% over CWT and 12.8% over SST.

Fig. 8(a)–(c) displays a series of scatter-plots of classification rates between NA-MEMD and the other analysis methods considered using all the classification results of two experiments in Tables III and IV. For equal performances between two methods, the values should lie on the diagonal—observe that most of values lie below the diagonal, indicating an enhanced performance of NA-MEMD compared to the other algorithms. For rigor, the difference in classification rate between NA-MEMD and the other methods was also analyzed using the one-tailed t-test. The corresponding p-values in Fig. 8(a)–(c) shows NA-MEMD performing significantly better than the other algorithms (p-values less than 0.01).

We next examined the relationship between the number of EEG channels analyzed and the MEMD performance. Table V shows classification results obtained by using the following.

1) IMFs for four channels (C3, C4, CCP3, and CCP4 for "BCI Competition IV Dataset I," and C3, C4, CP3, and CP4 for "Physiobank MMI database") of the original 11-channel MEMD decomposition denoted by $\mathrm{IMF}_{(4\subset 11)}$.
2) IMFs for four channels C3, C4, CCP3, and CCP4 for "BCI Competition IV Dataset I," and C3, C4, CP3, and CP4 for "Physiobank MMI database") of a 4-channel MEMD decomposition denoted by $\mathrm{IMF}_{(4\subset 4)}$.

The number of spatial filters, m in (13), was $m = 2$. The best performance for 9 subjects out of 14 was obtained using the $\mathrm{IMF}_{(4\subset 11)}$ approach, exhibiting on the average a 1.4% improvement compared to "$\mathrm{IMF}_{(4\subset 4)}$." A series of scatter-plots of these classification rates is displayed in Fig. 8(d). Most of the dots are distributed below the diagonal, which means that simulations based on 4-channel IMFs from an 11-channel MEMD decomposition outperformed those from 4-channel MEMD decomposition. These significant improvements were also confirmed by the one-tailed p-values of the t-test (less than 0.05).

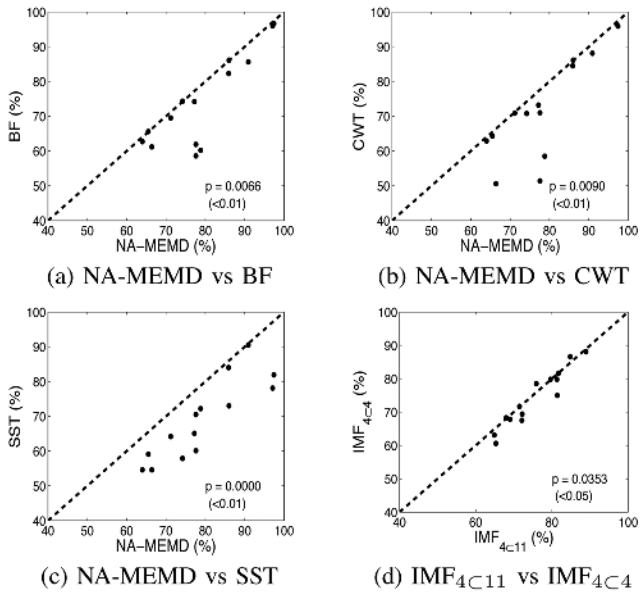


Fig. 8. Scatter-plot of classification results of "BCI Competition IV dataset I" and "Physiobank MMI database" for NA-MEMD versus the other methods based on Tables III and IV [(a), (b), and (c)], and 4-channel IMFs from an 11-channel MEMD decomposition ($\mathrm{IMF}_{4\subset 11}$) versus those from a 4-channel MEMD decomposition ($\mathrm{IMFs}_{4\subset 4}$) based on Table IV (d). The dots below the diagonal are cases where NA-MEMD outperforms the other algorithm in (a), (b), and (c), and $\mathrm{IMF}_{4\subset 11}$ outperforms $\mathrm{IMF}_{4\subset 4}$ in (d). The significance of these improvements are confirmed by the one-tailed p-values of the t-test (less than 0.05).

TABLE V
CLASSIFICATION RATES (IN %) OF "BCI COMPETITION IV DATASET I" AND "PHYSIOBANK MMI DATABASE" USING 4-CHANNEL IMFS OBTAINED USING MEMD FOR $m = 2$ [SEE (13)]. "$\mathrm{IMF}_{(4\subset 11)}$" DENOTES THE 4-CHANNEL IMFS FROM AN 11-CHANNEL MEMD DECOMPOSITION. "$\mathrm{IMF}_{(4\subset 4)}$" IS THE 4-CHANNEL IMFS FROM A 4-CHANNEL MEMD DECOMPOSITION. m DEFINES THE NUMBER OF SPATIAL FILTERS

Subject	IMFs	$\mathrm{IMF}_{(4\subset 11)}$	$\mathrm{IMF}_{(4\subset 4)}$
a	$c_2(t) - c_3(t)$	**81.3 ± 4.5**	79.8 ± 5.3
b	$c_2(t) - c_3(t)$	**65.3 ± 6.0**	60.6 ± 5.8
f	$c_2(t) - c_3(t)$	68.0 ± 5.1	**68.3 ± 4.8**
g	$c_2(t) - c_3(t)$	**88.9 ± 3.3**	88.1 ± 4.1
1	$c_3(t) - c_4(t)$	**81.7 ± 8.4**	81.6 ± 8.0
2	$c_3(t) - c_4(t)$	84.8 ± 8.0	**86.6 ± 8.9**
4	$c_3(t) - c_4(t)$	**72.1 ± 10.6**	67.5 ± 10.8
7	$c_3(t) - c_4(t)$	**81.4 ± 8.5**	75.0 ± 11.6
12	$c_3(t) - c_4(t)$	**72.2 ± 10.7**	69.4 ± 11.4
13	$c_3(t) - c_4(t)$	**64.9 ± 12.2**	63.1 ± 12.1
15	$c_3(t) - c_4(t)$	79.6 ± 8.8	**79.8 ± 10.6**
25	$c_3(t) - c_4(t)$	**69.0 ± 9.5**	67.8 ± 11.7
26	$c_3(t) - c_4(t)$	71.5 ± 11.9	**71.7 ± 12.2**
29	$c_3(t) - c_4(t)$	75.9 ± 10.3	**78.5 ± 10.7**
mean		**75.5**	74.1

V. DISCUSSION

Analysis of the motor imagery response in Section IV-B illustrates the high level of accuracy that is achievable using MEMD/NA-MEMD for time-frequency analysis. While the SST gave a similar performance, it inherited the problems of high frequency resolution associated with wavelets and was not sensitive to frequency components above 20 Hz, critically ignoring the full range of the beta band (13–30 Hz). In addition to the spectrogram analysis, the accurate estimation of synthetic mu and $beta$ rhythms using NA-MEMD in the time domain was also investigated.

Power spectra of MEMD/NA-MEMD contained in the IMF $c_1(t)$ in Fig. 7 have been shown to include some parts of the $beta$ band (20–30 Hz). However, the classification results obtained using IMF $c_1(t)$ were always outperformed by features that included IMF $c_2(t)$ and $c_3(t)$ only. This implies that the components $c_2(t)$ and $c_3(t)$ obtained using MEMD/NA-MEMD accurately reflected $beta$ and mu rhythms, which occupy respectively between 9 and 14 Hz and between 18 and 26 Hz [2].

The motor imagery classification experiments were conducted with the aim of showing that when IMFs are obtained from a greater number of data-bearing channels, this yields more accurate classification rates [see Table V and Fig. 8(d)]. This is also supported by the simulation results in Section III-B for synthetic data. This is explained by the enhanced decomposition accuracy that is enabled by processing data-bearing channels simultaneously—in this instance the mu and $beta$ rhythms. Observe that by introducing extra noise channels into the MEMD framework, the noise-assisted MEMD algorithm was furthermore able to alleviate the effects of mode mixing and mode misalignment in multivariate IMFs, as shown in Fig. 2 (time domain) and Fig. 7 (frequency domain).

In our analysis the IMF relevance was estimated empirically, however the high classification rates, obtained with significant improvements for the two motor imagery datasets (see Fig. 8), suggest that the MEMD spectrogram estimates for the motor imagery response approximate the ground truth. Unlike standard single-channel EMD, the robust natures of the MEMD and NA-MEMD ensure that the information contained in multivariate IMFs remains the same across the data channels, trials, and subjects.

EMD-based algorithms, including ensemble EMD (EEMD) and multivariate EMD (MEMD), require large computational resources [20], [25]; for instance, in our case it takes around 100 s for MEMD to process a quadrivariate white Gaussian noise (WGN) data of length $N = 100$. For NA-MEMD, the computational requirements are even larger since added noise channels (in NA-MEMD) also need to be processed. However, real-time and online computation of EMD-based operations has attracted considerable attention recently, which along with the advancements in computational hardware is expected to alleviate the computational limitations of EMD.

VI. CONCLUSION

It has been illustrated how the multivariate empirical mode decomposition (MEMD) algorithm provides robust time-frequency analysis for multichannel signals where the narrowband nonstationary signal of interest is buried in broadband noise. In a BCI study based on motor imagery EEG responses, the MEMD algorithm facilitated multicomponent extraction of the mu and $beta$ rhythms of interest. Unlike the standard single channel EMD algorithm, MEMD and, in particular its noise-assisted variant NA-MEMD allowed a more stable estimate of the time-varying frequency responses from multichannel EEG, providing physical meaning to the intrinsic data modes. The performance of the proposed approach for motor imagery classification via the common spatial patterns algorithm has been compared to those of conventional methods of a kind, including the short-time Fourier transform, wavelet transform and the

synchrosqueezed wavelet transform, exhibiting significant improvements.

APPENDIX A
SYNCHROSQUEEZED WAVELET TRANSFORM

The synchrosqueezed wavelet transform (SST) [36] is an extension of the CWT, given by

$$\bar{W}_s(\dot{a},\dot{b}) = \int s(t)\,\dot{a}^{-1/2}\,\psi\left(\frac{t-\dot{b}}{\dot{a}}\right)\,dt. \tag{5}$$

The CWT convolves the original signal with a wavelet (a finite energy oscillation) for different scales and time shifts, thus effectively projecting the original signal from the time domain to the time-scale domain. The CWT spreads the energy contained in a narrow range of frequencies in a signal around a particular scale, by $\dot{a} = \bar{\omega}_0/\bar{\omega}$ ($\bar{\omega}_0$—a frequency around which the applied wavelet is concentrated, $\bar{\omega}$—a pure harmonic signal); this also introduces redundancy in the wavelet coefficients. To address this issue, the synchrosqueezed wavelet transform (SST) reallocates the wavelet coefficients from the time-scale domain to the time-frequency domain through the procedure known as synchrosqueezing, ensuring that a wavelet in the positive spectrum is selected, from which the instantaneous frequency of the resulting CWT coefficients (5) is calculated as

$$\bar{\omega}_s(\dot{a},\dot{b}) = -j\bar{W}_s(\dot{a},\dot{b})\,\frac{\partial\bar{W}_s(\dot{a},\dot{b})}{\partial\dot{b}} \tag{6}$$

for each time-scale $(\dot{b},\dot{a})$. The next step is to perform frequency binning of the wavelet coefficients by synchrosqueezing. In other words, all the wavelet coefficients are combined from the same time-frequency bin to reduce the redundancy[13] in the CWT. Mathematically, synchrosqueezing (in the discrete case) can be formulated as

$$\bar{T}_s(\bar{\omega}_l,t) = \sum_{\dot{a}_k:|\bar{\omega}_s(\dot{a}_k,\dot{b})-\bar{\omega}_l|\le\Delta\bar{\omega}/2} \bar{W}_s(\dot{a}_k,\dot{b})\,\dot{a}^{-3/2}\,\Delta\dot{a}_k \tag{7}$$

where $\bar{\omega}_l$ is the central frequency of a selected frequency bin, $\Delta\dot{a}_k$ a difference of successive discrete scales and $\Delta\bar{\omega}$ a difference between successive frequency bin centers.

APPENDIX B
COMMON SPATIAL PATTERNS

A single trial EEG data is represented as an $N \times T$ matrix $\bar{\mathbf{X}}$, where N is the number of channels and T the number of samples per channel. The normalized spatial covariance of $\bar{\mathbf{X}}$ can be calculated from

$$\bar{\mathbf{C}} = \frac{\bar{\mathbf{X}}\bar{\mathbf{X}}^T}{tr(\bar{\mathbf{X}}\bar{\mathbf{X}}^T)} \tag{8}$$

where $(\cdot)^T$ denotes the matrix transpose operator and $tr(\mathbf{X})$ is the sum of the diagonal elements of $\mathbf{X}$. The spatial covariance $\bar{\mathbf{C}}_{d\in[a,b]}$ for a task, a or b, is obtained by the averaged covariance matrix of the task trials. The CSP analysis seeks to find a matrix $\mathbf{W}$ and diagonal matrices $\boldsymbol{\Lambda}_a$ and $\boldsymbol{\Lambda}_b$ ($\boldsymbol{\Lambda}_a + \boldsymbol{\Lambda}_b = \mathbf{I}$, the identity matrix) with elements $d \in [a,b]$ such that

$$\mathbf{W}^T\bar{\mathbf{C}}_a\mathbf{W} = \boldsymbol{\Lambda}_a,\ \mathbf{W}^T\bar{\mathbf{C}}_b\mathbf{W} = \boldsymbol{\Lambda}_b. \tag{9}$$

The composite spatial covariance is given as

$$\mathbf{C}_c = \bar{\mathbf{C}}_a + \bar{\mathbf{C}}_b \tag{10}$$

where $\mathbf{C}_c$ is factored as $\mathbf{C}_c = \mathbf{U}_c\boldsymbol{\Lambda}_c\mathbf{U}_c^T$, where $\mathbf{U}_c$ is the matrix of eigenvectors, and $\boldsymbol{\Lambda}_c$ is the diagonal matrix of eigenvalues. Using the whitening transformation, $\mathbf{P} = \sqrt{\boldsymbol{\Lambda}_c^{-1}}\mathbf{U}_c^T$, the variances in the space spanned by $\mathbf{U}_c$ are equalized, which makes all the eigenvalues of $\mathbf{P}\mathbf{C}_c\mathbf{P}^T$ equal to unity. Secondly, let $\mathbf{S}_a = \mathbf{P}\bar{\mathbf{C}}_a\mathbf{P}^T$ and $\mathbf{S}_b = \mathbf{P}\bar{\mathbf{C}}_b\mathbf{P}^T$, then $\mathbf{S}_a$ and $\mathbf{S}_b$ share the common eigenvector matrix, that is

$$\begin{aligned}\mathbf{B}^T\mathbf{S}_a\mathbf{B} &= \boldsymbol{\Lambda}_a,\ \mathbf{B}^T\mathbf{S}_b\mathbf{B}\\ &= \boldsymbol{\Lambda}_b \quad (\boldsymbol{\Lambda}_a + \boldsymbol{\Lambda}_b = \mathbf{I}).\end{aligned} \tag{11}$$

Since we assume the eigenvalues $\boldsymbol{\Lambda}_a$ are sorted in a descending order, the final spatial filter that satisfies (9) is given by $\mathbf{W}^T = \mathbf{B}^T\mathbf{P}$. This allows us to project the EEG signals as

$$\mathbf{Z} = \mathbf{W}^T\mathbf{X}. \tag{12}$$

Each column vector $\mathbf{w}_j$ $(j = 1,\ldots,N)$ of $(\mathbf{W}^T)^{-1}$ is called a spatial filter, or simply a filter. For discriminating between two motor imagery tasks, the variances of the spatially filtered signals using (12) are used as a feature. The row vectors $\mathbf{z}_p$ ($p = 1,\ldots,m$ ($p = 1,\ldots,m$ and $N-m+1,\ldots,N$) from $\mathbf{Z}$ that maximize the difference in the variance between the two groups are associated with the largest eigenvalues in $\boldsymbol{\Lambda}_a$ and $\boldsymbol{\Lambda}_b$. These signals are contained in the m first and last rows of $\mathbf{Z}$ in (12), due to the calculation of $\mathbf{W}^T$. The features of interest can be obtained as

$$f_p = \texttt{log}\left(\frac{\texttt{var}(\mathbf{z}_p)}{\sum_{i=1,\ldots,m \text{ and } N-m+1,\ldots,N}\texttt{var}(\mathbf{z}_i)}\right) \tag{13}$$

where the symbol $\texttt{var}(\cdot)$ denotes the variance.

[13]The implementation of the synchrosqueezed transform is detailed in [37].

REFERENCES

[1] A. Akrami, S. Solhjoo, A. Motie-Nasrabadi, and M. Hashemi-Golpayegani, "EEG-based mental task classification: Linear and nonlinear classification of movement imagery," in *Proc. 27th IEEE Eng. Med. Biol. Conf.*, Sep. 2005, pp. 4626–4629.

[2] G. Pfurtscheller, C. Neuper, D. Flotzinger, and M. Pregenzer, "EEG-based discrimination between imagination of right and left hand movement," *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, no. 6, pp. 642–651, 1997.

[3] D. J. McFarland, L. A. Miner, T. M. Vaughan, and J. R. Wolpaw, "Mu and beta rhythm topographies during motor imagery and actual movements," *Brain Topogr.*, vol. 12, no. 3, pp. 177–186, 2000.

[4] G. Pfurtscheller and F. H. L. da Silva, "Event-related EEG/MEG synchronization and desynchronization: Basic principles," *Clin. Neurophysiol.*, vol. 110, no. 11, pp. 1842–1857, 1999.

[5] H. Yuan, A. Doud, A. Gururajan, and B. He, "Cortical imaging of event-related (de)synchronization during online control of brain-computer interface using minimum-norm estimates in frequency domain," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 16, no. 5, pp. 425–431, Oct. 2008.

[6] V. V. Nikouline, K. Linkenkaer-Hansen, H. Wikstrom, M. Kesaniemi, E. V. Antonova, R. J. Ilmoniemi, and J. Huttunen, "Dynamics of mu-rhythm suppression caused by median nerve stimulation: A magnetoencephalographic study in human subjects," *Neurosci. Lett.*, vol. 294, no. 3, pp. 163–166, 2000.
[7] G. Pfurtscheller, "Graphical display and statistical evaluation of event-related desynchronization (ERD)," *Electroencephalogr. Clin. Neurophysiol.*, vol. 43, no. 5, pp. 757–760, 1977.
[8] G. Pfurtscheller and W. Klimesch, "Functional topography during a visuoverbal judgment task studied with event-related desynchronization mapping," *J. Clin. Neurophysiol.*, vol. 9, no. 1, pp. 120–131, 1992.
[9] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, Dec. 2000.
[10] N. E. Huang, Z. Shen, S. R. Long, M. L. Wu, H. H. Shih, Z. Quanan, N. C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. R. Soc. A*, vol. 454, no. 1971, pp. 903–995, 1998.
[11] C. Park, D. Looney, P. Kidmose, M. Ungstrup, and D. P. Mandic, "Time-frequency analysis of EEG asymmetry using bivariate empirical mode decomposition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 19, no. 4, pp. 366–373, Aug. 2011.
[12] Z. Wang, A. Maier, N. K. Logothetis, and H. Liang, "Single-trial classification of bistable perception by integrating empirical mode decomposition, clustering, and support vector machine," *EURASIP J. Adv. Signal Process.*, vol. 2008, no. 119, pp. 1–8, 2008.
[13] C. M. Sweeney-Reed and S. J. Nasuto, "A novel approach to the detection of synchronisation in EEG based on empirical mode decomposition," *J. Computat. Neurosci.*, vol. 23, no. 1, pp. 79–111, 2007.
[14] D. Looney, L. Li, T. Rutkowski, D. P. Mandic, and A. Cichocki, "Ocular artifacts removal from EEG using EMD," in *Proc. 1st Int. Conf. Cognitive Neurodynam.*, 2007, pp. 831–835.
[15] T. Rutkowski, D. P. Mandic, A. Cichocki, and A. Przybyszewski, "EMD approach to multichannel EEG data analysis—The amplitude and phase components clustering analysis," *J. Circuits Syst. Comput.*, vol. 19, no. 1, pp. 215–229, 2010.
[16] D. Looney and D. P. Mandic, "Multi-scale image fusion using complex extensions of EMD," *IEEE Trans. Signal Process.*, vol. 57, no. 4, pp. 1626–1630, Apr. 2009.
[17] D. Looney, C. Park, P. Kidmose, M. Ungstrup, and D. P. Mandic, "Measuring phase synchrony using complex extensions of EMD," in *Proc. IEEE Workshop Stat. Signal Process.*, 2009, pp. 49–52.
[18] G. Rilling, P. Flandrin, and P. Gonçalves, "On empirical mode decomposition and its algorithms," in *Proc. IEEE-EURASIP Workshop Nonlinear Signal Image Process.*, 2003, vol. 3, pp. 8–11.
[19] L. Cohen, "Instantaneous anything," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 1993, vol. 5, pp. 105–108.
[20] Z. Wu and N. E. Huang, Ensemble empirical mode decomposition: A noise-assisted data analysis method Center for Ocean-Land-Atmosphere Studies, Tech. Rep, 2004.
[21] Z. Wu, N. E. Huang, and X. Chen, "The multi-dimensional ensemble empirical mode decomposition method," *Adv. Adaptive Data Anal.*, vol. 1, no. 3, pp. 339–372, 2009.
[22] M. U. B. Altaf, T. Gautama, T. Tanaka, and D. P. Mandic, "Rotation invariant complex empirical mode decomposition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2007, vol. 3, pp. 1009–1012.
[23] T. Tanaka and D. P. Mandic, "Complex empirical mode decomposition," *IEEE Signal Process. Lett.*, vol. 14, no. 2, pp. 101–104, Feb. 2007.
[24] G. Rilling, P. Flandrin, P. Gonçalves, and J. M. Lilly, "Bivariate empirical mode decomposition," *IEEE Signal Process. Lett.*, vol. 14, no. 12, pp. 1–4, Dec. 2007.
[25] N. Rehman and D. P. Mandic, "Multivariate empirical mode decomposition," *Proc. R. Soc. A*, vol. 466, no. 2117, pp. 1291–1302, 2010.
[26] J. Cui and W. Freeden, "Equidistribution on the sphere," *J. Sci. Comput.*, vol. 18, no. 2, pp. 595–609, 1997.
[27] N. E. Huang, M. L. Wu, S. R. Long, S. S. P. Shen, W. Qu, P. Gloersen, and K. L. Fan, "A confidence limit for the empirical mode decomposition and Hilbert spectral analysis," *Proc. R. Soc. A*, vol. 459, no. 2037, pp. 2317–2345, 2003.
[28] N. Rehman and D. P. Mandic, "Filter bank property of multivariate empirical mode decomposition," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2421–2426, May 2011.
[29] D. P. Mandic, M. Golz, A. Kuh, and D. Obradovic, *Signal Processing Techniques for Knowledge Extraction and Information Fusion*, T. Tanaka, Ed. New York: Springer, 2008.
[30] B. Weng and K. E. Barner, "Optimal and bidirectional optimal empirical mode decomposition," in *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Process.*, 2007, vol. III, pp. 1501–1504.
[31] B. Blankertz, G. Dornhege, M. Krauledat, K. Muller, and G. Curio, "The non-invasive berlin brain- computer interface: Fast acquisition of effective performance in untrained subjects," *NeuroImage*, vol. 37, no. 2, pp. 539–550, 2007.
[32] H. H. Jasper, "The ten-twenty electrode system of the international, federation," *Electroencephalogr. Clin. Neurophysiol.*, vol. 10, pp. 371–375, 1958.
[33] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: A general purpose brain-computer interface (BCI) system," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1034–1043, Jun. 2004.
[34] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C.-K. Peng, and H. stanley, "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, pp. E215–220, 2000.
[35] I. Daubechies and S. Maes, *Wavelets in Medicine and Biology*, A. Aldroubi and M. Unser, Eds. Boca Raton, FL: CRC Press, 1996.
[36] I. Daubechies, J. Lu, and H. T. Wu, "Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool," *Appl. Computat. Harmonic Anal.*, vol. 30, no. 2, pp. 243–261, 2011.
[37] E. Brevdo, N. S. Fuckar, G. Thakur, and H. Wu, "The synchrosqueezing algorithm: A robust analysis tool for signals with time-varying spectrum," *CoRR*, 2011.
[38] S. Makeig, "Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones," *Electroencephalogr. Clin. Neurophysiol.*, vol. 86, pp. 283–293, 1993.
[39] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clin. Neurophysiol.*, vol. 110, no. 5, pp. 787–798, 1999.
[40] H. Drucker, C. J. C. Burges, L. Kaufman, A. J. Smola, and V. Vapnik, *Support vector regression machines*, M. C. Mozer, M. I. Jordan, and T. Petsche, Eds. Cambridge, MA: MIT Press, 1997.
[41] S. Canu, Y. Grandvalet, V. Guigue, and A. Rakotomamonjy, "SVM and kernel methods MATLAB toolbox," in *Perception Syst. et Inform., INSA de Rouen*, Rouen, France, 2005.
[42] G. R. Müller-Putz, R. Scherer, C. Brunner, R. Leeb, and G. Pfurtscheller, "Better than random? a closer look on BCI results," *Int. J. Bioelectromagnet.*, vol. 10, no. 1, pp. 52–55, 2008.

Cheolsoo Park received the B.Eng. degree in electronic engineering from Sogang University, Seoul, South Korea, the M.Sc. degree in biomedical engineering department from Seoul National University, Seoul, South Korea, and the Ph.D. degree in adaptive nonlinear signal processing from Imperial College London, London, U.K., in 2012. Currently, he is working as a postdoctoral researcher in University California, San Diego.

His research interests are mainly in the area of brain–computer interface, time-frequency analysis, and statistical signal processing.

David Looney received the B.Eng. degree in electronic engineering from University College Dublin, Dublin, Ireland. In 2011, he received the Ph.D. degree in signal processing from Imperial College, London, U.K., where he is currently a Research Associate.

His research interests are mainly in the areas of data fusion, exploratory data analysis, and wearable solutions for health monitoring.

Naveed ur Rehman received the Ph.D. degree in nonlinear data driven time-frequency algorithms from Imperial College London, London, U.K., in 2011.

Currently, he is working as an Assistant Professor in the Department of Electrical and Electronic Engineering, COMSATS Institute of Information Technology Islamabad, Pakistan. His research interests include the design and applications of data-driven multivariate (multi-channel) time-frequency algorithms, and nonlinear time-series methods.

Alireza Ahrabian received the M.Eng degree in electrical and electronic engineering from Imperial College London, London, U.K., in 2010. Currently he is a research student working towards the Ph.D. degree in signal processing.

His research interests include time-frequency methods, brain–computer interface systems, and statistical signal processing.

Danilo P. Mandic is a Professor of signal processing at Imperial College London, London, U.K. He has been working in the area of nonlinear and multivariate adaptive signal processing and nonlinear dynamics. He has published two research monographs titled Recurrent Neural Networks for Prediction and Complex Valued Nonlinear Adaptive Filters. He has been a guest Professor at KU Leuven, Belgium, a frontier researcher at RIKEN Japan.

Dr. Mandic has been a member of the IEEE Technical Committee on Signal Processing Theory and Methods, and an Associate Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING and IEEE TRANSACTIONS ON NEURAL NETWORKS. He has won best paper awards in the area of neurotechnology in ICIC'09 and ISNN'10, and his work on Ear-EEG has been nominated for the Annual BCI Award in 2012.