SUBSPACE DENOISING OF EEG ARTEFACTS VIA MULTIVARIATE EMD

David Looney, Valentin Goverdovsky*, Preben Kidmose[†] and Danilo P. Mandic**

*Imperial College London, U.K.   [†]Aarhus University, Denmark

ABSTRACT

The components obtained using the time-frequency algorithm empirical mode decomposition (EMD) enable unique advantages in the context of noise removal. In this paper, recent EMD-based denoising methods are reviewed and similarities with a more conventional class of techniques – subspace denoising – are illustrated. Standard subspace approaches which are based on the factorisation of covariance matrices are unsuitable for nonstationary data. By comparison, EMD facilitates a signal representation which enables denoising using short spatio/temporal windows. It is highlighted how the EMD property of local orthogonality can be extended via multivariate operations, and a denoising scheme is proposed and compared with standard methods in electroencephalogram (EEG) artefact-removal using a novel multimodal sensor.

Index Terms— empirical mode decomposition, electroencephalography, signal denoising, multimodal sensors

1. INTRODUCTION

The miniaturisation of recording platforms and battery technology has greatly expanded the opportunities for electroencephalogram (EEG) monitoring in real-world environments (wearable EEG [1]) critical for e.g. monitoring seizures and diagnosing sleep disorders. Despite advances, EEG artefacts are still a major problem and cause significant signal degradation making real-time analysis in uncontrolled environments challenging. A primary source of artefacts is motion between the skin and the electrode caused by e.g. user movement. Potentials from motion artefacts are often much larger in amplitude than the observable potentials from the brain.

Methods that perform artefact removal primarily cater for scenarios where many channels of EEG are recorded at multiple locations across the scalp [2, 3, 4]. In this way, there are sufficient data statistics available to identify and discard the artefact using e.g. wavelet decompositions and independent component analysis [4]. However, recent advances in wearable EEG [5, 6] have highlighted the practical appeal of platforms which utilise only a single (or small number of) electrode(s). To this end, we have developed a multimodal sensor that records both electrical and mechanical activity from the same site. The mechanical output is designed to detect motion only at the location of the sensor; this enables unique insight into any interference within the electrical output (voltage fluctuations proportional to skin-electrode motion). We seek a denoising scheme that is capable of performing single-channel EEG denoising, obtained via the electrical modality of the sensor, based on the mechanical modality.

The concept of subspace denoising is ideally suited to the problem at hand; the goal is to perform a decomposition of the single-channel electrical component, or some representation of the signal, which enables a separation of the desired EEG and the electrical interference or noise. Those decomposition components spanning the noise subspace may be determined via the mechanical output of the sensor. Subspace approaches based on well-known matrix factorisation techniques are the most established, and are widely used in speech processing applications in particular [7]. The disadvantage of standard techniques is that they are unsuitable for nonstationary signals as they are typically based on analysis of the covariance matrices and require a sufficient number of data samples to operate.

The components generated via empirical mode decomposition (EMD), originally developed for time-frequency analysis [8], pose a number of advantages for denoising such as local orthogonality and a suitability for nonstationary data. In this paper, we highlight some similarities between EMD denoising schemes and conventional subspace methods. It is shown how a property unique to multivariate extensions of EMD enables a novel denoising algorithm which can outperform standard subspace methods, and is suitable for emerging multimodal sensor technologies.

2. A NOVEL MULTIMODAL SENSOR

We have developed a novel sensor which is capable of generating two outputs of different modalities from the same location. This was achieved by insulating an inertia-free mechanical transducer – a condenser microphone (Sonion 9723 GX) – with a thin layer of sealant, and depositing a thin layer of conductive material over the sound hole of the microphone. In this way, the conductive material acts as an electrode for measuring electrical potential (modality a), while the transducer component measures motion between the electrode and skin (modality b). Compared to recent work on EEG denoising with the aid of an accelerometer [9], the inertia-free nature of the transducer means that it only detects activity directly linked to electrode/skin motions and disregards large-scale whole-body movements which do not cause artefacts.

Preliminary recordings demonstrate that the mechanical component is sensitive to motion artefacts caused by subject movement and muscle activity. Fig. 1 shows the operation of the sensor, where the black line in the first panel is the electrical output with the sensor located on the forehead.[1] The electrical ground truth (desired EEG) was established by placing four standard electrodes around the sensor at equal distances (3 cm) and obtaining the averaged waveform,[2] this is plotted in the same panel as a red line. A motion interference was induced in the electrical output of the sensor from time 7 s by pushing the sensor against the skin (motion artefact). This interference is evident from the first panel in Fig. 1, where there is a clear difference between the electrical output and the ground truth from time 7 s, and in the second panel where the error signal is plotted. The mechanical output of the sensor is shown in the third panel; observe a clear similarity with the error (interference).

---

[1]The sensor was referenced to the left earlobe and the sampling frequency was 256 Hz.

[2]Each of the four standard electrodes were referenced to the right earlobe and the sampling frequency was 256 Hz.

3. PROBLEM FORMULATION

Consider recordings of the same process via two different modalities

$$\mathbf{y}_a = \mathbf{x}_a + \mathbf{v}_a \tag{1}$$

$$\mathbf{y}_b = \mathbf{x}_b + \mathbf{v}_b \tag{2}$$

where $\mathbf{y}_a$, $\mathbf{x}_a$ and $\mathbf{v}_a$ are $L \times 1$ column vectors which denote respectively the recorded signal, the desired signal and the interference signal for modality a. Similarly $\mathbf{y}_b$, $\mathbf{x}_b$ and $\mathbf{v}_b$ are the respective signals for modality b. In the considered scenario, modality a reflects electrical activity and b mechanical activity. Specifically, $\mathbf{y}_a$ denotes the electrical signal obtained from the scalp via a bipolar recording (difference in potential between two electrodes). The desired signal $\mathbf{x}_a$ is the EEG and the interference $\mathbf{v}_a$ reflects any underlying voltage fluctuations caused by motion between the electrode and skin. The signal $\mathbf{y}_b$ is of a different modality (mechanical) but obtained at the same location as the compromised electrode. We assume that there exists a component in the recording ($\mathbf{x}_b$) which is approximately equal, after linear scaling, to the interference within the electrical modality ($\mathbf{v}_a$) as well as some additional mechanical noise component ($\mathbf{v}_b$). In summary, we assume that $\mathbf{v}_a \approx \alpha \mathbf{x}_b$ where α is a scalar. If the mechanical noise is low ($E\{\mathbf{x}_b^T \mathbf{x}_b\} > E\{\mathbf{v}_b^T \mathbf{v}_b\}$, where $E\{\cdot\}$ is the expected value and assuming zero mean) then $\mathbf{v}_a \approx \alpha \mathbf{y}_b$. In the sequel the algorithms are derived on the assumption that $\alpha = 1$, that is, it is assumed that some prior knowledge exists to scale the b modality appropriately before denoising.
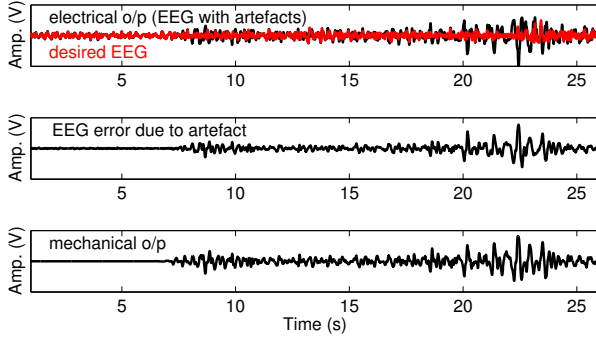

Fig. 1. Operation of multimodal sensor. The electrical output diverges from desired signal (panels 1,2) with motion-interference from 7 s which is also reflected by the mechanical output (panel 3).

3.1. Subspace Denoising

Subspace methods rely on a decomposition of the noisy signal, or some representation of the signal (covariance matrix), to obtain a subspace that is primarily occupied by the desired/clean signal and a subspace that is primarily occupied by the noise/interference signal; the desired signal can thus be obtained by nulling the noise subspace component. We let

$$\widehat{\mathbf{x}}_a = f_{\mathbf{y}_b}(\mathbf{y}_a) \tag{3}$$

denote an estimate of $\mathbf{x}_a$, the desired signal from modality a, where $f_{\mathbf{y}_b}(\mathbf{y}_a)$ is a linear operation on $\mathbf{y}_a$, or some decomposition of $\mathbf{y}_a$, based on knowledge of $\mathbf{y}_b$ using the same decomposition method.[3] The method must be appropriate for different noise types (coloured, white) and for data nonstationarity.

---

[3]The mechanical component is not treated as a regressor (i.e. enabling the mixing of electrical and mechanical components directly) as (i) they are of different modalities (the mechanical modality has not yet been fully characterised) and (ii) the mechanical component may exhibit additional noise.

4. CLASSIC SUBSPACE APPROACHES

The most established subspace methods are those based on matrix factorisation techniques, in particular singular value decomposition (SVD) or eigenvector-eigenvalue decomposition (EVD), and we here consider an approach based on the EVD of the signal covariance matrices. Typically, a matrix W is sought which performs linear estimation of the desired signal as follows

$$\widehat{\mathbf{x}}_a = f_{\mathbf{y}_b}(\mathbf{y}_a) = W \mathbf{y}_a. \tag{4}$$

The error of the estimator is given by $\mathbf{e} = \widehat{\mathbf{x}}_a - \mathbf{x}_a = W \mathbf{y}_a - \mathbf{x}_a$ and the well-known solution (Wiener) which minimises the mean square error ($E\{\mathbf{e}^T \mathbf{e}\}$) is given by [7]

$$W_{opt} = R_{\mathbf{x}_a}(R_{\mathbf{x}_a} + R_{\mathbf{y}_b})^{-1} \tag{5}$$

where $R_{\mathbf{x}_a}$ denotes the covariance matrix of $\mathbf{x}_a$ (i.e. $R_{\mathbf{x}_a} = E\{\mathbf{x}_a \mathbf{x}_a^T\}$) and $R_{\mathbf{y}_b}$ the covariance matrix of $\mathbf{y}_b$. Recall that unlike conventional approaches developed for single modalities [7], we are assuming that $\mathbf{y}_b$ approximates the interference in modality a ($R_{\mathbf{y}_b} \approx R_{\mathbf{v}_a}$). Of course, $R_{\mathbf{x}_a}$ is not known in practice and various algorithms have been proposed which seek to exploit properties of the covariance matrices ($R_{\mathbf{y}_a} = R_{\mathbf{x}_a} + R_{\mathbf{y}_b}$). One solution to (5) suitable for coloured noise is based on the observation that $R_{\mathbf{x}_a}$ and $R_{\mathbf{y}_b}$ can be diagonalised in the following way [10, 11]

$$V^T R_{\mathbf{x}_a} V = \Delta, \quad V^T R_{\mathbf{y}_b} V = I \tag{6}$$

where V and Δ are the eigenvector and eigenvalue matrices for $\Sigma = (R_{\mathbf{y}_b})^{-1} R_{\mathbf{x}_a} = (R_{\mathbf{y}_b})^{-1}(R_{\mathbf{y}_a} - R_{\mathbf{y}_b})$ respectively and I is the identity matrix. Thus, the optimal estimator can be rewritten as [11]

$$W_{opt} = V^{-T} \Delta (\Delta + I)^{-1} V^T \tag{7}$$

where the eigenvalues in Δ corresponding to eigenvectors in V spanning the noise subspace have been set to zero.[4] The above algorithm is typically applied in an overlapping window fashion (window size N). That is, the matrix W_{opt} is calculated from the P-order covariance matrices estimated for $\mathbf{y}_a$ and $\mathbf{y}_b$ within the window, and overlapping estimates of $\mathbf{x}_a$ within sub-windows (length P) are combined. Larger values for N and P enable the observation of more statistics about the signals and better estimates of covariance matrices, but hinder performance when the data is nonstationary.

5. EMD-BASED DENOISING APPROACHES

5.1. The EMD Algorithm and Multivariate Extensions

The univariate EMD algorithm [8] was proposed as a means of enabling highly localised time-frequency analysis of nonlinear and nonstationary signals. Unlike projection-based schemes (e.g. Fourier) which must sacrifice time localisation for frequency localisation, or vice versa, the Hilbert transform obtains a physically meaningful estimate of frequency at each time instant (instantaneous frequency) if the input satisfies narrow-band criteria. The principle of the EMD operation is to decompose any vector into a set of AM/FM components which satisfy such criteria. These components, called intrinsic mode functions (IMFs), represent the oscillation modes embedded in the data. The EMD algorithm decomposes an input as $\mathbf{x} = \sum_{i=1}^{M} X_i$, where $\{X_i, i = 1, \ldots, M\}$

---

[4]The solution in eqn. (7) is a realisation of the time-domain-constrained algorithm (with $\mu = 1$) described in [11], assuming $R_{\mathbf{y}_b} \approx R_{\mathbf{v}_a}$.

denotes the IMFs (the column vectors of the IMF matrix X) extracted via the sifting process (refer to [8]). The lowest index IMFs (e.g. X_1) contain the highest frequency dynamics and the higher index IMFs (e.g. X_M) contain the lowest frequency dynamics (trend). We define the IMF space as the vector space spanned by the IMFs.

5.1.1. *Multivariate EMD (MEMD) & Noise-Assisted MEMD*

A drawback associated with univariate EMD is that IMFs obtained for different signals with similar statistics may differ in their frequency content or even their number (problem of uniqueness [12]), this is a consequence of the data-driven nature of the algorithm and makes IMF-by-IMF comparisons unfeasible. Although ensemble EMD [13] alleviates the problem, this comes at the cost of great computational complexity. Instead, two or more signal components can be combined as a single bivariate/multivariate entity and decomposed using bivariate [14] or multivariate [15, 16] extensions of EMD; studies have shown [12, 16] how this enables mode alignment – the phenomenon whereby similar modes within multiple channels are aligned in frequency at the same IMF level. The operation can be further enhanced by placing white Gaussian noise (WGN) in adjoining channels and, owing to the mode alignment property of MEMD, the noise channels enforce dyadic structure within the signal channels (noise-assisted MEMD (NA-MEMD), refer to [17]).

5.2. Review of Existing EMD Denoising Methods

An approach that utilised the EMD for denoising was proposed in 2004 [18]. Following the behaviour of EMD as a dyadic filter bank for WGN [19, 20], and that the IMF log-variance follows a simple linear model controlled by the Hurst exponent, the method employed confidence intervals to discard IMFs dominated by noise. Although effective, the approach was not based on an optimality criterion. A solution which combined the IMFs to obtain the best estimate of the original signal in the least squares sense (Wiener solution) was proposed in 2007 [21]. This was followed by a method that combined the IMFs in a more adaptive fashion based on a standard adaptive filtering architecture in 2008 [22]. It was shown that, within the IMF domain, short window sizes were sufficient to obtain high performances. Based on wavelet-thresholding algorithms and EMD filterbank properties for WGN, a class of soft and hard thresholding approaches were developed for EMD which are suitable for WGN scenarios in 2009 [23]. In recent years EMD has been combined with conventional denoising methods. For instance it was used to generate, from a univariate signal, a set of components to perform independent component analysis (ICA) [24] and canonical component analysis (CCA) [25] to enhance the performance of the algorithms for single-channel scenarios in EEG artefact removal. In 2012, it was shown how MEMD could be used to identify and discard IMFs dominated by electrooculogram (EOG) artefacts [26]. It was illustrated in [27] how coloured noises, particularly ones which are difficult to isolate using standard subspace methods (pink, babble), are better separated using EMD in the context of speech denoising.

5.3. MEMD-Based Subspace Denoising

Denoising approaches based on EMD share some similarities with standard subspace algorithms. Just as a subset of the eigenvectors obtained from the covariance matrices are assumed to span the desired signal subspace, it can be assumed (under some conditions) that a subset of the IMFs span a similar space. A difference is that the eigenvectors can be orthogonal (e.g. if obtained from a real

and symmetric matrix), while the IMFs are approximately orthogonal over short spatio/temporal windows (local orthogonality). The EMD method was designed to obtain a sparse time-frequency signal representation so that the number of IMF components is typically very small compared to the signal length. Given the small number of IMFs and the property of local orthogonality, it is expected that the method will perform denoising optimally over short windows. In summary, the IMFs obtained via standard EMD enable the following advantages for denoising:

- **Suitability for nonstationary data**. By design, the method caters for data nonstationarity. This was highlighted in [22] where it was shown that optimal prediction and denoising in the IMF space can be achieved over short temporal windows.
- **Local orthogonality**. While the IMFs are not strictly orthogonal, they reflect data-driven independent frequency components and are locally orthogonal [8].
- **Physical interpretability**. Standard subspace methods have an established relationship with standard spectral estimation theory, but each of the IMF components has a well defined interpretation based on instantaneous frequency – a framework more suitable for nonlinear and nonstationary signal content [8].

Consider two signals $(\mathbf{x}, \mathbf{y})$, in which the IMFs have been obtained via MEMD (or NA-MEMD) $\{X_i, Y_i, i = 1, \ldots, M\}$, we can now revise the local orthogonality property as follows:

- **Multivariate local orthogonality**. Exploiting the mode alignment property of MEMD (and NA-MEMD), we can state that an independence exists, in the EMD sense, not only between the IMFs of $\mathbf{x}$, but also between IMFs of $\mathbf{x}$ and $\mathbf{y}$. Letting $X_i \perp_\mathsf{L} Y_j$ denote the estimation of the local orthogonality between X_i and Y_j we can state that $X_i \perp_\mathsf{L} Y_j = 0, \forall\, i, j = 1, \ldots, M (i \neq j)$.

5.3.1. *The Proposed Algorithm*

Let Y_a and Y_b denote respectively the $L \times M$ IMF matrices for $\mathbf{y}_a$ and $\mathbf{y}_b$ obtained via NA-MEMD, and X_a denote the (unknown) IMF matrix for $\mathbf{x}_a$. We propose the following scheme to estimate $\mathbf{x}_a$

$$\widehat{\mathbf{x}}_a = f_{\mathbf{y}_b}(\mathbf{y}_a) = Y_a \mathbf{w} \tag{8}$$

where $\mathbf{w} = [w_1, \ldots, w_M]^T$ denotes a $M \times 1$ vector. Due to the property of multivariate local orthogonality, a solution for each IMF level can be sought independently. That is, we can assume that the overall error can be minimised by solving for $\mathbf{e}_i = Y_{ai} w_i - X_{ai} = Y_{ai} w_i - (Y_{ai} - Y_{bi})$, $i = 1, \ldots, M$. We seek the solution which minimises the mean square error ($E\{\mathbf{e}_i^T \mathbf{e}_i\}$). Taking the derivative of $E\{\mathbf{e}_i^T \mathbf{e}_i\}$ with respect to w_i and setting it to zero gives

$$2\left[w_i(Y_{ai}^T Y_{ai}) - ((Y_{ai}^T Y_{ai}) - (Y_{ai}^T Y_{bi})) \right] = 0. \tag{9}$$

Rearranging and applying constraints gives the solution

$$w_i = \min\left[\frac{\max\left[(Y_{ai}^T Y_{ai}) - (Y_{ai}^T Y_{bi}), 0\right]}{(Y_{ai}^T Y_{ai})}, 1 \right]. \tag{10}$$

The weights w_i in (10) are continuous in the range $[0, 1]$, we also consider a binary weighting scheme $\tilde{\mathbf{w}} = [\tilde{w}_1, \ldots, \tilde{w}_M]^T$ where

$$\tilde{w}_i = \begin{cases} 1 & \text{if } w_i \geq \delta \\ 0 & \text{if } w_i < \delta \end{cases} \tag{11}$$

where δ is some threshold. In simulations the proposed algorithm, with both weighting schemes, was applied in a similar fashion to conventional subspace methods. A single NA-MEMD operation was applied to $\mathbf{y}_a$ and $\mathbf{y}_b$, but within the resulting IMF space $\widehat{\mathbf{x}}_a$ was calculated in an overlapping window fashion (window size N).

A key difference between the proposed algorithm and previous EMD-based methods is how it exploits the property of multivariate local orthogonality, essential for IMF-by-IMF comparisons between two or more signals. Unlike other works which rely on this property, such as the binary weighting scheme in [26], the proposed approach is based on an optimality criterion and enables both binary and continuous weighting schemes. Also, unlike schemes which utilise the properties of EMD to condition the input to other algorithms and thus enhance their performance, e.g. ICA [24] or CCA [25], the aim here is to explore the denoising advantages of the multivariate IMF space via a conceptually simple algorithm.

6. SIMULATION RESULTS

6.1. Removal of Simulated Artefacts

To recreate noisy EEG, segments of clean EEG were linearly mixed with artefacts. Ten EEG signals of length $50\,\mathrm{s}$ were obtained from an electrode placed at a frontal position of the head (AFz referenced to mastoid) while the subject remained still and made no eye movements or blinks, see first panel of Fig. 2. Separately, artefacts were recorded (vertical EOG), and were used to generate random artefact signals with some constraints (e.g. no overlapping of artifact segments), see second panel of Fig. 2. The artefact signals were linearly mixed with a randomly selected segment of EEG to obtain compromised EEG signals where the ground truth was known.[5]
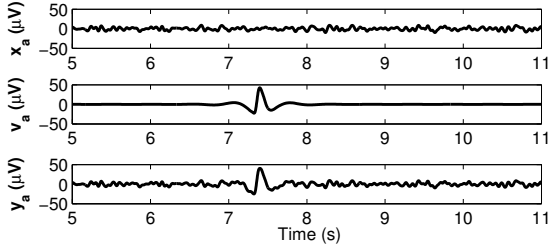
Fig. 2. Clean EEG recording (panel 1), an artificial interference signal generated from recorded EOG artefacts (panel 2), a mixture of the EEG and EOG (panel 3).

In the considered context, $\mathbf{x}_a$ is the clean EEG signal and $\mathbf{v}_a$ is the artefact signal (see Fig. 2). We approximate the mechanical modality by letting $\mathbf{x}_b = \mathbf{v}_a$ and approximate mechanical noise by letting $\mathbf{v}_b$ be WGN. Both the standard (EVD-based) and MEMD approaches were applied to estimate $\mathbf{x}_a$ based on the overlapping window scheme described previously, with varying levels of simulated mechanical noise set by the signal-to-noise-ratio[6] (SNR). The mean squared error (MSE) for the EVD and MEMD approach with continuous weights, based on (10), for high SNRs were similar, both outperforming the MEMD method with binary weights based on (11); however, the MEMD approaches were found to perform optimally for smaller values of N. The MSE for a low mechanical SNR of $-2\,\mathrm{dB}$, averaged over 50 realisations for $\mathbf{y}_a$, is plotted in Fig. 3, showing how the MEMD approaches achieved better performances and with smaller window sizes (N). The better performance of the

---

[5]Both the EEG and artefact signals were bandpass filtered within the range $1-30\,\mathrm{Hz}$ (standard EEG range of interest) before mixing.

[6]The mechanical SNR is given by $10\log_{10}\sum_n (x_b(n))^2 / \sum_n (v_b(n))^2$ where $\mathbf{v}_b$ is WGN, and $x_b(n)$ and $v_b(n)$, $n = 1,\ldots,L$, denote the elements of $\mathbf{x}_b$ and $\mathbf{v}_b$ respectively.

MEMD approaches for low SNRs is as expected as the weights are based on the cross-correlation of the modalities (within the IMF domain), a feature for which the EVD approach does not account. The results also show that the more restricted form of the MEMD denoising approach using binary weights was again outperformed by the one with continuous weights. This indicates that some IMFs span both the signal and noise space and simply nulling MEMD components is not sufficient for optimal denoising.
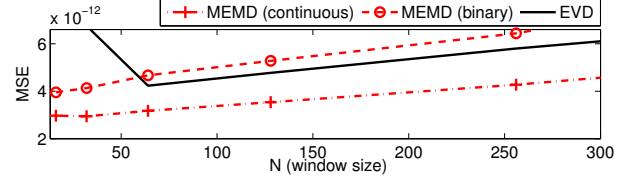
Fig. 3. Removal of simulated artefacts. The MSE of the EVD- and MEMD-based algorithms for different window sizes (N).

6.2. Removal of Motion Artefacts Using Multimodal Sensor

EEG artefact removal was performed for a recording obtained using the multimodal sensor, similar to that depicted in Fig. 1. Prior knowledge of the variance of $\mathbf{x}_a$ was assumed to enable the correct scaling of the mechanical component (α) before applying each of the algorithms. For a $60\,\mathrm{s}$ recording, which contained approximately $30\,\mathrm{s}$ of motion artefacts, the EVD and MEMD (continuous weights only) approaches achieved a similar MSE, segments of the denoised signals are shown in Fig. 4. Crucially, optimal performance was achieved for the EVD approach with $N = 128$ compared with $N = 32$ for the MEMD approach – highlighting its appeal in nonstationary analysis. This result conforms with those of previous studies for univariate EMD [22].
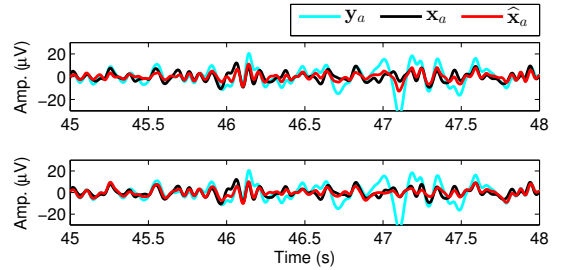
Fig. 4. Removal of motion artefacts using multimodal sensor. For a compromised EEG recording obtained using the electrical component ($\mathbf{y}_a$), the upper and lower panels show, respectively, the artefact-free estimates using the EVD- and MEMD-based (continuous weights) approaches (denoted by $\widehat{\mathbf{x}}_a$).

7. CONCLUSION

We propose a method based on multivariate extensions of empirical mode decomposition (MEMD) which utilises observations of an interference process to perform denoising of physiological data. Compared with a standard subspace algorithm, the proposed scheme can better cater for nonstationarity and for conditions where the interference observation is itself compromised. Its operation is illustrated for the removal of simulated and real-world motion artefacts based on EEG recordings obtained from a novel multimodal sensor.

8. REFERENCES

[1] A. J. Casson, D. Yates, S. Smith, J. S. Duncan, and E. Rodriguez-Villegas, "Wearable electroencephalography," *IEEE Engineering in Medicine and Biology Magazine*, vol. 29, no. 3, pp. 44–56, 2010.

[2] K. H. Ting, P. C. W. Fung, C. Q. Chang, and F. H. Y. Chan, "Automatic correction of artifact from single-trial event-related potentials by blind source separation using second order statistics only," *Medical Engineering & Physics*, vol. 28, no. 8, pp. 780 – 794, 2006.

[3] Y. Li, Z. Ma, W. Lu, and Y. Li, "Automatic removal of the eye blink artifact from EEG using an ICA-based template matching approach," *Physiological Measurement*, vol. 27, no. 4, pp. 425–436, 2006.

[4] N. Mammone, F. La Foresta, and F. C. Morabito, "Automatic artifact rejection from multichannel scalp EEG by wavelet ICA," *IEEE Sensors Journal*, vol. 12, no. 3, pp. 533–542, 2012.

[5] D. Looney, P. Kidmose, C Park, M. Ungstrup, M. L. Rank, K. Rosenkranz, and D. P. Mandic, "The in-the-ear recording concept: User-centered and wearable brain monitoring," *IEEE Pulse*, vol. 3, no. 6, pp. 32–42, 2012.

[6] P. Kidmose, D. Looney, M. Ungstrup, M. L. Rank, and D. P. Mandic, "A study of evoked potentials from ear-EEG," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, pp. 2824–2830, 2013.

[7] P. C. Loizou, *Speech Enhancement: Theory and Practice (Signal Processing and Communications)*, CRC, 1 edition, 2007.

[8] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis," *Proceedings of the Royal Society A*, vol. 454, pp. 903–995, 1998.

[9] I. Daly, M. Billinger, R. Scherer, and G. Muller-Putz, "On the automated removal of artifacts related to head movement from the EEG," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 21, no. 3, pp. 427–434, 2013.

[10] S. B. Searle, *Matrix Algebra Useful for Statistics*, New York: Wiley, 1982.

[11] Y. Hu and P. C. Loizou, "A generalized subspace approach for enhancing speech corrupted by colored noise," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 4, pp. 334–341, 2003.

[12] D. Looney and D. P. Mandic, "Multiscale image fusion using complex extensions of EMD," *IEEE Transactions on Signal Processing*, vol. 57, no. 4, pp. 1626–1630, 2009.

[13] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method," Tech. Rep. 193, Center for Ocean-Land-Atmosphere Studies, 2004.

[14] G. Rilling, P. Flandrin, P. Goncalves, and J. M. Lilly, "Bivariate empirical mode decomposition," *IEEE Signal Processing Letters*, vol. 14, no. 12, pp. 936–939, 2007.

[15] N. Rehman and D. P. Mandic, "Empirical mode decomposition for trivariate signals," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1059–1068, 2010.

[16] N. Rehman and D. P. Mandic, "Multivariate empirical mode decomposition," *Proceedings of the Royal Society A*, vol. 466, no. 2117, pp. 1291–1302, 2010.

[17] N. Rehman, C. Park, N. E. Huang, and D. P. Mandic, "EMD via MEMD: Multivariate noise-aided computation of standard EMD," *Advances in Adaptive Data Analysis*, vol. 5, no. 2, pp. 1350007, 2013.

[18] P. Flandrin, P. Goncalves, and G. Rilling, "Detrending and de-noising with empirical mode decompositions," in *Proceedings of the European signal processing conference (EUSIPCO)*, 2004, vol. 2, pp. 1581–1584.

[19] Z. Wu and N. E. Huang, "A study of the characteristics of white noise using the empirical mode decomposition method," *Royal Society of London Proceedings Series A*, vol. 460, pp. 1597–1611, 2004.

[20] P. Flandrin, G. Rilling, and P. Goncalves, "Empirical mode decomposition as a filter bank," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 112–114, 2004.

[21] B. Weng and K. E. Barner, "Optimal and bidirectional optimal empirical mode decomposition," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2007, vol. 3, pp. 1501–1504.

[22] D. Looney and D. P. Mandic, "A machine learning enhanced empirical mode decomposition," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2008, pp. 1897–1900.

[23] Y. Kopsinis and S. McLaughlin, "Development of EMD-based denoising methods inspired by wavelet thresholding," *IEEE Transactions on Signal Processing*, vol. 57, no. 4, pp. 1351–1362, 2009.

[24] B. Mijovic, M. De Vos, I. Gligorijevic, J. Taelman, and S. Van Huffel, "Source separation from single-channel recordings by combining empirical-mode decomposition and independent component analysis," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 9, pp. 2188–2196, 2010.

[25] K. T. Sweeney, S. F. McLoone, and T. E. Ward, "The use of ensemble empirical mode decomposition with canonical correlation analysis as a novel artifact removal technique," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 1, pp. 97–105, 2013.

[26] M. K. I. Molla, T. Tanaka, and T. M. Rutkowski, "Multivariate EMD based approach to EOG artifacts separation from EEG," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 653–656.

[27] M. Koh and E. Rodriguez-Marek, "Speech enhancement of color noise using empirical mode decomposition," in *Proceedings of the Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, 2012, pp. 1688–1692.