35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Application of a Variation of Empirical Mode Decomposition and Teager Energy Operator to EEG Signals for Mental Task Classification

M. Kaleem, A. Guergachi, S. Krishnan

Abstract— **This paper presents a simple and effective methodology for mental task classification using a novel variation of the empirical mode decomposition (EMD) algorithm and the Teager energy operator applied to electroencephalography (EEG) signals. EEG signals corresponding to various types of mental tasks performed by human subjects are decomposed using the variation of EMD, called Empirical Mode Decomposition-Modified Peak Selection (EMD-MPS), which allows direct separation of the signals into a de-trended component, and a trend, according to a frequency separation criterion. Teager energy operator is then applied to calculate the average energy values of both components obtained after signal decomposition using EMD-MPS. These energy values are used to construct feature vectors, and one-versus-one classification of mental tasks is performed using a simple classifier, namely the 1-NN classifier. An average correct classification rate of 87% is obtained, improving on previous results and thereby also demonstrating the effectiveness of the methodology.**

I. INTRODUCTION

In the context of various actions that can be taken by a brain-computer interface, such as computer cursor movement or selection of a letter for a typing task, identification of patterns in EEG signals corresponding to various types of mental tasks performed by human subjects can have important applications [1]. Given the non-stationary and non-linear nature of EEG signals, signal processing techniques such as empirical mode decomposition (EMD) [2] lend themselves well to EEG signal analysis, including analysis for mental task classification [3][4][5].

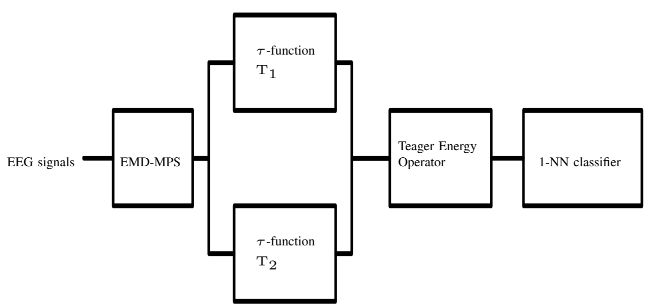

Fig. 1. Schematic representation of the methodology described in this paper.

The EMD algorithm decomposes a signal into intrinsic mode functions (IMFs), which are obtained by adaptive extraction of all the oscillatory modes present in a signal using a process called sifting [2]. The sifting process considers the oscillatory modes in the signal at the most local time-scale, which is defined by two consecutive extrema (peaks). Hence identification of all the extrema in the signal is an important part of the sifting process. This also means that a change in the choice of extrema will result in limiting the time-scale over which the sifting process allows an oscillatory mode in the signal to pass un-decomposed.

Exploiting the idea of selective extrema selection, we previously proposed a modification to the EMD algorithm [6], which we now call empirical mode decomposition-modified peak selection (EMD-MPS) [7]. In the EMD-MPS method, the sifting process uses intelligent peak selection in short-time windows of length τ. Based on different values of τ, different decompositions of a signal into what we term as τ-functions are possible. Therefore the short-time window acts as an operator which allows separation of different frequency components in a signal into τ-functions, as determined by the length τ of the short-time window. We have previously established a relation between the frequency components decomposed and the value of τ, and have shown that using an appropriate selection of values of τ allows a novel time-scale based de-trending of signals [6][7].

Another useful tool for non-stationary signal analysis comes in the form of the Teager-Kaiser energy operator (TEO) [8]. The TEO is a non-linear operator that can be used for energy estimation of amplitude-frequency (AM-FM) modulated representations of a non-stationary signal. There is precedent of using EMD in conjunction with the TEO for non-stationary signal analysis, also in context of mental task classification [4].

In this paper, we present a novel method for mental task classification based on application of EMD-MPS and TEO on EEG signals. EMD-MPS is used to decompose the EEG signals into two τ-functions representing the trend and a de-trended component. TEO is applied to both τ-functions to obtain values of the average energy, which are used to form the feature vectors. One-versus-one classification of mental tasks using these feature vectors is performed with a 1-NN classifier [9]. The overall methodology is represented in the diagram shown in Fig. 1.

The novel method presented in this paper achieves an average correct classification rate of 87% for a one-versus-one classification scheme, improving on previous results using similar methodologies on the same EEG signals [3][4]. At the same time, this method is characterized by the simplicity of the decomposition, feature extraction, as well as classification. The next sections will provide more details

M. Kaleem (corresponding author) and S. Krishnan are with the department of Electrical Engineering, Ryerson University, Toronto, Canada. m2kaleem@ryerson.ca

A. Guergachi is with the Ted Rogers School of Information Technology Management, Ryerson University, Canada.

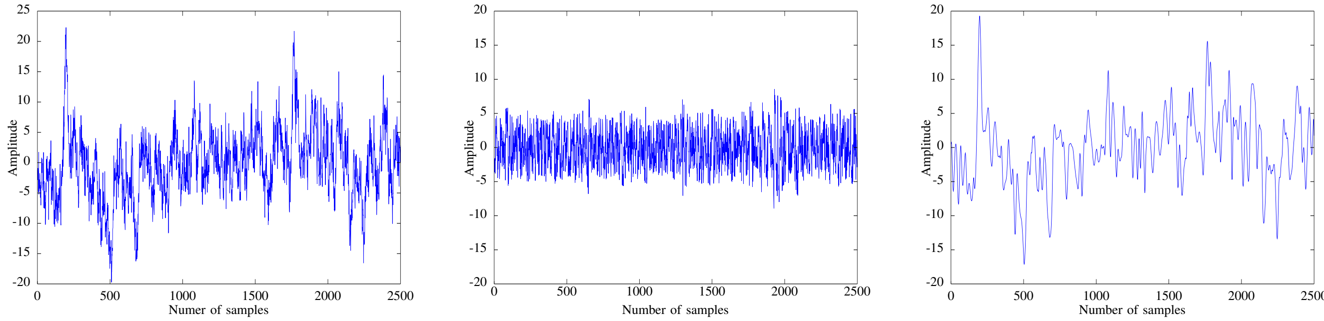

Fig. 2. (Left) EEG signal from Subject 1 performing Task 2. (Middle) τ-function T_1, representing de-trended signal. (Right) τ-function T_2, representing the trend. These τ-functions have been obtained using a value of τ=31.25 ($\hat{\tau}$=14), corresponding to a frequency separation value of F=8 Hz.

of the proposed method.

II. Empirical Mode Decomposition-Modified Peak Selection

EMD-MPS uses the sifting process to decompose a signal. However, a criterion for choosing the extrema based on short-time windows of length τ is used, instead of using a time-scale based on successive extrema, as is done in the case of EMD. Let us define an operator $W_i^\tau(\cdot)$, $i = 1...k, i \in \mathbb{Z}$, $0 < \tau < L, L \in \mathbb{R}$, which, given a signal $x[n]$ of length L, produces the i-th τ-function T_i, such that $\mathrm{T}_i[n] = W_i^\tau(x[n])$. This can be explained as:

1) For a given signal $x[n]$, a short-time window length denoted by τ is chosen, and for each interval τ over the whole signal, the highest/lowest from among the maxima and minima within τ are selected.
2) The upper and lower envelopes, $E_{n(U)}$ and $E_{n(L)}$, are calculated by using a cubic spline to connect all the maxima/minima identified (one maxima/minima or peak per τ).
3) The mean $E_{n(mean)}$ of the upper and lower envelopes is calculated, and $x[n]$ is updated by subtracting the mean from it $x[n] \leftarrow x[n] - E_{n(mean)}$.

Similar to EMD-based sifting, step 3 is continued till a stopping criterion is met, at which point $x[n]$ is reduced to a τ-function. The τ-function is subtracted from $x[n]$ to get a residue, which is then taken as the starting point instead of $x[n]$, and previous steps of the algorithm are repeated to find all the τ-functions T_i in the signal. Unlike IMFs extracted by the EMD algorithm, the coarse-grained τ-functions may contain different coexisting modes of oscillation, each superimposed on the other. This happens since the short-time window τ sets an upper limit on the periods of the oscillations that can be included in any given τ-function obtained using the EMD-MPS method. This limit is determined by :

$$F = \frac{F_s}{\tau} \tag{1}$$

where F_s represents the sampling frequency.

As an example for this relation, a value of τ=25 (in samples) corresponds to a frequency value F=40 samples/second for F_s=1000 samples/second. Using this value of τ, only one peak (maxima and minima each) in each 25 sample interval will be used in the envelope formation, and the sifting process should then decompose all $F \leq 40$ samples/second oscillatory components, and let all components with $F > 40$ samples/second pass through un-decomposed in one τ-function. Due to the non-linear nature of decomposition and mode-mixing phenomenon [2], the frequency separation does not represent a sharp cut-off. Additionally, in practice the value of τ is qualified by a scaling constant k, such that $\hat{\tau} = k\tau$, and $0 < k \leq 1$. The relation in Eq. 1 and the scaling constant k have been empirically validated using fractional Gaussian noise in our previous works [7].

III. Method

The next subsections will present different aspects of the methodology for mental task classification.

A. Data

The data used in this paper is in the form of EEG signals recorded on 7 subjects, who performed five tasks each. These tasks were 1) *baseline task*, which required subjects to relax and think of nothing in particular; 2) *multiplication task*, involving mental multiplication of 2 multi-digit numbers; 3) *letter task*, in which subjects mentally composed a letter without vocalizing; 4) *rotation task*, requiring subjects to visualize a three dimensional object being rotated; finally, 5) *counting task*, in which subjects imagined a blackboard with numbers being written on it. The subjects performed the tasks in ten trials each over two days, and for each trial, EEG was recorded from six electrodes at positions $(C_3, C_4, P_3, P_4, O_1, O_2)$ for ten seconds at a frequency of 250 samples per second. For the experiments in this paper, five trials of each task performed by all subjects were used. Further details about the data are available in [1], and the data is also available for public download[1].

B. Decomposition using EMD-MPS

EMD-MPS is used for a time-scale based de-trending of EEG signals, such that each EEG signals is decomposed into two τ-functions, one representing the de-trended signal containing the higher frequency components, and the other

[1] http://www.cs.colostate.edu/eeg/

representing the low frequency trend of the signal. This is done by appropriate selection of a value of τ according to Eq. 1, and does not require estimation of a trend model for model-based de-trending, or knowledge of the statistical properties of IMFs, as is the case of EMD-based de-trending approaches proposed in literature, e.g. [10][11].

For de-trending of EEG signals used in this study, we used two different values of τ, to obtain two different sets of τ-functions, containing frequency components separated according to Eq. 1. The first value of τ corresponds to a frequency value F=8 Hz, such that the τ-function T_1 contains frequency components with frequency values greater than 8 Hz, and the τ-function T_2 represents the trend containing frequencies lower than 8 Hz. Using Eq. 1 and sampling frequency value $F_s = 250$ Hz, the value of τ obtained is given by τ=31.25. However, for decomposition, we have to use the value $\hat{\tau}$, which is τ scaled by a constant k as described in Section II. A good estimate for the value of k is given by $k \approx 0.44$ [6][7], such that τ=31.25 corresponds to $\hat{\tau}$=14. Similarly, the other set of τ-functions are obtained corresponding to F=4 Hz, such that τ=62.5, with the corresponding $\hat{\tau}$=28.

All EEG signals used in this study were decomposed using EMD-MPS using these two values of $\hat{\tau}$. In this regard, Fig. 2 shows an example EEG signal, and the τ-functions T_1 and T_2 obtained with a value of $\hat{\tau} = 14$.

C. Feature Extraction and Classification

After the decomposition of EEG signals (6 EEG signals per subject per task, with 5 trials of each task), the TEO was applied to each of the two τ-functions obtained per EEG signal to estimate the average Teager energy of each τ-function.

The TEO was developed from the point of view of the energy required to generate a signal. This non-linear energy-tracking operator Ψ is given in its discrete form [8] as: $\Psi(x[n]) = x^2[n] - x[n+1] \cdot x[n-1]$.

The TEO is nearly instantaneous, given that only three samples are required for computing the energy at a given time instant. This also makes the operator easy to implement efficiently. The average Teager energy, e_i, for a τ-function T_i is calculated as:

$$e_i = \frac{1}{N} \sum_{n=1}^{N} \mid \Psi[\mathrm{T}_i(n)] \mid \tag{2}$$

where N is equal to the number of samples in the τ-functions, $\Psi[\cdot]$ is the discrete-time TEO and $i = 1, 2$.

The average Teager energy e_i is calculated for each τ-function according to Eq. 2. This way, for each EEG signal, we have two values of e_i corresponding to each τ-function T_i. Given that 6 EEG signals per task are used, the feature vector contains 12 elements consisting of the average Teager energy values. This is a significant reduction in feature vector dimension compared to our previous work [4], and other related works [3][5].

The feature vectors obtained for the tasks were used in a one-versus-one classification scheme using a 1-NN classifier. The classification accuracy for each task combination was estimated using the tenfold cross-validation method. The classification results are discussed in the next section.

IV. Results

The classification results obtained with the methodology presented in this paper are shown in Table I. The classification accuracy for different task combinations listed in Table I has been obtained using τ-functions extracted with both values of $\hat{\tau}$ used in the analysis. It can be seen from Table I that the classification accuracy for most of the task combinations is greater than 80%, with many task combinations having a classification accuracy of 100%. This is true for both values of $\hat{\tau}$ used for decomposition using EMD-MPS. However, for subjects 3 and 6, the average classification accuracy for all task combinations is considerably higher for a value of $\hat{\tau}$=14, compared to a value of $\hat{\tau}$=28. For other subjects, the difference in average classification accuracy over all tasks using both values of $\hat{\tau}$ is not large. Overall, the average classification accuracy obtained with $\hat{\tau}$=14 is 86.8%, which is about 3% better than that obtained with $\hat{\tau}$=28, and is higher than previously reported works [3][4][5].

The average classification accuracy for subject 3 is relatively low, having a value of 70%. This is due to three task-combinations having a low classification accuracy of 50%. For these task combinations, the following approach was used as a remedial measure. For both values of $\hat{\tau}$, the feature vectors obtained from τ-functions T_1 and T_2 were used independently for classification. For the *math-count* task combination, the classification accuracy increased from 50% to 80% when only the feature vector obtained from T_2 extracted with $\hat{\tau}$=14 was used. There was no change in the classification accuracy for the other two tasks, using either τ-function obtained with either value of $\hat{\tau}$. Therefore, decomposition was performed using a lower value of $\hat{\tau}$=7, which corresponds to a frequency separation value of F=16 Hz. For this value of $\hat{\tau}$, there was no change in classification accuracy for the task combinations *rot-count* and *letter-count* using feature vectors from both T_1 and T_2 together, or the feature vector from T_2 individually. However, using only the feature vector obtained from T_1, the classification accuracy increased to 70% and 60% for the *rot-count* and *letter-count* task combinations respectively.

This shows that for *rot-count* and *letter-count* task combinations in case of subject 3, the de-trended component T_1 containing the higher frequency components of the signals leads to better classification accuracy, whereas the lower frequency signal trend contained in T_2 is more relevant for the *math-count* task combination. Importantly, the used τ-functions in this case were obtained with different values of $\hat{\tau}$. This also demonstrates the flexibility of the method, whereby decomposition can be adapted to different time-scales, and either of the de-trended component, or the trend, can be used to extract discriminatory features for classification.

TABLE I

CLASSIFICATION ACCURACY FOR ONE-VERSUS-ONE CLASSIFICATION OF MENTAL TASKS FOR ALL SUBJECTS USING 1-NN CLASSIFIER AND 10-FOLD CROSS-VALIDATION

Task combination	Subject 1		Subject 2		Subject 3		Subject 4		Subject 5		Subject 6		Subject 7		Mean	
	$\hat{\tau}$=14	$\hat{\tau}$=28	$\hat{\tau}$=14	$\hat{\tau}$=28	$\hat{\tau}$=14	$\hat{\tau}$=28	$\hat{\tau}$=14	$\hat{\tau}$=28	$\hat{\tau}$=14	$\hat{\tau}$=28	$\hat{\tau}$=14	$\hat{\tau}$=28	$\hat{\tau}$=14	$\hat{\tau}$=28	$\hat{\tau}$=14	$\hat{\tau}$=28
base-count	100%	100%	90%	90%	80%	60 %	100%	100%	70%	70%	90%	80%	90%	90%	**88.6%**	**84.3%**
base-letter	90%	90%	100%	100%	90%	90%	90%	100%	90%	70%	90%	50%	90%	70%	**91.4%**	**81.4%**
base-math	90%	100%	80%	90%	90%	70%	80%	100%	70%	80%	90%	90%	80%	90%	**82.8%**	**88.6%**
base-rot	100%	100%	90%	100%	60%	60%	80%	80%	80%	80%	90%	100%	100%	100%	**85.7%**	**88.6%**
letter-count	100%	100%	60%	90%	50%	50%	100%	100%	70%	70%	90%	80%	100%	80%	**81.4%**	**81.4%**
letter-rot	100%	100%	100%	100%	60%	70%	80%	60%	90%	90%	100%	90%	100%	100%	**90%**	**87.1%**
math-count	100%	100%	70%	60%	50%	50%	80%	70%	60%	60%	100%	100%	100%	100%	**80%**	**77.1%**
math-letter	100%	100%	100%	100%	80%	80%	100%	90%	100%	100%	100%	100%	100%	100%	**97.1%**	**95.7%**
math-rot	100%	100%	90%	90%	90%	60%	80%	60%	80%	90%	100%	70%	80%	90%	**88.6%**	**80%**
rot-count	80%	70%	100%	80%	50%	40%	80%	90%	70%	70%	100%	100%	100%	100%	**82.8%**	**78.6%**
Average	**96%**	**96%**	**88%**	**90%**	**70%**	**63%**	**87%**	**85%**	**78%**	**78%**	**95%**	**86%**	**94%**	**92%**	**86.8%**	**84.3%**

Tasks: base=baseline; math=multiplication; rot=rotation; count=counting;

The average classification accuracy of nearly 87% obtained in this work improves on the classification accuracy of 85% reported in our previous work [4]. However, the methodology presented in this paper (denoted by **A**) improves on our previous approach (denoted by **B**) in the following ways:

1) **B** uses EMD to decompose the EEG signals into at most $\log_2(N)$ IMFs [2], where N is the length of the signals. On the other hand, **A** uses EMD-MPS to decompose the signals into two τ-functions, which makes the decomposition computationally more efficient.
2) **B** obtains the de-trended part of the signal and the signal trend by partial reconstruction based on checking a criterion. This criterion needs to be checked for every set of signals, and will change for different sets, e.g. if signal lengths are different. **A**, however, uses time-scale based de-trending of signals, which separates the trend according to a frequency separation criterion. Not needing to check a criterion, and not having to perform partial reconstruction after decomposition, makes **A** more computationally efficient and, importantly, portable to different signal types.
3) **B** uses two feature vectors of dimensions 30 and 18, whereas both feature vectors used in **A** have dimension 12. The improvement in classification accuracy in **A** is obtained with a reduced dimension feature vector.
4) Importantly, **A** allows flexibility in feature extraction, by allowing different time-scale based decompositions, to deal with difficult classification cases, as demonstrated earlier in this section. This is a unique feature of the methodology.

V. CONCLUSIONS

This paper presented a simple and effective method for mental task classification using a novel variation of empirical mode decomposition in conjunction with the Teager energy operator. Our method is distinguished by the simplicity and flexibility of the decomposition, as well as the low dimensions of the feature vector. The efficacy of the feature vectors obtained is demonstrated by the high classification accuracy for one-versus-one mental task classification using a simple classifier. The average classification accuracy obtained is higher than previously reported results using similar approaches on the same data.

REFERENCES

[1] C. W. Anderson and J. A. Bratman, "Translating Thoughts Into Actions by Finding Patterns in Brainwave", *Fourteenth Yale Workshop on Adaptive and Learning Systems*, Yale University, New Haven, CT, USA, June 2008, pp. 1-6.

[2] 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 nonlinear and non-stationary time series analysis, *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, Vol. 454, Issue. 1971, March 1998, pp. 903-995.

[3] P.F. Diez, V. Mut, E. Laciar, A. Torres and E. Avila, "Application of the empirical mode decomposition to the extraction of features from EEG signals for mental task classification", *31st Annual International Conference of the IEEE EMBS*, Minneapolis, Minnesota, USA, September 2-6, 2009, pp. 2579-2582.

[4] M.F. Kaleem, L. Sugavaneswaran, A. Guergachi and S. Krishnan, "Application of Empirical Mode Decomposition and Teager Energy Operator to EEG Signals for Mental Task Classification", *32nd Annual International Conference of the IEEE EMBS*, Buenos Aires, Brazil, Aug. 31 - Sept. 4, 2010, pp. 4590-4593.

[5] C. Park, D. Looney, P. Kidmose, M. Ungstrup and D.P. Mandic, "Time-Frequency Analysis of EEG Asymmetry Using Bivariate Empirical Mode Decomposition", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 19, Issue 4, 2011, pp. 366-373.

[6] M.F. Kaleem, A. E. Cetin, A. Guergachi and S. Krishnan, "Using a Variation of Empirical Mode Decomposition To Remove Noise From Signals", *Proceedings of the 21st International Conference on Noise and Fluctuations (ICNF)*, Toronto, Canada, 2011, pp. 123-126.

[7] M.F. Kaleem, A. Guergachi and S. Krishnan, "A Variation of Empirical Mode Decomposition with Intelligent Peak Selection in Short Time Windows", *Accepted for publication in Proceedings of 38th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vancouver, Canada, 2013.

[8] P. Maragos, J.F. Kaiser, and T.F. Quatieri, "On amplitude and frequency demodulation using energy operators", *IEEE Transactions on Signal Processing*, Vol. 41, Issue. 4, 1993, pp. 1532-1550.

[9] R.O. Duda and P.E. Hart and D.G. Stork, *Pattern Classification*, John Wiley and Sons, Inc., Second Edition, Ch. 4.

[10] P. Flandrin, P. Goncalves and G. Rilling, "Detrending and denoising with empirical mode decompositions", *12th European Signal Processing Conference*, Vienna, Austria, September 6-11, 2004, pp. 1581-1584.

[11] A. Moghtaderi, P. Borgnat and P. Flandrin, "Trend Extraction for Seasonal Time Series Using Ensemble Empirical Mode Decomposition", *Advances in Adaptive Data Analysis*, Vol. 3, Issue. 1 & 2, 2011, pp. 41-61.