

“© 2017 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

# EEG Emotion Recognition Using Reduced Channel Wavelet Entropy and Average Wavelet Coefficient Features with Normal Mutual Information Method

Henry Candra\*, Mitchell Yuwono, Rifai Chai, Hung T. Nguyen, and Steven Su

**Abstract**—Recognizing emotion from EEG signals is a complicated task that requires complex features and a substantial number of EEG channels. Simple algorithms to analyse the feature and reduce the EEG channel number will give an indispensable advantages. Therefore, this study explores a combination of wavelet entropy and average wavelet coefficient (WEAVE) as a potential EEG-emotion feature to classify valence and arousal emotions with the advantage of the ability to identify the occurrence of a pattern while at the same time identify the shape of a pattern in EEG emotion signal. The complexity of the feature was reduced using the Normalized Mutual Information (NMI) method to obtain a reduced number of channels. Classification with the WEAVE feature achieved 76.8% accuracy for valence and 74.3% for arousal emotion, respectively. The analysis with NMI shows that the WEAVE feature has linear characteristics and offers possibilities to reduce the EEG channels to a certain number. Further analysis also reveals that detection of valence emotion with reduced EEG channels has a different combination of EEG channels compared to arousal emotion.

## I. INTRODUCTION

Recognizing emotion from physiological signals mainly with electroencephalography (EEG) has gained attention by researchers in recent years [1]. Optimally identifying emotions from EEG signals requires a few important factors including features and number of channels used.

Many features have been developed [2], among the features, wavelet is most widely used for EEG signal analysis. This relates to the advantages of wavelet features including [3]:

- The wavelet transform (WT) is both a band-pass filter and a denoiser for decomposing and isolating EEG signals to obtain desired subbands, for example, extracting only alpha, beta, and gamma frequencies as they are more related to emotion elicitation [4].
- Wavelet entropy estimates inter-segment regularity. It can be used to identify the occurrence of a pattern.
- Wavelet approximation coefficient is a proxy for identifying the shape of a pattern.

\*Henry Candra is with the Electrical Engineering Department, Faculty of Industrial Technology, Trisakti University, Jakarta, Indonesia, and Centre for Health Technologies, Faculty of Engineering and Information Technology, University of Technology, Sydney, New South Wales, Australia (henry\_candra@yahoo.com). Mitchell Yuwono, Rifai Chai, Hung. T. Nguyen, and Steven W. Su are with the Centre for Health Technologies, Faculty of Engineering and Information Technology, University of Technology, Sydney, New South Wales, Australia (mitchellyuwono@gmail.com, [Rifai.Chai, Hung.Nguyen, Steven.Su]@uts.edu.au).

Another implementation of wavelet features is using wavelet coefficient average value to predict the area in the human brain associated with motor control using Near-Infrared Spectroscopy (fNIRS). The wavelet coefficient average value provides the magnitude property of each EEG channel [5].

Using many numbers of EEG channels initiate complex features. Therefore, efficient channel reduction algorithm is needed to simplify the features while also reduce the computation complexity and increase performance by reducing unnecessary channels [6].

Each application applies different approaches to channel reduction; for example, in EEG-emotion classification, asymmetric variance ratio and amplitude asymmetric ratio were used to assess channel locations associated with a certain emotion to reduce the channels from 63 to 4 using fuzzy c-means clustering, [6].

Mutual Information has been applied for image registration over the decades based on the marginal and joint entropies [7]. However, its implementation has been applied in data mining to group similar data records [8]. Recently, Normalized Mutual Information (NMI) has been proposed for feature selection with the advantage of reduced complexity of features [9]. The implementation of NMI algorithm in EEG signal processing offers the possibilities to measure the relation between features in EEG channels and their represented emotions [4].

In this paper, a combination of wavelet entropy and average wavelet coefficient called WEAVE is proposed as an EEG-emotion feature to classify low/high state of valence and arousal emotions with the advantage of the ability to identify the occurrence of a pattern while at the same time identify the shape of a pattern in EEG emotion signal. Further investigation with NMI algorithm was then conducted with the main objective to analyse and reduce the number of EEG channels. First the relation between NMI values and the features (channels) was analysed, and then channel reduction was conducted based on the NMI analysis. The effect of channel reduction on the classification results with Support Vector Machine (SVM) algorithm was also discussed together with the channel location related to valence and arousal emotions.

## II. MATERIALS AND METHOD OVERVIEW

### A. Materials

The experiment was conducted with publicly available Dataset for Emotion Analysis using electroencephalogram,

Physiological and Video Signals (DEAP) [10]. This dataset consist of 32 channels EEG-emotion signals collected from 32 participants while watching 40 video clips. The participants assessed and rated the videos using standardized Self-Assessment Manikins (SAM) [11] which has the disadvantage of a difficulty to relate the SAM pictorial score to the emotional state so that a rather similar EEG pattern may be translated into very contradicting emotions which leads to degradation of the classification reliability [12]. Therefore, to reduce the mistranslation effect, a subject grouping was conducted by applying the RBF kernel function [13] to calculate the transformation matrix of each person's EEG features related to each person's own emotion rating. The result is fitted as a logistic regression. The whole results of all participants were concatenated to form the matrices for clustering. These mapping matrices were clustered based on how closely related each person was to another and put in groups with dendrogram. A group of participants was then selected for the experiment.

### B. Method Overview

The block diagram of the proposed feature analysis and channels reduction with NMI is illustrated in Fig. 1. The proposed process consist of 5 main steps: 1. Segmentation of preprocessed EEG-emotion signals with 6 second window; 2. Wavelet features extraction and WEAVE formation; 3. NMI value computation and arrangement; 4. Feature analysis and channel reduction with NMI; 5. Classification with SVM.

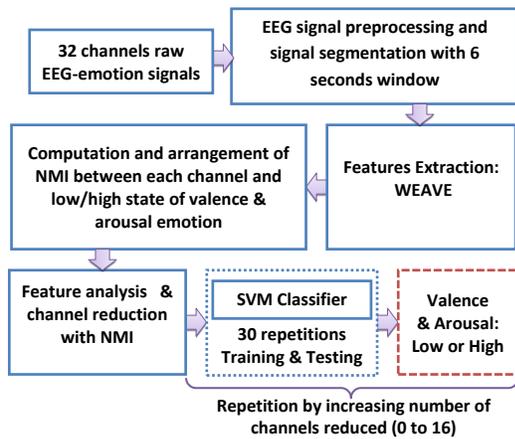


Fig. 1. The block diagram of proposed analysis and channel reduction with NMI.

## III. METHOD

### A. Segmentation of Preprocessed EEG-Emotion Signals with 6 second Window

First, the dataset was pre-processed as in [10] to remove artifacts.

Next, the preprocessed EEG-emotion signals were segmented with 6 second window as suggested in [14]. All the results were combined as one array. The procedure was conducted for each EEG channel.

### B. Wavelet Features Extraction and WEAVE formation

This process consists of 5 steps: 1) Computation of wavelet coefficients; 2) Wavelet entropy calculation; 3) Wavelet coefficients average value computation; 4) WEAVE formation in 3 bands; 5) Array formation of 32 channels WEAVE in 3 bands. The detail is as follows:

1) *Computation of wavelet coefficient*: Using dyadic scales Discrete Wavelet Transforms (DWT) can be computed as [14]:

$$\text{DWT}(m, n) = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{2^m}} \psi\left(\frac{t - 2^m n}{2^m}\right) dt \quad (1)$$

where  $2^m$  and  $2^m n$  are scale and time localization, and  $\psi(t)$  is the mother wavelet function. For multi resolution signal decomposition DWT, the signal  $y(t)$  is passed through a low pass filter (LPF) and high pass filter (HPF) and down sampled by 2. LPF output gives approximation  $A_m$  between band frequency  $f_k/2$  to  $f_k$ , whilst HPF output provides detail  $D_m$  between band frequency  $f_k/2$  to  $f_k$ , with  $f_s$  as frequency subband of sampling frequency of the original signal and  $p$  as decomposition level,  $f_k = f_s/2^{(1+p)}$ .

2) *Wavelet entropy calculation*: Daubechies 5 (db5) of wavelet function was selected in this experiment [14]. Decomposition of EEG signals in 5 bands with sampling rate 128 sample/sec delivers:  $\delta$  (3-4Hz);  $\theta$  (5-8Hz);  $\alpha$  (9-16Hz);  $\beta$  (17-32Hz); and  $\gamma$  (33-64Hz). The Energy of wavelet coefficients both approximation and detail can be represented as:

$$E_n = \sum_{m=1}^N |C_{m,n}|^2 \quad (2)$$

where  $C_{m,n}$  is the approximation/detail,  $N$  is the number of wavelet coefficients at each decomposition level. The energy of wavelet coefficient is compared to total energy of signal to obtain probability mass function. The relative wavelet entropy is then calculated with:

$$G(n) = -\frac{E(n)}{\sum_{m=1}^N E(n)} \log \frac{E(n)}{\sum_{m=1}^N E(n)} \quad (3)$$

3) *Wavelet coefficients average value computation*: Take wavelet approximation  $A_m(k)$ , the wavelet coefficient average value  $\mu_m$  can be calculated with:

$$\mu_m = \frac{1}{N} \sum_{i=1}^N A_m(k) \quad (4)$$

where  $N$  is the number of coefficients of each wavelet approximation with  $i = 1, \dots, N$ .

4) *WEAVE formation in 3 bands*: To form the WEAVE features, the wavelet entropy  $G(n)$  and wavelet coefficient average value  $\mu_m$  both were combined as one array according to the order of the EEG channels. Following the recommendation in [4] only 3 subbands were employed utilizing only  $\alpha$ ,  $\beta$ , and  $\gamma$  bands to reduce the length of array which also an initial reduction of features complexity.

5) *Array formation of 32 channels WEAVE in 3 bands:* All the WEAVE features from each channel were then arranged as one long array of 32 channels. Each array represents one segment of EEG-emotion signals.

### C. Mutual Information (MI) and Normalized Mutual Information (NMI)

Mutual Information (MI) is the measurement of mutual dependence (amount of information) between a random variable X towards another random variable Y. This measurement is an indication of coherence between two distributions that generate the variables (vectors). The MI can be defined as [7]:

$$MI(X, Y) = H(X) + H(Y) - H(X, Y) \quad (5)$$

The MI is normalized (NMI) to obtain a value between 0 (independence) and 1 (strong dependence) with the equation being rewritten as [15]:

$$NMI(X, Y) = \frac{MI(X, Y)}{\sqrt{H(X)H(Y)}} \quad (6)$$

### D. Classification with SVM

Multi-class SVM with a Radial Basis Function (RBF) kernel is applied for the EEG-emotion classification. Using Ensemble Rapid Centroid Estimation (ERCE) [16], [17] the kernel radius  $R_{SVM}$  is estimated from the training data with the estimation process as in [4]. Sequential Minimal Optimization (SMO) algorithm is implemented in the classification process to train the SVM [18].

## IV. RESULTS AND DISCUSSION

The NMI values of the WEAVE features are computed as the representation of Low/High state of valence/arousal emotion. The representation is shown in Fig. 2.

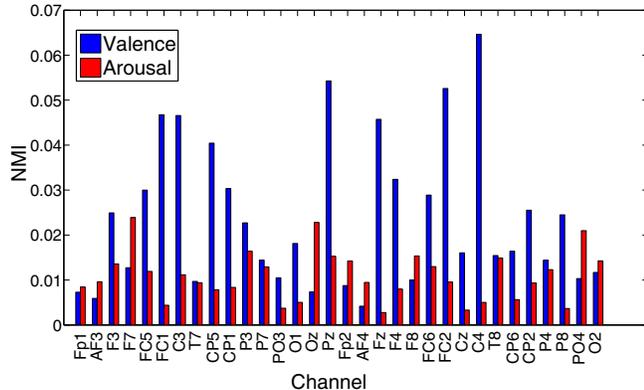


Fig. 2. The NMI values of Low/High state of valence and arousal emotion for each individual EEG channel.

In Fig. 2 almost all NMI magnitudes of valence are greater than arousal emotion in every channel which indicates that valence emotion is represented by the features better than arousal emotion.

The graph in Fig. 2 is split into two separated graphs. The NMI values were also rearranged from minimum to the maximum as shown in Fig. 3.

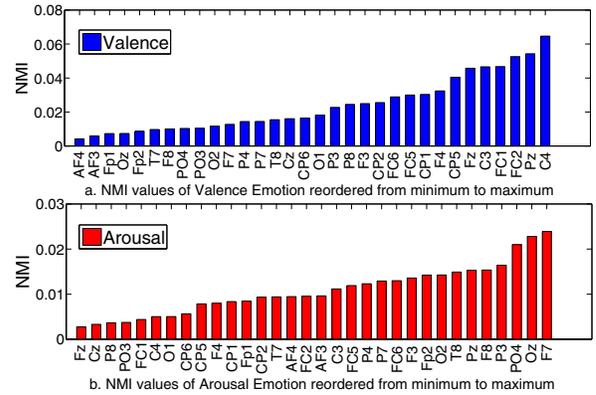


Fig. 3. NMI graphs of valence and arousal emotion with rearranged NMI values from minimum to maximum.

Fig. 3 reveals 2 main aspects. First, the difference between the maximum and minimum NMI value for valence (0.07 - 0.005 = 0.065) is greater than arousal (0.02 - 0.005 = 0.015). Therefore, valence has a steeper graph than arousal. Second, valence and arousal emotion each has a different sequence of EEG channels corresponding to the NMI values. Both aspects contribute significant effects in the channel reduction and classification results as discussed below.

Channel reduction was then conducted by following the sequence of the NMI values for each emotion. We began without channel reduction, continued with channel reduction by removing the channel with minimum NMI value. Each emotion has a different sequence of channel reduction. The reduction is limited to a minimum of 16 channels to maintain the reduction of classification results less than 8%. Each classification was repeated 30 times.

The classification results of valence and arousal emotion using the WEAVE feature with 32 EEG channels is compared to other researches using DEAP dataset with other features as listed in Table. I. The comparison shows that the WEAVE feature outperforms.

TABLE I  
COMPARISON OF CLASSIFICATION VALENCE & AROUSAL EMOTION WITH 32 EEG CHANNELS

Method	Feature	Accuracy $\pm$ SD (%)	
		Valence	Arousal
[19]	Spectral Power (Unsupervised)	70.9 $\pm$ 11.4	67.1 $\pm$ 14.2
[14]	3 Band Wavelet Entropy	65.1 $\pm$ 1.2	64.8 $\pm$ 1.2
This paper	WEAVE	<b>76.8 <math>\pm</math> 0.7</b>	<b>74.3 <math>\pm</math> 0.9</b>

Analysis of the classification results after channel reduction using NMI is illustrated in Fig. 4. The graph shows that valence has higher accuracy (76.8% to 73.5%) than arousal (74.3% to 68.6%). This result is supported by the fact that valence is represented better by the WEAVE feature compared to arousal as evidenced in the previous NMI analysis in Fig. 3. Furthermore, a channel with higher NMI value contains better feature representation and vice versa. When the channels with smaller NMI value were removed, the classification results of valence remained high, only about 3% degradation occurs. In contrast, the classification

results of arousal emotion were reduced more than 5%. However, because of the linear characteristic of WEAVE feature across all channels, substantial reduction of channels will be followed by the reduction of classification results for both valence and arousal emotions.

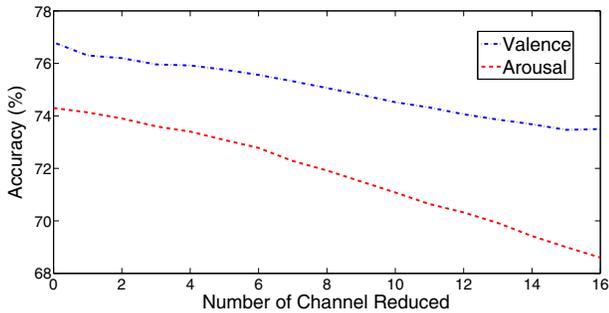


Fig. 4. Classification results of valence and arousal emotion after channel reduction using NMI method.

The location of 16 EEG channels for valence and arousal emotions after channel reduction using NMI method is illustrated in Fig. 5 which reveals that recognition of valence and arousal emotion involves a different combination of EEG channels. Valence emotion applies: F3, F4, Fz, FC1, FC2, FC5, C3, C4, CP1, CP2, CP5, P3, P8, Pz, O1 channels, which is more related to the middle left and right hemispheres. On the other hand, arousal emotion uses: Fp2, AF3, F3, F7, F8, FC5, FC6, C3, T8, P3, P4, P7, Pz, PO4, O2, Oz channels, which coincide with frontal and parietal lobes of the brain.

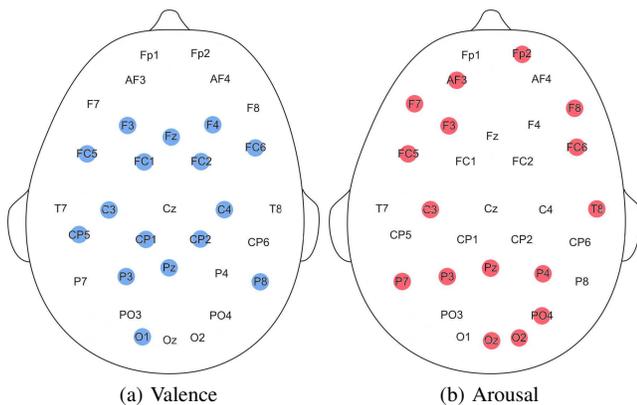


Fig. 5. Location of 16 EEG channels for valence and arousal emotion recognition after channel reduction with NMI.

## V. CONCLUSION AND FUTURE DIRECTION

EEG emotion recognition using WEAVE feature has been demonstrated with accuracy of 76.8% for valence and 74.3% for arousal. WEAVE has the advantage of the ability to identify the occurrence and the shape of a pattern in EEG signals. Feature analysis and channel reduction using the NMI algorithm has also been conducted. Using the NMI method, the EEG channels can be reduced from 32 to 16 with less than 8% reduction of accuracy. The use of 16

EEG channels reveals that valence is detected with a different combination of EEG channels compared to arousal. Future directions include the application of the NMI algorithm to examine other EEG-emotion features and implementation of reduced channels WEAVE for multimodal emotion analysis.

## REFERENCES

- [1] M. Soleymani, S. Asghari-Esfeden, Y. Fu, and M. Pantic, "Analysis of eeg signals and facial expressions for continuous emotion detection," *IEEE Trans. Affective Comput.*, vol. 7, no. 1, pp. 17–28, Jan 2016.
- [2] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from eeg," *IEEE Trans. Affective Comput.*, vol. 5, no. 3, pp. 1–1, 2014.
- [3] O. A. Rosso, S. Blanco, J. Yordanova, V. Kolev, A. Figliola, M. Schramm, and E. Baar, "Wavelet entropy: a new tool for analysis of short duration brain electrical signals," *Journal of Neuroscience Methods*, vol. 105, no. 1, pp. 65–75, 2001.
- [4] H. Candra, M. Yuwono, A. Handojoseno, R. Chai, S. Su, and H. T. Nguyen, "Recognizing emotions from eeg subbands using wavelet analysis," in *Proc. IEEE 37th Annu. Int. Conf. Eng. Med. Biol. Soc., Conference Proceedings*. IEEE, 2015, pp. 6030–6033.
- [5] T. N. Nguyen, T. H. Nguyen, and T. T. Nguyen, *Wavelet Coefficient Average Value for Prediction of Motor Control Area of Human Brain Using fNIRS*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, book section 4th International Conference on Biomedical Engineering in Vietnam, pp. 270–274.
- [6] T. Alotaiby, F. E. A. El-Samie, S. A. Alshebeili, and I. Ahmad, "A review of channel selection algorithms for eeg signal processing," *EURASIP Journal on Advances in Signal Processing*, vol. 2015, no. 1, pp. 1–21, 2015.
- [7] N. D. Cahill, "Normalized measures of mutual information with general definitions of entropy for multimodal image registration," in *Biomedical Image Registration*. Springer, 2010, pp. 258–268.
- [8] Z. He, X. Xu, and S. Deng, "k-anmi: A mutual information based clustering algorithm for categorical data," *Information Fusion*, vol. 9, no. 2, pp. 223–233, 2008.
- [9] P. A. Estévez, M. Tesmer, C. A. Perez, and J. M. Zurada, "Normalized mutual information feature selection," *IEEE Trans. Neural Netw.*, vol. 20, no. 2, pp. 189–201, 2009.
- [10] S. Koelstra, C. Muhl, M. Soleymani, L. Jong-Seok, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis using physiological signals," *IEEE Trans. Affective Comput.*, vol. 3, no. 1, pp. 18–31, 2012.
- [11] M. M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of behavior therapy and experimental psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.
- [12] M. Isomursu, M. Thi, S. Vinm, and K. Kuutti, "Experimental evaluation of five methods for collecting emotions in field settings with mobile applications," *Int. Journal of Human-Computer Studies*, vol. 65, no. 4, pp. 404–418, 2007.
- [13] J.-P. Vert, K. Tsuda, and B. Schölkopf, "A primer on kernel methods," *Kernel Methods in Computational Biology*, pp. 35–70, 2004.
- [14] H. Candra, M. Yuwono, R. Chai, A. Handojoseno, I. Elamvazuthi, H. T. Nguyen, and S. Su, "Investigation of window size in classification of eeg-emotion signal with wavelet entropy and support vector machine," in *Proc. IEEE 37th Annu. Int. Conf. Eng. Med. Biol. Soc., Conference Proceedings*. IEEE, 2015, pp. 7250–7253.
- [15] A. Strehl and J. Ghosh, "Cluster ensembles—a knowledge reuse framework for combining multiple partitions," *The Journal of Machine Learning Research*, vol. 3, pp. 583–617, 2003.
- [16] M. Yuwono, S. Su, B. Moulton, Y. Guo, and H. Nguyen, "An algorithm for scalable clustering: Ensemble rapid centroid estimation," in *IEEE Congress on Evolutionary Computation (CEC), 2014*, July 2014, pp. 1250–1257.
- [17] M. Yuwono, S. Su, B. Moulton, and H. Nguyen, "Data clustering using variants of rapid centroid estimation," *IEEE Trans. Evol. Comput.*, vol. 18, no. 3, pp. 366–377, June 2014.
- [18] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, May 2011.
- [19] X. Zhuang, V. Rozgi, and M. Crystal, "Compact unsupervised eeg response representation for emotion recognition," in *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, June 2014, pp. 736–739.