A Hybrid ICA - Wavelet Transform for Automated Artefact Removal in EEG-based Emotion Recognition

A. D. Bigirimana, N. Siddique, and D. Coyle
Intelligent Systems Research Centre
Ulster University
Londonderry/Derry, UK

***Abstract*— Removing artefacts from electroencephalographic (EEG) recordings normally increases their low signal-to-noise ratio and enables more reliable interpretation of brain activity. In this paper we present an evaluation of an automatic independent component analysis (ICA), a hybrid ICA - wavelet transform technique (ICA-W), for artefact removal from emotional-state correlated EEG. Frequency features in the brain rhythms and statistical measures were extracted to assess the performance of ICA-W against the regular ICA in terms of the accuracy of classifying emotional states from EEG. Support vector machines (SVM) were used to classify the emotional states, and the achieved accuracies on data from 14 subjects indicates that ICA-W performs better than traditional ICA in statistical and wavelet –based features from the brain rhythms whilst the best overall performance is achieved when combining ICA-W with statistical features, average accuracy of 73.50% for classifying four categories of emotion. ICA-W is therefore demonstrated to enhance EEG-based emotion recognition applications.**

Keywords—Independent component analysis, EEG, wavelet, emotion

I. INTRODUCTION

One of the current trends in EEG-based emotion recognition is to minimize the human intervention in the processing of the recorded EEG, automatizing emotion recognition systems as much as possible. This automatization should include noise and artefact removal techniques. The common artefacts in EEG include power line noise and physiological artefact mainly originating from heart activity – electrocardiography (ECG), eye movements or blinking – electrooculography (EOG), head and neck muscle activity – electromyography (EMG), and potentials from the brain (cephalic noise) not associated with the task [1]. Emotion related brain activity involves several processes, including processing the emotional stimulus, production of an affective state in response to the stimulus, and the regulation of the affective state [2]. With all these processes in the brain, additional artefacts can be detrimental to the discriminability of emotions from EEG.

Digital filters are widely used to reduce artefacts by extracting the relevant brain rhythms of interest, namely delta (0.1-4Hz), theta (4-8Hz), alpha (8-13Hz), beta (13-30Hz), and gamma (>30Hz), from the recorded data. Using appropriate bandpass filters artefacts can be reduced assuming the artefacts' frequencies do not overlap with the targeted band. EMG artefacts cover a wide frequency range (from 0 to >200Hz) and are predominantly present in EEG frequencies higher than 20Hz [2, 3], especially in the temporal electrodes. In some studies EOG artefacts are removed through adaptive filtering [5], however this method requires recording the EOG reference signal and may alter the non-artefactual EEG due to some EEG leaked in the artifact reference sensor. Other studies [6, 8] use independent component analysis (ICA), a blind source separation approach, to remove the EOG, EMG and ECG artefacts from EEG. ICA decomposes the signal into statistically independent components, and the artefactual components can be identified by visual inspection and removed. The cleaned signal is then obtained by recombining the remaining non-artefactual components.

Visually inspecting the independent components for artefacts requires some level of expertise, in identifying artefacts in the EEG signal, and can be time consuming when dealing with large datasets or data recorded from multiple subjects. To automate this inspection, various method including correlating the independent components with recorded reference artefacts [8] and thresholding based on high order statistics of the independent components [7] have been proposed. Castellanos et al [9] have proposed a wavelet transform based approach to filter out artefacts from ICA generated independent components, an approach that can be automated. Other automation of ICA artefacts removal include artefact detection approach based on joint spatial and temporal characteristics to identify artefactual components, a method also known as ADJUST [10]. In this paper we compare for the first time, the performance of artefact removal of standard ICA and the ICA-wavelet transform hybrid on emotion-related EEG signals classification. The paper is structured as follows: Section II presents the concept used in ICA-based noise removal, and section III describes the experimental setup and tools. The results are discussed in section IV, and section V concludes the paper.

II. ICA-BASED ARTEFACTS REMOVAL

The ICA algorithm assumes that the recorded EEG is a linear combination of temporal, independent and spatially fixed

signals, and the algorithm estimates these independent components. The artefactual components are removed, and the remaining components are recombined into cleaned EEG data. This approach has been effective in removing noise and artefacts from EEG data [6]. Assuming that EEG data were recorded using N electrodes, the ICA-based artefact removal can be summarized in these steps:

- Decompose the EEG data, X, into statistically independent components S. Each signal (signal from each EEG channel) in X has M samples. S contains n independent components (with $n \leq N$), each with M samples. This decomposition is done with the assumption of linear mixing of the components and through estimation of an un-mixing matrix A:

$$S = A * X \tag{1}$$

- Inspect the independent components for artefacts, and set artefactual components to zero
- Remix the resulting components with an inverse of the un-mixing matrix A:

$$X = A^{-1} * S \tag{2}$$

The ICA algorithms usually assume that the number of sources contributing to EEG data are less than (or equal to) data channels, so ICA can only generate a number of independent components less than or equal to the number of channels. If the channels are very limited, some cortical activity may be lost into artefactual components leading to alteration of neural content of the EEG. One of the solutions to this problem is to separate the leaked cortical signal from the noise in the generated components. This separation can be achieved by applying discrete wavelet transform (DWT) on the components, decomposing them into artefactual segments and residual segments [9], then thresholding the resulting wavelet coefficients; note that the wavelet thresholding is applied after ICA (ICA-W). The wavelet thresholding process assumes that each of the components, especially an artefactual component, is a sum of neural signal and Gaussian noise. Components are reconstructed by applying inverse DWT, and resulting components are remixed into cleaned EEG signal as in (2). The computation complexity in ICA-W can be reduced by using spatial constrained ICA followed by wavelet thresholding of the resulting independent components [11]. The first step in this constrained ICA-W is to define the spatial constraints, usually done based on the signal sensor topographies.

The ICA-W was previously compared with regular ICA on their effect on the neural content in EEG [9] using simulated data where ICA showed possible alteration in neural spectral content whereas ICA-W did not. In our evaluation, we compare the ICA-wavelet hybrid and regular ICA on their effect on classification of emotional states from real EEG data.

III. Methods

A. Data Acquisition

The data used in this study are taken from the database for emotion analysis using physiological signals, *DEAP* [12], acquired from Queen Mary University of London. In this dataset, 32 EEG electrodes placed according to 10-20 system were recorded at sampling rate of 512 Hz. The data were recorded from 32 healthy subjects while they watched selected videos. The videos were selected according to ratings attributed to them in terms of level of arousal and valence they stimulate in the subjects by volunteers prior the recordings. Forty videos were selected, 10 videos for each of the quadrants in the arousal-valence model [13], see Fig. 1, and each video was 60 seconds long. In the current study, each of the quadrants is considered as one class of emotion, i.e. high/low arousal-valence (HAHV, LAHV, LALV, and HALV). Besides, volunteers' ratings, after each trial, the participant ranked the video's valence and arousal on a continuous scale of 1 to 9. In order to reinforce trials labelling, only 14 participants who achieved correlation of 0.5 between their ratings and the volunteers' rating were considered in the current study.

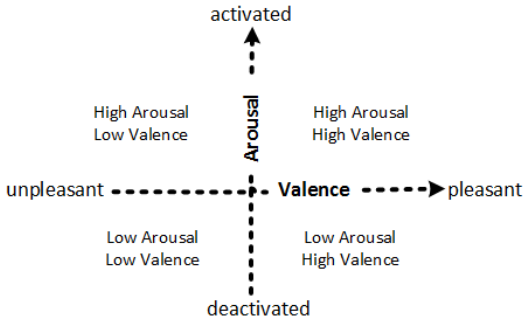

Fig. 1. Valence-Arousal model used to map the participants'video ratings

B. Preprocessing Units

The raw EEG data were preprocessed using EEGLAB toolbox [14] and the DSP Toolbox under MATLAB®. The signals were re-referenced to a common average reference, and then down-sampled to 128 Hz. The resulting signals were then filtered with a pass-band filter of 4 and 45 Hz cut-off frequencies. The band-passed data are used as reference data. Then ICA and ICA-W were applied on the band-passed signals. The *Runica function*, an ICA algorithm included in the EEGLAB toolbox, was used in both regular ICA and ICA-W methods to find independent components. In the regular ICA case, the independent components generated by *Runica* were visually inspected for artefacts, and identified artefactual components were removed. Independently of the visual inspection, ADJUST was also utilized to automate artefactual components selection in a non-manual regular ICA as in [10]. For ICA-W, a multiresolution analysis of each of the components was carried out. This analysis utilized the wavelet and scaling functions given by recursive functions in (3) and (4) respectively:

$$\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k) \quad (3)$$

$$\varphi_{j_0,k}(t) = 2^{-j_0}\varphi(2^{-j_0}t - k) \quad (4)$$

where j and k are scaling and shifting parameters respectively, with j_0 being the arbitrary starting scale. Both j and k are nonnegative integers, with $j_0 \le j \le L$; here L is the maximum decomposition level.

Each component is decomposed into approximation coefficients (cA) and details coefficients (cD), computed as in (5) and (6) respectively:

$$cA_{j_0,k} = \sum_t S_i(t)\varphi_{j_0,k}(t) \quad (5)$$

$$cD_{j,k} = \sum_t S_i(t)\psi_{j,k}(t) \quad (6)$$

where S_i is i^{th} independent component, with $i = 1, 2, ..., n$.

Details coefficients from of each decomposition level were thresholded before reconstructing the components through an inverse wavelet transform. We adopted a level-dependent adaptive threshold, proposed in [15]; this threshold is computed based on the estimated standard deviation, σ, of the noise in detail coefficients. The threshold τ, and standard deviation, σ, estimations in (7) and (8), respectively, were used in this work.

$$\tau = \sigma(2\log K)^{1/2} \quad (7)$$

$$\sigma = \left(\frac{median|cD_{jk}|}{0.6745}\right)^{1/2} \quad (8)$$

K is the length of the detail coefficients, $cD_{j,k}$.

The thresholded details of level 1 to L and the approximation coefficients at the level l were used to reconstruct the component. Debauchies wavelet, 'bd4', function was used in the wavelet transformations, and since the data were down-sampled to 128 Hz, 4 levels of decomposition are suitable to cover the main brain rhythms as shown in

Table 1. Decomposition of EEG signal into various frequency bands and corresponding wavelet decomposition levels on data sampled at 128 Hz. cD are detail coefficients whereas cA are approximation coefficients

Frequency band (Hz)	Wavelet level/coefficients
0—4 (≈Delta)	0 (cA)
4—8 (≈ Theta)	1 (cD)
8—16 (≈ Alpha)	2 (cD)
16—32 (≈ Beta)	3 (cD)
32—64 (≈ Gamma)	4 (cD)

C. Features Extraction and Classification

From the 60 second trial data features were extracted from sliding windows with widths between 2 to 24 s, every 2 seconds, with no overlap. A window starting 3s was used as a training and the segments later in the trial were used to test the trained classifier as shown in Fig. 2. For each training segment, all possible segments of the same size as the training segment were considered as testing segments. Also the distance between two successive testing segments varied from 2 to 8 seconds with 1 second increment; this allowed testing several portions of the trial beyond the training segment. Statistical measures of the preprocessed signals and spectral power in brain rhythms were then extracted from both training and testing segments, then a comparison between these features was conducted.

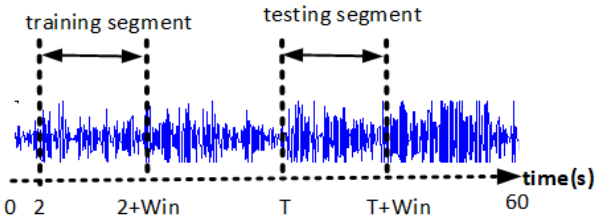

Fig. 2. Training and testing segments. *Win* here represents the segment size in seconds and *T* specifies where the testing segment begins in the trial

1) Statistical features

The common statistical measures used in emotion recognition include the EEG signal power, mean, standard deviation, first difference, normalized first difference, second difference, and normalized second difference [16]. The extracted statistical measures in this study were: the power (p_x), mean (μ_x), standard deviation (σ_x), and first difference (δ_x), and these values are computed as in (9) to (12) respectively:

$$p_x = \frac{1}{m}\sum_{t=1}^{m} |x(t)|^2 \quad (9)$$

$$\mu_x = \frac{1}{m}\sum_{t=1}^{m} x(t) \quad (10)$$

$$\sigma_x = \left(\frac{1}{m-1}\sum_{t=1}^{m}(x(t) - \mu_x)^2\right)^{1/2} \quad (11)$$

$$\delta_x = \frac{1}{m-1}\sum_{t=1}^{m-1}|x(t+1) - x(t)| \quad (12)$$

$x(t)$ is the signal sample at the time t, with t = 1, 2, 3… m; m being the number of samples in the window. The statistical measures were extracted from training and testing segments of each trial in the format below:

$$F_{Statistics}:\left\{\left[p_x,\mu_x,\sigma_x,\delta_x\right]\times 32channels\right\}$$

2) Spectral features

Spectral band power in the theta (4-8 Hz), alpha (8-16 Hz), beta (16-32 Hz), and gamma (32-45 Hz) bands was extracted from the preprocessed data. Another version of spectral band power was extracted by first applying the wavelet transform to subdivide the preprocessed data into frequency bands relatively close to the brain rhythms, as shown in Table 1, then extracting the band power from bands corresponding to Theta, Alpha, Beta, and Gamma. The spectral band power extracted from the wavelet-generated frequency bands is referred to as wavelet features in this paper. The band power and wavelet features were extracted from each trial in the formats shown below:

$$F_{Bandpower}:\left\{\left[theta,alpha,beta,gamma\right]\times 32channels\right\}$$

$$F_{Wavelet}:\left\{\left[theta,alpha,beta,gamma\right]\times 32channels\right\}$$

3) Classification

Regularized support vector classification of LIBSVM toolbox [17] was used for classification. LIBSVM uses *one-against-one* approach [18] for multiclass discrimination. The features extracted on the training segment were used to train SVM model. The features were mapped in high dimensional space using a polynomial kernel in (13), the coefficient r, was set to 0 (its default value in LIBSVM). Other parameters including the degree of the polynomial, z, regularization parameter, C, and kernel parameter γ were set through 10-folds cross-validation. C controls the trade-off between minimum classification error and model complexity [19], large C leads to model with many support vectors and possible overfitting whereas small C may lead to underfitting. The parameter γ controls the width of the kernel, very large γ leads to overfitting [20]. The trained model, for each segment size, was tested over the corresponding testing segments.

$$f(x_i,x_j)=(\gamma x_i^T x_j+r)^z \qquad (13)$$

IV. Results and Discusion

Fig. 3 shows a sample visual comparison of artefact-corrected data for each of the two ICA based methods. Both the ICA and ICA-W eliminated the strong artefacts found in the band-pass filtered data. The ICA-W seems to preserve the original signal trend while regular ICA cleaned data tend to be flat at the artefactual segment.

The three artefact removal methods for each of the three feature types were evaluated for different windows. The classification accuracy (CA) of the three feature-types, i.e. statistics measures, band power and wavelet features corresponding to band-pass filtering, ICA, ADJUST and ICA-W for artefact removal methods are reported in Fig. 4. The reported CA are averaged across all the subjects for each segment size. The CA in the case of regular ICA (with visual inspection) were slightly higher than automated regular ICA (ADJUST). CA increased as feature window size increased for wavelet and statistical features in the ICA-W cleaned data whereas the CA first slightly increases then slowly decreases in the other methods. The band power features performed poorly (CA < 57%) and remained nearly the same across all the segments in all the three methods. Slight improvement in band power performance was observed in both the ICA and ICA-W compared to band-pass filtering. Considering only the segment sizes with highest CA, the achieved high accuracies, averaged across all the 14 subjects, are 55.83%, 60.00%, 56.66% and 73.50% for pass-band filter, ICA, ADJUST, and ICA-W respectively, as shown in Fig. 5. Wilcoxon signed rank tests showed that ICA-W led to significant increases in classification accuracies over the regular ICA and band filtering methods in statistical features ($p < 0.001$). Also the ICA-W led to higher CA compared to ICA in the case of wavelet features ($p < 0.005$) and band-pass filtering ($p < 0.001$).

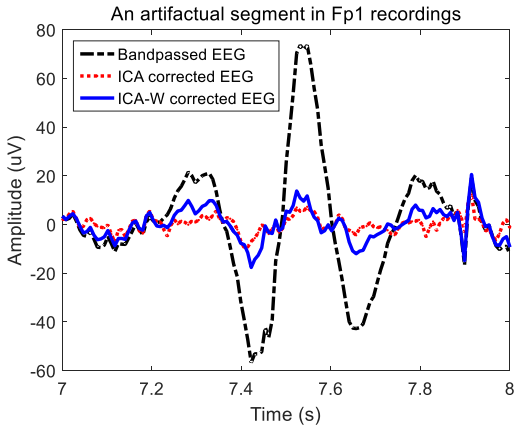

Fig. 3. A sample visual comparison of the two ICA methodss on artefactual segment recorded at the channel Fp1 during one of the trials of subject 3 (in the orginal 32 subjects of DEAP dataset)

The achieved CAs in ICA-W method for statistical features are significantly higher than in previous studies in which frequency band filtering [21] and visual inspection [22] were used as artefact removal methods. The observed higher CAs in the case of ICA-W method compared to traditional ICA suggests that ICA-W is more efficient for artefact removal than traditional ICA. Furthermore, statistical features were projected in two dimensions by t-distributed stochastic neighbor embedding (t-SNE) projection [23], and the distribution of features for each class in the case of band-passed data, ICA and ICA-W filtered data as shown in Fig. 6 for one of the subjects. Looking at the features distribution, the features for LALV and HALV on one side, are not separable enough, and on the other side LAHV and HAHV's features are not significantly separable. The overall best features separability seems to be in the ICA-W case.

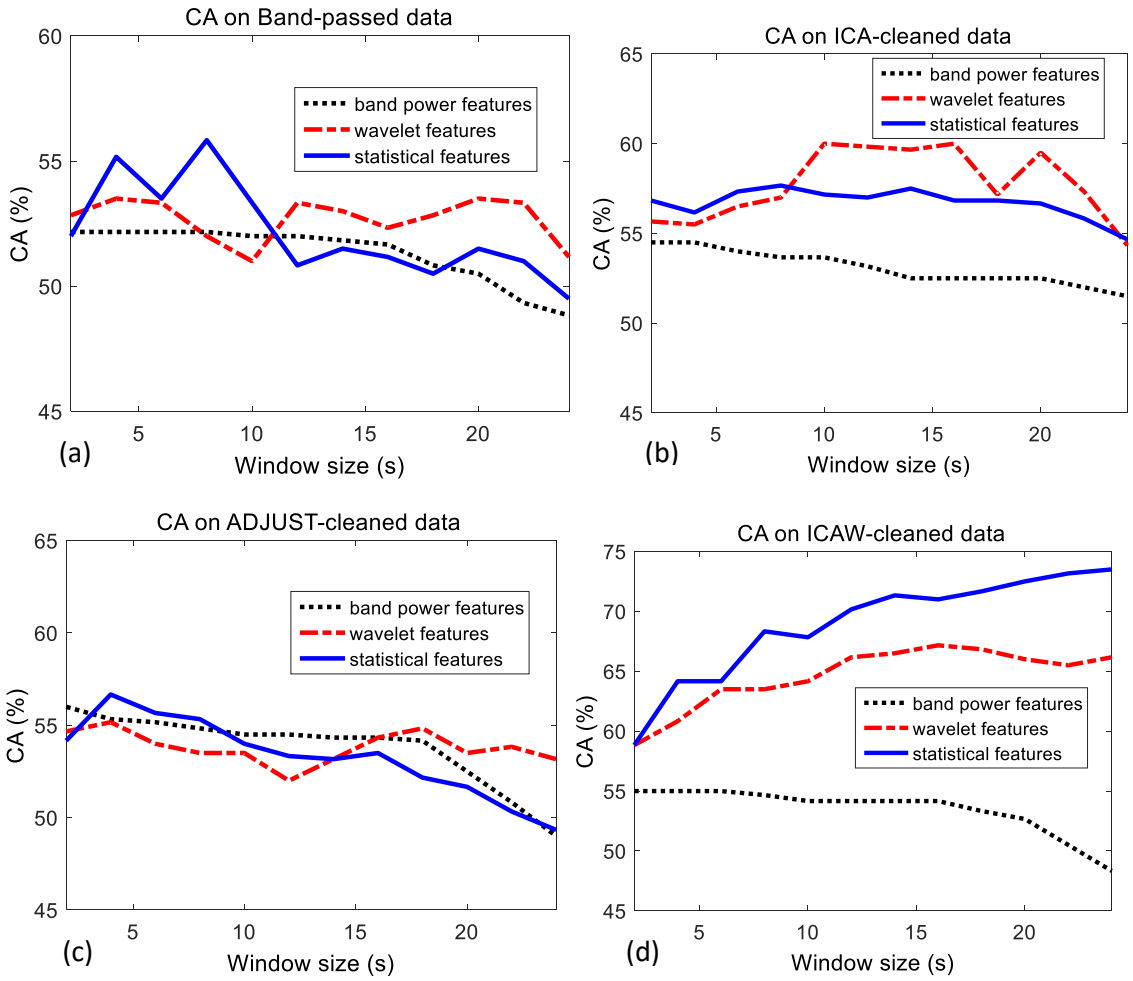

Fig. 4. Classification accuracies (CA) over various segment sizes in the considered noise removal methods: (a) case of band-passed data, (b) case of ICA cleaned data, (c) case of ADJUST, and (d) case of ICAW cleaned.

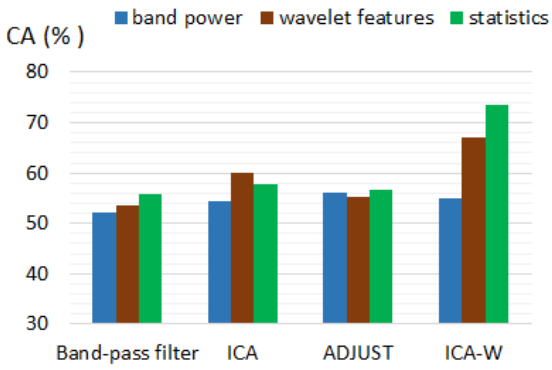

Fig. 5. Classification accuracies of the three features averaged across the participants using the best segment size for each of the considered denoising methods

V. Conclusion

The paper has presented a comparison of ICA and the automated version of ICA (ICA-W) as an artefact removal methods using emotion-related EEG datasets. Performance improvement and process automation of emotion recognition were investigated in this study. In additional to the automation advantage provided by ICA-W, the results suggest that EEG-based emotion recognition is improved when ICA-W is used to preprocess the EEG data compared to the regular ICA. Significant increase were found when using statistical features for both methods with ICA-W significantly outperforming all other methods. An analysis of the feature distributions shows some improvement into features separability. Future work will focus on features selection and comparing the performance of ICA-W to other spatial filters and applying the approach in real-time emotion recognition based brain-computer interface (BCIs).

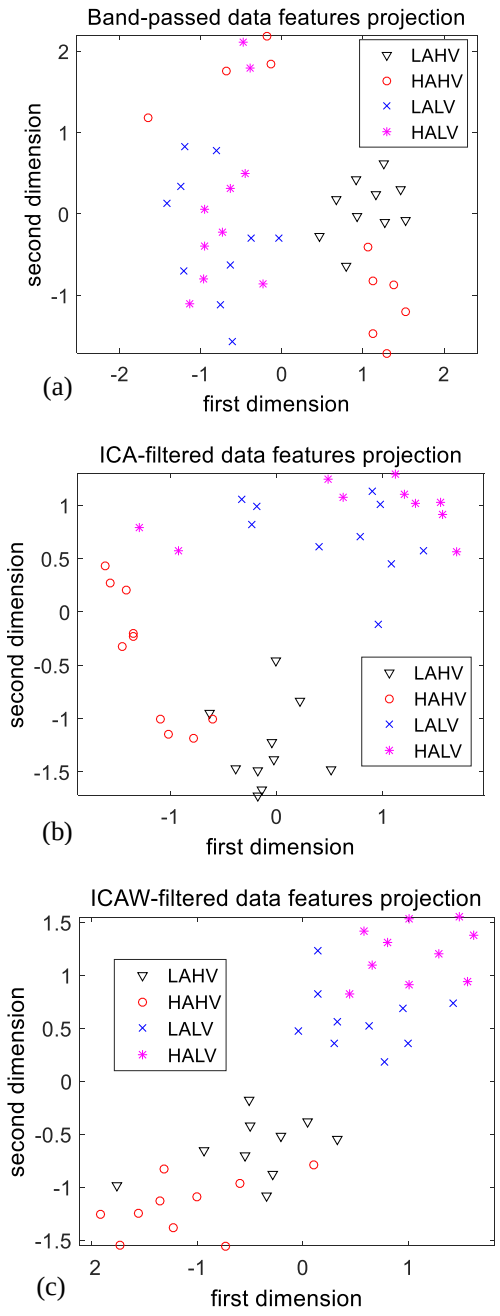

Fig. 6. Two-dimension t-sne projection for features of four emotions extracted from (a) band-passed data, (b) ICA-cleaned data, and (c) ICA-W cleaned data.

ACKNOWLEDGMENT

The authors would like to acknowledge the *DEAP* dataset's authors for providing the dataset used in this work.

REFERENCES

[1] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors*, vol. 12, pp. 1211–1279, 2012.

[2] M. L. Phillips, W. C. Drevets, S. L. Rauch, and R. Lane, "Neurobiology of emotion perception I: The neural basis of normal emotion perception," *Biol. Psychiatry*, vol. 54, pp. 504–514, 2003.

[3] J. O. Whitham, Emma M. Pope, Kenneth J. Fitzgibbon, Sean P. Lewis, Trent Clark, C. Richard Loveless, Stephen Broberg, Marita Wallace, Angus DeLosAngeles, Dylan Lillie, Peter Hardy, Andrew Fronsko, Rik Pulbrook, Alyson Willoughby, "Scalp electrical recording during paralysis: Quantitative evidence that EEG frequencies above 20 Hz are contaminated by EMG," *Clin. Neurophysiol.*, vol. 118, no. 8, pp. 1877–1888, 2007.

[4] I. I. Goncharova, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "EMG contamination of EEG: Spectral and topographical characteristics," *Clin. Neurophysiol.*, vol. 114, no. 9, pp. 1580–1593, 2003.

[5] A. Jafarifarmand and M. A. Badamchizadeh, "Artifacts removal in EEG signal using a new neural network enhanced adaptive filter," *Neurocomputing*, vol. 103, pp. 222–231, Mar. 2013.

[6] T.-P. Jung, S. Makeig, C. Humphries, T.-W. Lee, M. J. McKeown, V. Iragui, and T. J. Sejnowski, "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology*, vol. 37, no. 2, pp. 163–178, Mar. 2000.

[7] a Delorme, T. Sejnowski, and S. Makeig, "Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis," *Neuroimage*, vol. 34, no. 4, pp. 1443–1449, 2007.

[8] W. Y. Hsu, C. H. Lin, H. J. Hsu, P. H. Chen, and I. R. Chen, "Wavelet-based envelope features with automatic EOG artifact removal: Application to single-trial EEG data," *Expert Syst. Appl.*, vol. 39, no. 3, pp. 2743–2749, 2012.

[9] N. P. Castellanos and V. a. Makarov, "Recovering EEG brain signals: Artifact suppression with wavelet enhanced independent component analysis," *J. Neurosci. Methods*, vol. 158, pp. 300–312, 2006.

[10] A. Mognon, J. Jovicich, L. Bruzzone, and M. Buiatti, "ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features," *Psychophysiology*, vol. 48, no. 2, pp. 229–240, 2011.

[11] M. T. Akhtar, W. Mitsuhashi, and C. J. James, "Employing spatially constrained ICA and wavelet denoising, for automatic removal of artifacts from multichannel EEG data," *Signal Processing*, vol. 92, no. 2, pp. 401–416, 2012.

[12] S. Koelstra, C. Mühl, M. Soleymani, J. S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis; Using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, 2012.

[13] J. A. Russell, "A circumplex model of affect.," *J. Pers. Soc. Psychol.*, vol. 39, pp. 1161–1178, 1980.

[14] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, 2004.

[15] D. L. Donoho, "Nonlinear Wavelet Methods for Recovery of Signals, Densities, and Spectra from Indirect and Noisy Data," in *Proceedings of Symposia in Applied Mathematics*, 1993, pp. 173–205.

[16] R. Jenke, A. Peer, and M. Buss, "Feature Extraction and Selection for Emotion Recognition from EEG," *IEEE Trans. Affect. Comput.*, vol. 5, no. 3, pp. 327–339, Jul. 2014.

[17] C. Chang and C. Lin, "LIBSVM : A Library for Support Vector Machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.

[18] S. Knerr, L. Personnaz, and G. Dreyfus, "Single-layer learning revisited: a stepwise procedure for building and training a neural network," in *Neurocomputing*, no. April, Berlin, Heidelberg: Springer Berlin Heidelberg, 1990, pp. 41–50.

[19] Y. Ren and G. Bai, "Determination of optimal SVM parameters by using GA/PSO," *J. Comput.*, vol. 5, no. 8, pp. 1160–1168, 2010.

[20] C. L. Huang and C. J. Wang, "A GA-based feature selection and parameters optimizationfor support vector machines," *Expert Syst. Appl.*, vol. 31, no. 2, pp. 231–240, 2006.

[21] K. Takahashi, "Remarks on emotion recognition from multi-modal bio-potential signals," *2004 IEEE Int. Conf. Ind. Technol. 2004. IEEE ICIT '04.*, vol. 3, pp. 95–100, 2004.

[22] X.-W. Wang, D. Nie, and B.-L. Lu, “EEG-Based Emotion Recognition Using Frequency Domain Features and Support Vector Machines,” in *Science*, 2011, pp. 734–743.

[23] L. J. P. Van Der Maaten and G. E. Hinton, “Visualizing high-dimensional data using t-sne,” *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, 2008.