Automatic EEG bad epoch and artifact removal using clustering

Elizabeth Hames and Mary Baker
liz.hames@ttu.edu and mary.baker@ttu.edu
Department of Electrical Engineering
Texas Tech University
Lubbock, TX, U.S.A.
Submitted for consideration for the 2014 BIOCOMP Conference

Abstract—This paper presents a method for automatically identifying and cleaning artifactual independent components from EEG data. Artifactual components are identified through first clustering components and then labeling artifactual clusters. Clustering allows for identification of many types of artifactual components, including eye movements, electromyogram signals, electromyogram signals, movement, and bad channels. The proposed artifact detection method is analyzed using EEG data acquired with an EGI 65-channel net. The benefits of cleaning components compared to removing components are tested with a simulated dataset. Results indicate that cleaning components is beneficial when potential cerebral signals are falsely identified as artifactual components or are mixed into artifactual components.

Keywords—EEG, artifact removal, ICA, clustering, components

I. INTRODUCTION

EEG data undergo several steps of preprocessing before analysis to remove non-cerebral, or artifactual, contributions to the data. Artifact removal can be difficult because the spectrum, topography, and waveform shape of many physiological artifacts are identical to cerebral EEG activity [1]. Independent component analysis (ICA) separates the EEG data from the artifactual signal, leaving the data analyst with the task of identifying and removing artifactual components [2]-[4]. Identifying artifactual components is not a simple process. It involves viewing component time courses, topographies, and spectra. This entire process can be time consuming since the number of components is usually equal to the number of channels. Some artifacts are very ambiguous and difficult to label, making component selection subjective and prone to error. Automated component selection solves the issues of manual component selection by providing consistency and speed. This work proposes a method for automatically identifying artifactual components using clustering implemented through an eight step process called ABEAR (automatic EEG bad epoch and artifact removal).

A. Previous Artifact Removal Methods

Many other automated methods exist for artifact removal, but this method combines qualities of each to meet the following goals: 1) Detect multiple types of artifacts; 2) Automate independent component selection; 3) Eliminate the need for a supplemental signal; 4) Clean components.

Many previous methods for EEG artifact removal focus on the removal of only one or two types of artifacts. For example, filtering only removes artifacts outside of the EEG frequency band, such as DC drift or power line noise. Wavelet analysis is only capable of detecting artifacts that resemble a mother wavelet. We propose clustering as a solution to the problem of limited artifact detection. Clustering separates cerebral data from artifactual data regardless of artifact type. Through clustering, we automatically identify five types of artifactual components: eye movements (EOG), electromyogram signals (EMG), electromyogram signals (ECG), artifacts caused by movement (MOV), and artifacts caused by bad channels (BAD).

If not automated, artifact removal methods such as wavelet analysis or ICA require manual component or coefficient selection [2]-[6]. Manual artifact selection is time-consuming, inconsistent, and prone to human error. In this work, artificial clusters are labeled through an outlier criterion to reduce human interaction, similar to [7]. It dramatically reduces the time spent on manually selecting artifactual components.

Many automated artifact removal methods, such as adaptive filtering or time/frequency regression, require supplemental signals, such as a measured ECG or EMG signal or even a synthetic signal [8]-[11]. Automated methods for component selection, such as correlation with a reference signal or supervised learning, also need training signals [12]-[14]. Additional signals are not always available or accurate for all participants. The artifact detection method presented in this work does not need any additional reference signals. The use of clustering for artifactual component identification is an unsupervised learning method that groups the data without training signals. Using an outlier criterion to label artifactual clusters eliminates the need for a reference signal, similar to [15].

Removing data, removing artifactual ICA components, or even wavelet thresholding, can result in throwing away useful information [16]-[18]. This work presents cleaning components as an alternative to completely removing them. The benefit to cleaning components is demonstrated in the Results section using a simulated dataset.

In summary, the proposed artifact detection method, ABEAR, separates EEG data into ICA components and classifies those components using an unsupervised clustering algorithm. Clustering allows the detection of five types of artifactual components: EOG, EMG, ECG, MOV, or BAD. Identified artifactual components are cleaned to reduce data loss.
B. EEG Artifactual Components

Each type of artifactual component has unique properties, such as frequency spectrum, topography, or waveform shape, that distinguish it from EEG data or other artifacts. This section describes properties of five types of artifactual components.

EOG artifacts are the result of blinks, horizontal eye movements, or vertical eye movements. They have various shapes depending on the movement of the eyes. Blinks are characterized by high amplitude impulses while horizontal and vertical eye movements have rectangular-like shapes. The simultaneous occurrence of blinks and eye movements results in unpredictable waveform shapes [19]. All EOG components are located in the frontal channels near the eyes. In a topography map of EOG components, blinks have a single pole centered between the eyes and while eye movements usually have two opposite poles. The spectrum of EOG artifacts is dominated by low frequencies, below 5 Hz. Many physiological actions can result in EOG artifacts: clenching muscles, movement of the head, neck, or shoulders, or squinting and twitching of the eyes. EOG artifacts can take two forms: white-noise-like pattern and “railroad-cross-tie” pattern [20]. On topography maps, EOG artifacts are usually located in the peripheral channels (around the neck, ears, and eyes). EOG artifacts have a uniform spectrum [20]. ECG artifacts have fairly periodic QRS spikes [21]. The ECG waveforms repeat about once per second. ECG topographies show a bi-polar pattern across the entire scalp and their spectrum has mainly low-frequency dominance. MOV artifacts refer to any artifact caused by movement that is not an EMG waveform, i.e., shoulder or neck movement or repositioning. MOV artifacts tend to have very low-frequency dominance similar to eye movements and their topographies show peripheral channel localization. BAD artifacts refer to artifactual components whose topographies indicate localization to a single channel. BAD artifacts can result from high impedance electrodes, electrodes disturbed by touch, or misplaced electrodes. BAD components do not have a common shape or spectrum.

II. METHODS

We employed an eight step process to detect and remove EEG artifactual components, as seen in Fig. 1. The following sections describe the eight steps in more detail.

A. Bad Epoch Removal

“Bad” epochs are time-segments with large amounts of artifacts across several components. Bad epochs are instances where an artifact was not separable into just one component. Removing time-segments with inseparable artifacts can improve ICA and artifactual component identification. ICA must be computed before bad epoch removal since independent components are used to identify bad epochs. Variance and spectrum of components are used in identifying bad epochs.

The steps for bad epoch removal are as follows. The variance, \( var \), and the spectrum, \( spect \), between .5 and 50 Hz are computed for each epoch of each component. Next, the mean spectrum, \( Savg \), and mean variance, \( Vavg \), for each component are determined by (1) and (2). The spectrum error, \( dS \), and variance error, \( dV \), given by (3) and (4), are calculated for each epoch and each component using \( Savg \) and \( Vavg \). Within each epoch, components are labeled “bad” if \( dS \) or \( dV \) are greater than a threshold, \( Error \). Finally, an epoch is labeled as “bad” if it contains more bad components than a threshold, \( perbad*ncomp \). To improve the detection of bad epochs, the selection process is repeated with adjusted means that exclude previously labeled bad epochs. ICA is recomputed on the EEG dataset after all bad epochs are removed. The newly generated components are the input to the component features step. The thresholds used for bad epoch detection were selected based on a performance comparison using manually labeled bad epochs.

\[
Savg(i) = \frac{\sum_{j=1}^{nepoch} spect(i,j)}{nepoch}, i = 1..ncomp
\]

(1)

\[
Vavg(i) = \frac{\sum_{j=1}^{nepoch} var(i,j)}{nepoch}, i = 1..ncomp
\]

(2)

\[
dS(i,j) = \frac{|spect(i,j) - Savg(i)|}{Savg(i)}, i = 1..ncomp, j = 1..nepoch
\]

(3)

\[
dV(i,j) = \frac{|var(i,j) - Vavg(i)|}{Vavg(i)}, i = 1..ncomp, j = 1..nepoch
\]

(4)

B. Component Features

Component features are measurements taken from each component used for clustering. Eight 1-dimensional features were generated for each component: topography histogram, spectrum fit 1, spectrum fit 2, frontal location score, peripheral location score, average auto-correlation, Lorentz threshold, and syn3 repeatability. The eight features were normalized and then combined into one 8-dimensional feature.

The eight features are similar to features tested by [15]. The features were chosen as a representative of component properties commonly used to manually identify artifactual components. The eight features are chosen over single features such as spectrum, topography, etc. for two reasons. First, they provide consistency across participants for the five types of artifactual components. Time courses, spectrums, and even topographies can vary across participants within a given
category of artifactual component. Representing the components with multiple features reduces dependency on a single measurement, providing more consistency across participants. Second, the single dimensional features allow clustering of multiple types of artifactual components. No one feature provides good separation of all five types of artifactual components. For example, spectrum would be excellent for separating EOG and EMG components, but not necessarily BAD or MOV components since their spectrums are inconsistent. The eight one-dimensional features address the properties of each type of artifactual component.

The histogram of a component’s topography explains its scalp power distribution. The first number of the histogram array is the number of channels for which the component is least distributed. If the first number of the histogram is large (>0.9*nchan), the component is localized to one channel. This is useful for detecting BAD or MOV artifacts.

The shape of a spectrum can be described using a second order polynomial, given by (5). The first coefficient, C1, describes concavity and the second coefficient, C2, describes the slope. The first and second coefficients of the second order polynomial spectrum fit are obtained as features, spectrum fit 1 and spectrum fit 2. The concavity and slope are very large for EOG artifacts and very small for EMG artifacts.

\[ S = C_1x^2 + C_2x + C_3 \]  

(5)

The frontal location score describes the amount of component power located in the frontal channels. Given a component’s topography, the frontal location score is the fraction of channels with a relative power greater than 40% of the maximum relative power that are located in frontal channels. Frontal channels are designated as channels within -55 to 55 degrees from center and with a radius greater than .4 from Cz. Similarly, given a component’s topography, the peripheral location score is the fraction of channels with a relative power greater than 40% of the maximum relative power that are located in peripheral channels. Peripheral channels are designated as channels that have a radius greater than .4 from Cz.

The Lorentz threshold is used to quantify the amount of component signal described by a sym3 wavelet. It has previously been used to despike signals, meaning it would be largest for EOG or ECG components [22]. The signal, C, used for computing the threshold, is the approximate wavelet coefficient for the sym3 wavelet, shown in (6).

\[ L = \sqrt{\frac{\sum_1^N C_i^2}{N}} \]  

(6)

The sym3 repeatability feature uses the Lorentz threshold to estimate the number of spikes in a component. The repeatability of a component is the number of values within the sym3 approximation coefficient that are greater than 2*L.

C. Clustering

Common clustering algorithms, such as k-means, require a user specified number of clusters. It is not always clear how many groups exist in a set of components and this number can change across datasets. The Isodata clustering algorithm iteratively finds the optimal number of clusters by splitting and merging clusters [23]. This is beneficial for clustering EEG components since the number of artifactual components changes across datasets. In Isodata, clusters are split based on a standard deviation threshold and merged based on a separation distance threshold. Although Isodata can automatically determine the number of clusters, there are many user defined thresholds that rely upon the properties of the data. A modified Isodata algorithm eliminates the need for these thresholds by using fuzzy membership functions [24]. The modified Isodata algorithm is used to cluster the component features. The details of the modified Isodata algorithm will not be discussed in this paper to conserve space. However, details can be obtained through the author.

Three parameters determine the initial conditions for clustering: initial cluster centers, maximum number of iterations, and the number of repetitions. The results presented in this paper are generated from the following initial conditions. Initial cluster centers are randomly selected from the feature vectors. The maximum number of iterations is set to 100 and final clusters are selected from the best of three repetitions of clustering.

D. Labeling Clusters

Isodata segments component features into unlabeled groups. The artifactual clusters must be distinguished from non-artifactual clusters. Artifactual clusters have outlying average 1-dimensional features, so we implement an outlier criterion to determine which clusters are artifactual. The 1-dimensional features are averaged for members of each cluster. The average 1-dimensional features, fi, are used to compute the percent difference of each cluster from the average of all clusters, given by (7). Clusters whose percent difference is greater than .75 are labeled as artifactual.

\[ \text{Percent Difference} = \frac{|f_i - \sum_{i=1}^m f_i|}{\text{standard deviation}(f)} \]  

(7)

E. Component Cleaning

Rather than removing artifactual components, we implement a cleaning process to prevent a loss of cerebral EEG data. All artifactual components undergo despiking and denoising as part of the cleaning process.

1) Despiking: Despiking removes large spikes within a signal, such as blinks or heart beats. Despiking can be accomplished through wavelet thresholding [22],[24]. Sym3 wavelet coefficients for four levels are computed for overlapping segments of length one second. All values of the coefficients are thresholded using the level-dependent threshold in (8) [26]-[27], where MAD is the median absolute deviation of the wavelet coefficient, x. Dividing by .6745
normalizes the threshold by the standard deviation for Gaussian white noise [28]. All values that exceed the threshold are replaced using a cubic interpolation. The thresholding process is repeated until all values in the coefficients are below the threshold.

\[
Threshold = \frac{4 \times \text{MAD}(|x|)}{0.6745}
\]  

(8)

2) Denoising: Denoising is performed on the same segments from despiking. Denoising is also accomplished through wavelet thresholding. Db8 wavelet coefficients are computed for each segment using four levels. All values of the coefficients are thresholded using the level-dependent threshold in (9) [28]-[29]. All values below the threshold are replaced using a soft thresholding.

\[
Threshold = \frac{\text{MAD}(|x|)}{0.6745}
\]  

(9)

III. EEG DATA

Both recorded data and simulated data are utilized to evaluate the artifact detection and removal process. The recorded data provides ground truth for clustering and component labeling. The simulated data provides ground truth for cleaning components. Both types of data are described in the following sections.

A. Recorded Data

In a previous, unrelated study, we acquired EEG data from 62 participants using a 65-channel net. The session consisted of 42 cognitive tasks. Each participant’s dataset was epoched at the start of each task for six seconds. All 62 datasets were filtered with a 1 Hz high-pass filter and a 50Hz low-pass filter and epoched around 23 event markers for six seconds (for a total of 138 seconds). Each dataset was decomposed into 65 components using Infomax ICA. An artifactual version and a manually cleaned version of the recorded data was saved for use in this work. The manually cleaned version was acquired from removing bad epochs followed by removing artifactual ICA components. EEGLab was used to perform ICA, view properties of the components, and remove artifactual epochs and components [30]. Manually selected artifactual components were used to test component classification rates. It is important to note that only one expert provided labeled components. In order to reduce bias in component labeling, further experts need to be consulted in the future. Also, only data acquired from an EGI net is used to verify the proposed artifact removal algorithm. Generalization cannot be made for other EEG systems until further testing is done.

B. Synthetic Data

Simulated EEG data are created by forward projecting six artifactual sources onto a scalp and adding background EEG noise to the resulting data. A blink source is created by convolving a single blink signal, extracted from a recorded EEG time series, with a series of randomly generated impulses. The impulses are spaced between .5 and 1 second apart. Similarly, an ECG source is created by convolving an ECG signal, simulated using MATLAB’s `ecg` function, with a series of randomly generated impulses [31]. The impulses are spaced between 1 and 1.7 seconds apart. Three EMG sources are created from white Gaussian noise. Lastly, a cross-tie EMG pattern is created with a train of Gaussian pulses [20]. A forward model of the sources is created using BESA EEG simulation software, available at [www.besa.de](http://www.besa.de). Dipoles are positioned at the front of the head, middle of the head, and temporal and parietal locations corresponding to the blink, ECG, and EMG signals respectively. A thirty-three channel artifactual dataset is generated from the forward model. EEG background noise is added to each channel of the artifactual data. The background noise is generated to mimic the frequency spectrum and amplitudes of EEG data [32]. A non-artifactual dataset is also created from the EEG background noise to test component cleaning.

IV. RESULTS AND DISCUSSION

A. Bad Epoch Removal

The selection of bad epochs is dependent upon two user specified thresholds: Error and perbad. Values for Error and perbad are selected as .15 and .8, respectively, to maximize the precision and recall as calculated by (10) and (11). The manually labeled bad epochs from the 62 recorded datasets are used to determine the true positives, false positives, and false negatives generated by the automatic labeling of bad epochs. The chosen thresholds result in precision and recall rates of about .7. While classification is not perfect, precision and recall rates are acceptable. Manual classification is very subjective, leading to inconsistency and possible mislabeling. Automatic selection of bad epochs provides consistency.

\[
\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}
\]  

(10)

\[
\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}
\]  

(11)

B. Component Features

Three datasets are randomly selected from the 62 recorded datasets to display examples of the 1-dimensional features, shown in Fig.2. The eight features in order are topography histogram, spectrum fit 1 and 2, frontal and peripheral location scores, average auto-correlation, Lorentz threshold, and sym3 repeatability. The features are averaged for each of the five types of artifactual components, as determined by manual labeling: EOG, EMG, ECG, BAD, and MOV. The 1-dimensional features demonstrate consistency across participants. The EOG components are best distinguished by their large spectrum fit 1 and 2 values and high EOG scores.
C. Clustering

Initial cluster centers are randomly selected from the feature vectors. The maximum number of iterations is set to 100 and final clusters are selected from the best of three repetitions of clustering. Clustering was performed for three initial values of $m$, the number of clusters: 13, 16, and 19. The final number of clusters for these three values of $m$, averaged across all 62 datasets, are 21, 23, and 26 respectively.

Precision and recall, as measured by (12) and (13), are used to evaluate how well Isodata clusters the components into artifactual and non-artifactual clusters. The manually labeled components from the 62 recorded datasets are used to determine which components are artifactual and non-artifactual. High precision means that a cluster contains mostly artifactual components, indicating a good separation of artifactual and non-artifactual components. High recall means that the artifactual components are highly concentrated in a cluster, indicating that the artifactual components are not spread across many clusters.

\[
\text{Precision} = \frac{\text{number of artifactual components in a cluster}}{\text{total number of components in a cluster}}
\]

(12)

\[
\text{Recall} = \frac{\text{number of artifactual components in a cluster}}{\text{total number of artifactual components}}
\]

(13)

Table 1 shows the percentage of clusters in all 62 datasets with a precision greater than .72 for each value of $m$. A large percentage of clusters having a precision greater than .72 does not necessarily indicate improved clustering; it could result from many single member clusters. Recall can provide additional information to evaluate clustering. Table 4 also shows the average recall for the clusters with a precision greater than .72 for each initial value of $m$. The modified Isodata results in a large percentage of clusters with a high precision as well as high recall rates for those clusters, and this outcome is not overly sensitive to the initial choice of $m$.

![Fig. 2. 1-dimensional features averages across five types of artifactual components for three participants](image)

<table>
<thead>
<tr>
<th>Types of Clusters</th>
<th>Types of components within each type of cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-artifactual</td>
</tr>
<tr>
<td>Non-artifactual</td>
<td>44.0</td>
</tr>
<tr>
<td>EOG</td>
<td>3.2</td>
</tr>
<tr>
<td>EMG</td>
<td>6.0</td>
</tr>
<tr>
<td>ECG</td>
<td>0.6</td>
</tr>
<tr>
<td>MOV</td>
<td>0.7</td>
</tr>
<tr>
<td>BAD</td>
<td>3.4</td>
</tr>
</tbody>
</table>

Table II. Average of Cluster Member Survey

<table>
<thead>
<tr>
<th>Types of Clusters</th>
<th>Types of components within each type of cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-artifactual</td>
</tr>
<tr>
<td>Non-artifactual</td>
<td>5.9</td>
</tr>
<tr>
<td>EOG</td>
<td>3.2</td>
</tr>
<tr>
<td>EMG</td>
<td>5.5</td>
</tr>
<tr>
<td>ECG</td>
<td>1.3</td>
</tr>
<tr>
<td>MOV</td>
<td>1.4</td>
</tr>
<tr>
<td>BAD</td>
<td>3.4</td>
</tr>
</tbody>
</table>

Table III. Standard Deviation of Cluster Member Survey

The numbers along the diagonal of Table 2 represent the average number of each type of artifactual, or non-artifactual.
component. The off diagonal elements indicate the number of mixed components. For example, the top right number is the number of BAD components located within clusters that contain non-artifactual components. It would be ideal for all the numbers in the top row and far left column, except the top left cell, to be zero. This would indicate that all artifactual components are separated from non-artifactual components. This however, is not the case. The top row indicates that very few artifactual components are mixed into non-artifactual clusters. The left column indicates that there are large numbers of non-artifactual components mixed into EMG clusters. It is not necessarily ideal for all other off-diagonal elements to be zero. For instance, if some EMG components have similar properties of BAD components, then the corresponding EMG components might be assigned to BAD clusters. In this case, however, most of the non-diagonal elements are near zero.

In Table 3, most off-diagonal standard deviations are very low. The highest standard deviations occur for the average number of non-artifactual and EMG components as well as the number of non-artifactual components located in EMG clusters. Some standard deviations are higher than their averages, indicating that one or two datasets may have extreme outliers, with most datasets having similar values.

D. Cluster Labeling

Equations (10) and (11) are used to evaluate the results of cluster labeling. Following cluster labeling, all components are labeled as artifactual or non-artifactual depending on what type of cluster they belong to. The manually labeled components from the 62 recorded datasets are used to determine the true positives, false positives, and false negatives generated by cluster labeling. Precision and recall are computed from all datasets. For example, the number of true positives is the total number of true positives across all 62 datasets. Recall is calculated for each type of artifact: EOG, ECG, EMG, MOV, or BAD. Final classification results are given in Table 4. The EMG and BAD components result in the lowest recall rates while MOV and EOG components result in the highest recall rates. Recall can be increased by lowering the outlier threshold. However, this threshold adjustment results in a lower precision. Overall, the clustering and cluster labeling processes result in very good recall rates with a decent precision. Since the components are cleaned as opposed to removed, a lower precision is acceptable.

E. Component Cleaning

1) Simulated Data: Seven components are manually identified as artifactual after the simulated dataset is decomposed using ICA. Fig. 3 shows a subset of the components after components cleaning. The EOG and ECG spikes are no longer present. EMG noise is reduced.

Both component removal and component cleaning are performed on the simulated data to observe the benefits provided by cleaning components. Channel spectra of reconstructed data from cleaning and removing components, as well as the original artifactual data, are compared to channel spectra of non-artifactual data using a distance measure, $dS$, given by (14). Fig. 4 shows the spectra comparisons for three frequencies. In (14), the “cleaned data spectra” is replaced by the removed component spectra or original data spectra to compare them to the non-artifactual data. Two sets of components are labeled as artifactual for the calculation of (14). The first set consists of seven truly artifactual components. The second set of components consists of the seven artifactual components with an additional ten randomly chosen components.

$$dS = |\text{Nonartifactual Data Spectra} – \text{Cleaned Data Spectra}|$$  \hspace{1cm} (14)

Very little difference can be seen in spectra between cleaning and removing only the seven artifactual components. However, when more components are selected than necessary, a large difference between the data resulting from component removal and the non-artifactual data is observed across all frequencies. Component cleaning results in a smaller difference between spectra, indicating that it is beneficial for instances when too many components, or unnecessary components, are identified as artifactual.

2) Recorded Data: Two datasets are randomly selected from the 62 recorded datasets to view the results of component cleaning. Fig. 5 shows a subset of components for the two datasets before and after cleaning. All large spikes caused by EOG or ECG artifacts are completely removed. EMG components present more of a challenge for cleaning than EOG or ECG artifacts. The EMG components are reduced, while not completely removed.

<table>
<thead>
<tr>
<th>TABLE IV. COMPONENT CLASSIFICATION RESULTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Precision</td>
</tr>
<tr>
<td>Average Recall</td>
</tr>
<tr>
<td>EOG Recall</td>
</tr>
<tr>
<td>EMG Recall</td>
</tr>
<tr>
<td>ECG Recall</td>
</tr>
<tr>
<td>MOV Recall</td>
</tr>
<tr>
<td>BAD Recall</td>
</tr>
</tbody>
</table>

Fig. 3. Simulated data components before (blue) and after(red) cleaning.
A simulated dataset demonstrated the benefits of cleaning versus removing components. When a large number of components are labeled as artifactual, as is common in practice, the channel spectrum resulting from cleaning components is more similar to the non-artifactual channel spectrum as compared to removing components. A possible explanation for the benefit of cleaning components is that even artifactual components contain some cerebral signal. ICA cannot separate noisy signals such as EMG, as well as PCA. For this reason, artifactual signals can be spread across several components that contain cerebral signals.

In general, the artifact removal method presented in this work successfully identified and removed artifactual contributions to EEG data acquired with a 65-channel EGI net according to classification rates obtained by one expert. In order to generalize this method to EEG data acquired on other systems, further testing must be completed and additional experts should provide components labels. The concept of clustering components is not limited to EEG data. It could be extended to other neuroimaging data such as fMRI or DTI. Features can also be customized to detect cerebral signals as well as artifactual, extending ABEAR to the use of detecting event related responses.

V. CONCLUSIONS AND FUTURE WORK

This work evaluates the effectiveness of clustering for automated identification of artifactual EEG components as well as the benefits of cleaning components compared to removing components. The generation of components is improved through the automatic removal of bad epochs prior to ICA. The 8 1-dimensional features used for clustering demonstrate consistency across datasets. They also improve clustering speed through their low dimensionality. The 1-dimensional features were selected to represent features commonly used in manual classification. They are only a subset of possible features. A number of new features could be generated and tested using an algorithm like Sequential Forward Floating Search (SFFS) [33]. SFFS searches for the best combination of features to improve classification rates.

In this paper, an outlier criterion was used to label artifactual clusters. While the outlier detection resulted in acceptable recall and precision rates, it is not perfect and is dependent upon a threshold. Further exploration into possible methods for labeling clusters could improve classification rates and remove the dependency on a threshold.

REFERENCES


Fig. 4. Simulated data difference in spectra, dS, across all channels for three frequencies. The value for dS represents the difference between reconstructed data spectra, by either component removal, component cleaning, or the original data, and the non-artifactual data.

Fig. 5. Original (blue) and cleaned (red) components for two example datasets. The left figure magnifies an ECG and an EMG component while the right figure magnifies EOG components.


