A Comparison of Feature Extraction Methods for EEG Signals[1]

A. Moura, S. Lopez, I. Obeid and J. Picone
The Neural Engineering Data Consortium
Temple University
andersongmoura@gmail.com, iobeid@temple.edu, joseph.picone@gmail.com

Feature extraction for automatic interpretation of EEGs has been extensively studied. A number of commercial approaches use exotic feature sets such as wavelets or nonlinear statistical measures such as fractal dimension. These choices of features were the results of evaluations and optimizations conducted on small research databases often collected under very controlled conditions. These approaches have not been extensively evaluated on big data or clinical applications using state of the art machine learning technology. Therefore, in this study, we compare performance of a number of standard feature extraction techniques on the publicly available TUH EEG Corpus using a state of the art classification system.

The TUH EEG Corpus is the largest publicly available corpus of EEG data. It comprises over 28,000 sessions collected from 2002-2015 at Temple University Hospital. It is entirely composed of clinical data, which means the data is representative of all the problems typically encountered in clinical settings, such as patient movement, artifacts due to eye blinks, talking, etc. Such data poses a much different challenge for machine learning systems since rejection of background noise becomes a critical issue.

The classification system used for this study, known as AutoEEG™, automatically recognizes specific events in the EEG data and generates annotations. AutoEEG™ is based on a hidden Markov model (HMM) approach to modeling the temporal evolution of the spectrum. A maximum likelihood (ML) approach is used to train standard three-state HMMs consisting of 8 Gaussian mixtures per state and diagonal covariance matrices.

The system detects three events of clinical interest (PLED, GPED and SPSW) and three events used to model background noise (ARTF, EYEM and BCKG). The current system uses an enhanced feature extraction approach based on Mel Frequency Cepstral Coefficients (MFCC's) together with differential energy, first and second derivatives. In this study, we evaluated 15 features, shown to the right, by augmenting the standard feature vector with one additional feature. These were evaluated on a subset of the TUH EEG Corpus designed to give rapid turnaround on experiments, yet correlate well with results on the full dataset. This particular set of 15 features was chosen based on analysis of historically significant publications in the field.

Feature extraction	Mathematical definition
Integrated EMG (IEMG)	$\mathrm{IEMG} = \sum_{n=1}^{N} \lvert x_n \rvert$
Mean absolute value (MAV)	$\mathrm{MAV} = \frac{1}{N} \sum_{n=1}^{N} \lvert x_n \rvert$
Modified Mean Absolute Value (MMAV)	$\mathrm{MMAV} = \frac{1}{N} \sum_{n=1}^{N} w_n \lvert x_n \rvert$; $w_n = \begin{cases} 1, & \textit{if } 0.25N \le n \le 0.75N \\ 0.5, & \text{otherwise} \end{cases}$
Simple Square Integral (SSI)	$\mathrm{SSI} = \sum_{n=1}^{N} \lvert x_n \rvert^2$
Variance of EMG (VAR)	$\mathrm{VAR} = \frac{1}{N-1} \sum_{n=1}^{N} x_n^2$
Root Mean Square (RMS)	$\mathrm{RMS} = \sqrt{\frac{1}{N} \sum_{n=1}^{N} x_n^2}$
v-Order 2 and 3 (V2, V3)	$\mathrm{V2} = \left(\frac{1}{N} \sum_{i=1}^{N} x_i^2 \right)^{\frac{1}{2}}; \mathrm{V3} = \left(\frac{1}{N} \sum_{i=1}^{N} \lvert x \rvert_i^3 \right)^{\frac{1}{3}}$
Log detector (LOG)	$\mathrm{LOG} = e^{\frac{1}{N} \sum_{n=1}^{N} \log(\lvert x_n \rvert)}$
Waveform length (WL)	$\mathrm{WL} = \sum_{n=1}^{N-1} \lvert x_{n+1} - x_n \rvert$
Average amplitude change (AAC)	$\mathrm{AAC} = \frac{1}{N} \sum_{n=1}^{N-1} \lvert x_{n+1} - x_n \rvert$
Difference absolute standard deviation value (DASDV)	$\mathrm{DASDV} = \sqrt{\frac{1}{N-1} \sum_{n=1}^{N-1} (x_{n+1} - x_n)^2}$
Maximum Fractal Length (MFL)	$\mathrm{MFL} = \log_{10} \left(\sqrt{\sum_{n=1}^{N-1} (x_n - x_{n+1})^2} \right)$
Myopulse percentage rate (MYOP)	$\mathrm{MYOP} = \frac{1}{N} \sum_{i=1}^{N} [f(x_i)]$; $f(x) = \begin{cases} 1, & \textit{if } x \ge \textit{threshold} \\ 0, & \text{otherwise} \end{cases}$
Total power (TTP)	$\mathrm{TTP} = \sum_{j=1}^{M} P_j$
Willison amplitude (WAMP)	$\mathrm{WAMP} = \sum_{n=1}^{N-1} f(\lvert x_n - x_{n+1} \rvert)$; $f(x) = \begin{cases} 1, & \textit{if } x \ge \textit{threshold} \\ 0, & \text{otherwise} \end{cases}$

The MFCC approach has been in use for speech recognition applications for several decades and is known to provide a robust characterization of the temporal and spectral properties of the signal. In general, our findings indicate that any of these features individually influence performance very little. This contradicts findings previously published, but was not unexpected. Clinical data is extremely challenging and quite different from most published EEG corpora. Maximum Fractal Length (MFL) provided the greatest reduction in error rate, though the improvements were not statistically significant.

In related work, we demonstrate that wavelets, which are often proposed as an alternative to MFCCs, also provide no gain in performance. Though literature suggests that EEG signals can be viewed as chaotic time series with significant amounts of nonlinearities, the features we investigated, which are designed to characterize such properties, add little value to our standard feature extraction approaches. Future research will be focused on better time-frequency representations of the signal based on correlation and coherence.

1. Research reported in this publication was supported by the National Human Genome Research Institute of the National Institutes of Health under Award Number U01HG008468.
2. This research was also supported by the Brazil Scientific Mobility Program (BSMP) and the Institute of International Education (IIE).

NEURAL ENGINEERING DATA CONSORTIUM
www.isip.piconepress.com

FEATURE EXTRACTION METHODS FOR EEG EVENT DETECTION

College of Engineering
Temple University

Anderson G. Moura, Dr. Iyad Obeid and Dr. Joseph Picone

The Neural Engineering Data Consortium, Temple University

Abstract

- The emergence of big data and deep learning is enabling the ability to automatically learn how to interpret EEGs from a big data archive.
- The AutoEEG™ is a system that automatically recognizes specific events in the EEG data and generates annotations.
- The system detects three events of clinical interest (PLED, GPLE and SPSW) and three events used to model background noise (ARTF, EYEM and BCKG).
- The current system uses an enhanced feature extraction approach based on Mel Frequency Cepstral Coefficients (MFCC's) together with differential energy, first and second derivatives.
- This study evaluated a range of features by augmenting the standard feature vector with one additional feature.
- Maximum Fractal Length (MFL) provided the greatest reduction in error rate, though the improvements were not statistically significant.
- None of the features improved performance over the baseline MFCC approach.

Introduction

- Electroencephalography (EEG) measures the electrical activity in the brain and is used to diagnose patients suffering from neurological disorders such as epilepsy and strokes.
- AutoEEG™ uses a speech recognition approach for classifying 1 second epochs of an EEG signal into one of events: generalized periodic epileptiform discharge (GPED), periodic lateralized epileptiform discharge (PLEDs), spike and sharp wave (SPSW), artifact (ARTF), eye movement (EYEM), and background activity (BCKG).

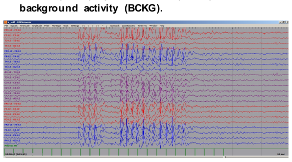


Figure 1: An example of a spike

- AutoEEG™ is based on a hidden Markov model (HMM) approach to modeling the temporal evolution of the spectrum.
- A maximum likelihood (ML) approach is used to train standard three-state HMMs consisting of 8 Gaussian mixtures per state and diagonal covariance matrices.
- A frame duration of 0.1 secs is used to model 1second epochs of the signal.
- An ML approach is used for classification.

Mel Frequency Cepstral Coefficients

- Machine learning algorithms based on hidden Markov models and deep learning are used to learn mappings of EEG events to diagnoses.
- The system accepts multichannel EEG raw data files as input. Desired output is a transcribed signal and a probability vector with various probable diagnoses.
- Currently a filter bank-based cepstral analysis (MFCC) is used to convert EEG signals to features.
- The signal is analyzed in 1 sec epochs using 100 msec frames. HMMs are used to map frames to epochs and classify epochs.

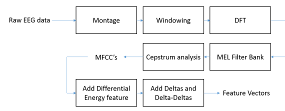


Figure 2: Feature Extraction Process

- A differential energy feature is defined as the difference between the maximum and minimum energy in a window (typically 9 secs in duration).
- The performance of a pattern recognition system can be greatly enhanced by adding time derivatives to the basic static parameters. Derivatives are calculated using a standard regression approach.
- The delta features are calculated using a window of 5 frames centered about the current frame.
- The delta-delta features (acceleration) are calculated in the same way as the delta coefficients, but over the delta coefficients instead of over the static coefficients.
- Derivatives accentuate spectral dynamics.

Feature Extraction Methods

- Feature extraction reduces the sampled data sequence to a sequence of vectors that contain the most relevant information for classification:

Feature extraction	Mathematical definition
Integrated EMG (IEMG)	$\mathrm{IEMG} = \sum_{n=1}^{N} \lvert x_n \rvert$
Mean absolute value (MAV)	$\mathrm{MAV} = \frac{1}{N}\sum_{n=1}^{N} \lvert x_n \rvert$
Modified Mean Absolute Value (MMAV)	$\mathrm{MMAV} = \frac{1}{N}\sum_{n=1}^{N} w_n \lvert x_n \rvert$; $w_n = \begin{cases} 1, & \text{if } 0.25N \le n \le 0.75N \\ 0.5, & \text{otherwise} \end{cases}$
Simple Square Integral (SSI)	$\mathrm{SSI} = \sum_{n=1}^{N} \lvert x_n \rvert^2$
Variance of EMG (VAR)	$\mathrm{VAR} = \frac{1}{N-1}\sum_{n=1}^{N} x_n^2$
Root Mean Square (RMS)	$\mathrm{RMS} = \sqrt{\frac{1}{N}\sum_{n=1}^{N} x_n^2}$
v-Order 2 and 3 (V2, V3)	$\mathrm{V2} = \left(\frac{1}{N}\sum_{i=1}^{N} x_i^2\right)^{\frac{1}{2}}$; $\mathrm{V3} = \left(\frac{1}{N}\sum_{i=1}^{N} \lvert x \rvert_i^3\right)^{\frac{1}{3}}$
Log detector (LOG)	$\mathrm{LOG} = e^{\frac{1}{N}\sum_{n=1}^{N}\log(\lvert x_n \rvert)}$
Waveform length (WL)	$\mathrm{WL} = \sum_{n=1}^{N-1} \lvert x_{n+1} - x_n \rvert$
Average amplitude change (AAC)	$\mathrm{AAC} = \frac{1}{N}\sum_{n=1}^{N-1} \lvert x_{n+1} - x_n \rvert$
Difference absolute standard deviation value (DASDV)	$\mathrm{DASDV} = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N-1}(x_{n+1} - x_n)^2}$
Maximum Fractal Length (MFL)	$\mathrm{MFL} = \log_{10}\left(\sqrt{\sum_{n=1}^{N-1}(x_n - x_{n+1})^2}\right)$
Myopulse percentage rate (MYOP)	$\mathrm{MYOP} = \frac{1}{N}\sum_{i=1}^{N}[f(x_i)]$; $f(x) = \begin{cases} 1, & \text{if } x \ge \text{threshold} \\ 0, & \text{otherwise} \end{cases}$
Total power (TTP)	$\mathrm{TTP} = \sum_{j=1}^{M} P_j$
Willison amplitude (WAMP)	$\mathrm{WAMP} = \sum_{n=1}^{N-1} f(\lvert x_n - x_{n+1} \rvert)$; $f(x) = \begin{cases} 1, & \text{if } x \ge \text{threshold} \\ 0, & \text{otherwise} \end{cases}$

Figure 3: Mathematical definitions for a variety of features evaluated in this study

Experimental Design

- A pilot study was conducted on a small data set of 12 EEG sessions for training and an independent set of 12 EEGs for evaluation. This data contains a rich variety of signal events.
- This small set was chosen so that parameter tuning experiments could be conducted quickly.
- The data was sampled at 250 Hz and analyzed using a frame duration of 0.1 secs and an analysis window duration of 0.2 secs (50 samples).

Methods

- The MFCC coefficients for each EDF file (EEG Signals) are stored in one HTK file per channel before the derivatives computation.
- The selected new feature is calculated in a per window basis over each channel of the EEG signals and added to the respective HTK file immediately after the MFCC's.
- The derivatives are then computed over each window (feature vector) resulting in a total of 30 features per vector.

Baseline Performance

- An error confusion matrix for the HMM-based system (MFCC's):

	SPSW	PLED	GPED	EYEM	ARTF	BCKG
SPSW	5.30%	23.48%	13.64%	32.58%	3.79%	21.21%
PLED	11.19%	53.73%	23.88%	2.99%	0.75%	7.46%
GPED	2.40%	28.80%	68.80%	0.00%	0.00%	0.00%
EYEM	0.00%	18.87%	9.43%	64.15%	7.55%	0.00%
ARTF	0.00%	0.00%	0.00%	0.48%	83.09%	16.43%
BCKG	4.47%	6.22%	0.76%	0.11%	14.83%	73.61%

- Accurate detection of the SPSW class is most important since it is the most important indicator of a potential neurological disorder.
- Additional analytics can be applied to data labeled as PLED or GPED.
- Collapsing the background noise classes into a single class gives this confusion matrix:

	BCKG	SPSW	GPED	PLED
BCKG	96.26%	0.25%	2.12%	1.36%
SPSW	35.61%	0.76%	40.91%	22.73%
GPED	4.00%	1.60%	53.60%	40.80%
PLED	11.19%	4.48%	18.66%	65.67%

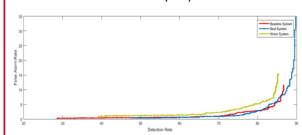
- The detection error rate for 6 classes is 33.2% and 17.8% for the collapsed 4 classes.
- Additional post processing steps are used to further improve performance, but these were not applied in this study.

Preliminary Results

- Detection error rates:

	6 Classes	4 Classes
MFCCs	33.2%	17.8%
+ IEMG	28.0%	19.2%
+ MAV	29.0%	19.9%
+ MMAV	27.0%	19.6%
+ SSI	26.9%	19.8%
+ VAR	26.6%	19.5%
+ RMS	26.3%	19.2%
+ V3	25.8%	18.6%
+ LOG	80.2%	23.5%
+ WL	25.8%	18.4%
+ AAC	26.5%	18.6%
+ DASDV	27.2%	18.9%
+ MFL	25.3%	17.6%
+ MYOP	32.4%	18.1%
+ WAMP	30.2%	17.8%
+ TTP	26.0%	19.5%
+ MDF	26.7%	19.3%

- A Detection Error Tradeoff (DET) curve:



- For low false alarm rates, which is the most important area of the DET curve for this application, performance is comparable.
- The additional feature typically increases computation time by 14%.

Summary

- The results presented here were obtained using a small pilot corpus that is designed to give rapid turnaround on experiments.
- Our preliminary results show that features such as the Modified Fractal Length and Willison amplitude can improve performance slightly.
- Additional experiments need to be run on the entire TUH EEG Corpus.
- Experiments investigating combinations of these features and optimal ways to weight these combinations will yield more insight into the potential benefits of an expanded feature set.
- Additional features based on frequency domain information (e.g., frequency ratio) will be explored.

Acknowledgements

- This research was also supported by the Brazil Scientific Mobility Program (BSMP) and the Institute of International Education (IIE).