Proceedings Paper:
Toth, J. and Arvaneh, M. (2017) Facial Expression Classification Using EEG and Gyroscope Signals. In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), July 11 to 15, 2017, Seogwipo, South Korea. IEEE , pp. 1018-1021.

https://doi.org/10.1109/EMBC.2017.8036999

Facial expression classification using EEG and gyroscope signals

Jake Toth[1] and Mahnaz Arvaneh[1]

Abstract—In this paper muscle and gyroscope signals provided by a low cost EEG headset were used to classify six different facial expressions. Muscle activities generated by facial expressions are seen in EEG data recorded from scalp. Using the already present EEG device to classify facial expressions allows for a new hybrid brain-computer interface (BCI) system without introducing new hardware such as separate electromyography (EMG) electrodes. To classify facial expressions, time domain and frequency domain EEG data with different sampling rates were used as inputs of the classifiers. The experimental results showed that with sampling rates and classification methods optimized for each participant and feature set, high accuracy classification of facial expressions was achieved. Moreover, adding information extracted from a gyroscope embedded into the used EEG headset increased the performance by an average of 9 to 16%.

I. Introduction

A brain-computer interface (BCI) is a system that allows an external device to be controlled by brain activities [1,2]. Thus, BCI enables severely disabled people such as tetraplegic / quadriplegic people to control assistive devices through their brain activities [2]. Electroencephalography (EEG) is a noninvasive technique that measures electrical brain activity through electrodes placed on the scalp. The majority of BCI systems use EEG for measuring brain activities due to its low cost and ease of use [1]. EEG based BCI is currently used for external device control through activities such as motor imagination [3] or P300 [4].

One of the challenges in EEG-based BCI is the high level of noise and artifacts present in EEG signals [5]. A bi-product of having electrodes on the scalp is that local muscle signals are inadvertently detected. The magnitude of the voltage from muscle signals is several orders of magnitude greater than that of brain signals [6]. As such muscle induced voltages in EEG systems are often treated as artifacts to be removed [7].

Hybrid BCI is a new approach to develop a more practical and accurate BCI by fusing data from other modalities with EEG signals [8]. In such a system disabled users are able to use more of their remaining functionalities as control possibilities in parallel with the BCI. There are 282,000 people with spinal cord injury in the United States. 58.3% of which suffer from tetraplegia [9]. Being tetraplegic / quadriplegic results in limited voluntary muscle movement below the neck. However, most of these people exhibit voluntary control of their facial muscles. Therefore, a hybrid BCI system that fuses brain activities in parallel with information extracted from facial muscle signals can be potentially used in controlling assistive devices. Muscle activities generated by facial expressions are seen in EEG data recorded from scalp. Using this already present EEG device to classify facial expressions allows for a new hybrid BCI without new hardware such as separate electromyography (EMG) electrodes.

Currently facial expressions can be detected through methods such as image processing [10] or EMG [11]. There are a number of inherent downsides to using image processing, namely the need for a camera facing the user's face any time the system wants to be used. Facial expression detection through image processing also has limited functionality if light and pose are suboptimal [10].

For the use of EMG, assuming non-invasive methods, the signal acquired is noisy, but with the correct electrode placement facial expressions are detectable [11]. Using the EEG headset electrodes on the forehead is effectively the same as using an EMG system. However in this case there will only be two electrodes on the forehead as opposed to many electrodes across the face covering all major muscle groups. In this paper we develop a facial expression classification method based on EEG and gyroscope data. For this purpose, data from an Emotiv Epoc+ was collected. The Epoc+ is relatively cheap compared to most other EEG recording devices [12]. The Epoc+ also contains a two axis gyroscope which allows for further degrees of control when movements such as head shaking is introduced. Among available Epoc+ electrodes, the frontal AF3 and AF4 electrodes are used as the most prone to interference from facial muscle signals.

For this paper both time and frequency feature sets are explored to extract information from the EEG signals. The output of different facial expressions in both of these domains is analyzed using data collected from four participants. During preprocessing, for each participant the optimal features are chosen with the intent of increasing the classification accuracy. Multiple machine learning classifiers are used with the optimal classification technique for each participant and feature set being chosen. In addition, the gyroscope data is investigated in similar manor, with both the time and frequency domain features of all facial expressions investigated. A technique that allows for the beneficial use of the gyroscope is explored.

II. Experiment

A. Participants

Four young adults between 20 and 22 years of age with no history of neurological illness of involuntary muscle movements participated in this study. Three males and one female participated with informed consent given.

[1]M. Arvaneh and J. Toth are with the Dept. of Automatic Control & Systems Engineering at the University of Sheffield. (email:jtoth1@sheffield.ac.uk, m.arvaneh@sheffield.ac.uk)

B. EEG data acquisition

An Emotiv Epoc+ [12] was used to collect EEG and gyroscope data. The Emotiv Epoc+ is a wireless EEG headset with 14 EEG channels and 2 reference channels. Compared to many other EEG headsets, the Epoc+ headset is cost effective, wireless, and non-gel based. Thus it represents a more practical and economic EEG acquisition device for daily applications out of the laboratory. Moreover, the two-axis gyroscope embedded in the headset provides information about head movements that can be used as complementary information in many EEG-based applications.

In this study, data from AF3, AF4 and the gyroscope were used for detecting different facial expressions. According to the international 10-20 electrode placement system channels AF3 and AF4 are placed on the upper forehead. These electrode positions are associated with the largest interference from facial muscle signals. The sample rate of the headset is 128 Hz. A saline solution was used on all relevant electrodes to ensure appropriate impedance levels [12].

C. Facial expression task

Participants were requested to perform six facial expressions while in an isolated room. These six facial expressions consisted of smile, clench, blink, eyebrow raise, headshake and neutral. Each facial expression was performed sixteen times. To allow for sufficient breaks, sessions were split into four blocks. These four blocks consisted of each facial expression being performed four times. Between each block there was a two to five minute break. Each action (i.e. facial expression) was preceded by a 4 second relaxation period, where the participant could move freely. Followed by a four second preparation period where participants saw both text and a visual cue for the upcoming action to perform. Finally, there was a two second period of being able to perform the action while a cross was shown on the screen (see Fig.1). During this period participants were asked to remain relaxed and minimize any body movement other than what was requested, and remaining focused on the screen. This was immediately followed by the next four second relaxation period, as the sequence repeated.

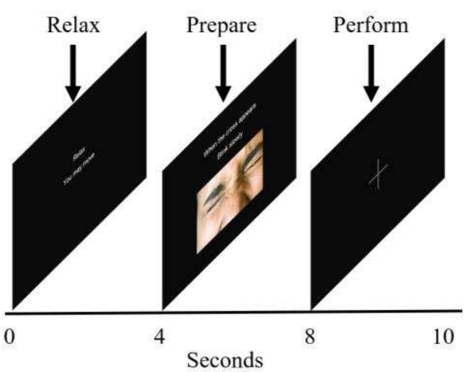


Figure 1. User interface for facial expression data collection

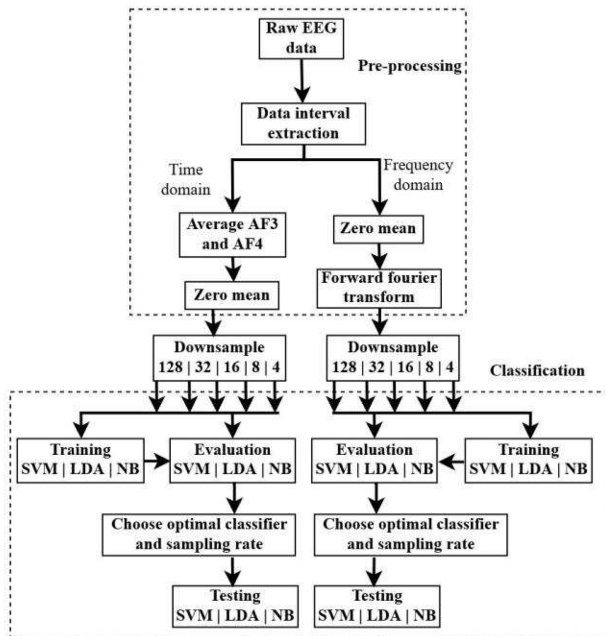
III. METHODOLOGY

A. Preprocessing and feature selection

To allow for optimal classification the raw EEG data was preprocessed and features were extracted. The two main feature sets were based on the time and frequency domains.

The process for the time domain went as follows; front channels AF3 and AF4 were averaged. This reduced the potential noise in any individual channel affecting performance. Zero mean was then applied to this data, followed by the application of down sampling at several different rates. Finally, the down-sampled data were used as the time-domain features.

For the frequency domain the zero mean was initially taken followed by a logarithmic forward fourier transform. Using the logarithmic frequency response allows for lower frequencies to be better represented. This data also went through down sampling at several different rates. All five sampling rates shown in Fig.2 were used for all participants. The optimal down sampling rates were chosen during the evaluation stage of classification. The optimal down sampling rates were chosen during the evaluation stage of classification.



Acronyms used: SVM (Support Vector Machine), LDA (Linear Discriminant Analysis) and NB (Naive Bayesian)

Figure 2. Schematic illustration of the applied EEG signal processing procedure

The gyroscope data was preprocessed using the method illustrated in Fig.3. Firstly the X axis gyroscope data had zero mean applied, followed by the absolute value of the data being taken. The mean of the resulting dataset was consequently used for classification.

B. Classification

Classification of the EEG time and frequency domain data went through three stages. Firstly, support vector machine, linear discriminant analysis and naive Bayesian classifiers were trained. This was achieved using seven out of the sixteen recorded trials of each facial expression. All five sampling rates were also used at this stage. This resulted in fifteen total trained classifiers for each participant and feature set.

The evaluation stage involved testing the performance of all fifteen trained classifiers on three new trials of each facial expression.

An average accuracy for each sampling rate and classifier combination was found, the combination with the greatest average classification accuracy was then chosen as the optimal combination for each participant for that particular feature set. Finally, using this optimal combination classifier accuracy was then tested by applying the final six performances of each facial expression.

Preprocessed gyroscope data was simply compared to a threshold value of 500mV. If the preprocessed data surpassed this threshold it was classified as a headshake, if not the EEG classification results were used to determine the class of the data. When the gyroscope data was in use the machine learning classifiers were no longer trained for the head-shake class.

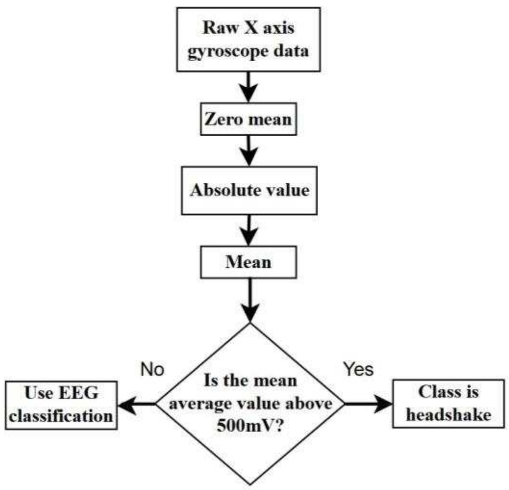
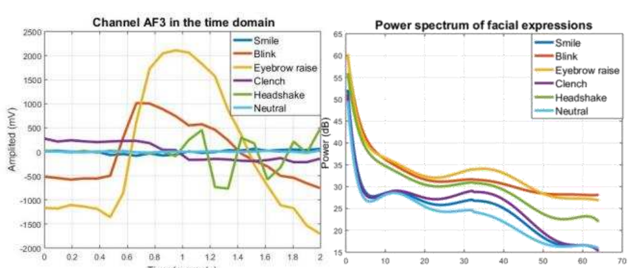


Figure 3. Schematic illustration of the applied gyroscope signal processing procedure

IV. RESULTS

A. EEG data in the time domain

Each facial expression had their respective time domain response, which is shown in Fig.4(a). Notably eyebrow raise had a distinctly greater amplitude than the other classes. This is likely due to the proximity of the muscles associated with eyebrow movements to the EEG electrodes AF3 and AF4. The smile class appears similar to the neutral class in the time domain. This similarity constitutes a lower performance for this feature set and these classes.



(a) (b)

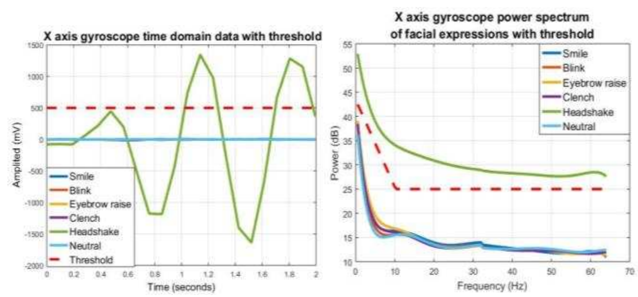
Figure 4. EEG data for 6 facial expressions
(a) Time domain (b) Frequency domain

B. EEG data in the frequency domain

The frequency domain response has visually clearer separations between classes as shown in Fig.4(b). Notably the 30Hz - 40Hz range has a clear separation of classes. However, headshake and blink or smile and clench are still relatively close. In the 20Hz - 30Hz range or 35Hz - 45Hz range these respective pairs of classes are further separated. As a result of this observation, and through testing, using the full range of frequencies produced a higher average classification accuracy as opposed to a smaller range of frequencies.

C. Gyroscope data

The gyroscope response of all facial expressions other than the head shaking was very similar. Classification of these facial expressions was therefore at or below chance level. As a result, head shaking gyroscope data was compared to all other facial expressions in a binary fashion. As seen in Fig.5(a) and Fig.5(b) both time and frequency responses have clear differences between head shaking and other facial expressions. A threshold could, therefore be applied to either domain. However, for the purpose of this paper only the time domain was used.



(a) (b)

Figure 5. Gyroscope data for 6 facial expressions with a threshold limit
(a) Time domain (b) Frequency domain

D. Optimal sampling rate and classification method

For each feature set and participant, the optimal sampling rate and machine learning classification method was found. During the evaluation stage all three classification methods and all five sampling rates were used, the combination with the greatest overall accuracy was then chosen and used in the testing stage.th

In Fig.6 the optimal sampling rates and machine learning classification methods chosen for all participants are shown. There was not a clear benefit to choosing a specific sampling rate and applying it universally. Sampling rates of 128Hz and 8Hz were optimal in 23% of cases, the least commonly optimal sampling rates were 16Hz and 4Hz which were optimal in 17% of cases. There also was not a clear best sampling rate for a particular feature set or participant. As a result, deciding on the best sampling rate on a participant and feature set basis was the most effective method.

The optimal machine learning method to use was linear discriminant analysis in 51% of cases followed by Naïve Bayesian in 43% of cases and support vector machine in only 6% of cases. Once again there was no clear correlation between any particular feature set or sampling rate and the optimal machine learning method.

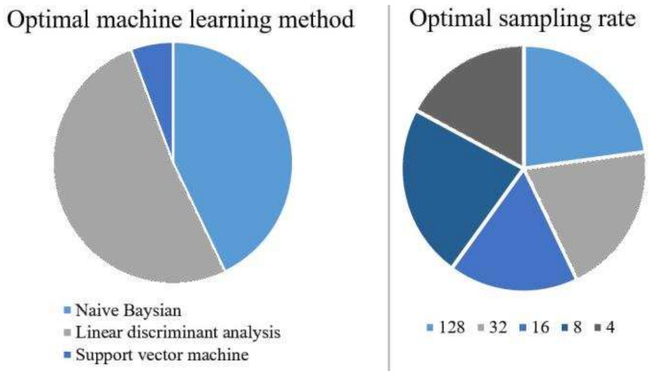


Figure 6. Optimal sampling rate and classification method
(a) Sampling rate (b) Machine learning methods

E. Classification

After applying the classification techniques described previously, and finding the mean average across all participants the results shown in Fig.7 were obtained. Notably the inclusion of the gyroscope increased classification accuracy by an average of 4.6%.

The classes with the least separation visible in Fig.4, as expected resulted in having the lowest classification accuracy, with the smile class averaging 50%. Classes with clear separation such as eyebrow raise and neutral classes had much higher performance with 83.4% and 82.3% respectively.

As shown in Fig.4 the separation between classes in the time domain was not as great as that of the frequency domain. This observation translated to results which reflect this characteristic. The time domain response without the gyroscope had a mean classification accuracy of 59.7% compared to 68.1% in the frequency. The inclusion of the gyroscope reduces this difference significantly with the time and frequency domain differences going from 8.4% to 0.7% with and without the gyroscope respectively. This is likely due to the false classification of other classes being classified as head shaking in the time domain. However, with the inclusion of the gyroscope, and the lack of classifier training for the headshake movement this false classification no longer occurred.

	Average classification accuracy (%)				
	EEG		**Gyroscope**		
Classes	**Time**	**Frequency**	**Time**	**Frequency**	**Mean**
Smile	54.2	37.5	62.5	45.8	50.0
Blink	33.3	54.2	62.5	66.7	54.2
Eyebrow raise	91.7	79.2	79.2	83.3	83.4
Clench	66.7	79.2	75	70.8	72.9
Headshake	54.2	62.5	100	100	79.2
Neutral	58.3	95.8	79.2	95.8	82.3

Table 1. Table of average classification accuracies across participants

V. Conclusion

In this study it was shown that using the frontal electrodes of an EEG headset and a gyroscope, average classification accuracy of over 75% with six classes is achievable. This performance was achieved using time and frequency domain feature sets with optimal sampling rates and machine learning classifiers for each participant and feature set. This performance was consistent over multiple participants with a relatively small training and evaluation data set of seven and three instances of each class respectively. The introduction of a gyroscope increased performance by an average of 4.6%, it is therefore a positive asset to have.

The system proposed in this paper could be used in a hybrid BCI system to allow for new modes of control of an assistive device without the introduction of new hardware.

References

[1] L. F. Nicolas-Alonso and J. Gomez-Gil, “Brain computer interfaces, a review,” Sensors, vol. 12, no. 2. pp. 1211–1279, 2012.

[2] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, “Brain-computer interfaces for communication and control,” Clin Neurophysiol, vol. 113, no. 6, pp. 767–791, 2002.

[3] G. Pfurtscheller, G. R. Müller, J. Pfurtscheller, H. J. Gerner, and R. Rupp, “‘Thought’ - Control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia,” Neurosci. Lett., vol. 351, no. 1, pp. 33–36, 2003.

[4] F. Nijboer et al., “A P300-based brain-computer interface for people with amyotrophic lateral sclerosis,” Clin. Neurophysiol., vol. 119, no. 8, pp. 1909–1916, 2008.

[5] D. J. McFarland, W. A. Sarnacki, T. M. Vaughan, and J. R. Wolpaw, “Brain-computer interface (BCI) operation: signal and noise during early training sessions,” Clinical Neurophysiology, vol. 116, no. 1, pp.56-62, 2005.

[6] S. D. Muthukumaraswamy, “High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations,” Front. Hum. Neurosci., vol. 7, no. April, p. 138, 2013.

[7] T.-P. JUNG et al., “Removing electroencephalographic artifacts by blind source separation,” Psychophysiology, vol. 37, no. 2, p. S0048577200980259, 2000.

[8] G. Pfurtscheller, B. Z. Allison, G. Bauernfeind, C. Brunner, T. Solis Escalante, R. Scherer, T. O. Zander, G. Mueller-Putz, C. Neuper, and N. Birbaumer, “The hybrid BCI,” Frontiers in neuroscience, vol. 4, p.3. 2010.

[9] NSCISC. (2016). Spinal cord injury (sci) facts and figures at a glance, [Online]. Available: https://www.nscisc.uab.edu/, Last Accessed: 11/22/2016

[10] R. Mittal, P. Srivastava, A. George, and A. Mukherjee, “Autonomous Robot Control Using Facial Expressions,” Int. J. Comput. Theory Eng., vol. 4, no. 4, pp. 631–635, 2012.

[11] C.-N. Huang, C.-H. Chen, and H.-Y. Chung, “The Review of Applications and Measurements in Facial Electromyography,” J. Med. Biol. Eng., vol. 25, no. 1, pp. 15–20.

[12] Emotiv, https://www.emotiv.com/, Last Accessed 11/02/2017