Classifying Eye and Head Movement Artifacts in EEG Signals

Neisha A. Chadwick, David A. McMeekin, Tele Tan
Digital Ecosystems and Business Intelligence Institute (DEBII)
Curtin University
Perth, Australia.
D.McMeekin@curtin.edu.au

Abstract—**Brain Computer Interfaces has some exciting prospects such as controlling devices at the speed of thought. However BCI technology is far from attaining this goal. A significant challenge the EEG-based system has is the interference of artifacts in the EEG generated by eye and head movement. This paper presents the use of machine learning techniques to classify artifacts in the EEG. Successful artifact classification was then be applied to improve existing artifact removal techniques. The experiment used a state-of-the-art EEG system to gather the classifier input. An eye tracker and motion sensor were also used to measure and provide the ground truth for the classification experiments. The data from these devices were captured using custom built software developed for this research. The classifiers tested showed potential to classify artifacts in the EEG when trained on a per-person basis. This research paves the way for further work to be carried out to explore subject-independent artifact classification.**

Keywords-component; eye movement; head movement; classifying, classifiers, eeg artifacts.

I. INTRODUCTION

A Brain Computer Interface (BCI) is a direct communication method between the brain and an external device. BCI research is targeted at developing new supplemental communication and control technology for people with severe motor impairments. This can be in the form of basic communication with a computer or controlling neuroprostheses. BCI technology can allow severely disabled people to function independently, giving a revolutionary improvement to their quality of life [1].

A BCI works by detecting changes of electrophysiological signals caused by mental activity that relies on electrodes to pick up the electrical signals generated through the scalp of the head. It then transforms that signal into a control signal that can be used by such a system.

Despite the increase human brain understanding over recent years, much still remains unknown. This lack of understanding makes detecting and interpreting these signals difficult.

What compounds this is that the electroencephalography (EEG) signal is weak and easily effected by noise. One of the most disruptive noise forms is an artifact generated by the subject. Artifacts are any undesired signal that interferes with the brain signals being measured. The most prevalent artifact cause is ocular and head movements. Artifacts introduce significant changes to the brain signals and even mimic control signals.

A useful way to handle artifacts is to remove them from the EEG. Several artifact removal methods have been developed [3], however they are not yet optimal and suffer from a variety of issues including (and not limited to) a dependency on electrooculography (EOG) and electromyography (EMG) measurements, distortion and/or erosion of the EEG signal, and the requirement of manual intervention.

This study focused on classifying eye and head movement-related artifacts solely within the EEG, without the use of EOG and EMG recordings rather than trying to remove artifacts from the EEG. This was attempted through an innovative approach using state of the art equipment. A current generation EEG cap and amplifier to gather EEG data as input to the classifier, a mobile eye tracker to detect the ground truth of where the eyes are moving, and wireless motion sensors to detect the ground truth head movement were used. The amalgamation of these instruments is made possible via the development of custom-built software.

From a practical perspective, the research presented in this paper lays a foundation for an artifact classification system that could be used in current and future BCI systems. With further extension it could be used to augment current and future artifact removal techniques.

II. BACKGROUND

A (BCI) is a system that allows a user to control a device, such as a computer, using non-muscular (neurological) communication. The most suitable and practical method for monitoring brain activity for use in a BCI system is using an EEG [13].

The building blocks of the brain are neurons. Neurons are individual nerve cells connected by the dendrites that receive signals, and the axon that sends signals. Functioning neurons generate small electrical signals that transfer between neurons. The EEG measures this electrical signal via electrodes placed on the scalp. In a BCI, the signal is interpreted and used for communication or to control a device.

One issue plaguing the BCI field is artifact presence in the EEG signal. Artifacts are any undesired potentials, mostly of

Table 1. Participant movement.

		Marker	
	Movement type	Origin	Destination
1	Eyes only	C	L1
2	Eyes only	C	L2
3	Eyes only	C	R1
4	Eyes only	C	R2
5	Eyes only	C	U1
6	Eyes only	C	U2
7	Eyes only	C	D1
8	Eyes only	C	D2
9	Head only	C	L1
10	Head only	C	L2
11	Head only	C	R1
12	Head only	C	R2
13	Head only	C	U1
14	Head only	C	U2
15	Head only	C	D1
16	Head only	C	D2
17	Eyes + head	C	Left
18	Eyes + head	C	Right
19	Eyes + head	C	Up
20	Eyes + head	C	Down
21	Blink	None	None

non-cerebral origin, that can change the brain signals, usually affecting the neurological phenomenon being studied. The most common and influential artifact source are electrical signals generated from EOG eye movements and EMG particularly of the head [3]. These potentials distort the electric fields over the scalp, contaminating the EEG. EMG artifacts are generated by the muscle electrical activity. This electrical activity propagates over the entire scalp surface, with the greatest effect seen at electrode sites closest to the origin.

Currently there are no ideal solutions to the artifact problem in EEGs. Three ways have been put forward to handle artifacts: avoidance, rejection or removal.

Avoidance: requires that the user to refrain from any actions that may cause artifacts, such as blinking, eye movement and body movement. It has been shown that such instructions can introduce additional cognitive load resulting in EEG changes [8,11].

Rejection: involves discarding trials that contain artifacts, either manually or automatically. Manual artifact rejection requires an expert to visually inspect the data for artifacts while automatic artifact rejection is performed off-line using pre-determined thresholds. Each method can result in a sampling bias and substantial data loss [7,12].

Removal: is a more practical artifact handling method and can also keep the related neurological phenomenon intact. A common artifact removal method involves subtracting the EOG signal (measured using electrodes placed around the eyes) from the contaminated EEG using linear combination and regression [4].

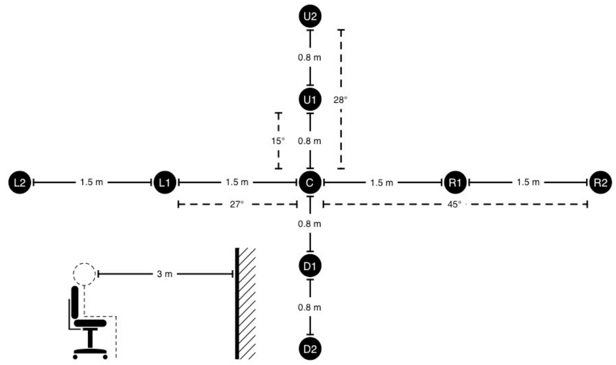


Figure 1. Marker configuration place on wall.

Commonly used in research into EOG-based control is the application of a high-pass filter, with varying cut-offs at frequencies between 0.05 and 0.2 Hz. This process removes the long-term drift inherent in all channels of the highly sensitive EEG amplifier [2]. Over short instances of time (e.g. less than 10 seconds) the drift is negligible given that the subject is at rest. During this phase, the variation in EOG is nearly collinear to the glance angle within the customary field of view [5,6].

III. EXPERIMENT DESIGN AND METHODOLOGY

The research objective was to classify artifacts in the EEG. To achieve this goal, a substantial amount of training data consisting of examples of the artifacts to classify was needed. Most of the EEG datasets available contain examples of motor imagery (imagination of motor functions such as moving left and right hands) [10], yet there are no EEG datasets containing artifact examples. As such, it was necessary to design an experiment to collect the required data. Whilst capturing the EEG data, an eye tracker and motion sensor was also used to provide ground truth.

Data quality is important because any patterns located in the signal can potentially be very subtle and easily drowned out by noise. This study's goal was to classify artifacts for use in BCI systems, hence the artifacts chosen needed to reflect commonly occurring artifacts.

The artifacts chosen to classify are generated from common movements a user would make whilst using a BCI system. These movements include:

- Eye movements only: Moving the eyes left, right, up and down whilst keeping the head still.
- Head movements only: Moving the head left, right, up and down whilst keeping the eyes still. This movement involves moving the eyes with the head in order to keep the eyes still.
- Joint Eyes and head movement: Keeping the eyes fixed in the center whilst moving the head left, right, up and down. This movement generates movement in both the eyes and head at the same time.
- Eye Blink: A natural blink, keeping the eyes and head still.

Table 2. Number of instances per class.

	Class description	# instances
1	Eyes only $C \rightarrow L1$	208
2	Eyes only $C \rightarrow L2$	209
3	Eyes only $C \rightarrow R1$	207
4	Eyes only $C \rightarrow R2$	210
5	Eyes only $C \rightarrow U1$	209
6	Eyes only $C \rightarrow U2$	207
7	Eyes only $C \rightarrow D1$	205
8	Eyes only $C \rightarrow D2$	209
9	Head only $C \rightarrow L1$	207
10	Head only $C \rightarrow L2$	208
11	Head only $C \rightarrow R1$	207
12	Head only $C \rightarrow R2$	211
13	Head only $C \rightarrow U1$	208
14	Head only $C \rightarrow U2$	210
15	Head only $C \rightarrow D1$	210
16	Head only $C \rightarrow D2$	210
17	Eyes + head $C \rightarrow Left$	211
18	Eyes + head $C \rightarrow Right$	211
19	Eyes + head $C \rightarrow Up$	208
20	Eyes + head $C \rightarrow Down$	210
21	Blink	837
	Total	5012

Markers were placed on the wall as a guide for the participants, a reference point for instructions, and to ensure the same amount and type of movement each time. Participants were instructed to move their eyes and head to specific markers depending on the artifact being recorded.

Figure 1 shows that there were two different degrees of each direction. For example, from the center in the left direction, the first marker L1 is a 27◦ movement left, and the second marker L2 is a 45◦ movement left. The reason behind this was to capture moderate and extreme movements. The varying degrees also allowed investigation into whether there was a difference between the two in the EEG. The outer marker placement was designed to be the furthest possible eye movement within a reasonable comfort level. This was to ensure the participant generated a noticeable muscle movement artifact.

Figure 1 also shows a wider horizontal than vertical range. This accounts for the panoramic field of view, which is approximately 180◦ horizontal range by 120◦ vertical range.

Each movement was performed starting from C, the center point, to ensure that there was no overlap between movements and each movement didn't contain components of other movements. This was also the reason that eye and head movements were performed separately.

The combination eyes + head movements were not performed to varying degrees, as the fine-grain adjustments on movement combinations were too difficult for the participant, and could introduce an undesired change in the signal due to concentration. In addition, the individual movements already covered the varying degrees.

Table 1 provides a list of the 21 unique movements measured. The test suite consisted of 24 tests in total to incorporate all of the movement. The activity sequence of was eye movement only, two blinks, head movement only, two blinks and joint eye and head movements. The blinks were inserted in the middle of changing between different movement types to aid in creating a distinction between participants' movement types. It also assisted in ensuring neighboring movement types did not influence each other.

The test suite was carried out 10 times by each participant. The reasons for this were:

1. It allowed for any recording errors. Having participants do each movement multiple times (e.g. ten) increased the chances of obtaining correct movement recordings.
2. It introduced enough data per participant to minimize the effect of outliers, or 'one-off' instances. In addition, there was enough data to have the choice to classify an individual's data, if needed.
3. Participants performed the movements better over time. Initially, the movements were unfamiliar and awkward, then the participant became comfortable with the movements and there was a period of stability. Towards the end participants became complacent and fatigue set in. Executing the test suite multiple times allowed to counter balance the task activities across the recording session.

Each trial took between 1.5 and 2 hours including the EEG cap fitting. A total of 26 participants were tested over a period of two and a half weeks. People's brain work differently, as well as their physiology and responses vary. Such variations may produce quite different results and introduce bias. Having a large sample and therefore a variety of participants reduces the possibility of misleading results unrepresentative of reality. Additionally, a classifier is more effective with a large amount of data.

IV. Eye and Head Movement Classification

The literature review of classifiers used in BCI research revealed that there was no recommended classifier(s) to use for EEG data. In addition, no surveyed paper dealt with artifact classification.

Feature selection is a difficult, complex process. The complexity is compounded when the data's characteristics are not well known, as is the case with artifacts. There are a myriad of different features and feature combination and data processing that can be evaluated, but this is a lengthy process.

A decision tree classifier was chosen due to its ability to select features. The decision tree chooses features based on what is useful to classify the data with. At the very least, an examination of the features chosen may reveal characteristics about the data.

However, decision trees cannot handle time-series data. The Hidden Markov Model (HMM) was chosen as a second classifier due to its ability to model the data's time component [9]. The time component may have additional information or patterns that can be used in classification. The HMM has also been successfully applied to many human-related classification tasks such as speech, handwriting and gesture recognition.

A. Alternatives Considered

To verify that non-obvious details had not been overlooked, preliminary testing with other classifiers was conducted using a small portion of the data collected. This allowed for an objective decision, reinforced with quantitative data, to be made, in the classifier selection for pursuit. The alternatives considered are described below.

Naive Bayes: a special form of Bayesian network based on conditional probabilities, applying Bayes' theorem with strong independent assumptions. This classification did not perform as well as the decision tree. This was because it took all features into account. To improve its performance, beneficial features would have had to be chosen and poor features manually removed.

Support Vector Machine (SVM): is a linear classifier that works by finding the optimum way to split the data into two parts using a hyperplane. In this experiment an extended version was used that allowed classifying more than two classes. However, this did not perform well when classifying artifacts in the EEG as it classified a large number of instances as blinks. This may have been due to the original SVM design for binary (two class) classification and does not apply well to multi-class problems.

Common Spatial Patterns (CSP): is a spatial filter that has been widely used in EEG data classification. CSP was applied to the data using the technique described in [10]. This method did not perform well for several reasons. Firstly, all data had to be of equal length in time. Ideally, a dynamic time warping method would be used to achieve this, but due to time constraints, a less sophisticated approach was used, truncating the data. Secondly, it relies on applying optimal frequency band filters, which were not feasible to implement. Lastly, CSP is usually applied to imagined motor imagery EEG data, which is also a binary problem.

B. *Verification Metrics*

To measure the classifiers performance, K-fold cross validation was used. This method reveals how a classifier performs when asked to classify unseen data. K-fold cross validation is a robust method designed to give a realistic indication of the performance of a classifier.

K-fold cross validation involves randomly dividing the data set into k subsets (folds), where one subset is used for validation, and the remaining for training. This ensures that the classifier hasn't seen the data it is being tested on, giving a good indication of how it will perform with new data. This process is then repeated k times. This ensures that performance measurement isn't just the result particular division of the data.

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	<-- classified as
55	30	31	8	10	3	5	3	11	11	8	3	2	0	4	2	3	6	0	6	7	a = Eyes only C -> L1
40	92	14	17	4	2	6	0	3	12	0	3	0	2	1	0	0	5	0	3	5	b = Eyes only C -> L2
34	8	54	27	25	6	10	1	4	5	10	1	1	0	1	1	7	0	0	1	11	c = Eyes only C -> R1
10	13	39	99	12	3	3	1	4	3	2	6	1	0	0	0	6	3	0	2	3	d = Eyes only C -> R2
6	1	7	3	43	34	51	10	4	1	4	0	2	0	5	2	0	0	0	2	34	e = Eyes only C -> U1
8	0	1	3	48	56	21	19	1	1	2	2	1	1	2	6	0	1	0	4	30	f = Eyes only C -> U2
8	1	5	2	39	21	24	37	7	0	3	3	2	2	9	4	1	0	0	1	36	g = Eyes only C -> D1
4	1	1	0	9	22	49	57	2	2	5	2	2	1	1	11	1	1	0	0	38	h = Eyes only C -> D2
13	4	5	1	7	3	8	4	21	37	23	10	6	3	7	2	10	2	1	2	38	i = Head only C -> L1
11	10	5	2	2	3	6	3	40	43	22	18	6	1	6	6	6	7	0	2	9	j = Head only C -> L2
5	1	17	3	3	2	4	6	28	21	31	38	2	1	5	9	7	4	0	2	18	k = Head only C -> R1
2	4	9	10	1	1	5	1	15	13	54	48	6	5	6	5	8	9	1	2	6	l = Head only C -> UR
2	1	0	0	2	2	2	2	5	9	3	9	55	64	6	6	7	3	10	3	17	m = Head only C -> U1
0	0	1	0	0	1	3	3	3	1	2	4	80	48	16	6	3	3	23	3	10	n = Head only C -> U2
4	3	3	1	3	4	13	1	2	7	10	7	6	8	46	37	3	2	5	11	34	o = Head only C -> D1
1	0	1	0	2	7	7	12	5	7	5	7	10	10	37	46	11	2	3	13	24	p = Head only C -> D2
3	2	10	10	5	0	4	1	10	12	13	19	5	4	3	8	59	16	6	3	18	q = Eyes + head C -> Left
4	11	2	0	1	0	0	1	5	12	6	10	8	3	4	6	24	93	4	14	3	r = Eyes + head C -> Right
1	1	2	0	1	1	0	0	2	1	3	2	16	29	4	7	7	4	120	6	1	s = Eyes + head C -> Up
5	2	2	1	7	6	3	5	2	2	3	4	5	3	20	13	10	10	9	89	9	t = Eyes + head C -> Down
3	2	9	2	14	24	17	20	19	6	7	9	4	4	18	11	2	2	1	4	659	u = Blink

Figure 3. Decision tree confusion matrix, all classes. The confusion between movements of the same direction is denoted by the boxed outline.

a	b	c	d	e	f	g	h	i	j	k	l	m	<-- classified as
62	14	6	7	5	20	3	6	15	29	10	19	15	a = Eyes only C -> Left
22	105	5	10	11	1	1	1	12	21	10	10	2	b = Eyes only C -> Right
8	11	110	10	0	0	2	1	2	2	50	10	2	c = Eyes only C -> Up
12	6	8	87	8	4	10	5	7	6	8	40	9	d = Eyes only C -> Down
4	5	0	1	223	63	21	24	39	8	2	5	22	e = Head only C -> Left
5	3	0	1	41	268	46	15	10	10	3	3	12	f = Head only C -> Right
2	1	0	3	15	25	197	109	10	6	5	3	40	g = Head only C -> Up
0	1	2	1	9	8	90	192	11	4	3	26	67	h = Head only C -> Down
9	8	1	3	40	17	11	19	146	87	11	17	46	i = Eyes + head C -> Left
21	4	1	1	7	26	10	15	78	190	15	28	22	j = Eyes + head C -> Right
4	2	41	3	2	2	10	7	13	14	262	34	24	k = Eyes + head C -> Up
15	2	2	19	4	4	18	23	28	22	37	192	54	l = Eyes + head C -> Down
7	4	1	6	5	4	30	55	24	9	7	30	655	m = Blink

Figure 2. Decision tree confusion matrix, using 13 broad classes. The strengthening of the diagonal (correctly classified instances) is denoted by the boxed outline.

Ten folds were used to validate the classifiers, as it is the standard number of folds to use due to more folds not making a substantial difference.

C. *Feature Selection*

Decision trees cannot accept time-series data, which is the EEG format. For this reason, it was necessary to select features that summarized the data. The features chosen were mean, median, minimum, maximum, standard deviation, range, and the same features taking the data's first derivative over time.

These features were chosen, as they are simple statistics that give an overall representation of the data. The same features were applied to the first derivative of the data for each channel. Taking the first derivative introduces a summary of how the signal changed over time, providing a way to quantify the time based nature of the data.

D. *Data Format*

The Wakaito Environment for Knowledge Analysis (Weka) [14] implementation of C4.5 + Adaboost was used. Weka uses ARFF (Attribute-Relation File Format) for specifying the data. An ARFF file is a text file that lists the instances according to what attribute they are.

E. *Data Pre-processing*

The HMM was given the 'raw' (synchronized) time-series data of all 36 channels, so feature selection was not required. A discrete-value HMM was used so the data had to be quantized into discrete bins. The full range of possible values of each channel was evenly divided into equal size bins. 10, 20 and 40 bins were tested, with 20 bins chosen as the optimal number. The data was then stored in a format that the HMM can read.

The HMM was trained using several different numbers of hidden states ranging from four to ten. Seven hidden states were chosen as a suitable number because it achieved the best results. No additional constraints were imposed such as forced left right and zero probabilities. The EM algorithm was used for learning the HMM.

V. DATA CAPTURE

A. *EEG*

The EEG (Electroencephalograph) is a medical instrument that records electric currents generated by the brain. The EEG

Table 3. Number of instances per broad class.

	Class description	# instances
1	Eyes only $C \rightarrow Left$	417
2	Eyes only $C \rightarrow Right$	417
3	Eyes only $C \rightarrow Up$	416
4	Eyes only $C \rightarrow Down$	414
5	Head only $C \rightarrow Left$	415
6	Head only $C \rightarrow Right$	418
7	Head only $C \rightarrow Up$	418
8	Head only $C \rightarrow Down$	420
9	Eyes + head $C \rightarrow Left$	211
10	Eyes + head $C \rightarrow Right$	211
11	Eyes + head $C \rightarrow Up$	208
12	Eyes + head $C \rightarrow Down$	210
13	Blink	837
	Total	5012

is used to record the participant's brain's electrical signals, measured in volts, for the purpose of being able to detect artifacts that occur.

The amplitude of the EEG signal is very small requiring amplification prior to being read. The signal is passed through a low-pass filter minimizing any distortion caused by aliasing that may occur when the analog signal is converted to digital. The Compumedic NuAmp EEG amplifier with the 36 electrodes Quik Cap was used.

B. Eye Tracker

The eye tracker calculated the gaze point by measuring eye position and movement. The eye tracker was used to record movement made by the eyes as ground truth to the EEG measurements. The eye tracker was a pair of glasses that has two cameras and an infrared emitter mounted over the right eye, and an adjustable monocle piece.

C. Motion Sensor

The motion sensor was used to record the participants' head movement so that the ground truth was captured with the EEG measurements. The system consisted of a sensor and a receiver/processor. It was an inertia-based orientation only sensor that measured three degrees of freedom.

D. Data Capture Framework

To collect the data from the three different devices simultaneously, a framework was built using each device's SDK. The framework was written in C# taking advantage of the .NET framework's integration with COM objects.

VI. Results and Analysis

The trials conducted resulted in a total of 5880 instances. After removing erroneous data, a total of 5012 instances remained. **Table 2** shows the breakdown of the number of instances per class.

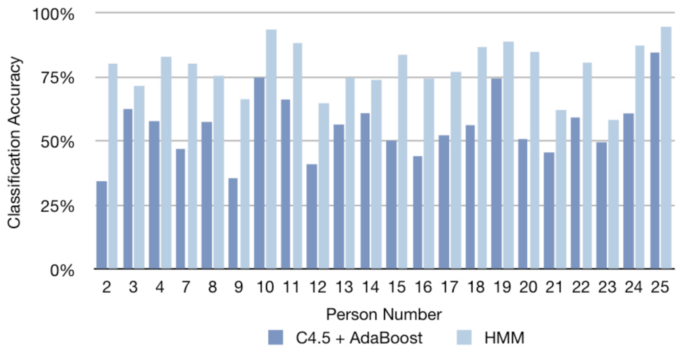


Figure 4. Comparison of classification accuracies of individuals using the decision tree and HMM.

Each instance was represented with the following summary statistics in its raw format and in its first derivative: mean, median, minimum, maximum, standard deviation and range. Each statistic was applied to every channel individually producing a total of 504 features.

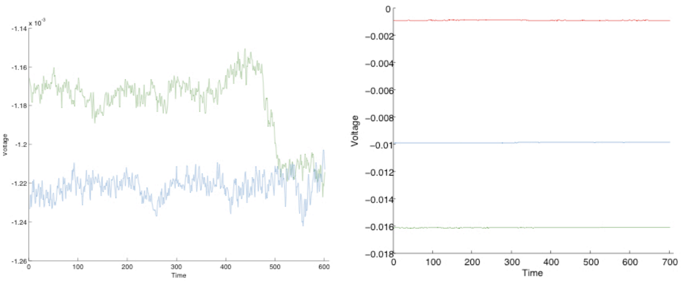
Examining the confusion matrix highlighted a consistent confusion between same direction movements. This is shown by the boxed outline in Figure 3. This observation is consistent with a logical deduction that the same physical movements will produce the same or similar artifact.

Hence, the 'fine-grained'/granular classes were grouped into broader classes where the different degrees of movements were grouped together as one class. For example, the classes 'eyes only C → L1' and 'eyes only C → L2' were put into one class called eyes only 'C → Left'. The resulting 13 'broad' classes are detailed in **Table 3**.

This dataset was classified the same way using C4.5 + AdaBoost. It returned 53.65% classification accuracy for the 15 'broad' class dataset. The improvement in accuracy indicated that degrees of the same movements produce artifacts that are too similar for differentiation. This is confirmed by the strengthening of the numbers along the diagonal of the confusion matrix, which indicate correctly classified instances, shown in **Figure 2**.

Although there is an accuracy improvement, there is still a prohibitive amount of confusion between classes. No clear explanation could be discerned from the confusion matrix. Further investigation into the cause was conducted. Examining the data instances revealed that the intra-class variability exceeded the inter-class variability (shown in Figure 6). This presented a sizeable problem, as the disparity between two people performing the same movement was greater than the disparity between two different movements.

In fact, it was remarkable that the decision tree was capable of classifying the data as well as it did. A possible reason for this unexpected result was due to the way that 10-fold cross validation works and the amount of data available per person. With the way that 10-fold cross validation splits the data, using 90% for training and 10% for testing, it is very probable that an instance of each class for each person was in the training set. Therefore, instead of learning the general signal pattern for all people, the classifier learned the specific pattern and frequency



(a) Two different movements (b) Three different participants performing the same movement

Figure 6. Intra-class variability > inter-class variability in the data (Note the difference in scale).

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	<-- classified as
98	12	18	5	3	1	0	0	25	6	12	2	2	2	3	0	1	13	0	5	0	a = Eyes only C -> L1
35	123	2	8	3	0	0	0	3	13	2	1	0	4	2	0	1	10	0	1	1	b = Eyes only C -> L2
24	0	102	17	2	0	0	0	5	2	28	10	1	2	0	0	11	0	0	3	0	c = Eyes only C -> R1
10	4	29	140	1	0	0	0	2	0	2	9	2	1	1	0	8	0	0	1	0	d = Eyes only C -> R2
2	0	0	0	120	24	11	2	6	2	10	2	6	4	2	3	2	1	0	6	6	e = Eyes only C -> U1
1	0	0	0	42	118	6	11	0	2	3	1	1	3	1	2	2	1	1	8	4	f = Eyes only C -> U2
0	0	0	0	38	4	101	24	4	1	4	1	2	2	8	7	3	0	2	2	2	g = Eyes only C -> D1
0	0	0	0	3	7	45	119	1	1	0	0	2	1	9	16	1	0	3	1	0	h = Eyes only C -> D2
21	3	0	0	20	1	10	0	64	22	25	0	4	2	11	1	7	3	1	4	8	i = Head only C -> L1
26	14	1	0	9	0	8	0	39	42	10	0	5	0	12	4	12	14	3	8	1	j = Head only C -> L2
2	0	16	0	13	2	3	2	29	1	78	26	5	0	9	4	7	1	0	6	3	k = Head only C -> R1
5	0	24	13	12	1	0	0	14	0	46	56	4	5	3	7	11	1	1	7	1	l = Head only C -> R2
1	0	0	0	54	3	8	0	8	0	13	2	41	32	13	5	3	0	10	12	3	m = Head only C -> U1
0	0	0	0	34	18	7	0	4	2	5	3	36	45	7	1	2	4	15	15	12	n = Head only C -> U2
0	0	1	0	31	2	43	2	10	5	9	1	11	9	36	35	2	0	0	9	4	o = Head only C -> D1
1	0	0	0	10	1	22	37	7	5	6	1	7	2	40	55	3	0	5	7	1	p = Head only C -> D2
2	0	9	8	10	2	4	0	7	4	50	28	1	2	7	4	59	1	2	9	2	q = Eyes + head C -> Left
6	21	0	0	0	0	0	0	26	30	0	2	1	5	8	1	1	93	5	11	1	r = Eyes + head C -> Right
0	0	0	0	2	0	24	3	1	1	5	1	12	17	7	14	1	4	103	13	0	s = Eyes + head C -> Up
0	0	0	0	36	8	3	0	6	1	8	2	11	15	12	3	3	2	5	83	12	t = Eyes + head C -> Down
3	0	1	0	177	36	6	1	42	4	40	9	20	21	9	8	14	0	5	23	418	u = Blink

Figure 5. HMM confusion matrix, all 21 classes. The confusion between movements of the same direction is denoted by the red outline.

for each person, and consequently used that specific pattern to classify any testing data that was similar.

The decision tree classifier was then trained on each participant individually, using the 13 broad classes. The results for each individual's overall classification accuracy are shown in Figure 6. These results show an increase in classification accuracy in some cases, with a maximum of 84.58%, but also a decrease in classification accuracy in some cases, with a minimum of 34.38%.

When reviewing the notes taken during the trials for each participant, it was noticed that individuals with high classification accuracy had low electrode impedances, were relaxed and coordinated with the movements. Individuals with low classification accuracies were not comfortable or found the movements difficult or had problems during the trial such as high electrode impedances or equipment complications.

The natural temporal ordering of the data indicated that the variation in the EEG signal over time was an important factor. In addition, the data may have been too complex to represent with summary statistics. Therefore, a HMM was also used to attempt to solve the classification problem.

HMM learning was performed identically to the decision tree, using each participant's individual datasets with the 13 broad classes. The results compared with the decision tree are shown in Figure 4. The HMM performed better, with consistently higher classification accuracies. This indicates that the HMM was able to use the raw data and temporal information to better classify the data. As with the decision tree, the highest classification accuracies were consistent with individuals with clean and correct data.

The HMM classifier was also executed using the whole dataset with the 21 fine-grained classes. In order to eliminate the differences between people, the raw data was baselined to the mean of all of the samples. The HMM classification accuracy was 40.64%, which was a small improvement over C4.5 + AdaBoost.

The confusion matrix in Figure 5 shows an even clearer confusion between degrees of the same movement. This further supports the hypothesis that these similar artifacts are the greatest source of confusion between the fine-grained classes.

Additionally, the HMM was executed using the entire dataset with the 13 broad classes. This returned a classification accuracy of 49.92% that is less than the decision tree.

This unexpected result was probably due to the decision tree classification accuracies being misleadingly high. To explain, the decision tree classifier was able to use the extra information of individual participant's absolute voltages. Whereas this information was not available to the HMM, as it was lost when the data was baselined to the mean. To substantiate this claim, decision tree learning was performed on the same dataset with- out features that could be used to identify an individual's signal, such as mean, median, minimum and maximum. The classification accuracy dropped from the original 53.65% to 22.51%, verifying this hypothesis.

VII. Conclusion

This research evaluated methods for classifying artifacts in the EEG. Two classifiers were chosen for in-depth study. In order to train the classifiers, a large dataset of EEG artifacts was required. This was accomplished by conducting human trials where the participants were instructed to perform 21 different artifact-generating movements whilst wearing an EEG. An eye tracker and motion sensor were also used to measure the ground truth. A data capture framework was developed to accommodate the special use of this equipment. This framework was developed with flexibility and ease of use as priority features so that future researchers using the equipment can use it.

The two classifiers implemented and evaluated were a decision tree classifier, C4.5 with AdaBoost, and a probabilistic classifier, the Hidden Markov Model. The HMM outperformed the decision tree, proving it to be the most suitable classifier to be used for classifying artifacts. The classifiers were trained using the 21 'fine-grained' classes, and then these were re-grouped into 13 'broad' classes. Both of the classifiers did not perform well, achieving a maximum classification accuracy of 54% using the broad classes. This low classification accuracy is due to the variability between participants performing the same movement being greater than the variability between different movements. The classifiers were also trained using individual participant's data, with the decision tree achieving a maximum of 85%, and the HMM achieving a maximum of 95%, with correct data.

This shows that while a general artifact classification system may not be feasible, on a per-person basis artifacts can be classified with a high degree of accuracy.

REFERENCES

[1] R. Barea, L. Boquete, M. Mazo, and E. Lopez, System for Assisted Mobility Using Eye Movements Based on Electrooculography, IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 10, no. 4, pp. 209-218; 2002.

[2] R. Barea, L. Boquete, M. Mazo, E. Lopez and L. Bergasa, E.O.G. Guidance of a Wheelchair using Neural Networks, International Conference on Pattern Recognition, vol. 4; 2000.

[3] M. Fatourechi, A. Bashashati, R. Ward, and G. Birch, EMG and EOG artifacts in brain computer interface systems: A survey, Clinical Neurophysiology 118, pp. 480-494; 2007.

[4] R. Croft and R. Barry, Removal of ocular artifact from the EEG: a review. Neurophysiologie Clinique 30(1), 5–19, 2000.

[5] L. Geddes, J. Bourland, G. Wise, and R. Steinberg, Linearity of the Horizontal Component of the Electro-oculogram, Medical and Biological Engineering, pp. 73-77; January 1973.

[6] D. Kumar and E. Poole, Classification of EOG for Human Computer Interface, Proceedings of the 2nd Joint IEEE Annual International Conference of the Engineering in Medicine and Biology Jointly with the 24th Annual Conference of the Biomedical Engineering Society, vol. 1, pp. 64-67; October 2002.

[7] J.R, Millan, M. Franz, J. Mouri, F. Cincotti, and F. Babiloni, Relevant EEG features for the classification of spontaneous motor-related tasks. Biological Cybernetics 86(2), 89–95, 2002.

[8] C. Ochoa and J. Polich, P300 and blink instructions. Clinical Neurophysiology 111(1), 93–98, 2000.

[9] L. Rabiner, A tutorial on Hidden Markov Models and selected applications in speech recognition, Proc. of the IEEE, 77(2): 257-286, 1989.

[10] Ramoser, H, Müller-Gerking, J & Pfurtscheller, G 2000, “Optimal spatial filtering of single trial EEG during imagined hand movement,” Rehabilitation Engineering, IEEE Transactions on, vol. 8, no. 4, pp. 441-446.

[11] R. Verleger, The Instruction to Refrain From Blinking Affects Auditory P3 and N1-Amplitudes. Electroencephalography And Clinical Neurophysiology 78(3), 240–251, 1991.

[12] G. Wallstrom, R. Kass, A. Miller, J. Cohn, and N. Fox, Automatic correction of ocular artifacts in the EEG: a comparison of regression-based and component-based methods. International Journal Of Psychophysiology 53(2), 105–119, 2004.

[13] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. Vaughan. Brain-computer interfaces for communication and control. Clinical Neurophysiology 113(6), 767–791, 2002.

[14] E. Frank, M. Hall, G. Holmes, R. Kirkby, B. Pfahringer, I. Witten, and L. Trigg. Weka - A machine learning workbench for Data Mining. Data Mining And Knowledge Discovery Handbook, 1305–1314, 2005.