34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

What does clean EEG look like?

Ian Daly[1], Floriana Pichiorri[2], Josef Faller[1], Vera Kaiser[1],
Alex Kreilinger[1], Reinhold Scherer[1] and Gernot Müller-Putz[1]

Abstract— **Lack of a clear analytical metric for identifying artifact free, clean electroencephalographic (EEG) signals inhibits robust comparison of different artifact removal methods and lowers confidence in the results of EEG analysis. An algorithm is presented for identifying clean EEG epochs by thresholding statistical properties of the EEG. Thresholds are trained on EEG datasets from both healthy subjects and stroke / spinal cord injury patient populations via differential evolution (DE).**

I. INTRODUCTION

The Electroencephalogram (EEG) is a frequently used technique for measuring electrical activity within the brain. The EEG may be used to answer neuroscientific questions related to the brains' response to a range of stimuli and conditions [1]. It may also be used to allow control of a Brain-computer interface (BCI), a device which allows direct control of a computer or device via the modulation of electrical activity in the brain [2].

However, the EEG is subject to a great deal of noise, both from internal and external sources of electrical interference. This can include, but is not limited to, subject generated noise such as the electrical responses to eye blinks and head movement, and external electrical noise such as the power line noise at either 50 or 60 Hz, cable movement, sweating, electrode movement etc. [3], [4]. It is therefore very important to ensure that the EEG is clean and free of these noise artifacts before it is used. This ensures that any analysis results may be attributed to brain function.

A large number of studies present a wide variety of methods for identifying artifacts in the EEG [5]–[9]. These methods may operate in either a fully automatic or semi-automatic manner and may look for one or several types of artifact. Artifacts are identified via a wide range of different features which characterise properties of the artifact as diverse as their time series topology, their spectral template, and/or statistical properties of either uni- or multivariate EEG. They may also be removed via, for example, independent component analysis (ICA) [5]. However, often data is cleaned by hand, a laborious imperfect process.

*This work was supported by the FP7 Framework EU Research Project ABC (No. 287774).

[1]I. Daly, J. Faller, V. Kaiser, A. Kreilinger, R. Scherer and G. Müller-Putz are with the Institute for Knowledge Discovery, Laboratory of Brain-Computer Interfaces, Institute for Knowledge Discovery, Graz University of Technology, Krenngasse 37/IV, 8010, Graz, Austria i.daly at tugraz.at

[2]F. Pichiorri is with the Clinical Neurophysiology, Neuroelectrical Imaging and BCI Lab, Fondazione Santa Lucia, IRCCS, Via Ardeatina, 306, 00179, Rome, Italy

The efficacy of artifact identification and removal methods may be evaluated in a number of ways. Frequently, visual inspection of the time series of the EEG is presented as sufficient evidence for the efficacy of the removal method [8], [10], [11]. Less subjective metrics may also be used.

Metrics may be defined as the reduction of certain undesirable properties of the EEG. Such reductions must be done in such a way that other, desirable, properties are maintained. For example, in [12] measures of sensitivity and specificity with regards the amount of artifact removed and the influence of this removal on the clean EEG are quantified and in [13] channels known to be contaminated with artifacts are used to quantify the accuracy of artifact detection methods.

This concentration on the removal of artifacts omits a very important question. Namely, what does clean EEG actually look like? The, often unspoken, assumption is that clean EEG comprises anything that does not contain artifacts. This definition omits the possibility of unrecognised artifacts. For example, an artifact removal method intended for blinks would assume the remaining EEG was clean. However, the EEG could still contain other artifacts, such as electrode pops, electromyography (EMG) noise from muscles etc [14].

There is therefore a need for an analytic definition of clean EEG. This could serve the purpose of allowing for a more rigorous evaluation of artifact removal methods. Currently artifact removal studies define their success in different ways. Thus, [15] defines success as the number of artifacts detected minus non-detected artifacts, divided by the total number of artifacts, while [12] defines success by measures of sensitivity and specificity with regards changes in artifactual and non-artifactual EEG components.

A metric for testing how clean an EEG epoch is, would allow rigorous comparisons between these studies to be made. For example, artifact removal methods could be evaluated by looking at the change in the EEG cleanliness metric before and after use of the method.

Efforts to measure EEG signal quality are often only based upon the impedance and the approximate number of artifacts observed. Additionally, in the case of BCI studies, the classification accuracy or information transfer rate (ITR) may be used as proxy measures for signal quality. Such measures are indirect and often specific to a particular study.

An approach by [16] measures the signal to noise ratio (SNR) between periods of EEG in which the subject was instructed to blink and periods when no blink was observed. This measure was used to compare the SNR (in relation to only blinks) between the EEG and electrocorticogram

978-1-4577-1787-1/12/$26.00 ©2012 IEEE

(ECoG). However, this approach relies on the subjects being able to generate blinks at will and does not account for other artifact types (e.g. movement artifacts). Furthermore, it does not allow a robust measure of measurement noise.

Recent work by [17] uses combined measures of EEG quality and classifier behaviour to determine when to switch between different control modes in a hybrid joystick-BCI controlled game. EEG quality is measured by the amount of EMG (determined by amplitude thresholding) and the accuracy of BCI control related to performance at the game. Hence, the metric is specific to this BCI application. Note, the increasing interest in hybrid BCIs [18] is an area for which measures of EEG quality would also be very useful.

By way of contrast [19] and [20] provide excellent descriptions of what normal EEG in the waking adult brain should look like. However, this work is intended for clinicians and related fields who visually inspect and diagnose the EEG. Therefore, translation of the visual descriptions of the EEG into analytic metrics of EEG quality remains to be done.

In this study we first describe a series of metrics based upon the definition of clean EEG [19]. We then attempt to identify the distribution and limits of these metrics in clean EEG (identified by independent raters). Datasets used contain both clean and artifact contaminated EEG.

EEG from healthy subjects, stroke patients and spinal cord injury (SCI) patients is investigated. EEG is known to vary significantly between these groups [19], [20]. Including patients allows for a broad characterisation of the EEG.

Thresholds are trained on the data to optimally identify clean epochs via differential evolution (DE). Finally, the optimal thresholds are presented and an algorithm is discussed for identifying clean EEG epochs.

II. METHODS

A. Metrics

A detailed description of key characteristics of clean EEG is provided in [19], [20].

1) Raw signal characteristics
   a) Amplitudes should typically range between 10 and 100 uV (mostly below 50 uV).
   b) The signal should generally exhibit rounded or arc shaped sinusoidal morphology.
2) Alpha rhythm characteristics
   a) The EEG between 8-12 Hz should exhibit a rounded or arc-shaped sinusoidal morphology.
   b) Additionally, amplitudes typically take values in the range of between 10 and 100 uV.
   c) Alpha rhythms are typically larger over paratial / occipital regions then frontal / central regions.
3) Beta rhythm characteristics
   a) The EEG between 13-35 Hz should exhibit a rounded sinusoidal morphology.
   b) Amplitudes are typically lower than 30 uV.
4) Power spectrum characteristics
   a) Low frequencies typically exhibit high power while high frequencies exhibit low power.

To check what range of values these quantities take in clean EEG a number of metrics are defined. Firstly, the EEG is grouped according to the region of the scalp the electrodes are located in. To account for spatial differences in amplitude distribution channels are grouped into 4 regions, frontal channels - channels labelled with an F in the 10/20 system, central channels - labelled with a C, temporal channels - labelled with a T and finally channels over the parietal and occipital regions - labelled with either a P or an O.

Metrics are calculated from the raw EEG in each region.

1) Maximum amplitude values.
2) Standard deviation of the amplitude values.
3) Kurtosis of the amplitude values.
4) Skewness of the amplitude values.

The kurtosis and skewness statistical measures attempt to provide some measure of the distribution of amplitude values (an indication of the signals morphological properties). They are also popular measures to identify artifacts [10], [21].

The EEG in each region is band-pass filtered (3rd order Butterworth filter) into the alpha (8 – 12 Hz) and beta (13 – 35 Hz) frequency bands. The following metrics are then calculated from each band in each region.

1) Mean power.
2) Standard deviation of power.
3) Maximum amplitude.
4) Standard deviation of amplitude.
5) Kurtosis of the amplitude.
6) Skewness of the amplitude.

Finally, the power spectral density is calculated in the range 0 – 40 Hz and the median power is calculated from 40 frequency windows, each of width 1 Hz.

B. Algorithm

An algorithm is defined to determine if a given portion of EEG falls within the thresholds of clean EEG. Each metric, as defined above, is calculated for each region available in the EEG dataset. The value of each metric is then checked against a set of threshold values (one threshold per metric). If all the metrics are lower than their corresponding thresholds the EEG epoch is labelled as clean, i.e., it falls within the expected limits of clean EEG.

C. Data

EEG with epochs labelled by independent raters as either artifact free or not are used to train thresholds to accurately identify clean epochs. Four datasets are used.

1) Dataset one: Data from a pilot study performed prior to the study described in [22] is used. EEG was recorded in an analogous manner to that in [22] but on ten healthy subjects who performed cued motor imagery (MI).

2) Dataset two: Data is recorded from twenty-nine stroke patients (14 female) from a rehabilitation hospital ward at Fondazione Santa Lucia (Rome, Italy). The mean patient age was 58 ± 15 years and the mean time between stroke incidence and participation in the study was 4 ± 4 months. The patients were instructed to perform motor imagery and execution tasks with either hand [23].

3) Dataset three: EEG was recorded from healthy subjects attempting to use a MI EEG BCI to control a car game as part of an experiment into MI and error potentials [24]. EEG was recorded from 10 healthy subjects (5 female). Mean participant age was 24.9 $\pm$ 2.3 years. EEG was recorded from 32 positions on the 10-20 system.

4) Dataset four: EEG was recorded from 5 SCI and 4 stroke patients as part of a study into an adaptive BCI [25]. Patients were instructed to perform cue based MI, subtraction and word association. EEG was acquired from 30 electrodes positioned chiefly over central and paratial regions.

D. Threshold identification

Differential Evolution (DE) is used to train the threshold values to separate clean and artifact contaminated EEG epochs. DE is an evolutionary meta-heuristic search technique, similar to a Genetic Algorithm, that has been shown to be highly successful in identifying near optimal solutions when applied to a wide range of different search spaces [26]. Fitness of potential solutions is measured via two criteria

1) Accuracy at identifying clean epochs is maximised.
2) The normalised mean threshold is minimised.

The dataset is shuffled to remove serial dependencies, then split evenly into training and testing sets. Thresholds are trained to optimally separate clean and contaminated epochs in the training set and verified on the testing set.

III. RESULTS

Figure 1 illustrates the values taken by each of the metrics across epochs measured from raw EEG. Box plots illustrate the distribution of typical values and outliers are plotted individually. Note, maximum amplitudes are typically below 50 uV and higher over the parietal and occipital channels (as expected from [20]). The thresholds for separating clean EEG from artifacts are illustrated as a solid horizontal line. Note, thresholds include all but the extreme outlying values.

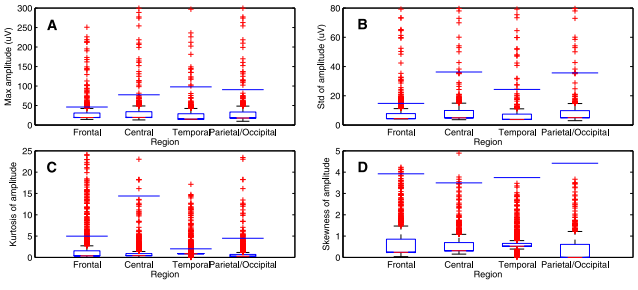

Fig. 1. Profile of raw, artifact free, EEG. A: maximum amplitude, B: standard deviation of amplitude, C: kurtosis of amplitude, and D: skewness.

The values taken by the metrics measured from the alpha band are illustrated in figure 2. Maximum amplitudes are considerably lower than calculated from raw EEG and again are largest over the parietal / occipital channels.

Figure 3 illustrates the distribution of metrics in the beta band. Largest amplitudes are in the parietal / occipital region.

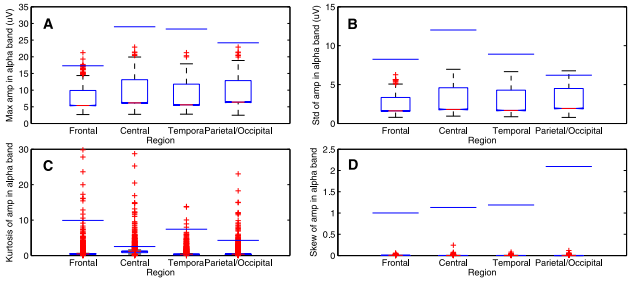

Fig. 2. Profile of artifact free EEG in the alpha band. A: maximum amplitude, B: standard deviation of amplitude, C: kurtosis, and D: skewness.

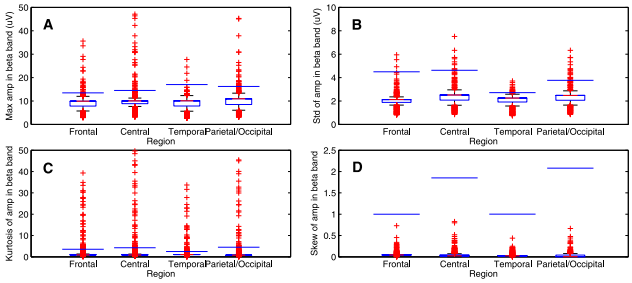

Fig. 3. Profile of artifact free EEG in the beta band. A: maximum amplitude, B: standard deviation of amplitude, C: kurtosis, and D: skewness.

The power spectral densities (PSD) in each of the channel regions are illustrated in figure 4. Note that, as expected, low frequencies exhibit greater power than higher frequencies.

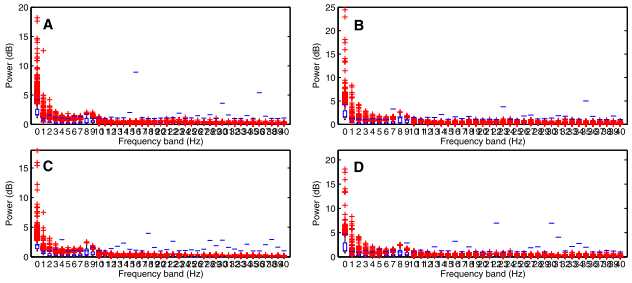

Fig. 4. PSD of artifact free EEG. A: frontal region, B: central region, C: temporal region, and D: occipital / parietal region.

Finally, the accuracy at differentiating clean EEG from artifacts in the left out testing set is 0.872. The false positive rate is 0.085 and the false negative rate is 0.169.

IV. DISCUSSION

This paper presents a simple algorithm to determine if a given portion of EEG may be labelled as clean of artifacts. Channels are grouped into regions of interest and metrics are extracted from each region. These metrics are then thresholded to determine if they fall within the bounds of clean EEG. Thresholds are trained on a large grouping of several datasets from both healthy subjects and patient populations performing a range of different cognitive tasks.

Note, the method may be applied to datasets containing different numbers of channels in different regions. Furthermore, it may be applied to single channel EEG, provided it

is known which region the channel was recorded on. Thus, the method could, potentially, be used to identify individual channels which do not contain clean EEG.

The identified thresholds encompass the majority of the maxima extracted from the clean trials but not all. There is some considerable overlap between clean EEG and artifact contaminated EEG. Note, the distribution of artifact contaminated EEG is not illustrated. This is because artifacts vary a great deal in their profile and in which particular metrics maxima may be found. For example, a blink artifact may cause a large amplitude in frontal channels but no significant change in occipital / parietal channels. Whereas, an electrode pop on an occipital channel will cause no discernible difference in the amplitude on frontal channels. Thus the distribution of metric values over all artifact types is not discernibly different from the distribution of clean EEG.

The high accuracy on the left out testing set indicates the algorithm, with the trained thresholds, is able to accurately differentiate clean and artifact contaminated EEG. Furthermore, the low false positive rate of 0.085 indicates only a very small proportion of clean trials are mistaken for artifacts. The false negative rate gives the proportion of artifacts which are flagged as clean and is slightly higher at 0.169. The thresholds may be adjusted to minimise either of these rates, by adjusting the DE fitness criteria, depending upon requirement, e.g. do we wish to avoid mistaking any artifacts for clean epochs at the risk of mistaking some clean epochs for artifacts or vice-versa?

There are, doubtless, many other cognitive tasks, patient and healthy populations and conditions which could also be investigated to determine more accurate threshold values. Furthermore, other metrics could be investigated and may allow more accurate identification of clean EEG. Future work will aim to further establish and test this metric to ensure it is robust across different paradigms and subject groups.

The thresholds and algorithm form a basis for the development of a measure of EEG signal quality. This may be applied to measure the quality of the EEG signal prior to analysis and is intended to firstly supplement and eventually replace visual inspection. Unlike visual inspection this metric is analytic and not subjective. It will therefore allow more direct comparisons to be made between different EEG recording systems, artifact removal methods and studies.

REFERENCES

[1] L. Schomer and F. Lopes de Silva, eds., *Niedermeyer's electroencephalography: Basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 6 ed., 2011.

[2] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control.," *Clinical neurophysiology*, vol. 113, pp. 767–791, June 2002.

[3] W. O. Tatum, B. A. Dworetzky, and D. L. Schomer, "Artifact and recording concepts in EEG.," *Journal of clinical neurophysiology*, vol. 28, pp. 252–63, June 2011.

[4] B. W. McMenamin, A. J. Shackman, L. L. Greischar, and R. J. Davidson, "Electromyogenic Artifacts and Electroencephalographic Inferences Revisited.," *NeuroImage*, vol. 54, pp. 4–9, Aug. 2010.

[5] J. T. Gwin, K. Gramann, S. Makeig, and D. P. Ferris, "Removal of movement artifact from high-density EEG recorded during walking and running.," *Journal of neurophysiology*, vol. 103, pp. 3526–34, June 2010.

[6] R. Magjarevic, Y. Peng, X. Weng, and W. Zhou, "Dipole Analysis of Eye Movement Artifacts from the EEG," in *7th Asian-Pacific Conference on Medical and Biological Engineering* (Y. Peng and X. Weng, eds.), vol. 19 of *IFMBE Proceedings*, (Berlin, Heidelberg), pp. 337–340–340, Springer Berlin Heidelberg, 2008.

[7] A. Schlogl, C. Keinrath, D. Zimmermann, R. Scherer, R. Leeb, and G. Pfurtscheller, "A fully automated correction method of EOG artifacts in EEG recordings," *Clinical Neurophysiology*, vol. 118, pp. 98–104, 2007.

[8] G. Gomez-Herrero, W. Clercq, H. Anwar, O. Kara, K. Egiazarian, S. Huffel, and W. Paesschen, "Automatic Removal of Ocular Artifacts in the EEG without an EOG Reference Channel," in *Proceedings of the 7th Nordic Signal Processing Symposium - NORSIG 2006*, pp. 130–133, IEEE, June 2006.

[9] P. S. Kumar, R. Arumuganathan, K. Sivakuma, and C. Vimal, "Removal of ocular artifacts in the EEG through wavelet transform without using an EOG reference channel," *Int. J. Open Problems Comptutaional Mathematics*, vol. 1, no. 3, pp. 188–200, 2008.

[10] A. Greco, N. Mammone, F. C. Morabito, and M. Versaci, "Semi-automatic artifact rejection procedure based on kurtosis , Renyi s entropy and independent component scalp maps," *Engineering and Technology*, pp. 22–26, 2005.

[11] C. A. Joyce, I. F. Gorodnitsky, and M. Kutas, "Automatic removal of eye movement and blink artifacts from EEG data using blind component separation," *Psychophysiology*, vol. 41, no. 2, pp. 313–325, 2004.

[12] B. W. McMenamin, A. J. Shackman, J. S. Maxwell, D. R. W. Bachhuber, A. M. Koppenhaver, L. L. Greischar, and R. J. Davidson, "Validation of ICA-based myogenic artifact correction for scalp and source-localized EEG.," *NeuroImage*, vol. 49, pp. 2416–32, Feb. 2010.

[13] D. Talsma, "Auto-adaptive averaging: detecting artifacts in event-related potential data using a fully automated procedure.," *Psychophysiology*, vol. 45, no. 2, pp. 216–28, 2008.

[14] W. Barry and G. M. Jones, "Influence of eye lid movement upon electro-oculographic recording of vertical eye movements," *Aerospace medicine*, vol. 36, pp. 855–8, Sept. 1965.

[15] A. Delorme, T. Sejnowski, and S. Makeig, "Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis.," *NeuroImage*, vol. 34, no. 4, pp. 1443–9, 2007.

[16] T. Ball, M. Kern, I. Mutschler, A. Aertsen, and A. Schulze-Bonhage, "Signal quality of simultaneously recorded invasive and non-invasive EEG.," *NeuroImage*, vol. 46, pp. 708–16, July 2009.

[17] A. Kreilinger, V. Kaiser, C. Breitwieser, J. Williamson, C. Neuper, and G. R. Muller-Putz, "Switching between manual control and brain-computer interface using long term and short term quality measures," *Frontiers in Neuroprosthetics*, vol. 5, no. 147, 2012.

[18] G. Pfurtscheller, B. Z. Allison, C. Brunner, G. Bauernfeind, T. Solis-Escalante, R. Scherer, T. O. Zander, G. Müller-Putz, C. Neuper, and N. Birbaumer, "The hybrid BCI," *Frontiers in Neuroprosthetics*, vol. 4, no. 30, 2010.

[19] E. Niedermeyer, "The normal EEG of the waking adult," in *Electroencephalography: Basic Principles, clinical applications and related fields1*, pp. 149–173, 1999.

[20] D. L. Schomer, A. S. Blum, and S. B. Rutkove, *The Clinical Neurophysiology Primer*. Totowa, NJ: Humana Press, 2007.

[21] L. Shoker, S. Sanei, and J. Chambers, "Artifact removal from electroencephalograms using a hybrid BSS-SVM algorithm," *IEEE Signal Processing Letters*, vol. 12, pp. 721–724, Oct. 2005.

[22] G. Pfurtscheller, P. Linortner, R. Winkler, G. Korisek, and G. Müller-Putz, "Discrimination of motor imagery-induced EEG patterns in patients with complete spinal cord injury.," *Computational intelligence and neuroscience*, vol. 2009, p. 104180, Jan. 2009.

[23] F. De Vico Fallani, F. Pichiorri, C. Di Lanzo, F. Ceccarelli, I. Pisotta, F. Cincotti, M. Molinari, F. Babiloni, and D. Mattia, "Assessment framework of functional brain networks during covert motor performance after stroke," in *5th Int.BCI con.*, pp. 184–186, 2012.

[24] A. Kreilinger, C. Neuper, and G. Müller-Putz, "Detection of error potentials during a car-game with combined continous and discrete feedback," in *5th Int. BCI con.*, pp. 204–207, 2011.

[25] J. Faller, C. Vidaurre, E. V. C. Friedrich, U. Costa, and E. Opisso, "Automatic adaptation to oscillatory EEG activity in spinal cord injury and stroke patients," in *TOBI workshop*, no. 1, pp. 7–8.

[26] R. Storn and K. Price, "Differential evolution - a simple and efficient adaptive scheme for global optimisation over continuous spaces," vol. 11, no. 12, pp. 1–15, 1995.