

# Choosing strategies to deal with artifactual EEG data in children with cognitive impairment

Ana Tost<sup>1</sup>, Carolina Migliorelli<sup>2,1,3</sup>, Alejandro Bachiller<sup>1,2,3\*</sup>, Sergio Romero<sup>1,2,3</sup>, Miguel A. Mañana<sup>1,2,3\*</sup>, M. Ángeles García-Cazorla<sup>4,3</sup>

<sup>1</sup> Department of Automatic Control (ESAI), Biomedical Engineering Research Centre (CREB), Universitat Politècnica de Catalunya (UPC), Barcelona, Spain.

<sup>2</sup> CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Spain.

<sup>3</sup> Institut de Recerca Sant Joan de Déu, Barcelona, Spain.

<sup>4</sup> Neurology Department, Neurometabolic Unit and Synaptic Metabolism Lab, Institut Pediàtric de Recerca, Hospital Sant Joan de Déu, metabERN and CIBERER-ISCIII, Barcelona, Spain.

## Abstract

*Rett syndrome is a disease that involves acute cognitive impairment and, consequently, a complex and varied symptomatology. This study evaluates the EEG signals of twenty-nine patients in order to implement an effective classification method to find the optimal artifact reduction strategy in each case. The classification has been made based on the mean and standard deviation (SD), allowing to differentiate patients with stereotyped and constant movements from those with a greater number of spasm or sudden movements. Since the various signal patterns may require diverse treatments, two artifact reduction methods have been analyzed. The first one is based on the distribution, using again the mean and SD, and the second one is based on an energy function which, theoretically, should be more robust to outliers and more stable to signal to noise ratio. The results confirm the existence of three groups of signals differentiated by having: low mean and low SD, high mean and low SD and high mean and high SD. However, despite finding three different patterns, the energy-based method is the one that works best for all them, offering adequate adaptation to each type of signal without losing robustness and stability. In conclusion, its implementation improves the detection of outliers without compromising artifact-free data segments, which allows to maintain the quality and quantity of the records.*

## 1. Introduction

Electroencephalogram (EEG) has been consolidated over the years as one of the main techniques to identify brain activity and behavior. Measurement of neurophysiological changes related to postsynaptic activity in the neocortex has proven to be a powerful tool for the study of complex neuropsychiatric disorders, since variations in EEG signals depict a definite type of brain abnormality [1]. Information must be extracted by designing and developing signal processing algorithms in order to use it for diagnosis, monitoring and treatment of the identified brain pathologies. It is fundamental to make sure that the signal that will be analyzed corresponds only to the brain activity, and therefore, results will be reliable. This is incompatible with the presence of artifacts, which lead to an obstacle to interpret EEG signals due to their high amplitudes [2].

The challenge comes when signals to be dealt with are from awake pediatrics patients with a severe neurodegenerative disease [2]. Literature covers all known artifact reduction methods but, mostly, under controlled circumstances. For example, during children's sleeping periods, where the sensitivity to artifacts is highly reduced because movement,

eye blinks and muscle artifacts mostly disappear [3]. Constrained signals without uncontrolled big artifacts are well studied and numerous solutions are already presented [4]. But, what happens when patients present movement disorders, seizures or autistic behavior? Or even all at the same time? This is what occurs in the case of patients with Rett Syndrome.

Rett Syndrome (RS) is a neurodevelopmental disorder caused by a mutation in the X-linked dominant MECP2 gene. The occurrence varies between 1/10.000 and 1/15.000 newborn females that follow a normal development during the first 6 to 18 months of age. After this period, patients will lose the abilities acquired until that moment and will initiate a degenerative process that is divided in 4 stages. Principal symptoms that will appear are: microcephaly, intellectual disability, seizures, autistic behavior, hyperventilation, spasticity, hyperreflexia and stereotyped movements [5]. Despite the severity and variability of the symptoms, it is possible to evaluate the level of cognitive performance objectively [6]. Analyzing the power of the EEG signals at the different frequency bands before, during and after visual stimulation tasks, allows to know the awareness, learning and comprehension capacities of patients. The combination of cognitive training and EEG is a powerful tool to better understand the disease and to incorporate communication improvements in the patient's day-to-day life.

For the purpose of the study and given its relevance, ensure the quality of the data constitutes the first step of the analysis. Previous studies with children with similar pathologies have shown difficulties in obtaining an EEG without artifacts. Many patients must be discarded due to excess artifacts, studies are based on signals no longer than 1 minute due to the lack of artifact-free epochs or visual inspections to manually remove artifacts must be done, which is highly time consuming [7], [8]. The most common solution to remove artifacts is to reject the data segments that contain those to get an artifact-free signal to work with. Thus, the aim of the study is to optimize this procedure to find the appropriate method to detect outliers automatically and accurately.

## 2. Methods

### 2.1. Participants, data acquisition and pre-processing

The database is formed by 29 patients with Rett Syndrome recruited from the pediatric neuroscience unit at the Sant Joan de Déu Hospital, Barcelona. A twenty-channel EEG (according to 10/20 system) was recorded with a Starstim 20 wireless device from Neuroelectronics, Barcelona, at a sampling frequency of 500 Hz. An accelerometer was placed at the back of the head to record the movement in x, y and z axes. During the session, the subject was using an eye-tracking device from Tobii Technology to record its visual response to a given stimulus on the screen. A total of 85 records containing basal and activity periods were registered. All the obtained signals are filtered using an elliptic bandpass filter between 1 and 50 Hz.

The study was approved by the local ethics committee following the Declaration of Helsinki.

### 2.2. Influence of patient's symptomatology in data variability

The clinical characteristics of the subjects are important to understand the high variability of the data: there are subjects who remain fairly stable during the session with slight random movements, patients who have a constant and continuous repetitive movement and patients who produce very strong and sudden movements at times but who remain mostly stable throughout the session. This is due to the symptoms of each patient as well as the stage of the disease in which they are. The degree of attention, understanding, motor control, autism or epilepsy varies in each child and influences their behavior during the activities.

In order to classify the data and group subjects according to their degree of movement, the mean and the standard deviation (*SD*) of the modulus of the derivative of the accelerometer signal in the 3 axes are analyzed. The derivative is used to maximize the slope changes of the accelerometer signal to better detect the head movements. Finding the relationship between the mean and the *SD* is the key point for ranking. It is expected to find a group with low mean and *SD*, due to the lack of sudden movements and high stability; another group with high mean but low *SD*, due to constant repetitive movements and, finally, a third group with high mean and *SD* due to sporadic strong movements. Figure 1 shows the mean-*SD* relationship found in the eighty-five recordings from the twenty-nine patients.

Two classification methods are used to analyze the mean and the *SD*: The Kernel Density Estimation (KDE), which represents the data using a continuous two-dimensional probability density curve that is analogous to a histogram, and the Gaussian Mixture Model (GMM), which depicts the density representation as the weighted sum of Gaussian distributions. The GMM algorithm is applied to the dataset for fitting three mixture-of-Gaussian models and to assign each record to the Gaussian it mostly belongs to [9]. As it is a probabilistic model, we can filter by probabilities to keep only those records with a probability (*p*) greater than 95% of belonging to its group and, therefore, obtain only those groups with clearly differentiated data.

### 2.3. Artifact rejection methods

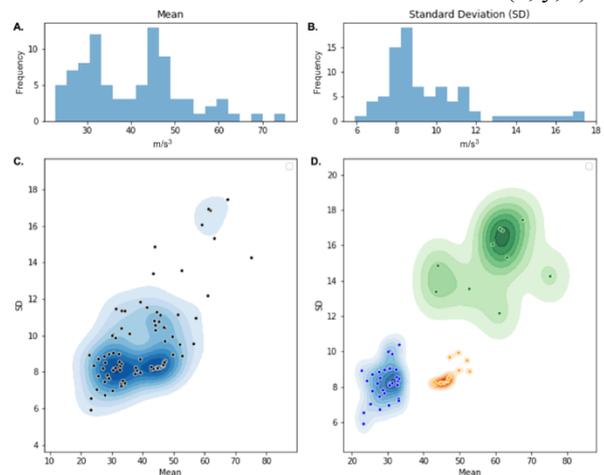
In this study, two artifact rejection methods will be analyzed.

The first one is based on the mean and the standard deviation and needs a *k-factor* to fit the data correctly. It is calculated as the mean + *k-factor* · *SD* of the filtered signal. In this study, 3 *k-factors* are tested: 3, 4 and 5. The second one is based on an energy function that has entropy as its main component. The purpose of the algorithm is to determine the baseline of the signal without the influence that movement peaks may add. It has the advantage of not having a threshold that depends on some events such as artifacts, as the first artifact detection method does by using the data distribution. Therefore, and in theory, this method will be more robust to outliers and more stable to signal-to noise ratio [9]. The methodology consists of: the filtered signal is divided into 100-ms segments and the wavelet entropy of the autocorrelation of each segment is calculated. Then, the baseline is computed as the 95% percentile of the entropy value distribution and the threshold is determined as the *k-value* % of the obtained baseline. For this method, five *k-values* are tried. The five *k-values* are 0.995, 0.996, 0.997, 0.998 and 0.999. As the threshold is determined by the *k-value*-quantile of the absolute value of the baseline, the higher the *k-value* is, the higher the threshold. In order to decide the *k-value*, the percentage of artifact epochs versus the *k-value* will be analyzed.

## 3. Results

### 3.1. Data classification

Figure 1 shows the mean and *SD* distribution of the modulus of the accelerometer derivative in the three axes (x, y, z).



**Figure 1.** *A.* Histogram showing the distribution of the means. *B.* Histogram showing the distribution of the standard deviations. *C.* Mean-*SD* ratio distribution using kernel density estimation (KDE). *D.* Density representation as the weighted sum of Gaussian distributions using the Gaussian Mixture Model (GMM). Three groups are differentiated: 1. LmLsd (blue): low mean and low *SD*. 2. HmLsd (orange): high mean and low *SD*. 3. HmHsd (green): high mean and high *SD*.

Figure 1a shows the histogram of the mean values of the accelerometer where three main peaks can be observed: the first, which is the second highest, has the lowest mean around 30 m/s<sup>3</sup>, the second, that is the highest, represents the group with an intermediate mean around 45 m/s<sup>3</sup>, and finally, the smallest peak shows a more limited group of records that have the highest mean around 60 m/s<sup>3</sup>. Figure 1b shows the histogram of the standard deviation (*SD*). There is one main peak around the value of 8 m/s<sup>3</sup> indicating that most records have a low *SD*, then, there is a lower concentration of records between 10 and 12 m/s<sup>3</sup>, and finally, the smallest part of the distribution between 14 and 18 m/s<sup>3</sup> that represents the registers with the highest *SD*.

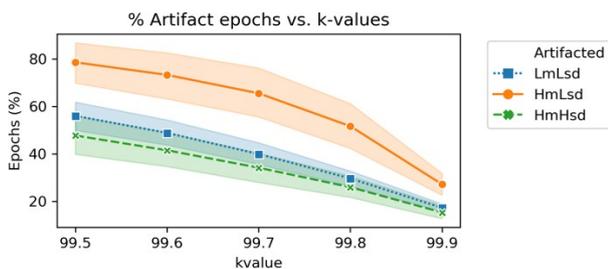
So far, and based on histograms, the presence of the three expected groups is observed. Figure 1c shows the mean- $SD$  ratio distribution using KDE. Three groups are differentiated according to the density distribution: the area of highest density with low mean and low  $SD$ , the second highest density area with a higher mean and the same low  $SD$  and the smallest group with high mean and high  $SD$ .

After recognizing those three distinct groups, Figure 1d shows the results from the application of the GMM algorithm: LmLsd (Low mean Low  $SD$ ), low mean due to stable behavior during the session with a low  $SD$  due to the lack of strong sudden movements. HmLsd (High mean Low  $SD$ ): high mean due to constant repetitive movement during the session but low  $SD$  due to the lack of strong sudden movements. HmHsd (High mean High  $SD$ ): both the mean and the  $SD$  are high because subjects present very strong and sporadic movements, which means that they remain mainly calm during the session but the outliers caused by these movements have extremely high amplitude, which increases the global mean and the  $SD$ .

By clustering with GMM, three groups could be obtained according to the behavior of the children during the sessions. The variability of the signals can cause each group to require a different solution when calculating the threshold to detect artifacts, for this reason, in this study two different methods of artifact rejection will be evaluated.

### 3.2. Selection of the $k$ -value for the energy threshold

The energy threshold was obtained as the  $k$ -value% of the baseline. Five different values have been tested to find the optimum. Figure 2 shows the relationship between the  $k$ -values and the percentage of artifact epochs for the three groups classified in section 3.1.

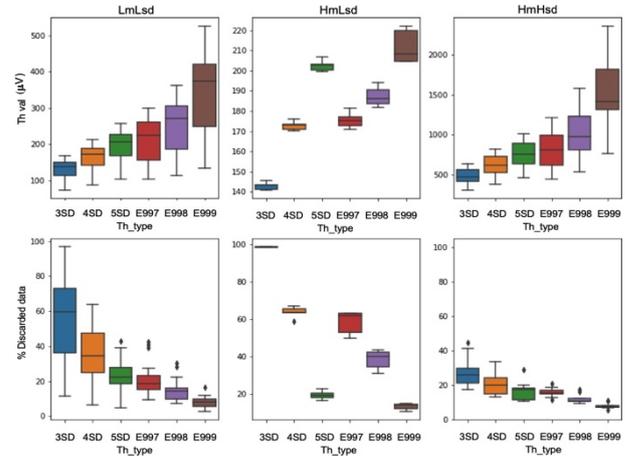


**Figure 2.**  $K$ -values ( $x$  axis) versus the percentage of epochs with artifacts ( $y$  axis) for the three groups LmLsd, HmLsd and HmHsd.

Both LmLsd and HmHsd, show a linear decrease in the percentage of artifactual epochs as  $k$ -value increases, while in the case of HmLsd, the variation in the percentage of artifactual epochs is more abrupt and is quantized with notable changes in the slope between each  $k$ -value. The HmLsd is the one that presents a differentiated behavior and the response to each  $k$ -value implies a greater change in the final selection of data. Before taking a decision, the evaluation will continue in the next section 3.3, where the thresholds with  $k$ -values from 99.7 to 99.9% will be studied in depth, as they are the ones that present the most abrupt changes in the HmLsd slope.

### 3.3. Threshold examination: mean- $SD$ vs. energy

Once all procedures are defined, it is time to see how the different thresholds fit the data and what percentage of artifacts are detected in each case.



**Figure 3.** The figure shows the threshold values and the percentage of discarded data, both versus each type of threshold. The mean plus  $SD$  method is tested with three “ $k$ -factors” of 3, 4 and 5 and the energy-based threshold is tested for three of the five  $k$ -values shown in Figure 2, due to preliminary results. The  $k$ -values are: 0.997, 0.998 and 0.999.

The observation and analysis of Figure 3 in detail is crucial to understand how both methodologies are working and how they affect data. The first column contains the information regarding the LmLsd group. It can be seen how all three energy-based thresholds have higher median values (from 220 to 380  $\mu V$ ) than the mean- $SD$  thresholds, which increase more smoothly and linearly and go from 120 to 200  $\mu V$ . Therefore, the mean plus  $SD$  method seems to have a lower variation of the median threshold values for the different  $k$ -factors, and it also shows less variability between records (the range of values between quartiles of 25 and 75% is smaller), however, its influence on the percentage of discarded data is the opposite. It ranges from 60% of the data discarded with 3SD to 20% in 5SD, showing in this case a greater variation of the median threshold values and a greater variability between records than the energy-based threshold. This second one has a greater disparity between  $k$ -values in the median threshold calculation, but it is more precise selecting outliers and shows less variability between records, going from 20% of the discarded data with  $k=0.997$  to 10% with  $k=0.999$ .

The second column, which describes the behavior of the HmLsd group, shows that the energy-based thresholds vary from 175 to 205  $\mu V$  while for the mean- $SD$  method go from 142 to 203  $\mu V$ . In this case, the energy threshold presents a smaller variation with respect to the calculation of the median thresholds, unlike the LmLsd group, but greater variability between registers, as in the previous group. Following the same line, the percentage of discarded data is again more consistent for the energy thresholds, which vary from 60% for the lowest  $k$ -value to 15% for the highest, while for the 3SD threshold the percentage of discarded data is almost the 100% and, for the 5SD, it is around 20%. Therefore, the mean plus  $SD$  method is less consistent when setting the thresholds and discarding the data.

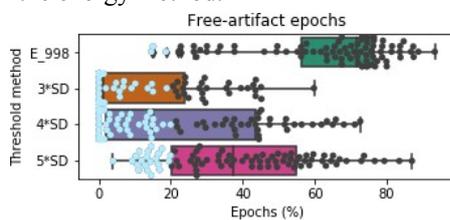
Finally, the third column shows the boxplots for the HmHsd group, whose behavior is very similar to LmLsd. All three energy-based thresholds are greater than the mean- $SD$  ones with respect to the threshold's median values (750 to 1450  $\mu V$  versus 500 to 750  $\mu V$ , respectively), and again, this second method is more consistent calculating the median threshold values for the different  $k$ -factors.

Likewise, the variability between registers when calculating the thresholds is greater for the energy-based threshold. On the contrary, and also coinciding with the first group, the variability between records in the percentage of discarded data is greater for the mean- $SD$  method. Regarding the median values of the percentage of discarded data, both methods show a similar behavior, going from 26 to 18  $\mu V$  for the mean- $SD$  method and from 17 to 8  $\mu V$  for the energy-based method.

In summary, the variability in the threshold calculation is greater for the energy-based method while it is also more consistent discarding data. For this reason, it would be the most suitable method for all three groups. The last decision to make is which  $k$ -value to use. Observing Figure 3 and analyzing Figure 2 again, the decision will be made based on the calculation of the slopes between the  $k$ -values of the HmLsd group (due to its differentiated behavior). It is observed that the steepest change in the slope occurs between 99.8% and 99.9%, so the  $k$ -value that will be used to define the energy-based threshold is 0.998.

Figure 4 translates these results to observe the effect of the threshold chosen as optimal (energy threshold with  $k$ -value of 0.998) against the three thresholds of mean plus  $SD$ , which are used as the reference classification because were already used in the previous study to obtain the preliminary results that lead to the actual one.

For the purpose of the study, it is considered that at least 20% of the data must be free of artifacts. In Figure 4 all those records that conserve less than this 20% are shown in blue, while the ones that conserve enough data are shown in black. The difference is visible between the two main methods, energy and mean- $SD$ . For the energy threshold, only 3 records would be eliminated, while for the  $3SD$  method more than 50% of the registers would be removed, specifically 56 out of 85. In the case of the  $4SD$ , a total of 47 registers would be eliminated (also over 50%) and finally, for the  $5SD$ , 21 records would be lost. So, for the less restrictive mean- $SD$  method, the percentage of records deleted is 24.7% versus 3.5% for the energy method.



**Figure 4.** The figure presents a boxplot for each threshold method (x-axis) with the information regarding the percentage of free-artifact epochs (y-axis) for all records.

#### 4. Conclusions and limitations

The complexity and variability of symptoms in patients with Rett Syndrome require an equally complex and diverse data analysis. The initial idea of the study about the existence of different groups of patients, due to their clinical stage and behavior, has been exhibited by using the mean and standard deviation. Three groups of registries have been identified, showing the importance of knowing the origin of movement, since it influences the performance of the different artifact detection methods. However, after evaluating both methods, if a single solution has to be offered for all types of data, the energy-based method is the optimal for all them.

The greater consistency discarding data together with a greater variability in the calculation of the thresholds, implies a more precise adaptation to each signal, obtaining more robustness in the selection of artifacts. This helps to reduce the loss of artifact-free data by removing only outliers, leading to a reduction in the percentage of records deleted.

The main limitations of the study lie in the signal recordings, since the artifacts must be annotated in real time to contrast automatically with the classification of the data and the detection of outliers. So far it has been done with video recordings, which is a manual method that can include more errors and involves a greater investment of time.

#### 5. Acknowledgements

We would like to acknowledge specific funding support from the Spanish Patient Associations Mi Princesa Rett and Rettando al Síndrome de Rett. This project has received also funding from Torrons Vicens and the Ministry of Economy and Competitiveness (MINECO), Spain, under contract DPI2017-83989-R. CIBER-BBN is an initiative of the Instituto de Salud Carlos III, Spain. Alejandro Bachiller is a Serra Hünter Fellow.

#### 6. References

- [1] J. Wang, J. Barstein, L. E. Ethridge, M. W. Mosconi, Y. Takarae, and J. A. Sweeney, "Resting state EEG abnormalities in autism spectrum disorders," *Journal of Neurodevelopmental Disorders*, vol. 5, no. 1, pp. 1–14, Dec. 2013, doi: 10.1186/1866-1955-5-24.
- [2] M. Dovgialo *et al.*, "Assessment of Statistically Significant Command-Following in Pediatric Patients with Disorders of Consciousness, Based on Visual, Auditory and Tactile Event-Related Potentials," *International Journal of Neural Systems*, vol. 8, no. 3, p. 1850048, 2019, doi: 10.1142/S012906571850048X.
- [3] A. Kaminska, F. Cheliout-Heraut, M. Eisermann, A. Touzery de Villepin, and M. D. Lamblin, "EEG in children, in the laboratory or at the patient's bedside," *Neurophysiologie Clinique*, vol. 45, no. 1, pp. 65–74, Mar. 2015, doi: 10.1016/j.neucli.2014.11.008.
- [4] D. W. Klass, "The continuing challenge of artifacts in the EEG," *American Journal of EEG Technology*, vol. 35, no. 4, pp. 239–269, Dec. 10, 1995, doi: 10.1080/00029238.1995.11080524.
- [5] N. P. Verma, R. L. Chheda \*\*, M. A. Nigro, and Z. H. Hart, "Electroencephalographic findings in Rett Syndrome" 1986.
- [6] P. M. Baptista, P. M. Baptista, M. T. Mercadante, E. C. Macedo, and J. S. Schwartzman, "Correspondence: Cognitive performance in Rett syndrome girls: a pilot study using eyetracking technology," *Journal of Intellectual Disability Research*, doi: 10.1111/j.1365-2788.2006.00818.x.
- [7] R. A. Fabio, L. Billeci, G. Crifaci, E. Troise, G. Tortorella, and G. Pioggia, "Cognitive training modifies frequency EEG bands and neuropsychological measures in Rett syndrome," *Research in Developmental Disabilities*, vol. 53–54, pp. 73–85, 2016, doi: 10.1016/j.ridd.2016.01.009.
- [8] K. J. Roche, J. J. Leblanc, A. R. Levin, H. M. O'Leary, L. M. Baczewski, and C. A. Nelson, "Electroencephalographic spectral power as a marker of cortical function and disease severity in girls with Rett syndrome," *Journal of Neurodevelopmental Disorders*, vol. 11, no. 1, p. 15, Jul. 2019, doi: 10.1186/s11689-019-9275-z.
- [9] C. Migliorelli *et al.*, "SGM: a novel time-frequency algorithm based on unsupervised learning improves high-frequency oscillation detection in epilepsy," *J. Neural Eng.*, vol. 17, p. 26032, 2020, doi: 10.1088/1741-2552/ab8345.