







# Performance of Different Average Methods for the Automatic Detection of Evoked Potentials

Idileisy Torres-Rodríguez<sup>1</sup> , Carlos Ariel Ferrer-Riesgo<sup>1</sup> ,  
Juan Carlos Oliva Pérez<sup>2</sup> , and Alberto Taboada-Crispi<sup>1</sup> 

<sup>1</sup> Informatics Research Center, Universidad Central “Marta Abreu” de Las Villas, UCLV, 54830 Santa Clara, Cuba

{itrodriguez, cferrer, ataboada}@uclv.edu.cu

<sup>2</sup> Electronic and Telecommunications Department, Universidad Central “Marta Abreu” de Las Villas, UCLV, 54830 Santa Clara, Cuba  
jpoliva@uclv.cu

**Abstract.** The evoked potentials can be auditory, visual or somatosensory. Noise reduction is the first step in most biomedical signal processing systems. The quality and accuracy of the rest of the operations carried out on the signal depend to a large extent on the quality of the noise reduction algorithms that have been used in the preprocessing of the signal. The method commonly used to enhance the signal of interest is the coherent average, however, this technique has some limitations that justify the search for alternatives to detect or extract the characteristics of these signals. The weighted average is a possible alternative; however, this is still not appropriate enough when the signal may have outliers within the epoch. Trimmed average techniques have a better solution when in the presence of impulsive noise. The modified trimmed average is adapted for use in auditory evoked potentials. In this work, we compare different techniques of trimmed average, with the weighted average and the coherent average; using quality measures in the frequency domain for detect Auditory Brainstem Evoked Potentials. The results showed that the proposed method and his variants are superior to the rest of the used ones. The Q-Sample Modified measure offers the best result.

**Keywords:** Auditory evoked potentials · Ensemble average · Weighted average · Trimmed average · Q-sample uniform · Q-Sample Modified · Watson Q-sample

## 1 Introduction

### 1.1 Evoked Potentials and Their Classification

The specific neural activity that arises from acoustic stimulation as a pattern of voltage fluctuations lasting approximately half a second is an auditory evoked potential [1]. Depending on the type and placement of the electrodes, the amplification of the signal, the selection of the filters and the post-stimulation period, it is possible to detect the

neuronal activity that arises from different structures that span from the auditory nerve to the cerebral cortex [2, 3]. Noise reduction is the first step in most biomedical signal processing systems. The quality and accuracy of the rest of the operations carried out on the signal depend to a large extent on the quality of the noise reduction algorithms that have been used in the preprocessing of the signal. The coherent average (CA) or arithmetic mean, as it is also known, can be calculated from the ensemble matrix that is formed with the evoked responses. Where the response to the  $i$ -th stimulus is assumed as the sum of the deterministic component of the signal or response evoked  $s$  plus a random noise  $r_i$  (see Eq. 1), which is asynchronous with the stimulus. Where the noise in progress is assumed to be stationary, with zero mean. Consequently, the variance of the noise must be fixed and equal in all the potentials. The average is a simple and direct method. The estimated signal can be modeled as the sum of the deterministic component plus the attenuated noise by a factor of  $1/M$ , as show in Eq. 2.

$$p_i = s + r_i \quad (1)$$

$$\hat{s} = \frac{1}{M} \sum_{i=1}^M p_i(n) \quad (2)$$

Considering the presence of several types of noise which cause degradation in the performance of the average, the development of new methods that correctly handle these problems is justified. One of these methods that have been proposed in the literature is the weighted average. Several criteria have been used in order to determine the vector of weights that best fits the problem. One of these criteria (minimization of the mean square error) is based on the noise variance of all the cycles. A potential with a high noise level is assigned a lower weight than one with a lower noise level [4–8].

The average ensemble and the weighted average represent linear techniques, and consequently they perform very well when the noise is of Gaussian type. However, in the case that out-of-range artifacts appear occasionally, with large amplitude values, these techniques are limited. The ensemble average and the ensemble median can be seen as special cases within a broad family of existing estimators known as trimmed means, within which are the trimmed average, the Winsorized average, the L-trimmed average, or mean TL and the average Tanh [9–12].

## 1.2 Automatic Detection of Evoked Potentials

Methods for the objective detection of Auditory Brainstem Evoked Potentials (ABR) can be characterized as template-based methods and non-template based methods, as reviewed by [6] and referenced in [13]. Most used methods for detection in the frequency domain. The original test of uniform sample scores  $q$  (Q-sample uniform) [14] is a nonparametric test that uses the phase ranges of the Fourier components of Q spectral bands to test whether the phases share the same distribution. The test only uses the phase angles in the form of their ranges and rejects the spectral amplitudes.

The most powerful test in the frequency domain according to [15] is Q-sample uniform. This test uses only the phase angles in the form of their ranges while the spectral information amplitude is rejected. In [16] a modification is introduced where the spectral amplitude is also considered, this test has come to occupy a better position than its predecessor in its use for the detection of auditory evoked potentials. Another test [17] that can be considered as a special type of Q-sample test is the Watson Q-sample. This test also uses both phase angles and spectral amplitudes.

## 2 Methods

Two versions of the Modified Trimmed Mean [18] are adapted for their use in Brainstem Auditory Evoked Potentials (BAEP). It also analyzes the characteristics of the database and how the comparison between the different methods will be carried out.

### 2.1 Modified Trimmed Mean Adapted to the BAEP

The Trimmed Mean Modified in [18], proposes to determine the cut-off factor  $t$  as  $2 * \sigma_{\Gamma}$ , where  $\sigma_{\Gamma}$  is the standard deviation of the average background noise, estimated from the isoelectric segment of the electrocardiographic signal as posted in [19–21]. In the case of evoked potentials, they do not have an isoelectric segment. However, in the literature consulted several ways of estimating the variance of background noise have been found. Following this approach, the trimming factor could be estimated as 3 or 2 standard deviations of noise, based on the fact that the variance is the square of the standard deviation. In several articles also reviewed, the estimation of background noise has been proposed as the average of the variances of several unique points, used in the estimation of the Fmp [22, 23], as a better approximation to the noise of the signal. In this paper we propose to determine the cut-off factor  $t$  as three standard deviations of the estimated background noise using the Fmp.

Another variant of modifying the cut average in order to determine the cut-off factor  $t$  was the use of the interquartile range (IQR), defined as the difference between the third quartile  $Q_3$  and the first quartile,  $Q_1$ . The interquartile range is a robust measure, because it only takes into account 50% of the data.

$$[Q_1 - 1.5 * IQR, \quad Q_3 + 1.5 * IQR] \quad (3)$$

The range given in this equation depends on the factor 1.5, this is an arbitrary value, but it finds its justification according to a normal distribution. In this way, two versions of the Modified trimmed mean are obtained.

### 2.2 Data

The database used in this study consists of Transient Auditory Evoked Potentials registered in 39 neonatal patients between 1–3 months of age born in Hospital Materno

Ramón González Coro, in Havana, Cuba [27]. The signals were recorded with an AUDIX electroaudiometer. A click stimulus with duration 0.1 ms was provided at different intensities (100, 80, 70, 60, 30 dBnHL and 0 dBpSPL) via insert earphones (EarTone3A) [28, 29]. Ag/AgCl dry electrodes were used, which were fixed with electrolytic paste on the forehead (positive), ipsilateral mastoids (negative) and contralateral mastoids (earth). The impedance values were maintained below 5 k $\Omega$ . The sampling frequency used was 13.3 kHz, and the analysis windows to form the ensemble matrix  $P$  (Eq. 5) and calculate the coherent average were of approximately 15 ms, that is about 200 samples per window ( $N = 200$ ). From this database, only records obtained at 100 dBnHL (78 signals) were used, where it was confirmed by specialists that a response was present. These signals were used in order to guarantee the maximum values of the quality measures for this database.

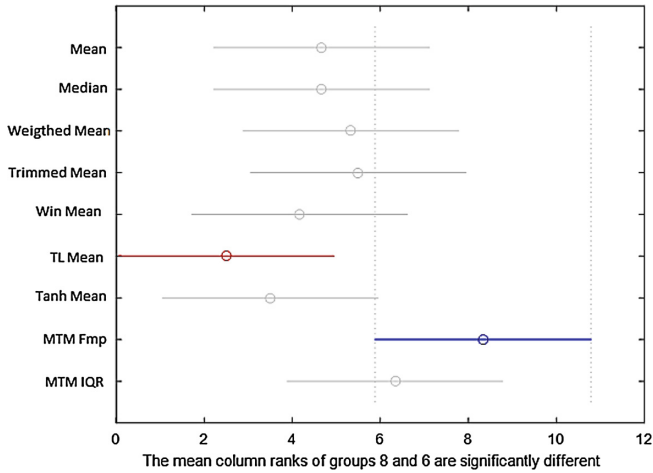
### 2.3 Description of the Experiment

In order to carry out the experiment, set matrices of each of the registers were formed using the time between stimuli of 15 ms as reference. With each record an ensemble matrix of approximately 2000 epochs was formed by 200 samples on average. Each ensemble matrix was averaged using the average methods described in previous sections. The automatic detection measures used to establish the comparison were measured in the frequency domain; they were the Q-Sample Uniform, the Modified Q-Sample and the Q-Sample of Watson U2.

To establish the comparison of the different methods using measures in the frequency domain, first, the records were transformed into set matrices. From each set matrix, sets of 250 epochs are randomly taken 100 times and averaged using each of the average methods that were described. A new matrix ( $100 \times 200$ ) composed of the average vectors obtained in the previous step was formed trying to simulate a Monte Carlo experiment. The matrices obtained for each of the average methods and for each of the different sizes of  $X$  are transformed to the frequency domain using the fft (Fast Fourier Transform, from its acronym in English), from which it is obtained an array of dimension  $100 \times 200 \times 2$ , which contains the phase and amplitude values of the result of applying the transform. This arrangement is to which the different quality measures are calculated. The design of the experiment is based on the design of the experiment proposed in [32] for the comparison of detection measures in the frequency domain.

## 3 Results

To evaluate the results obtained, a Friedman test was performed. The multipair comparison returns an interactive plot that allows to visually determining which of the methods have differences between them as show the Fig. 1. In all cases, the test resulted in a value of  $p < 0.05$ , which suggests that there are significant differences between at least two methods. In order to identify the methods in which the differences existed, a post-hoc test was developed using the Bonferroni method.



**Fig. 1.** Multiple comparison of methods using all quality measures

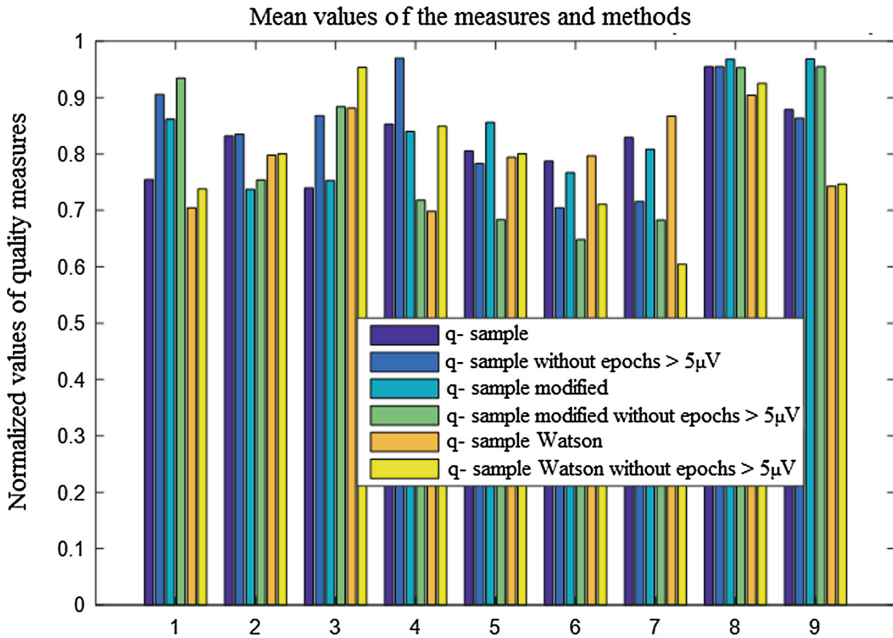
Table 1 gives the values of the average range returned by the Friedman test, giving a higher score to the method with the best performance. In this case, although the significant differences are only between the Average TL and the MTM Fmp, according to the Range the best method was the MTM Fmp, followed by the MTM IQR., Which constitute the proposals of this work.

**Table 1.** Values of the average ranges given by the Friedman test to each of the methods.

Mean	Median	Weighted average	Trimmed mean	Mean Winsorized	Mean TL	Mean Tanh	MTM Fmp	MTM IQR
4.667	4.667	5.333	5.500	4.167	2.500	3.500	8.333	6.333

Another analysis that is important to consider is to determine which method presented minor differences when it acted with the clean database and with the data without artifacts, values outside the range of  $\pm 5 \mu\text{V}$ . Figure 2 is a total view of all the methods under analysis, in them it can be seen that for the coherent average there are visible differences in their behavior, while the proposed methods prove to be more robust in behaving in a similar way before the two variants of the data.

Next, the signals obtained from a random subject are shown using the different methods. In Fig. 3 it can be seen how the MTM Fmp method is the one that comes closest to the expected while other methods move away from the expected signal and present higher noise levels.



From left to right, 1 is Mean, 2 is Median, 3 is Weighted mean, 4 is Trimmed mean, 5 is Win mean, 6 is TL mean, 7 is Tanh mean, 8 is MTM Fmp and 9 is MTM IQR

Fig. 2. Mean values of the measures and methods.

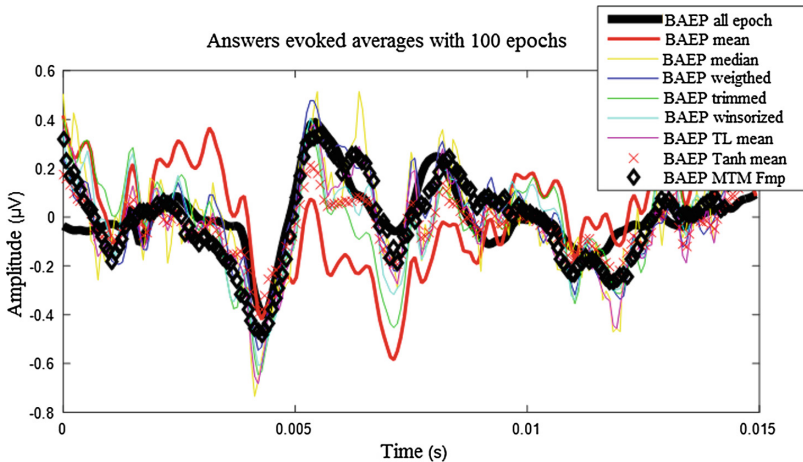


Fig. 3. Answers evoked averages with 100 epochs

## 4 Conclusions

Despite the variety of the proposed methods, this paper describes the proposals for modifications made to the Modified Trimmed Media in order to adapt it for the reduction of noise of auditory evoked potentials. The Modified Trimmed Media is a method that proposes to combine the solutions to the main drawbacks of the coherent average. The Friedman test used showed that the best method was the MTM using Fmp, however, the differences were significant only with respect to the Trimmed Media TL. From the review of the bibliography, the measurements in the frequency domain were selected to establish the comparison between the results obtained by each method. The measure that behaved better was modified Q-sample.

As a consequence of the analysis carried out, it is recommended that the experiment be developed with a data with higher noise levels, which is carried out in a more exhaustive way, checking the results in a Monte Carlo experiment for smaller steps in the formation of the set matrix trying to simulate a real recording environment. It is also recommended that the results obtained be compared with similar results, but in the time domain.

## References

1. Sharma, A., Dorman, M.F., Kral, A.: The influence of a sensitive period on central auditory development in children with unilateral and bilateral cochlear implants. *Hear. Res.* **203**(1–2), 134–143 (2005)
2. Eggermont, J.J.: On the rate of maturation of sensory evoked potentials. *Electroencephalogr. Clin. Neurophysiol.* **70**(4), 293–305 (1988)
3. Kraus, N., Nicol, T.: Auditory evoked potentials. In: Binder, M.D., Hirokawa, N., Windhorst, U. (eds.) *Encyclopedia of Neuroscience*. Springer, Heidelberg (2008). <https://doi.org/10.1007/978-3-540-29678-2>
4. Sornmo, L., Laguna, P.: *Bioelectrical Signal Processing in Cardiac and Neural Applications* (2005)
5. Stapells, D.R.: Frequency-specific threshold assessment in young infants using the transient ABR and the brainstem ASSR. In: *Comprehensive Handbook of Pediatric Audiology*, pp. 409–448 (2011)
6. Suárez, A., De Armas, J.L., Aznielle, T., Charroó, L., Hernández, R.: SenseWitness: Un nuevo software para potenciales evocados (2003)
7. Nodarse, E.M., et al.: Cribado auditivo neonatal con potenciales evocados auditivos de estado estable a múltiples frecuencias. *Acta Otorrinolaringol. Esp.* **62**(2), 87–94 (2011)
8. Torres, I.: *Estimadores Robustos de Promedio para la Detección de Potenciales Relacionados con Eventos*. Universidad Central “Marta Abreu” de Las Villas (2014)
9. Elamir, E.A.H., Seheult, A.H.: Trimmed L-moments. *Comput. Stat. Data Anal.* **43**(3), 299–314 (2003)
10. Leonowicz, Z., Karvanen, J., Shishkin, S.L.: Trimmed estimators for robust averaging of event-related potentials. *J. Neurosci. Methods* **142**(1), 17–26 (2005)
11. Hampel, F., Ronchetti, E., Rousseeuw, P., Stahel, W.: *Robust Statistics: The Approach Based on Influence Functions*. Wiley, Hoboken (2011)
12. Valderrama, J.T., Alvarez, I., De La Torre, A., Segura, J.C., Sainz, M., Vargas, J.L.: A portable, modular, and low cost auditory brainstem response recording system including an algorithm for automatic identification of responses suitable for hearing screening. In: *IEEE EMBS*

- Special Topics Conference on Point-of-Care Healthcare Technologies Synergy Towards Better Global Healthcare, PHT 2013, pp. 180–183 (2013)
13. Laciár, E., Jane, R.: An improved weighted signal averaging method for high-resolution ECG signals, pp. 69–72 (2001)
  14. Chesnaye, M.A., Bell, S.L., Harte, J.M., Simpson, D.M.: Objective measures for detecting the auditory brainstem response: comparisons of specificity, sensitivity and detection time. *Int. J. Audiol.* **57**, 1–11 (2018)
  15. Lins, O., Picton, P., Picton, T.: Auditory steady-state responses to tones amplitude-modulated at 80–110 Hz. *J. Acoust. Soc. Am.* **97**, 3051–3063 (1995)
  16. Congedo, M., Korczowski, L., Delorme, A., Lopes da Silva, F.: Spatio-temporal common pattern: a companion method for ERP analysis in the time domain. *J. Neurosci. Methods* **267**, 74–88 (2016)
  17. Stürzebecher, E., Cebulla, M., Wernecke, K.D.: Objective response detection in the frequency domain: comparison of several q-sample tests. *Audiol. Neuro-Otology* **4**(1), 2–11 (1999)
  18. Taboada-Crispi, A., Lorenzo-Ginori, J.: Adaptive line enhancing plus modified signal averaging for ventricular late potential detection. *Electronics* **35**, 1293–1295 (1999)
  19. Heckenlively, J., Arden, G.: *Principles and Practice of Clinical Electrophysiology of Vision*. MIT Press, Cambridge (2006)
  20. Leski, J.M.: Robust weighted averaging. *IEEE Trans. Biomed. Eng.* **49**(8), 796–804 (2002)
  21. The BC Early Hearing Programme: Audiology Assessment Protocol (BCEHP) (2012). <http://www.phsa.ca/health-professionals/clinical-resources/hearing/BCEarlyHearingProgram:Audiologyassessmentprotocols.pdf>
  22. Mijares, E., et al.: Sensitivity of the automated auditory brainstem response in neonatal hearing screening. *Pediatrics* **3**(3), 117–128 (2014)
  23. Da Silva, I.G.A.: Objective estimation of loudness growth using tone burst evoked auditory responses. *Diss. Abstr. Int. B Sci. Eng.* **70**(8), 4987 (2010)
  24. Paradis, J.: The development of English as a second language with and without specific language impairment: clinical implications. *J. Speech Lang. Hear. Res.* **59**(2), 171–182 (2016)
  25. Pander, T.: A new approach to robust, weighted signal averaging. *Biocybern. Biomed. Eng.* **35**(4), 317–327 (2015)
  26. Levit, Y., Himmelfarb, M., Dollberg, S.: Sensitivity of the automated auditory brainstem response in neonatal hearing screening. *Pediatrics* **136**(3), e641–e647 (2015)
  27. Rodrigues, G.R.I., Ramos, N., Lewis, D.R.: Comparing auditory brainstem responses (ABRs) to toneburst and narrow band CE-chirp® in young infants. *Int. J. Pediatr. Otorhinolaryngol.* **77**(9), 1555–1560 (2013)
  28. Kotas, M., Pander, T., Leski, J.M.: Averaging of nonlinearly aligned signal cycles for noise suppression. *Biomed. Signal Process. Control* **21**, 157–168 (2015)
  29. Pander, T., Pietraszek, S., Przybyła, T.: A novel approach to robust weighted averaging of auditory evoked potentials. In: Piętka, E., Kawa, J., Wicławek, W. (eds.) *Information Technologies in Biomedicine, Volume 4. AISC*, vol. 284, pp. 321–332. Springer, Cham (2014). [https://doi.org/10.1007/978-3-319-06596-0\\_30](https://doi.org/10.1007/978-3-319-06596-0_30)
  30. Valderrama, J.T., et al.: Automatic quality assessment and peak identification of auditory brainstem responses with fitted parametric peaks. *Comput. Methods Programs Biomed.* **114**(3), 262–275 (2014)
  31. Berninger, E., Olofsson, A., Leijon, A.: Analysis of click-evoked auditory brainstem responses using time domain cross-correlations between interleaved responses. *Ear Hear.* **35**(3), 318–329 (2014)
  32. Tsurukiri, J., Nagata, K., Kumasaka, K., Ueno, K., Ueno, M.: Middle latency auditory evoked potential index for prediction of post-resuscitation survival in elderly populations with out-of-hospital cardiac arrest. *Signa Vitae* **13**, 80–83 (2017)