

# An Offline/Real-Time Artifact Rejection Strategy to Improve the Classification of Multi-Channel Evoked Potentials

Hyunseok Kook, Lalit Gupta, Dennis Molfese, Heikki Lyytinen, & K.C. Fadem

*Abstract*—The primary goal of this paper is to improve the classification of multi-channel evoked potentials (EPs) by introducing a new temporal domain artifact detection strategy and using this strategy to evaluate how the performance of classifiers is affected by artifacts in offline and real-time classification experiments. The temporal domain artifact detection tests include: a within-channel standard deviation (STD) test that can detect signals with little or no variations in each channel and also detect faulty channels, a within-channel clipping (CL) test to detect amplitude clipped EPs in each channel, and a multi-channel EP median distance (MC-MED) test to detect atypical signals not identified by the STD and CL tests. The performance of the artifact detection strategy is demonstrated on real single-trial EP ensembles and it is shown that the strategy is quite effective in identifying atypical EPs. The classification performance is evaluated on several real EP ensembles with artifact contaminations in the training and test sets, in only the training set, and on ensembles that are free of artifacts in both the training and test sets. It is shown that the improvement in classification accuracy through the incorporation of the artifact detection strategy can be quite significant especially if artifacts are removed from both the training and test sets in offline and in real-time classification trials. Furthermore, the generalized formulation of the artifact rejection classification strategy makes it adaptable to various other multi-class, multivariate signals, of multiple sensors.

## I. INTRODUCTION

THE general procedure for designing pattern classifiers is to collect an ensemble of patterns from each pattern category, partition the ensembles into a training set and a test set, use the training set to estimate the classifier parameters, and use the test set to evaluate the classifier performance. Although efforts are made to carefully collect the patterns in practice, the ensembles are often contaminated with artifacts. In this study, we define artifacts as patterns in the training set that lead to inaccurate estimation of classifier parameters and patterns in the test set that yield misleading performance evaluations. In real-time classification, such artifacts can give inaccurate test results which can have serious consequences, for example, inaccurate clinical diagnosis. It is clearly more prudent to repeat a test if an artifact can be detected rather than to have

an inaccurate test result. The study in this paper, therefore, focuses on the following questions: (a) Can artifacts be detected in off-line and in real-time classification experiments? and (b) How do artifacts affect the estimation of classifier parameters and the classifier performance? The first question will be answered by introducing a new temporal domain artifact detection strategy which can not only accurately detect artifacts in offline ensembles but can also detect artifacts accurately in real-time. The second question will be answered by evaluating and comparing the performance of a selected classifier using (a) training and test sets contaminated with artifacts, (b) an artifact-free training set and a test set contaminated with artifacts, and (c) artifact-free training and test sets. To our knowledge, such a study investigating the effects of artifacts on offline and real-time classifier performance is quite unique.

We will specifically focus on the classification of evoked potentials (EPs) and the ultimate objective is to improve the performance of EP classifiers by detecting and rejecting artifacts. Artifacts in EP/EEG brain waveform recordings typically result from voltage changes due to eye blinks, eye movements, muscle activities, and power line noise. Artifact detection in EPs has received considerable attention because artifacts are known to frequently occur in EP acquisition and various techniques have been developed to detect artifacts. Examples include techniques based on regression [1], PCA [2], and ICA [3].

We introduce a new temporal domain artifact detection strategy consisting of a series of generalized tests to detect artifacts, if any, in an ensemble of multi-channel single-trial EPs. The performance of the artifact detection strategy is demonstrated on real EP data sets of 2-different paradigms: match-mismatch and mismatch negativity. The classification performance without and with artifact detection using our artifact detection strategy is demonstrated by incorporating it into our recently developed multi-channel decision fusion strategy [4].

## II. ARTIFACT DETECTION STRATEGY

The temporal domain artifact detection strategy involves 2 steps: detecting the easily identifiable Type A artifacts in the first step and detecting the more difficult Type B artifacts in the second step. In Step 1, the Type A artifacts are rejected by first removing signals with little or no amplitude variations using a standard deviation (STD) test and removing signals whose samples have been clipped using a

H. Kook and L. Gupta are with the Department of Electrical & Computer Engineering, Southern Illinois University Carbondale, IL 62901 USA.

D. Molfese is with the Department of Psychological & Brain Sciences, University of Louisville Louisville, KY 40292.

H. Lyytinen is with the Department of Psychology, University of Jyväskylä, Finland

K. C. Fadem is with Neuronetrix, Inc., Louisville, KY 40208.

clipping test (CL) test. The STD and CL tests are within-channel tests designed to detect artifacts in each channel. In this step, we implicitly assume that if an artifact occurs in one channel, then the responses of all the channels are also artifacts. Therefore, for a given trial, if a Type A artifact is detected in any one or more channels, the single-trial data of all the channels for that trial are removed. Note that the STD test can be used to identify faulty “stuck-at” recording channels, that is, channels that always give the same readings. If a channel has a stuck-at fault, the EPs of that channel are discarded. The Type B artifacts are detected in Step 2 using a multi-channel approach in which a single-trial multi-channel EP (MCEP) is formed by concatenating the single-trial EPs of all channels. Instead of estimating the underlying MCEP from the Type A-free ensemble using conventional averaging, we use the median MCEP as the representation of the ensemble of each channel. The more difficult Type B artifact detection problem then becomes one of identifying single-trial responses that are not similar to the median MCEP using an appropriate measure for determining similarity. We prefer to use this multi-channel approach at the second stage because we do not want to decrease the ensemble size by discarding the EPs of all channels just because one or few channels may appear to have problematic EPs. Furthermore, if a single trial MCEP is truly an artifact, it would appear as problematic across the majority of the channels.

The 3 tests are described next using  $Z_{m/c;n}$  to represent single trial EP  $n, n=1,2,\dots,N$ , in the ensemble of class  $c, c=1,2,\dots,C$ , recorded at channel  $m, m=1,2,\dots,M$ , where,  $N$  is the number of single-trial EPs in each ensemble (assumed equal across channels and classes),  $C$  is the number of brain activity categories (classes), and  $M$  is the number of channels. The  $c$ -class ensemble of EPs collected at channel  $M$  will be referred to as the  $m/c$ -ensemble. The  $n$ th single-trial MCEP formed by concatenating the  $M$  single-trial responses at trial  $n$  is represented by  $Z_{c;n}$  [5]. That is

$$Z_{c;n} = Z_{1/c;n} \nabla Z_{2/c;n} \nabla Z_{3/c;n} \dots \nabla Z_{M/c;n} \quad (1)$$

where  $\nabla$  represents the concatenation operation. If the dimension of each single trial EP is  $K$ , then the dimension of MC-EP will be  $M \times K$ .

#### A. The Standard Deviation (STD) Test

The standard deviation  $\sigma_{m/c;n}$  of the samples of a single-trial response  $Z_{m/c;n}$  in the  $m/c$ -ensemble is compared with a threshold  $\tau_\sigma$ . If the standard deviation of a single-trial EP  $Z_{m/c;n}$  is below the threshold, then, the  $n$ th single-trials for all  $M$  channels are regarded as artifacts and are discarded from each  $m/c$  ensemble. The threshold selected is close to zero in order to detect responses that are relatively constant

(flat) over the entire duration of the ERP. The threshold  $\tau_\sigma$  is not a function of class  $c$  because the STD test is the same across all  $C$  classes.

#### B. The Clipping (CL) Test

This CL test is designed to exclude single-trials whose amplitudes have been clipped. An EP will be detected as a clipped signal if more than  $\lambda$  samples have the same maximum or minimum values. If  $Z_{m/c;n}$  is clipped for one or more values of  $m$ , then, the MCEP  $Z_{c;n}$  is regarded as clipped and removed from the ensemble of class  $c$ . Also note that the parameter  $\lambda$  is not a function of  $c$  because the CL test is the same across all  $C$  classes

#### C. The Multi-Channel Median Distance (MC-MED) Test

This iterative test is designed to identify the remaining atypical responses in the ensembles after the STD and CL test have been applied. Instead of using an average biased by artifacts [6], we, use a median-based approach to exclude atypical EPs. For convenience, we assume that the number of single-trial MCEPs  $N^0$  after removing Type A artifacts is equal across all  $c$  ensembles. In the first iteration, the multi-channel median MCEP  $\bar{Z}_c^0$  of ensemble  $c$  is given by

$$\bar{Z}_c^0 = \text{Median}[Z_{c;1}, Z_{c;2}, \dots, Z_{c;N^0}] \quad (2)$$

where the  $j$ th sample of  $\bar{Z}_c^0$  is given by

$$\bar{z}_c^0(j) = \text{Median}[z_{c;1}(j), z_{c;2}(j), \dots, z_{c;N^0}(j)], \quad (3)$$

$$1 \leq j \leq (M)(K)$$

where  $z_{c;n}(j)$  is the  $j$ th sample of  $Z_{c;n}$ . The distance between median MCEP  $\bar{Z}_c^0$  and  $Z_{c;n}$  is given by

$$d_{c;n} = \| \bar{Z}_c^0 - Z_{c;n} \|, n=1,2,\dots,N^0 \quad (4)$$

where  $\| \cdot \|$  is the Euclidean norm. The MC-EPs  $Z_{c;n}$  can be re-ordered from 1 to  $N^0$  such that

$$d_{c;n} \leq d_{c;n+1}, \quad n=1,2,\dots,(N^0-1). \quad (5)$$

The median distance of the MCEPs to  $\bar{Z}_c^0$  is given by

$$d_c^0 = \text{Median}[d_{c;1}, d_{c;2}, \dots, d_{c;N^0}]. \quad (6)$$

The distance between the distance  $d_{c;a}$  of the reordered MCEP  $Z_{c;a}$  and the median distance  $d_c^0$  is given by

$$\omega_{c;a} = (d_c^0 - d_{c;a}) \quad (7)$$

In the re-ordered set in the current iteration, a MCEP  $Z_{c;n}$  is identified as an artifact if

$$(d_{c;n} - d_c^0) > \omega_{c;a}, \quad n > (N^0 + 1) / 2 \quad (8)$$

The artifact region determined by the MC-MED test in the first iteration is illustrated in Figure 1. By increasing the region trimming parameter  $a$ , the artifact-free region shrinks and the artifact region increases. The iterations are repeated until convergence, that is, no further artifacts are detected. To keep the notations simple, we have not included the iteration count parameter on  $Z_{c;n}$ ,  $d_{c;n}$ ,  $\omega_{c;1}$ , and  $\omega_{c;a}$ .

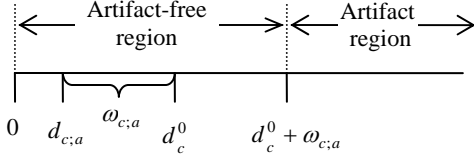


Fig. 1. Artifact region of the MC-MED test for Type B artifacts

### III. REAL-TIME ARTIFACT DETECTION

The parameters of the STD and CL tests are independent of the EP classes and can, therefore, be applied quite easily to detect Type A artifacts during testing. Detecting Type B artifacts, however, is not quite straight-forward because the parameters of the MC-MED test depend on the EP classes and the class of the test EP is, obviously, unknown. We propose a novel semi-class dependant “pre-testing” approach to detect Type B artifacts in real-time using our MC-MED test. Each incoming single-trial MCEP  $Z$  is pre-tested by comparing it with the median MCEPs  $\bar{Z}_c^L$ ,  $c=1,2,\dots,C$  to initially determine the closest class  $c^*$ ,  $c^* \in c=1,2,\dots,C$ , where,  $L$  is the last iteration number. If  $\rho_c$  is the distance of  $Z$  from  $\bar{Z}_c^L$ , then, the closest class  $c^*$  is determined from

$$c^* = \arg \min_c [\rho_c]. \quad (9)$$

We can now use the Type B artifact detection parameters for class  $c^*$  to determine if  $Z$  is an artifact. That is,  $Z$  is a Type B artifact if

$$(\rho_{c^*} - d_{c^*}^L) > \omega_{c^*}^L, \text{ where,} \quad (10)$$

$$\omega_a^L = \text{Min}[\omega_{1,a}^L, \omega_{2,a}^L, \dots, \omega_{C,a}^L] \quad (11)$$

Although the left-hand-side of Equation (11) depends on the class  $c$ , the artifact region threshold on the left-hand-side is not class dependent. By selecting the minimum, we select the smallest size of the  $c$  artifact free regions around each central median MCEP  $\bar{Z}_c$ . Selecting a common size for all

classes is perfectly valid given that the noise in the EPs is not class-dependent [5]. Therefore, the scatter in the classes can be assumed to be the same across all  $c$  classes.

### IV. ARTIFACT DETECTION RESULTS

A total of 3 real EP ensemble sets, borrowed from other studies, were used to evaluate the performance of the artifact detection strategy and to demonstrate the improvement in classification using artifact detection. The sets consisted of a 6-channel match-mismatch (MM) set [7], a 14-channel match mismatch-set [5], and an 11-channel mismatch negativity (MMN) set [8]. Trials in which the peak-to-peak

Table I  
Number of Type A and Type B artifacts detected in the 3 ensembles set using  $a=3$

Ensemble type	Ensemble classes	Number of EPs	Type A	Type B
6-channel	Match	280	0	6
	Match/Mismatch	Mismatch	280	0
14-channel	Match	71	0	2
	Match/Mismatch	Mismatch	71	0
11-channel MMN	Standard	350	27	12
	Deviant	350	27	18

amplitude exceeded  $100 \mu V$  in the MM experiments and  $60 \mu V$  in the MMN experiments, in any one electrode channel or in the eye channel were regarded as artifacts and rejected. The artifact detection strategy was applied using  $\lambda=5$  and  $a=3$  and the results are presented in Table I. Examples of Type A single channel artifacts and Type B multi-channel artifacts detected in the ensembles are shown in Figure 2 and 3, respectively.

### V. CLASSIFICATION RESULTS

In order to show the effects of artifacts on classifier parameter estimation and classifier performance, we selected our most recent multi-channel decision fusion (MCDF) classifier. Briefly, the MCDF classification strategy introduced in [4] dynamically exploits the information from multiple channels. The strategy is dynamic in the sense that different channels are selected at different time instants. The EP samples are ranked, across stimuli and across channels at different time instants, according to their nearest-mean classification accuracies which are pre-determined from a training set. During testing, the EP samples of the selected stimulus-channel-time combinations with the highest ranks

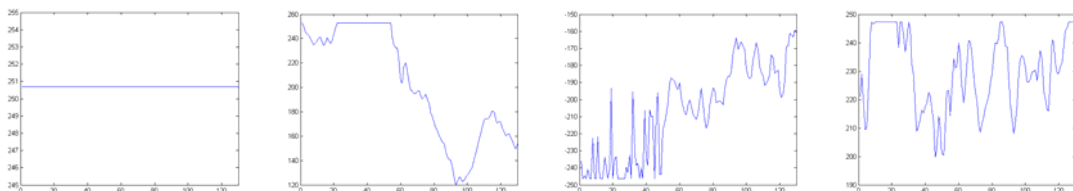


Fig. 2. Examples of Type A artifacts in the 11-channel MMN ensembles.

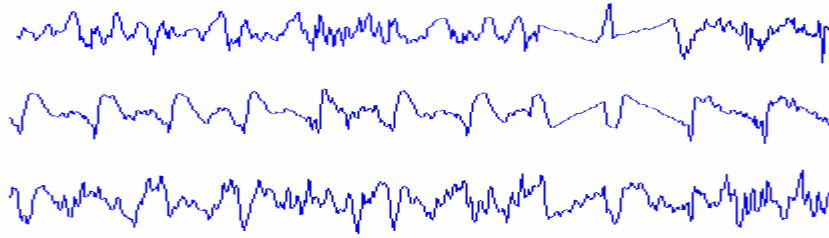


Fig. 3. Examples of Type B artifacts in (a) 6-channel, (b) 14-channel, (c) 11-channel ensembles

are classified, independently, and the decisions are fused into a single decision fusion vector. The resulting decision fusion vector is classified using a discrete Bayes classifier.

The first set of experiments (E1) used the ensembles directly without artifact rejection, the second set (E2) used trimmed artifact-free training ensembles but the test sets were not preprocessed to remove artifacts, the third set (E3) used artifact-free training and test ensembles, and the fourth set also used trimmed training and test ensembles, however, the EPs removed were selected randomly. The number of EPs removed in the fourth set (E4) was exactly the same as the number of artifacts removed in the third set of experiments. The fourth set of experiments were designed to show that the improvement in performance in the third set of experiments was not simply due to the reduction in the ensemble sizes. Through random partitioning of the ensembles into training and test sets [4], [5], each classification experiment was repeated 200 times. The results of EPs averaged over  $r=4$  single-trials [5],[7] and  $a=3$  are shown in Table II. Each result presented was averaged across 200 classification experiments. We were unable to finish the second set of experiments (E2) in time, however, the results for E2 along with extensive real-time test results be presented at the conference. The improvement in performance through artifact rejection is quite clear, especially for the MMN ensembles because of the relatively high artifact contaminations.

Table II  
Classification accuracies (%) with and without artifacts

Experiment	6-channel MM	14-channel MM	11-channel MMN
E1	89.729	91.611	55.221
E3	93.379	93.848	65.026
E4	90.131	91.259	55.334

## VI. CONCLUSIONS

The primary objective of the study described in this paper was to improve the performance of EP classifiers by identifying and rejecting artifacts. A new offline/real-time temporal domain artifact detection strategy with the following notable features was introduced: single-trial EPs are not removed from an artifact-free ensemble, artifacts can

be identified in individual and multiple channels, the ensembles may belong to multiple classes, artifacts can be identified offline and in real-time, the strategy is independent of any particular classification algorithm.

The classification results showed that the performance improved significantly when artifacts were removed from the training and test set ensembles. This result is an indication of the performance that can be expected if our artifact detection strategy is used to detect and reject artifacts in real-time classification trials. A novel pre-testing approach was proposed for real-time artifact detection.

In conclusion, it is clear that the performance of EP classifiers can be significantly improved by incorporating the artifact detection strategy developed in this paper. Also noteworthy is that the generalized formulations of the artifact detection strategy and the multi-channel decision fusion strategy make them both easily adaptable to various multi-sensor, multivariate, and multi-class signal classification problems.

## REFERENCES

- [1] R. J. Croft & R. J. Barry, "Removal of ocular artifact from EEG: a review," *Clinical Neurophysiology*, 30(1), 5-19, 2000.
- [2] T. D. Lagerlund, F. W. Sharbrough, N. E. Busacker, "Spatial filtering of multichannel electro-encephalographic recordings through principal component analysis by singular value decomposition," *Journal of Clinical Neurophysiology*, 14, 73-82, 1997.
- [3] T.-P. Jung, S. Makeig, C. Humphries, T.-W. Lee, M.J. Mckeown, V. Iragui, & T.J. Sejnowski, "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology*, No. 37, 163-178, 2000.
- [4] L. Gupta, H. Kook, D. L. Molfese, & K. C. Fadem, "Multi-stimuli multi-channel data and decision fusion strategies for dyslexia prediction using neonatal ERPs," *Pattern Recognition*, Vol. 28, No. 11, 2174-2184, 2005.
- [5] L. Gupta, B. Chung, M.D. Srinath, D. L. Molfese, & H. Kook "Multi-channel fusion models for the parametric classification of differential brain activity," *IEEE Transactions on Biomedical Engineering*, Vol. 52, No. 11, 1869-1881, 2005.
- [6] L. Gupta, D. L. Molfese, R. Tammana, and P. G. Simos, "Non-linear alignment and averaging for estimating the evoked potential," *IEEE Transactions on Biomedical Engineering*, vol. 43, No. 4, 348-356, 1996.
- [7] L. Gupta, J. Phegley, & D. L. Molfese, "Parametric classification of multichannel evoked potentials," *IEEE Transactions on Biomedical Engineering*, vol. 49, No. 8, 905-911, Aug. 2002 (vol. 49, No. 9, 1070, September 2002).
- [8] T. Huttunen, A. Halonen, J. Kaartinen, & H. Lyytinen, "Does mismatch negativity show differences in reading disabled children as compared to normal children and children with attention deficit? Department of Psychology, University of Jyväskylä, Finland (work in progress).