Supervised Segmentation of Microelectrode Recording Artifacts Using Power Spectral Density

Eduard Bakštein, Jakub Schneider, Tomáš Sieger, Daniel Novák, Jiří Wild and Robert Jech

Abstract— **Appropriate detection of clean signal segments in extracellular microelectrode recordings (MER) is vital for maintaining high signal-to-noise ratio in MER studies. Existing alternatives to manual signal inspection are based on unsupervised change-point detection. We present a method of supervised MER artifact classification, based on power spectral density (PSD) and evaluate its performance on a database of 95 labelled MER signals. The proposed method yielded test-set accuracy of 90%, which was close to the accuracy of annotation (94%). The unsupervised methods achieved accuracy of about 77% on both training and testing data.**

I. Introduction

Extracellular micro-electrode recording (MER, sometimes micro EEG or μEEG) has been in the neuroscientific toolbox for decades, being the principal method to study single and multi-unit activity. With growing amount of data being stored and processed in experiments, there is a still-growing need for automatic methods, allowing automatic processing of the extensive databases. One of the problems faced in the early stages of μEEG processing is dealing with noise and different artifacts, introduced during the recording process. Due to the small dimensions of the electrode tip used for micro recording (commonly around 1 or several μm), the system is susceptible to motion artifacts. Other sources of interference, affecting the signal between the recording site and the A/D converter of the recording system may include the mains (multiples of 50 or 60 Hz) and other electrical appliances in the recording room. Some artifacts caused by these sources may be of short duration (mechanical, interference of electrical devices in the vicinity, motors etc.), some may be stationary and affect the whole recording (e.g. mains noise) [1][2] – several examples of commonly seen artifacts can be found in Figure 1. Many of these undesirable effects can be removed either by precautions: switching off all non-vital electrical devices in the vicinity and leaving some safe time period before starting recording (both of which may be difficult or impossible in clinical conditions when recordings are done in the operating theatre, e.g. during deep brain stimulation surgery); or post-processing: e.g. notch-filtering the mains noise. Some artifacts may also be removed by common μEEG preprocessing, which includes high-pass filtering at a frequency of ca 200-500 Hz and low-pass filtering removing frequencies above ca 3-6 kHz [1]. Despite all precautions and preprocessing, it is common that the resulting signals contain artifacts that can affect further processing. Especially vulnerable is the spike detection, which is in most cases based on voltage threshold [3][4][5] and can therefore lead to confusing artifact peaks for spikes (false positives) or overlooked spikes (false negatives) when a higher than appropriate threshold is calculated from the contaminated signal [4]. In this perspective, looking for artifacts in the recorded signals prior to further processing seems like a reasonable step, preventing noise in the results. Apart from manual signal checking, which can be done in small studies, large databases, as well as automatic processing and classification systems call for an automatic artifact segmentation method.

This paper focuses on artifact identification in single-electrode recordings, where there is just one electrode or the electrodes are further apart (in the order of mm), so that interference between channels is very low. In this case, blind source separation suggested and discussed e.g. in [4] cannot be used. As single-electrode recording scenario is very common especially in human studies, we consider this contribution needed. There are several single-unit segmentation methods proposed in literature, which will be roughly described in the following. Artifacts related to simultaneous electrical stimulation are not considered here, as they have very specific character and their filtering is supported by common toolboxes [6] and sufficiently studied [2].

Existing detection methods

The default way of artifact detection used in most studies is manual, consisting of visual inspection of the recorded signals. Although this method is able to detect most of the most harmful artifacts that clearly alter amplitude of the signal, artifacts that are present throughout the whole signal may easily be overlooked. Also, the ability to detect extraneous events in the time course depends on the level of detail and other visualization parameters. This can be overcome by auditory inspection (listening to the signal may reveal interfering frequencies in the otherwise noisy character of the signal) or additional visualizations, such as the power spectral density, spectrogram and other. However, proper manual artifact annotation usually requires interacting with

*Research supported by grant of the Czech Technical University in Prague, no. SGS13/203/OHK3/3T/13

EB, JS, TS, DN and JW are with the Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, Czech Republic.

R.J. and TS are with the Department of Neurology and Center of Clinical Neuroscience, First Faculty of Medicine and General University Hospital, Charles University in Prague, 128 21, Prague, Czech Republic;

Correspondence should be addressed to E.B., e-mail eduard.bakstein@fel.cvut.cz or J.S., e-mail jakub.schneider@fel.cvut.cz.

978-1-4244-9270-1/15/$31.00 ©2015 IEEE

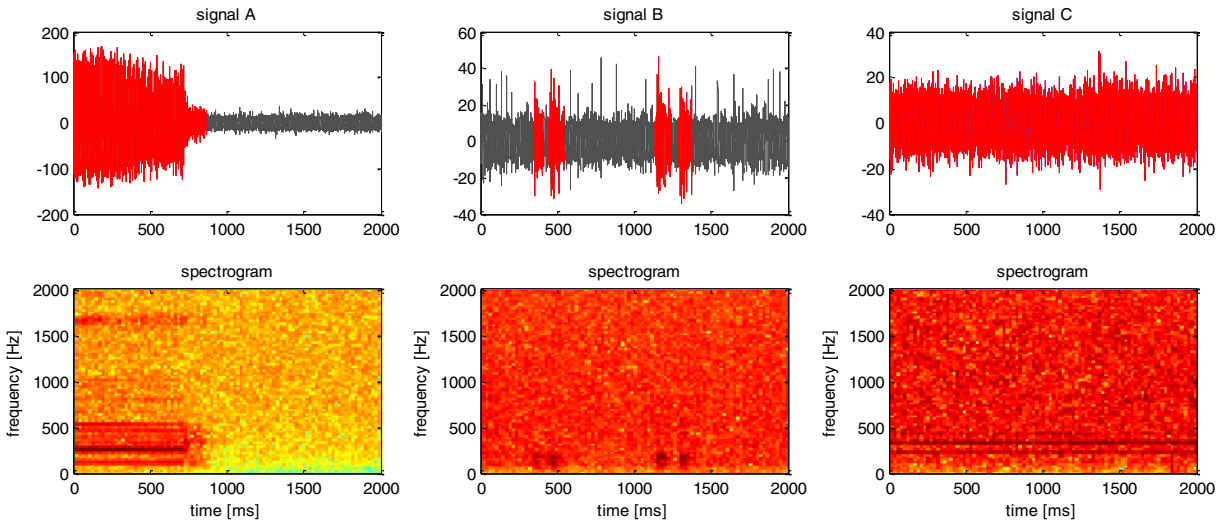

Figure 1. Example portions of MER signal with artifacts of different severity (top row, artifacts shown in red) together with corresponding spectrograms (bottom row). Columns A) and B) show transient artifacts affecting only a small part of the signal, column C) shows artifact of constant frequency, present throughout the signal. Sampling frequency is fs=24kHz, spectrograms created with FFT length equal to window length 1024, 50% window overlap. Frequency range is truncated for illustrative purposes in the spectrogram images.

the visualization at different levels of detail and is relatively time consuming. The rest of this paper is thus focused on automatic artifact identification methods

An example of the simplest automatic methods for identification of contaminated signals is approach implemented in [7]. First, recordings with very high maximum amplitude are rejected. Then the authors use ANOVA to statistically test consistency of signal energy distribution at the beginning and at the end of a signal. If the null hypothesis of equal distributions is rejected, the whole signal is excluded from further processing. This simple approach is easy to implement but is not capable of detecting stationary artifacts and artifacts in the middle of a signal.

A more sophisticated method presented by Falkenberg and Aboy in [8][9] divides the signal into a sequence of non-overlapping segments. Variance of the autocorrelation function is then estimated for each segment and variances of consecutive segments are divided to form a ratio, which is used as a distance measure (See Methods section and Eq. (3) and (4) for details). If the ratio of two neighboring segments exceeds a preset threshold, the algorithm marks given point as a transition. In the last step, the longest signal segment without any transitions is returned for further processing. Similar approach is adopted in [10], with the difference that variance is calculated from stationary wavelet transform of the signal, rather than from autocorrelation function.

The main problem of these unsupervised segmentation methods is that the algorithms use no prior information about clean signal nor artifact. Our inspections on a database of real MER signals indicate that the main type of problem faced in real-life data may be different than abrupt changes of amplitude in an otherwise clean signal. The unsupervised techniques for change point detection will work well on signals with short instationarities (such as short-time movement artifacts), but will fail from definition on stationary artifacts, present in the signal most of the time.

In this paper, we present a simple supervised method based on power spectral density, and compare it to the methods mentioned above.

II. METHODS

A. *MER artifact detection using normalized power spectral density*

Our proposed algorithm for supervised detection of MER artifacts can be summarized in the following steps:

a. *Mean PSD calculation*

The input to the algorithm is formed by a set of MER signals $X_1, X_2 \dots X_n$ with manually labelled artifacts in corresponding annotation vectors $A_1, A_2 \dots A_n$, where $A_i = 1$ if X_i contains clean signal and 0 otherwise. The mean normalized spectrum $mPSD$ is then calculated according to

$$mPSD = \frac{1}{\sum_{i=0}^{n} A_i} \cdot \sum_{j=1}^{n} A_j \frac{psd(X_j)}{\sum psd(X_j)} \quad (1)$$

Where $psd(X_j)$ is power spectral density (or periodogram) of segment X_j, estimated using Welch method and $\sum psd(X_j)$ sum of the periodogram for given segment. The normalization ensures that shape of the power spectrum is extracted independently of signal power

b. *Comparison to a signal*

To calculate the distance between the $mPSD$ and a signal Z, a spectrogram $\bar{P}$ is calculated for the classified signal. In case of the presented algorithm, this is equivalent to dividing the signal into segments with 50% overlap, each of length equal to the number of points used in the FFT algorithm (2048 in case of the reported results). The distance is calculated according to:

$$d_k = max \left| \frac{\bar{P}_k}{\sum \bar{P}_k} - mPSD \right| \quad (2)$$

Where $\bar{P}_k$ is k-th slice (spectrum) of the spectrogram with $\sum \bar{P}_k$ being its sum (i.e. total power) and max | | denotes

maximum element in absolute value of the resulting vector. The length of the spectrum is the same as in the calculation of mean PSD. When the segments to be compared are longer than the number of FFT points in the spectrogram, distance d_k is calculated for all slices corresponding to the desired segment time and maximum distance is taken as result for the whole segment.

c. Learning and Classification process

In the learning phase, $mPSD$ is first estimated from clean segments of labelled signals according to Eq. (1). Next, the $mPSD$ is compared to all signals in the training set. The classification threshold value is set to maximize classification accuracy on training data. In the prediction phase, the distance from mean spectrum is calculated for given signal and compared to the threshold.

B. Extension of existing methods

We compare our method to algorithms for stationary segment identification based on autocorrelation function from [8][9] and on stationary wavelet transform from [10]. These methods first divide the signal X into m non-overlapping segments $X_1, X_2 \dots X_n$ and transform the segments using autocorrelation function and stationary wavelet transform, respectively. In the next step, variance of each transformed segment is calculated according to

$$v_i = var\{\gamma(X_i)\}, \; i \in \langle 1, m \rangle, \tag{3}$$

where γ represents the transformation function – either autocorrelation or Stationary Wavelet Transform. Variances of neighboring segments are then compared according to:

$$d_{ij} = \frac{\max(v_i, v_j)}{\min(v_i, v_j)}, \quad i \in \langle 1, m-1 \rangle, j = i+1 \tag{4}$$

Divisions between segments with distance statistic d_{ij} exceeding a manually pre-chosen threshold θ then determine breakpoints between stationary segments. The longest stationary segment is found and returned. We further extend this method by computing distance between all possible segment pairs, forming a distance matrix

$$D = \begin{pmatrix} 0 & d_{12} & \dots & d_{1,m-1} \\ d_{21} & 0 & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ d_{m-1,1} & d_{m-1,1} & \dots & 0 \end{pmatrix} \tag{5}$$

with zero diagonal. Values d_{ij} exceeding the classification threshold θ are replaced with ones, other with zero. The resulting matrix is then scanned for the longest uninterrupted segment (sequence of zeros) using greedy algorithm.

With these modifications, the algorithm may return even a non-contiguous subset of the signal, which is much more suitable for comparison with manual annotation. This approach can also be used in cases where no contiguity is requested and amount of unnecessarily removed data can be minimized – such as in background activity feature calculation etc.

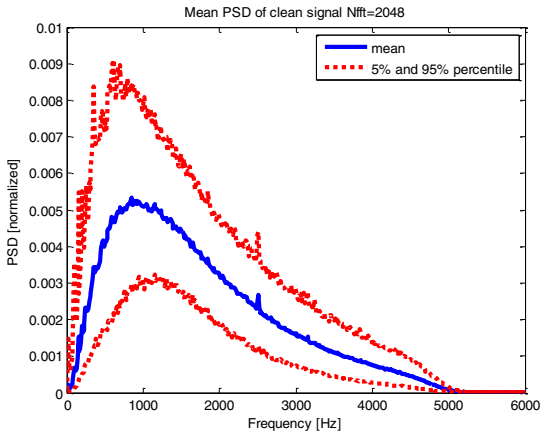

Figure 2. Mean PSD spectrum from training data with 5% and 95% percentile bounds. Note similarity in general shape of the spectrum in all curves, induced by spectral formation of the mEEG signal and filtering (see [11] for details).

C. Data

To test the algorithm, we used a database consisting of 95 signals of length 10s from 18 different Parkinson's disease patients (mean 5.3 signals per patient), obtained during microrecording and targeting process in deep brain stimulation surgery. The length of all signals was 10s, sampling frequency was 24 kHz, signals were band-pass filtered in the range 500-5000 Hz. Annotation was done for each second of the signals according to judgement based on visual inspection of time course, spectrogram and listening to the signal using headphones by a team of 5 trained evaluators. All 5 parallel annotations were then joined using majority voting to form the resulting signal annotation. The overall agreement of individual team members with the majority voting annotation was 93.5%. The data set yielded 255 seconds of clean signal and 695 seconds of contaminated signal, according to majority voting results. In order to estimate each methods' performance on unseen data, the data set was divided into training set of 60 signals and test set of 35 signals.

The experimental procedures involving human subjects described in this paper were approved by the Institutional Review Board.

III. Results

The mean power spectrum of clean data, necessary for the proposed algorithm was first computed on the test set, according to Eq. (1) – see Figure 2 for the result. Next, ROC curve was computed for each method by shifting the detection threshold, see Figure 3. Optimal threshold was chosen for each method according to F1 score (geometric mean of precision and recall as suggested method by [10]) and the threshold was used to perform classification on training and testing set – see TABLE I. for results. Due to the unexpectedly high performance drop of the unsupervised methods on the test set and after evaluating the ROC curves, we performed additional classification using threshold selected by maximum accuracy on the training set. The improved results are also presented in the table.

TABLE I. CLASSIFICATION RESULTS

Method	TRAIN SET accuracy		TEST SET accuracy	
	Max acc. threshold[a]	*Max(F1) threshold)*[b]	*Max acc. threshold*[a]	*Max(F1) threshold*[b]
COV	76.7%	72.5%	76.3%	64.0%
SWT	77.3%	71.7%	70.6%	64.4%
PSD	91.7%	91.7%	89.7%	89.7%

Classification threshold set for a) maximum train-set accuracy b) maximum train-set F1 score

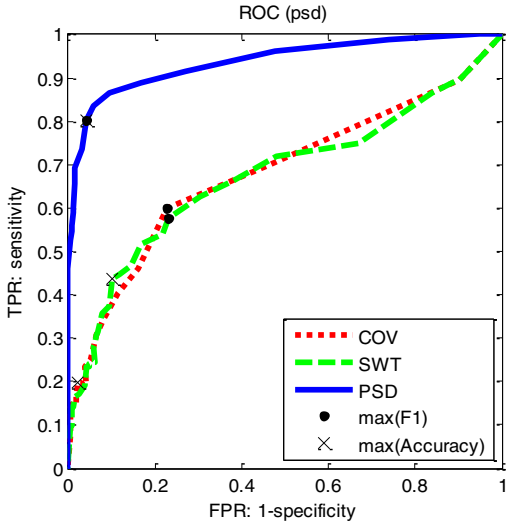

Figure 3. Receiver-operating characteristic for setting threshold on the training data. Points corresponding to thresholds set based on maximum F1 score and maximum accuracy are shown for each method: Autocorrelation (COV), Stationary Wavelet transform (SWT) and proposed Power spectral density (PSD) method.

IV. DISCUSSION

When comparing results of the proposed supervised method and the unsupervised segmentation algorithms, the supervised method achieved much higher accuracy of 91.7% / 89.7% accuracy on the training/testing set, which is close to the accuracy of the annotation itself (93.5%). The unsupervised methods achieved around 77% on both training and testing set.

The unsupervised stationary segmentation methods can identify the break points well, but possess no mechanism to mark the resulting segments as either clean signal or artifact. Therefore, such methods fail on signals with long-term or very stationary artifacts (such as mains noise etc.). The results indicate that a supervised approach may be a worthy step in real μEEG preprocessing, even despite the necessity to manually select clean signals for mean spectrum calculation. The strain of manual classification step is further eased by the fact that some sort of visual inspection will always be present in a proper data-evaluation pipeline and is also necessary for threshold selection in the unsupervised methods as well.

It has to be noted that autocorrelation function, used in [8][9] and PSD of a signal is clearly related through a single step of Fourier transformation and will therefore provide similar results. However, we chose the power spectral density due to its more intuitive interpretation – by visualization of difference from the mean PSD, dominant artifact frequencies may be identified and may possibly indicate the cause of the artifact.

V. CONCLUSION

The presented supervised method for artifact detection based on power spectral density showed highly improved results over unsupervised change point detection models on real MER data. The price paid is the necessity to manually mark artifacts in the first stage, which can possibly be eased by an appropriate graphical tool in the future. A full classification system, based on more signal features can then be a good solution to detection of MER artifacts.

ACKNOWLEDGMENT

We acknowledge the helpful comments by Jiri Anyz from the CTU and help of Pavel Vostatek from CTU during data annotation. The authors also thank Cristian Guarnizo, author of [10] for kindly providing source codes for his SWT method. The research presented was supported by grant of the Czech Technical University in Prague, no. SGS13/203/OHK3/3T/13 and the Czech Science Foundation: grant project 309/09/1145

REFERENCES

[1] Israel, Z, Burchiel K, "Microelectrode Recording in Movement Disorder Surgery", Thieme Medical Publishers, NY, 2004

[2] M. E. J. Obien, K. Deligkaris, T. Bullmann, D. J. Bakkum, and U. Frey, "Revealing neuronal function through microelectrode array recordings," Front. Neurosci., vol. 8, no. January, pp. 1–30, 2015.

[3] J. Wild, Z. Prekopcsak, T. Sieger, D. Novak, and R. Jech, "Performance comparison of extracellular spike sorting algorithms for single-channel recordings," J. Neurosci. Methods, vol. 203, no. 2, pp. 369–376, Jan. 2012.

[4] I. Gligorijevic, M. Welkenhuysen, D. Prodanov, S. Musa, B. Nuttin, W. Eberle, C. Bartic, and S. Van Huffel, "Neural signal analysis and artifact removal in single and multichannel in-vivo deep brain recordings," *Neurosurgery*, pp. 8–11, 2009.

[5] R. Q. Quiroga, Z. Nadasdy, and Y. Ben-Shaul, "Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering.," Neural Comput., vol. 16, pp. 1661–1687, 2004.

[6] U. Egert, T. Knott, C. Schwarz, M. Nawrot, a. Brandt, S. Rotter, and M. Diesmann, "MEA-Tools: An open source toolbox for the analysis of multi-electrode data with MATLAB," J. Neurosci. Methods, vol. 117, no. 1, pp. 33–42, 2002.

[7] A. Moran, I. Bar-Gad, H. Bergman, and Z. Israel, "Real-time refinement of subthalamic nucleus targeting using Bayesian decision-making on the root mean square measure.," Mov Disord, vol. 21, no. 9, pp. 1425–1431, Sep. 2006.

[8] J. H. Falkenberg and J. McNames, "Segmentation of extracellular microelectrode recordings with equal power," in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2003, vol. 3, pp. 2475–2478.

[9] M. Aboy and J. H. Falkenberg, "An automatic algorithm for stationary segmentation of extracellular microelectrode recordings," Med. Biol. Eng. Comput., vol. 44, no. 6, pp. 511–515, Jun. 2006.

[10] C. Guarnizo, A. a. Orozco, and G. Castcllanos, "Microelectrode signals segmentation using stationary wavelet transform," Biomed. Eng. Informatics New Dev. Futur. - Proc. 1st Int. Conf. Biomed. Eng. Informatics, BMEI 2008, vol. 2, pp. 450–454, May 2008.

[11] J. Martinez, C. Pedreira, M. J. Ison, and R. Quian Quiroga, "Realistic simulation of extracellular recordings," J. Neurosci. Methods, vol. 184, pp. 285–293, 2009.