Segmentation algorithm of roadheader vibration signal based on the stable distribution parameters

Jacek Wodecki[1], Pawel Stefaniak[2], Agnieszka Wylomanska[3], Radoslaw Zimroz[4], Grzegorz Zak[5]

[1, 2, 4]KGHM CUPRUM R&D Ltd., Sikorskiego 2-8, 53-659 Wroclaw, Poland

[1, 2, 4]Diagnostics and Vibro-Acoustic Science Laboratory, Wroclaw University of Technology, Wroclaw, Poland

[3]Hugo Steinhaus Center, Department of Mathematics, Wroclaw University of Technology, Wroclaw, Poland

[5]Diagnostics and Vibro-Acoustic Science Laboratory, Wroclaw University of Technology, Wroclaw, Poland

[2]Corresponding author

E-mail: [1]*jwodecki@cuprum.wroc.pl*, [2]*pkstefaniak@cuprum.wroc.pl*, [3]*agnieszka.wylomanska@pwr.edu.pl*, [4]*rzimroz@cuprum.wroc.pl*, [5]*grzegorz.zak@pwr.edu.pl*

Abstract. In the real signal analysis the main problem is the non-stationarity of given data. The non-stationarity can be manifested in different ways. One of the possibility is the assumption that the signal is a mixture of different processes that exhibit different statistical properties. Thus before the further analysis the observed data should be segmented. In this paper we propose an automatic segmentation method which is based on α-stable distribution approach. In the proposed procedure we estimate the parameters of stable distribution for consecutive sub-signals of given length and then by using expectation-maximization algorithm we classify the parameters. The obtained classes correspond to different segments of the signal. The proposed procedure we apply to the real vibration signal from roadheader working in mining industry. As a final result we obtained segments of real signal which constitute samples of different behaviors and are related to different modes of operation of the machine.

Keywords: segmentation, stable distribution, expectation-maximization algorithm, roadheader.

1. Introduction

Analysis and modeling of real signals measured by advanced data acquisition systems provide possibility to obtain information about the considered process. Very often observed time series exhibits strong non-stationary behavior resulting from the fact that it consists of not one but multiple processes that coexist or occur one after the other. In most of the cases those processes cannot be analyzed by the same tools because they exhibit completely different character and have different statistical properties. Sometimes the processes can be described by the same or similar models but their parameters differ. Therefore before the modeling of given signal its preliminary segmentation should be performed.

One of the most fundamental reason for segmentation is extraction of parts (segments) of the signal with homogenous properties. Segmentation is also related to finding such time points where the signal changes the properties and switches to the other regime [1-4].

There are many segmentation methods and during recent years number of papers in this field has increased. Segmentation techniques are very often dedicated to applications for which they were used. Some of the methods are based on special behavior of the signal in time domain [1, 3, 5] however there are also methods which are based not on the raw signal but on its transformations to other domains [6].

Signal segmentation has been applied in many areas. It is especially crucial in condition monitoring (to isolate shocks related to damage) [6], machine performance analysis (to find when machine operates under overloading, idle mode etc.) [4], experimental physics [3], biomedical signals (like ECG signals) [7], speech analysis [8] and in seismic signal analysis [9]. In this paper we are especially interested in the first two mentioned applications.

In this paper we propose the segmentation method which is based on the modeling of given signal by α-stable distribution and on the expectation-maximization algorithm. The stable distributions are especially important in the context of modeling of data with visible peaks but it should be mentioned that for α parameter equal to 2 the stable distribution reduces to Gaussian one. There are many applications of such distributions like finance, environmental engineering or condition monitoring [10]. The expectation-maximization algorithm was invented for finding maximum likelihood or maximum a posteriori estimates of parameters in statistical models, where the model depends on unobserved latent variables. However this algorithm very often is applied in the classification problem. In the proposed technique for the application to raw two-channel signal first we estimate the appropriate parameters of stable distribution and then by using expectation-maximization algorithm and Silhouette criterion we classify the data. We apply this procedure to the real vibration signal from roadheader working in the mining industry. Investigated machine is one of the most crucial components of the technological process in underground mining industry. It eliminates blasting operations in exploitation area and ensures continuous mechanical drilling. There exist many factors having influence on vibration signal profile (various physical and mechanical properties of deposit, design features of the machine, depth of drilling, motion parameters and wear level of machine elements). Appropriate segmentation of vibration signal allows for identification of spectrum domain components corresponding to different processes.

2. Methodology

The segmentation procedure of roadheader vibration signal is based on the alpha-stable distribution approach. As it was mentioned, the α-stable distribution is an extension of Gaussian one, namely for some values of parameters ($\alpha = 2$) it reduces to normal distribution. A random variable X is an α-stable distributed if its characteristic function takes the following form:

$$E e^{itX} = \begin{cases} \exp\left(-\Sigma^{\alpha}|t|^{\alpha}\left(1 - \beta \text{sgn}(t)\tan\left(\frac{\pi\alpha}{2}\right)\right) + i\mu t\right), & \alpha \neq 1, \\ \exp\left(-\Sigma|t|\left(1 + i\beta \text{sgn}(t)\frac{2}{\pi} + i\mu t\right), & \alpha = 1, \end{cases} \tag{1}$$

where α ($0 < \alpha \leq 2$) is stability parameter, β ($-1 \leq \beta \leq 1$) is asymmetry parameter, Σ ($\Sigma > 0$) is scale parameter and μ ($\mu \in R$) is location parameter. Of the four parameters defining the family of stable distribution, most attention has been focused on the stability parameter α. The α-stable distributed random variable has heavy tails, i.e. its cumulative probability density function decays with power law. Therefore, there is a high probability of the variable having extreme values which is useful in modelling of impulsive signal [10]. The other distribution that can be useful in impulsive signal modelling is presented in [11].

In our segmentation procedure first the raw signal is divided into consecutive sub-signals of length N samples. The sub-signals overlap by K %. Next, for each sub-signal the empirical standard deviation (σ) is calculated. Because the sub-signals exhibit behavior related to heavy tailed data (visible peaks), then we propose to model the data by stable distribution and next parameters taken to the classification are α and Σ. In the literature one can find different estimation methods that can be used here [12, 13]. In this paper we apply the regression method which is based on the characteristic function of the considered distribution (see Eq. (1)).

As a final step, the mentioned parameters (σ, α, Σ) are used in the expectation-maximization algorithm to the classification. We normalize the data before the further analysis. We should mention that in the analysis we consider two-channel signal and for sub-signals for each of them we estimate the parameters therefore in the expectation-maximization algorithm we take under consideration six parameters.

In statistics, an expectation-maximization (EM) algorithm is an iterative method for finding

maximum likelihood or maximum a posteriori estimates of parameters in statistical models, where the model depends on unobserved latent variables. The EM iteration alternates between performing an expectation (E) step which creates a function for the expectation of the log-likelihood evaluated using the current estimate for the parameters, and a maximization (M) step which computes parameters maximizing the expected log-likelihood found on the E step. These parameter-estimates are then used to determine the distribution of the latent variables in the next E step [14-17]. EM is frequently used for data clustering in machine learning and computer vision. In natural language processing, two prominent instances of the algorithm are the Baum-Welch algorithm and the inside-outside algorithm for unsupervised induction of probabilistic context-free grammars. In our procedure we also propose to estimate number of clusters with Silhouette criterion [18, 19] for limited range of number of clusters k (in our application $k = 1:5$) with the measure of distance set to Euclidean.

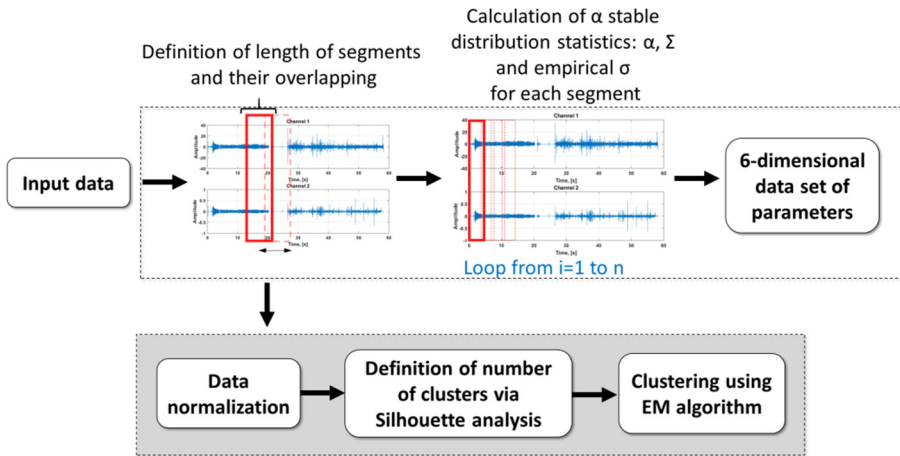In Fig. 1 we present the scheme of our segmentation procedure.

Fig. 1. Scheme of the segmentation procedure

3. Real signal analysis

The analyzed real data set represents vibration signal of a roadheader. The sampling frequency $Fs = 25$ kHz. Measurement was performed by using two accelerometers placed orthogonally on the mining head's arm. Duration of measurement was equal to 60 seconds. The analyzed two-channel data set is presented in Fig. 2.
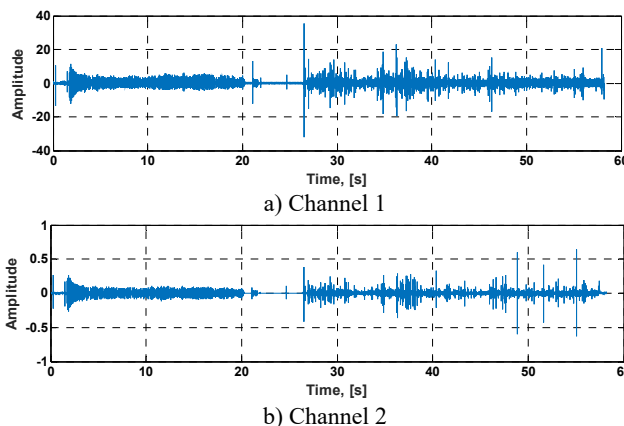

a) Channel 1

b) Channel 2
Fig. 2. The two-channels raw roadheader vibration signal

According to the presented segmentation procedure we divide the two-channel raw signal into sub-signals consisting of 10000 observations with overlap 50 %. Next, for each sub-signal the standard deviation and two parameters related to α-stable distributions are calculated. In Fig. 3 we present the estimated parameters for sub-signals corresponding to both analyzed channels.
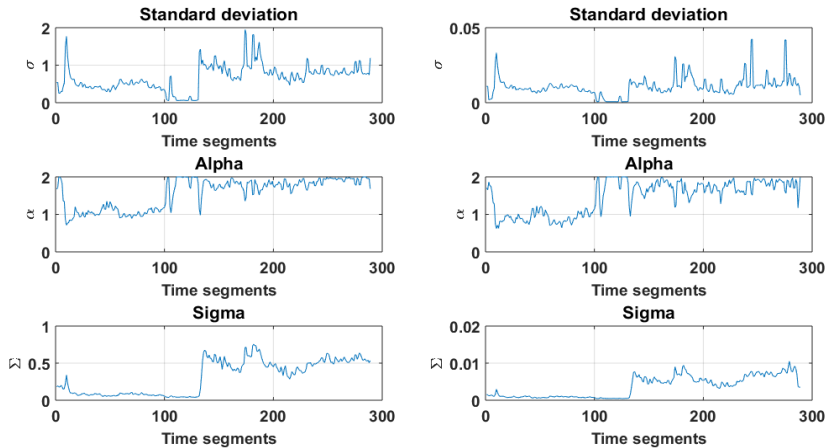

Fig. 3. The estimated parameters σ, α, Σ for sub-signals (segments) related to two-channels vibration signal

These six statistics represent set of N points χ in six-dimensional feature space, within which we perform clustering. Dataset χ is then centered and variance over each dimension is set to 1. This operation allows to achieve more reliable and efficient cluster amount analysis.

After obtaining the set of parameters we estimate number of clusters with Silhouette criterion for limited range of number of clusters $k = 5$. When optimal number of clusters k is estimated, EM algorithm performs clustering of the dataset χ splitting it into k classes. For the analyzed signal we obtain that the optimal number of clusters is equal to four. In Fig. 4 we present the division of the raw signal from the first channel and second channel into four clusters.
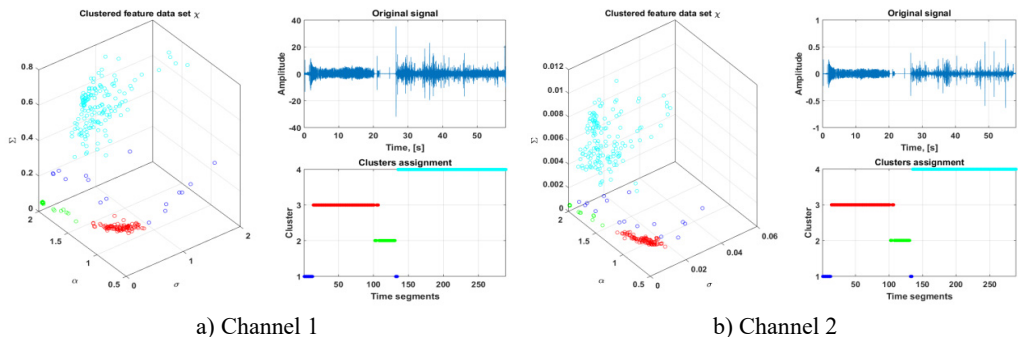

a) Channel 1   b) Channel 2

Fig. 4. Result of data clustering. Data points in the left part are displayed as features of signal corresponding to appropriate channel in 3-dimensional space for presentational purpose, however they were analyzed and clustered in 6 dimensions. Data is also presented in the form before centering and variance normalization for clear visual interpretation; top – analyzed raw signal; bottom – association of data points to clusters in the function of time for comparison with original signal above it

As a result of the presented procedure we obtain segments belonging to certain clusters. We can apply the technique once again to data from one clustering in order to identify processes inside

it, i.e.to make the so-called "next-layer analysis". As an example, we consider part of the third cluster presented in red color in Fig. 4. For this cluster we observe that there are regions of denser and more sparse impulses that suggests the segment is a mixture of different processes with different statistical properties, see Fig. 5. By using presented methodology we separate the data into classes.

For the signal related to the third cluster we apply the introduced technique. In this case the Silhouette evaluation estimated two clusters and results of EM clustering are presented in Fig. 6.
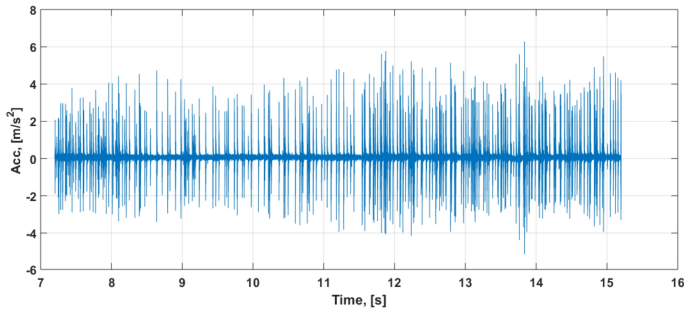

Fig. 5. Third cluster (marked in red in Fig. 4) from channel 1

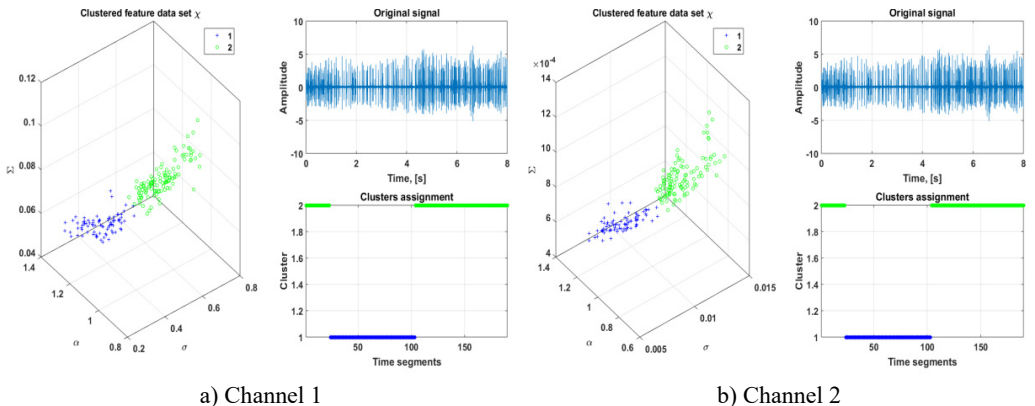a) Channel 1   b) Channel 2
Fig. 6. Second-layer analysis of the third cluster. Distinction of two separate groups is clearly visible

In practice we can proceed with as many layers as we wish, since number of layers can be provided as a parameter to the implemented algorithm. Although it is a good practice to limit search space of Silhouette estimation to about 2-3 clusters for deeper layers, and proceed with no more than 2-3 layers of analysis. If we go past those recommendations, results will make no sense since algorithm will greedily try to find clusters even as single points.

4. Summary

In this paper we have proposed a new segmentation technique which can be applied to the vibration signal from roadheader working in the mining industry. Very harsh mining conditions, presence of many sources of interferences and many different operation modes have impact on the difficulty of signal analysis, for example in the diagnostic point of view. Therefore it is necessary to develop a novel technique for signal segmentation in order to identify different operational modes. The proposed methodology is automatic and based on the stable distribution approach, Silhouette criterion and expectation-maximization algorithm applied in last step of the procedure, i.e. to the classification of the appropriate parameters. As a result we obtain the time periods for which the analyzed signal has homogenous structure. The obtained segments

correspond to different modes of operation of the machine. Furthermore, detection and parameterization of such events might help to improve efficiency of machine usage, for example to minimize number of segments and their duration for machine operation under idle mode or overload mode.

Acknowledgements

References

[1] **Lopatka M., Laplanche C., Adam O., Motsch J.-F., Zarzycki J.** Non-stationary time-series segmentation based on the Schur prediction error analysis. 13th Workshop on Statistical Signal Processing, 2005, p. 251-256.

[2] **Wyłomańska A., Zimroz R., Janczura J.** Identification and stochastic modelling of sources in copper ore crusher vibrations. Journal of Physics: Conference Series, Vol. 628, 2015, p. 012125.

[3] **Gajda J., Sikora G., Wyłomańska A.** Regime variance testing – a quantile approach. Acta Physica Polonica B, Vol. 44, Issue 5, 2013, p. 1015-1035.

[4] **Wyłomańska A., Zimroz R.** Signal segmentation for operational regimes detection of heavy duty mining mobile machines – a statistical approach. Diagnostyka, Vol. 15, Issue 2, 2014, p. 33-42.

[5] **Makowski R., Zimroz R.** A procedure for weighted summation of the derivatives of reflection coefficients in adaptive Schur filter with application to fault detection in rolling element bearings. Mechanical Systems and Signal Processing, Vol. 38, 2013, p. 65-77.

[6] **Obuchowski J., Wyłomańska A., Zimroz R.** The local maxima method for enhancement of time-frequency map and its application to local damage detection in rotating machines. Mechanical Systems and Signal Processing, Vol. 46, 2014, p. 389-405.

[7] **Azami H., Mohammadi K., Bozorgtabar B.** An improved signal segmentation using moving average and Savitzky-Golay filter. Journal of Signal and Information Processing, Vol. 3, 2012, p. 39-44.

[8] **Makowski R., Hossa R.** Automatic speech signal segmentation based on the innovation adaptive filter. International Journal on Applied Mathematics and Computer Science, Vol. 24, 2014, p. 259-270.

[9] **Popescu T. D.** Signal segmentation using changing regression models with application in seismic engineering. Digital Signal Processing, Vol. 24, 2014, p. 14-26.

[10] **Żak G., Obuchowski J., Wyłomańska A., Zimroz R.** Application of ARMA modeling and alpha-stable distribution for local damage detection in bearings. Diagnostyka, Vol. 15, Issue 3, 2014, p. 3-11.

[11] **Stefaniak P., Wyłomańska A., Obuchowski J., Zimroz R.** Procedures for Decision Thresholds Finding in Maintenance Management of Belt Conveyor System Statistical Modeling of Diagnostic Data. Lecture Notes in Production Engineering, Springer, 2015, p. 391-402.

[12] **Allen J. B.** Short term spectral analysis, synthesis, and modification by discrete Fourier transform. IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 25, Issue 3, 1977, p. 235-238.

[13] **Samorodnitsky G., Taqqu M. S.** Stable Non-Gaussian Random Processes. Chapman and Hall, New York, 1994.

[14] **Dempster A. P., Laird N. M., Rubin D. B.** Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B, Vol. 39, Issue 1, 1977, p. 1-38.

[15] **Sundberg R.** Maximum likelihood theory for incomplete data from an exponential family. Scandinavian Journal of Statistics, Vol. 1, Issue 2, 1974, p. 49-58.

[16] **Neal R., Hinton G.** A View of the EM Algorithm That Justifies Incremental, Sparse, and Other Variants. Learning in Graphical Models, MIT Press, Cambridge, MA, 1999, p. 355-368.

[17] **Hastie T., Tibshirani R., Friedman J.** 8.5 the EM algorithm. The Elements of Statistical Learning. Springer, New York, 2001, p. 236-243.

[18] **Kaufman L., Rousseeuw P. J.** Finding Groups in Data: An Introduction to Cluster Analysis. John Wiley & Sons, Hoboken, NJ, 1990.

[19] **Rousseeuw P. J.** Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, Vol. 20, Issue 1, 1987, p. 53-65.