



## SIMPLE METHOD OF FAILURE DETECTION OF ROTARY MACHINES

Andrzej BIELECKI<sup>1</sup>, Marzena BIELECKA<sup>2</sup>, Adam JABŁOŃSKI<sup>3</sup>, Wojciech STASZEWSKI<sup>3</sup>

<sup>1</sup> Institute of Computer Science, Faculty of Exact and Natural Sciences, Pedagogical University in Kraków, Podchorążych 2, 30-084 Kraków, Poland; [azbielecki@gmail.com](mailto:azbielecki@gmail.com)

<sup>2</sup> Chair of Geoinformatics and Applied Computer Science, Faculty of Geology, Geophysics and Environmental Protection, AGH University of Science and Technology, Mickiewicza 30, 30-059 Kraków, Poland, [bielecka@agh.edu.pl](mailto:bielecka@agh.edu.pl)

<sup>3</sup> Chair of Robotics and Mechatronics, Faculty of Mechanical Engineering and Robotics, AGH University of Science and Technology, Mickiewicza 30, 30-059 Kraków, Poland, [ajab@agh.edu.pl](mailto:ajab@agh.edu.pl), [wojciech.staszewski@agh.edu.pl](mailto:wojciech.staszewski@agh.edu.pl)

### Abstract

In the paper a simple unsupervised monitoring method of rotary machines is proposed. The method consists of three stages – multi-reference preliminary analysis of the vibration signals, auto-reference preliminary analysis and probabilistic analysis of the signals. The method was tested by using signals from eight simulated machines. The efficiency of the method has been positively verified.

Keywords: rotary machine; intelligent monitoring; early failure detection

### PROSTA METODA WYKRYWANIA USTEREK W MASZYNACH WIRNIKOWYCH

#### Streszczenie

W niniejszej publikacji prosta proponujemy prostą metodę nienadzorowanego monitoringu maszyn wirnikowych. Metoda jest trzystopniowa – wieloreferencyjna wstępna analiza sygnałów drganiowych, autoreferencyjna wstępna analiza oraz probabilistyczna analiza sygnałów. Metoda została przetestowana na ośmiu symulowanych maszynach. Skuteczność metody została pozytywnie zweryfikowana.

Słowa kluczowe: maszyny wirnikowe; inteligentny monitoring; wczesne wykrywanie usterek

## 1. INTRODUCTION AND PROBLEM FORMULATION

Rotary machines monitoring is a crucial part of their maintenance [6,8,10,14,15]. The automatic monitoring is based on various complex systems such as SCADA [5] or systems of artificial intelligence [2,3,4,7]. Systems of this type, apart from the obvious advantages, have some drawbacks, however. First of all, the complexity of the monitoring system can cause problems with ensuring real-time work. Furthermore, most of artificial intelligence systems requires a tedious learning process. Moreover, after restarting the machine after fault, the learning process usually has to be repeated because of change of characteristic of the working parameters that, even if are tiny, can have influence at monitoring process. The fact that the mentioned monitoring is connected with a problem of on-line processing and analysis of the big data stream is an additional difficulty – see [1,9,10] and the review paper [11]. Therefore, there is a need for extremely simple but effective rotary machinery monitoring systems that can be used as complementary to systems based on more complex algorithms. In this paper such system is proposed.

The problem is formulated in the following way. There are a group of identical rotary machines that start working at the same time. All of them are monitored with the same frequency. The monitoring is conducted on the basis of vibrodiagnostics. A certain initial working time of the machines is trouble-free operation. The detection of the moment of a pre-failure, a failure or a fault of each machine as early as possible is the aim of the monitoring. Furthermore, the detecting algorithm should be as simple as possible. The approach was tested by using a simulated model – see Section 3.

## 2. DESCRIPTION OF THE METHOD AND ITS APPLICATION

The proposed method consists of three stages – multi-reference preliminary analysis, auto reference preliminary analysis and probabilistic analysis of the signals. Both preliminary analyzes include, among others, method calibration. The multi-reference analysis can be applied if at least two identical machines are monitored; on the other case, this stage has to be omitted.

The multi-reference preliminary analysis consists in comparison of the signals from two machines. We compare the average signals from each waveform according to the formula

$$y_{jm}^{(k)} = \sqrt{\frac{1}{N} \sum_{i=1}^N (z_{ij}^{(k)} - z_{im}^{(k)})^2} \quad (1)$$

where the index  $i$  numbers the measurement in a given run,  $k$  numbers the run, whereas  $j$  and  $m$  numbers machines. The superscript bracket denotes that the number is an index, not a power. The number  $N$  denotes the total number of the measurements in the run. Thus,  $z_{ij}^{(k)}$  denotes the value of the  $i$ th signal in the  $k$ th run in the  $j$ th machine. The number  $y_{jm}^{(k)}$  denotes the  $k$ th run mean-square error of the  $j$ th and  $m$ th machine. The multi-reference preliminary analysis allows to determine the threshold value above which the signals should be considered as different. In addition, the method enables the extension of the reference range and allows for the initial indication of the first defect in some machines.

In the auto-reference preliminary analysis the sequential signals are compared with the first  $s$  runs of the same machine that are treated as the reference ones. The obtained value, calculated from the formula (2) below, is the average difference between the given  $k$  run and the reference runs, minimized after the reference runs:

$$y^{(k)} = \min_{n \in \{1, \dots, s\}} \frac{1}{N} \sqrt{\sum_{i=1}^N (z_i^{(n)} - z_i^{(k)})^2} \quad (2)$$

where  $n$  ( $1 \leq n \leq s$ ) is the number of reference run,  $k$  is the number of the measured run ( $1 \leq k \leq 250$ ),  $i$  is the number of the signal in the run ( $1 \leq i \leq 250000$ ). In order to calibrate the method, the pairwise differences between the reference runs were calculated using the formula (2) according to the formula:

$$y^{(l)} = \max_{n \in \{1, \dots, s\}; n \neq l} \frac{1}{N} \sqrt{\sum_{i=1}^N (z_i^{(n)} - z_i^{(l)})^2} \quad (3)$$

The auto-reference preliminary analysis enables to supplement the conclusions obtained from the previous method, allowing for a more accurate determination of the threshold, the moment of the first failure. However, the previous method allows faults to be distinguished between machines – see the next section.

The probabilistic stage consists in using the chi-square test for checking whether two different runs have the same distribution. The results of the measurements in each run were treated as sample of distribution of the feature. Chi-square test for homogeneity was used for the comparison. This test is convenient for usage because the assumptions are minimal. It is only assumed that the statistical samples are large and the number of elements in each bin is not less than five. Furthermore, bin

widths have to be equal. However, the variable can be both continuous and discrete. Furthermore, the distribution function needs not to be continuous. The statistics is given by the formula

$$\chi^2 = \frac{(n_1 + n_2)^2}{n_1 n_2} \left[ \left( \sum_i \frac{n_{1i}^2}{n_{1i} + n_{2i}} \right) - \frac{n_1^2}{n_1 + n_2} \right] \quad (4)$$

where  $n_1$  i  $n_2$  are the numbers of elements in the first and in the second sample, respectively. Variables  $n_{1i}$ ,  $n_{2i}$  denote the numbers of elements in the  $i$ -th bin of the first and the second statistical sample. The results of measurements in each run are a single sample. The probabilistic stage allows precisely determine the moment of the first failure.

### 3. RESULTS

Simulated model consists of two elements - a synthetic model of vibration signal and synthetic model of development of particular fault of rotary machinery. Signal is defined as a sampled time waveform of acceleration of vibrations. Depending on the simulated fault mode, consecutive vibration signals are modified differently. Each vibration signal is constructed as a phenomenological-behavioral model with GAD [13] shaft components (AM-FM harmonics) and gearbox components (AM-FM harmonics with multiple double sidebands) as well as and GATD [12] rolling-element bearings components (AM-FM cyclo-nonstationary components with additional phase-locked amplitude modulation). Fault development is modeled as a combination of linear, 2nd order polynomial or exponential growth of amplitudes of individual signal components with relatively low or relatively high variance. The model is based on [9], chapter 8.

Thus, a group of eight rotary identical engines were simulated. For each machine 250 runs were monitored. Each run lasted 10 seconds and was monitored with frequency 25 000 Hz which means that one run consisted of 250 000 measurements. The first 50 seconds of each machine operation is trouble-free and, therefore, were used as referential runs.

#### 3.1. Results obtained by multi-reference preliminary analysis

According to the parameters of simulation specified above, the range of the indices in formula (1) are following:  $1 \leq i \leq 250000$ ,  $1 \leq k \leq 250$ ,  $1 \leq j, m \leq 8$ ,  $N = 250\ 000$ ,  $s = 5$ .

Taking into account the fact that the first five runs are reference on each machine, they were used to calibrate the method. The greatest difference measured by the formula (1) for the reference waveforms is between machine 1 and machine 8 and is 0.03. Thus, it was assumed that the value of 0.05 is the limit value above which the signals should be considered as significantly different. The following results were obtained. Machines 1 and 2

show no significant difference in signals during the first 230 runs, then their signals start to diverge and this is an upward trend, but the differences remain below 0.1 - see Fig.1. Likewise, machine 4 shows no appreciable difference with both machine 1 and machine 2 for the first 200 runs; then their signals start to diverge more and more. Finally, the difference is 0.125.

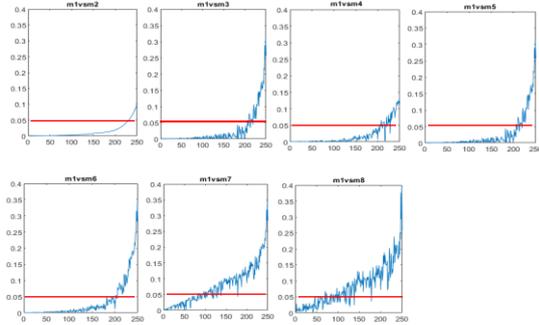


Fig. 1. Comparison of the operation of 1 machine with each of the other seven. Description in headings according to the convention: the first number corresponds to the index  $j$  in formula (1), and the second to the index  $m$  (i.e. m1vam2 means the comparison of the first machine with the second one). On the horizontal axis, the value of the parameter  $k$  from the formula (1) – number of the run, on the vertical axis – the value of  $y_k^{jm}$ . The red line denotes the limit value equal to 0.05.

This means that machines 1 and 2 either continue to run normally, or one of them has a minor fault after 230 runs. Machine 4 malfunctions after 200 runs. In addition, the first 200 runs of machines 1, 2 and 4 can be considered as referential in relation to the rest of the machines. A comparison of machines 1, 2 and 4 with machine 3 (see Fig. 1, Fig. 2, Fig. 3) shows the failure of machine 3 at approximately 180 runs. Likewise, the comparison of machines 1, 2 and 4 with machine 5 (see Fig. 1, Fig. 2, Fig. 4) also shows a failure of machine 5 at roughly 180 runs. Moreover, a comparison of machine 3 with machine 5 (see Fig. 3) shows that they perform the same. The comparison of machines 1, 2 and 4 with machine 6 (see Fig. 1, Fig. 2, Fig. 5)

4) shows the failure of machine 6 also around 180 runs. However, the behavior of machine 6 differs from that of machines 3 and 5 - see Fig. 3 and Fig. 5. A comparison of machines 1, 2 and 4 with machines 7 and 8 (see Fig. 1, Fig.2 and Fig.4) shows a failure of approximately 70 run. Machines 7 and 8 behave the same - see Fig.7.

To sum up, on the basis of multi-reference preliminary analysis the following conclusions were obtained.

Machines 1 and 2 work without failure or one of them has a fault of approximately 230 run.

Machine 3 and 5 behave the same. The fault occurs in them around 180 run.

Machine 4 has approximately 200 run fault.

Machine 6 has approximately 180 run fault. This is a different fault than on machines 3 and 5.

Machine 7 and 8 behave the same. The fault occurs in them about 70 runs. This is a different fault than the rest of the machines.

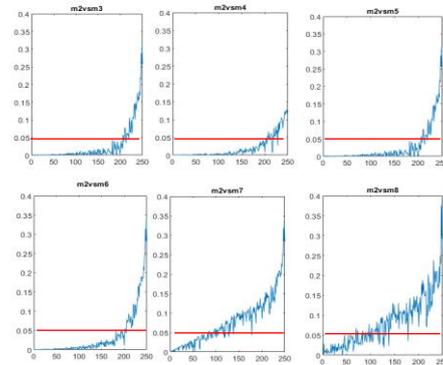


Fig. 2. Comparison of the operation of machine 2 with the others. Comparison with the first one - see Fig. 1.

The naive method is only suitable for detecting the first fault, assuming that at least two identical machines are tracked in parallel. If we only have two machines, we cannot decide which of them the fault occurred in, nevertheless, we are able to conclude that a failure has occurred in one of them.

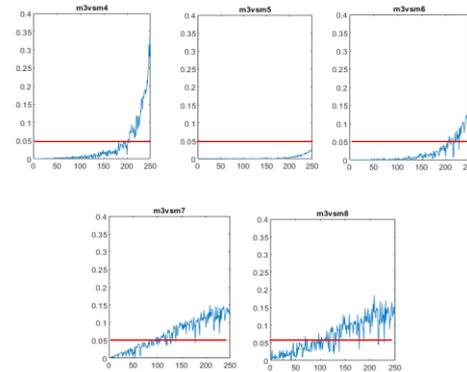


Fig.3. Comparison of the operation of machine 3 with the others. Comparison with the first and the second ones - see the previous figures.

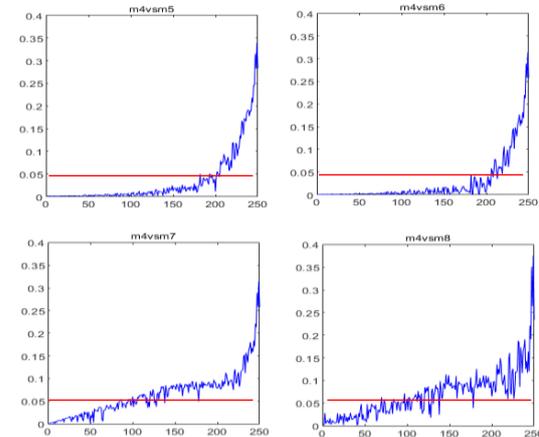


Fig. 4. Comparison of the operation of machine 4 with machines 5,6,7 and 8. Comparison with the others - see the previous figures.

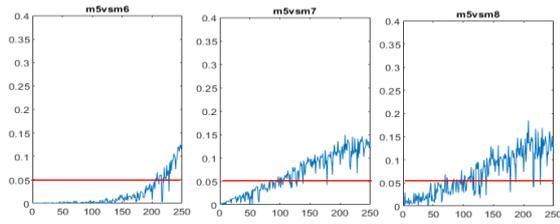


Fig. 5. Comparison of the operation of machine 5 with machines 6, 7, and 8. Comparison with the others - see the previous figures.

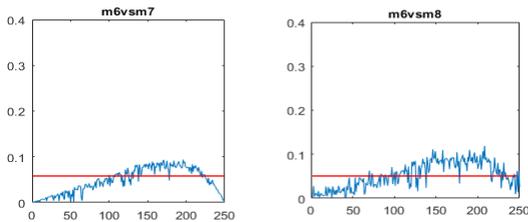


Fig. 6. Comparison of the operation of machine 6 with machines 7 and 8. Comparison with the others - see the previous figures.

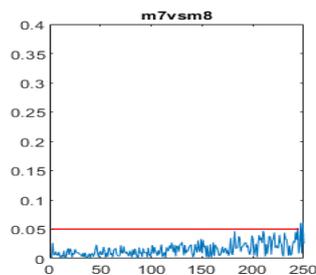


Fig. 7. Comparison of the operation of machines 7 and 8. Comparison of the machine 7 with the others - see the previous figures.

### 3.2. Results obtained by auto-reference preliminary analysis

According to the parameters of simulation specified at the beginning of Section 3, the values of the indices in formulae (2) and (3) are following:  $l=5$ ,  $s=5$ . The ranges of values of the remaining parameters are the same as for the multi-reference preliminary analysis.

For each of the eight machines, the obtained value of  $y_l$  was approximately 0.2. Therefore 0.25 was taken as the threshold above which the signal should be treated as different from the reference. The results - see Fig. 8 - confirmed the predictions of the multi-reference preliminary analysis, but they refined them - both machine 1 and 2 worked flawlessly all the time. The failure of machines 3, 5 and 6 should be moved to approximately 200 run, while machines 7 and 8 to approximately 150 run.

The multi-reference preliminary analysis and the auto-reference preliminary analysis should be treated as complementary. The auto-reference analysis gives a more precise moment of failure, because it refers to its own pattern. However, only the multi-reference analysis allowed to distinguish

between the failures of machines 3 and 5 and machine 6.

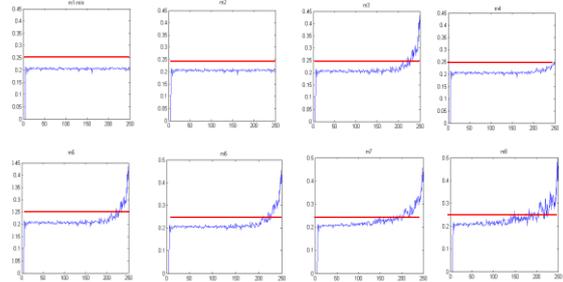


Fig. 8. Compare the operation of each machine with its own reference runs, which were the first 5 runs. On the horizontal axis, the value of the  $k$  parameter, on the vertical axis, the difference values, calculated by using formula (2). The red line denotes the limit value equal to 0.25.

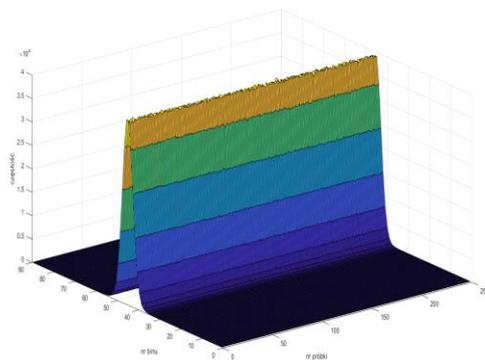


Fig.9. Distribution of measurement results for the machine 1. On the horizontal axes - run number and bin number. On the vertical axis, the number of cases in a given bin.

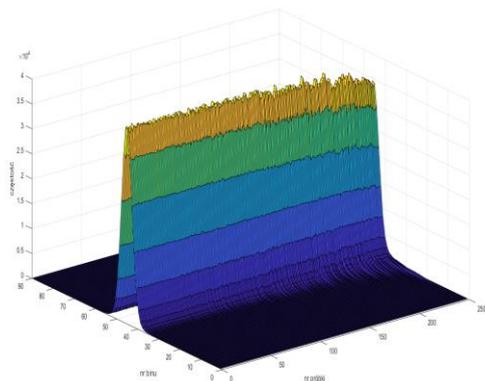


Fig.10. Distribution of measurement results for the machine 8. On the horizontal axes - run number and bin number. On the vertical axis, the number of cases in a given bin.

The probabilistic stage gave the following results.

Machines 1 and 2 all the time operates below the critical value which means that the distribution in any run does not differ statistically significantly from any reference distribution.

Machine 3 - up to 101st run operates below critical value. In the range of 102-130 runs oscillation

occurred – the value of statistics was partly above, partly below critical value.

Starting from the 131st run the value of statistics was permanently above critical value.

Machine 4 - up to 205 run below critical value, 206-229 - oscillations, from the 230th run - permanently above critical value.

Machine 5 operates exactly in the same way as machine 3.

Machine 6 - up to the 86th run below the critical value, oscillations 87-206, from the 207th run - permanently above the critical value.

Machine 7 - up to the 84th mileage below the critical value, 85-139 oscillations, from the 140th run - permanently above the critical value.

Machine 8 - single above critical value for runs 6, 17 and 27, 84-139 oscillations, from the 140th run - constantly above critical value.

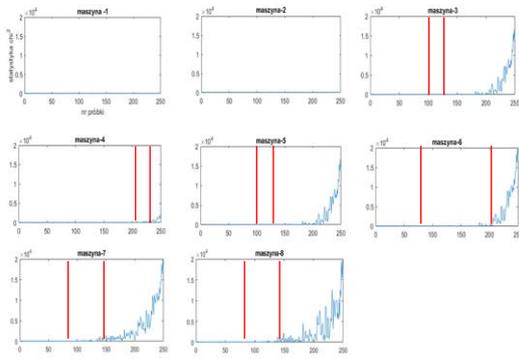


Fig.11. Graphs of the statistics of successive runs for all machines – see formula (4). On the vertical axis the value of the statistics – see formula (4). Up to the first vertical red line values of the statistics are less than critical value, the area of oscillations lies between the vertical red lines, whereas to the right of the right vertical line the values of the statistics are greater than critical value.

It can be observed, that both preliminary analyzes give very similar results, but are slightly complementary. Probabilistic stage enables to distinguish the period of transitional operation of a machine – oscillations – after which always occurs permanent operation above critical value of the statistics. This means that the proposed probabilistic approach enables to detect the failure at the very early stage. As a consequence, the onset of the fault should be taken at the beginning of the oscillation.

#### 4. CONCLUDING REMARKS

In the paper the simple method of early detection of failures was proposed. The method enables to detect the failure at the very early stage of its occurrence. According to its simplicity, the method is extremely easy for implementation and very fast. The last is important in the real-time systems, to which systems of intelligent monitoring belong.

#### SOURCE OF FUNDING

The paper was partially supported from the National Centre for Research and Development (NCBiR) under the grant no. POIR.04.01.04-00-0080/19.

#### REFERENCES

1. Aggarwal C. Outlier analysis, in: Data Mining, Springer 2015; 237–263.
2. Bielecki A, Wójcik M. Hybrid system of ART and RBF neural networks for online clustering. Applied Soft Computing. 2017;58:1-10. <https://doi.org/10.1016/j.asoc.2017.04.012>
3. Bielecki A, Wojcik M. Hybrid AI system based on ART neural network and mixture of Gaussians modules with application to intelligent monitoring of the wind turbine. Applied Soft Computing 2021; 108:107400. <https://doi.org/10.1016/j.asoc.2021.107400>
4. Bielecki A, Barszcz T, Wójcik M, Bielecka M. Hybrid system of ART and RBF neural networks for classification of vibration signals and operational states of wind. Lecture Notes in Artificial Intelligence. 2014;8467:3-11. [https://doi.org/10.1007/978-3-319-07173-2\\_1](https://doi.org/10.1007/978-3-319-07173-2_1).
5. Dao PB, Staszewski WJ, Barszcz T, Uhl T. Condition monitoring and fault detection in wind turbines based on cointegration analysis of SCADA system. Renewable Energy 2018; 116: 107-122. <https://doi.org/10.1016/j.renene.2017.06.089>
6. Dinardo G, Fabbiano L, Vacca G. A smart and intuitive machine condition monitoring in the Industry 4.0 scenario. Measurement 2018; 26. <https://doi.org/10.1016/j.measurement.2018.05.041>
7. Gan M, Wang C. Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings. Mechanical Systems and Signal Processing. 2016;72-73:92-104. <https://doi.org/10.1016/j.ymsp.2015.11.014>
8. Górski J, Jabłoński A, Heesh M, Dziendzikowski M, Dworakowski Z. Comparison of novelty detection methods for detection of various rotary machinery faults. Sensors. 2021;21(10):3536; <https://doi.org/10.3390/s21103536>
9. Jabłoński A. Condition Monitoring Algorithms in MATLAB, Springer 2021.
10. Randall RB. Vibration-based Condition Monitoring: Industrial, Automotive and Aerospace Applications. John Wiley and Sons, Ltd., 2<sup>nd</sup> edition 2021.
11. Rogalewicz M, Sika R. Methodologies of knowledge discovery from data and data mining methods in mechanical engineering. Management and Production Engineering Review 2016; 7(4): 97-108. <https://doi.org/10.1515/mper-2016-0040>.
12. Urbanek J, Barszcz T, Jabłoński A. Application of angular-temporal spectrum to exploratory analysis of generalized angular-temporal deterministic signals. Applied Acoustics 2016; 109: 27–36. <https://doi.org/10.1016/j.apacoust.2016.03.004>
13. Urbanek J, Barszcz T, Strączkiewicz M, Jabłoński A. Normalization of vibration signals generated under highly varying speed and load with application to signal separation. Mechanical Systems and Signal

Processing 2017;82:13–31.

<https://doi.org/10.1016/j.ymsp.2016.04.017>

14. Wodecki J, Michalak A, Zimroz R, Wyłomańska A. Separation of multiple local-damage-related components from vibration data using Nonnegative Matrix Factorization and multichannel data fusion. *Mechanical Systems and Signal Processing* 2020; 145:106954.  
<https://doi.org/10.1016/j.ymsp.2020.106954>
15. Wodecki J, Stefaniak P, Obuchowski J, Wyłomańska A, Zimroz R. Combination of principal component analysis and time-frequency representations of multichannel vibration data for gearbox fault detection. *Journal of Vibroengineering*. 2016; 18:2167-2175.  
<https://doi.org/10.21595/jve.2016.17114>



#### Wojciech

**STASZEWSKI** received his M.S. degree from AGH University of Science and Technology in 2015. He is currently a PhD student researching the area of system identification and development of Low Cost

Condition Monitoring Systems.

Received 2021-06-15

Accepted 2021-10-04

Available online 2021-10-13



Prof. dr hab. **Andrzej BIELECKI** received the M.Sc. degree in Physics and Mathematics from the Jagiellonian University in 1985 and 1992 respectively, Ph.D. in Mathematics in 1999, D.Sc. (habilitation) in Mathematics in 2009 and

professorship in 2020 in Computer Science. Dynamical systems theory, artificial intelligence, cybernetics and philosophy of science are the topics of his scientific interest. He is an author of one book and over 120 scientific papers.



Dr hab. **Marzena BIELECKA** received PhD in Computer Science and D.Sc. (habilitation) in Computer Science as well. She is an expert in soft computing and statistics and their applications in engineering and medicine. She is an author of over 50 scientific papers.



Dr hab. **Adam JABŁOŃSKI** graduated from AGH University of Science and Technology, Poland and Central Texas College, USA. For many years, he has been involved in developing customized condition monitoring systems and he

is a full-time researcher at AGH University of Science and Technology. His main area of research interest is signal processing for diagnosis purposes.