

Article

Identification, Decomposition and Segmentation of Impulsive Vibration Signals with Deterministic Components—A Sieving Screen Case Study

Karolina Gašior ¹, Hanna Urbańska ¹, Aleksandra Grzesiek ^{2,*}, Radosław Zimroz ² 
and Agnieszka Wylomańska ¹

¹ Faculty of Pure and Applied Mathematics, Hugo Steinhaus Center, Wrocław University of Science and Technology, Wyspiańskiego 27, 50-370 Wrocław, Poland; 242966@student.pwr.edu.pl (K.G.); 243070@student.pwr.edu.pl (H.U.); agnieszka.wylomanska@pwr.edu.pl (A.W.)

² Faculty of Geoengineering, Mining and Geology, Wrocław University of Science and Technology, Na Grobli 15, 50-421 Wrocław, Poland; radoslaw.zimroz@pwr.edu.pl

* Correspondence: aleksandra.grzesiek@pwr.edu.pl

Received: 7 August 2020; Accepted: 30 September 2020; Published: 2 October 2020



Abstract: Condition monitoring is a well-established field of research; however, for industrial applications, one may find some challenges. They are mostly related to complex design, a specific process performed by the machine, time-varying load/speed conditions, and the presence of non-Gaussian noise. A procedure for vibration analysis from the sieving screen used in the raw material industry is proposed in the paper. It is more for pre-processing than the damage detection procedure. The idea presented here is related to identification and extraction of two main types of components: (i) deterministic (D)—related to the unbalanced shaft(s) and (ii) high amplitude, impulsive component randomly (R) appeared in the vibration due to pieces of ore falling down of moving along the deck. If we could identify these components, then we will be able to perform classical diagnostic procedures for local damage detection in rolling element bearing. As deterministic component may be AM/FM modulated and each impulse may appear with different amplitude and damping, there is a need for an automatic procedure. We propose a method for signal processing that covers two main steps: (a) related to R/D decomposition and including signal segmentation to neglect AM/FM modulations, iterative sine wave fitting using the least square method (for each segment), signal filtering technique by subtraction fitted sine from the raw signal, the definition of the criterion to stop iteration by residuals analysis, (b) impulse segmentation and description (beginning, end, max amplitude) that contains: detection of the number of impulses in a decomposed random part of the raw signal, detection of the max value of each impulse, statistical analysis (probability density function) of max value to find regime-switching), modeling of the envelope of each impulse for samples that protrude from the signal, extrapolation (forecasting) envelope shape for samples hidden in the signal. The procedure is explained using simulated and real data. Each step is very easy to implement and interpret thus the method may be used in practice in a commercial system.

Keywords: mixture of signals; source separation; shock description; sieving screen; raw materials industry

1. Introduction

A vibration signal measured on machine housing is usually a non-stationary one, with a complex, time-varying frequency structure. It may consist of deterministic, random, and transients (impulsive) components. All these components can be modeled, analyzed, and might be a source of important

information about machine condition or process performance; however, all these components should be treated using different tools. Therefore, understanding of the signal's nature and extraction of each component for further analysis is a critical pre-processing step in the condition monitoring.

In engineering applications, the condition monitoring is related to the tracking of amplitudes and frequencies of deterministic components related to shafts, meshes, etc. In such a case, any random, transient components are considered to be unwanted disturbances [1–5].

There is another serious class of solutions that plays with random signals assuming that local damage in gearboxes or bearings “produces” random signals with two important properties: impulsiveness and periodicity [6,7]. From that perspective, cyclostationary analysis is probably the most important solution [6–9].

As mentioned in [6], before cyclostationary analysis—an appropriate tool for random processes—the deterministic components should be removed. It is the so-called Wold decomposition [10]. In fact, this kind of signal decomposition inspired us to prepare this paper.

Our machine, a sieving screen, generates vibrations with three main components: deterministic—related to the unbalanced shaft to excite the upper deck to classify raw materials; random, impulsive, and non-cyclic component related to falling down or moving along the deck the oversized pieces of raw materials and—in the case of local damage in rolling element bearings—an impulsive, cyclic, weak random signal. The mixture of these components definitely requires separation for further analysis. The ultimate goal in condition monitoring of the sieving screen is related to damage detection in bearings. Bearings used in the sieving screen are subjected to overloading related to mentioned falling down pieces of ore as well as intentionally introduced unbalance of the shaft. Therefore, in fact, we have two mentioned phenomena: varying load (due to varying amount of ore to classify and unbalance of the shaft) and impulsive load considered to be impulsive background noise. As discussed in research works published already [11,12], for this class of machines the problem of local damage detection is challenging and may be defined as the detection of cyclic impulsive signal in presence of non-cyclic impulsive noise [12–19].

The presence of non-Gaussian noise plays a crucial role here. If the level of impulsive disturbances is relatively small, one may use classical diagnostic techniques. When impulsive, non-Gaussian noise becomes a dominating component (the amplitude of a single impact is higher 20–30 times than cyclic impulses), most of the classical techniques used for diagnostics are non-effective. In the considered case, the situation is very dynamic. If the mined material delivered to the screen is well fragmented (depends on geology, blasting procedure, etc.), it may not contain the oversized pieces, so high amplitude impulses will not be present. However, if the oversized pieces of material are present in the material stream, a high level of disturbance will complicate diagnostic procedures. Thus, the key issue is to recognize if we have unwanted, non-cyclic impulses in the signal.

The motivation of this paper is to identify these non-cyclic impulses with high amplitude and to use information about their presence and properties in decision making related to damage detection. In other words, we would like to detect high energy impulses in the raw signal, segment them, and remove them from the diagnosis process. To detect high amplitude impulses, to estimate their duration, and to remove such a part of the signal may be seen as a very easy task. Unfortunately, due to a specific use case, a simple problem becomes more complicated and requires a special procedure.

A raw signal contains high energy deterministic components, so decomposition into random and deterministic signals should be performed first. Without this step, the precise localization of the impulse may be difficult.

Due to the complexity of the mining process, the input materials stream is time-varying (volume/mass is changing in time). It leads to a time-varying load for the sieving screen drive unit and, in consequence, it will bring slight frequency modulation for deterministic components (rotational speed will be time-varying). From a signal processing point of view, it is frequency modulation when a modulating signal is both deterministic (sine function due to shaft unbalance) and random process (material volume variation).

A decomposed random signal is then processed for impulse properties estimation. We need information related to the beginning, amplitude, and end of the impulse. Finding the maximum value of the impulse is in our case easy. However, precise detection of the beginning and especially the end of impulse may be challenging as they may be still hidden in background noise. Thus, we propose to identify impulse behavior using statistical analysis, envelope analysis, and simple curve approximation (fitting).

As was mentioned, in our case there are several well-known tasks in signal processing, namely, signal decomposition, modelling, detection, and segmentation. They are fundamental in signal processing so one may find many references related to them.

Signal decomposition in our context means decomposition into random and deterministic components (the Wold Decomposition). As we know, the source of determinism in the signal (unbalance of the shaft) can be approximated by a sine-wave function. When the deterministic signal is estimated, then the residual signal could be obtained by simple subtraction of the deterministic component from the raw signal. It is a very popular technique used in many applications, see [10,20–22]. In the case of several sine components, it could be done in an iterative way or the model could contain several sine components (we have proposed criteria for such iteration). It is worthy to mention that due to slight frequency modulation we use sine function fitting for a short segment of the signal (it does not work for the whole signal). It should be also highlighted that we “model” deterministic components to remove it from the signal, we do not pay attention to them later.

Obviously, the problem of deterministic part extraction is well described in the literature and many advanced techniques could be found as time-synchronous averaging, adaptive filters based on linear prediction theory, Vold-Kalman filters, etc. Even more papers are focused on non-stationary signal modelling [23–28]. Mostly they considered time-frequency domain (STFT, PWVD, wavelets, chirplets, etc.) [29–32]. A special case of the non-stationary signal is AM/FM modulation (this is also our case) exists in the vibration signal due to changing operating conditions or existing damage [33–35]. However, as our signal is not very complicated (at least the deterministic part), we decided to use the simplest solution for modelling of this component.

After deterministic components removal, in theory the purely random signal appears. In our case, it consists of the impulsive disturbances related to falling rocks and noise. Our next step is to identify each impulse in an automatic way taking into consideration the specific nature of the impulses. Impulse detection is a classical task in signal processing. In condition monitoring, it is related mostly to local damage detection and a variety of detectors can be used. In our context, the problem is simplified as the amplitude of the impulse is high and it is clearly visible in the raw signal, so simple statistical detectors can be used (peak, kurtosis, etc.) [36].

However, we would like to get precise information about impulses (when appeared, how strong it is, when disappeared in background noise). For such a task one may use signal segmentation. The signal segmentation is considered for random signals (this is a reason we need deterministic/random components separation). The segmentation is usually applied for a signal with essential changes of its properties (changes of the regime) [37,38] or extraction a specific pattern [39]. It is commonly used in many applications (biomedical signals, experimental physics, speech analysis, econometrics, seismic signal analysis, machine performance analysis, aircraft noise, physical sciences, etc.) [40–46].

In our case, we propose a procedure for segmentation that is based on hybrid statistical modelling. First, we detect the beginning of impulse using analysis of detected maximum values in decomposed signal, second, we exploit a concept of damped vibration and we fit the envelope (Hilbert-based) to the signal for samples that protrude from the signal. Knowing the shape of the envelope (exponential function), we can extrapolate it (forecast) for samples hidden in the signal.

The paper is organized as follows: in Section 2 we describe the machine, the experiment, and the data. In Section 3 we present the methodology used in the data processing which consists of four steps including iterative removal of deterministic components, determining the number of impulses,

and identifying their location in the signal. In Section 4 we demonstrate the efficiency of the proposed methodology for simulated data. In Section 5 we analyze a real signal. The last section concludes the paper.

2. Experiment and Data Description

2.1. Machine Description

In the considered machine (see Figures 1 and 2), a sieving screen is designed to separate the input raw material into three classes: less than 40, 40–110, and more than 110 mm. The material stream is coming into the screen (falling on the deck) from a conveyor located on the top. The screen supports consist of four sections with three springs in each corner. The screen is driven by two electric motors and individual belt transmissions. Special spherical roller bearings are used on the shafts of unbalanced exciters, see [12] for more details.



Figure 1. Vibrating screen machine (source: <https://kghm.com/pl/w-pracy>).



Figure 2. Vibrating screen machine (source: <https://kghm.com/pl/w-pracy>).

2.2. Experiment Description

An experiment was a passive one, we have measured vibration during normal operation of the machine with various types and quality of input material. As the screen is a critical component in the production line, no modification or stoppage was possible. During the experiment, many times the oversized pieces of ore hitting the deck have been noticed. It is a normal situation as blasting technology used in the mine provides various sizes of ore pieces. We have used typical devices for vibration measurements.

2.3. Data Description

During the experiment, many vibration signals from several channels have been recorded. Sampling frequency was set to 50 kHz to assure a wide range of frequency for transient impulsive

data. The duration of the signal was c.a. 2 min to avoid too big to handle data files with multi-channel signals. Due to NDA, we are not allowed to provide more detailed information.

3. Methodology

In this section, we present the methodology used in real data analysis. The main goal of this procedure is to identify the location of the impulses in the acoustic signal presented in Section 2. The proposed methodology can be divided into four stages:

- iterative removal of deterministic components from the raw data,
- determining the number of impulses in a cleansed signal,
- detection of the moments when the impulses begin,
- detection of the moments when the impulses end.

A scheme illustrating the overall procedure is presented in Figure 3 and the methods used in the subsequent stages are described in detail in the following subsections. Moreover, the performance of the proposed procedure is verified in Section 4, where a simulated signal is analyzed.

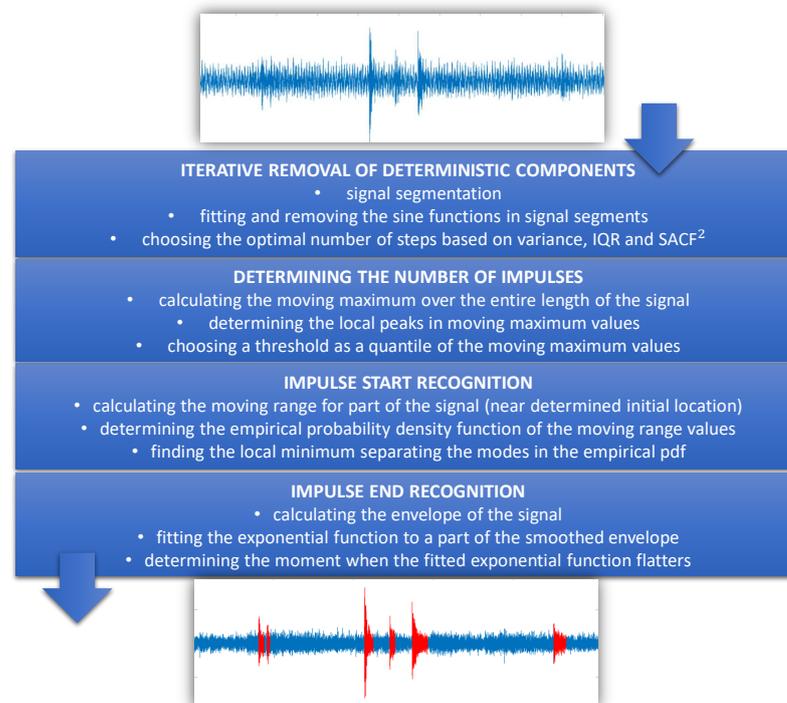


Figure 3. A scheme presenting the overall procedure. More details are provided in the Sections 3.1–3.4.

3.1. Iterative Removal of Deterministic Components

On the raw data plot in Figure 4a and in the signal spectrum in Figure 4b one can recognize the periodic trends that are present along the entire length of the signal. However, the frequencies of the periodic functions observed in the data change in time because, as it was already mentioned in the Introduction, the material stream coming into the sieving screen affects the engine operation. Therefore, to clean the data from the deterministic components, first, we divide the signal into segments. To choose the number of segments, we propose to compare the frequencies of the sine functions fitted to each segment of the signal for all segment lengths changing from 10,000 to 55,000 with a step equal to 5000 (by the segment length we mean the number of samples in one segment). As the best length, we choose the one for which the sine frequencies in the subsequent segments have the smallest variance.

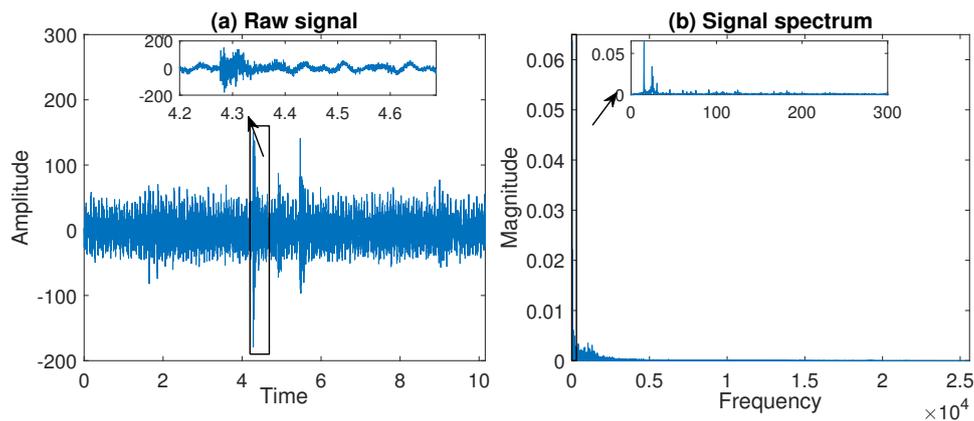


Figure 4. The trajectory of a raw signal—panel (a), and the signal spectrum—panel (b) for the real data analyzed in Section 5.

Now, let m and n indicate the number of segments in the signal and the length of one segment, respectively. Then, $\mathbf{x}_i = (x_{i,1}, x_{i,2}, \dots, x_{i,n})$ denotes the observations in the i -th segment where $i = 1, 2, \dots, m$. In the first step, we fit a sine function to each segment \mathbf{x}_i , $i = 1, 2, \dots, m$ separately, using the least square method, and subtract the fitted function from the raw data. Let $\mathbf{r}_{i,1}$, $i = 1, 2, \dots, m$ indicate the residual time series obtained after removing the first sine function from the data. Next, we repeat the above step three more times and finally compare the raw signal and the obtained residual time series by calculating some statistics of \mathbf{x}_i , $\mathbf{r}_{i,1}$, $\mathbf{r}_{i,2}$, $\mathbf{r}_{i,3}$, and $\mathbf{r}_{i,4}$ for the subsequent segments $i = 1, 2, \dots, m$. More precisely, we compare the variances, the interquartile ranges and the sums of squared auto-correlation functions computed for all segments. Let us mention here that since for the observations $\mathbf{y} = (y_1, y_2, \dots, y_n)$ the empirical auto-correlation function at k takes the following form [10]

$$\widehat{\text{ACF}}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)}, \quad (1)$$

where

$$\hat{\gamma}(k) = \frac{1}{n} \sum_{j=1}^{n-k} (y_{j+h} - \bar{y})(y_j - \bar{y}), \quad (2)$$

for $k = 0, 1, \dots, k_{max}$ and $0 < k_{max} \leq n$, the mentioned sum of squared auto-correlation function is defined as follows

$$\text{SACF}^2 = \sum_{k=0}^{k_{max}} \left(\widehat{\text{ACF}}(k) \right)^2. \quad (3)$$

Since the auto-correlation function quantifies the interdependence in the data, after removing the deterministic components the value taken by SACF^2 given in Equation (3) should decrease in comparison to the statistic computed for raw signal because the strong dependence arising from the presence of deterministic components disappears. Therefore, based on the values taken by SACF^2 and other mentioned statistics, we decide which of the cleaned time series, $\mathbf{r}_{i,1}$, $\mathbf{r}_{i,2}$, $\mathbf{r}_{i,3}$, or $\mathbf{r}_{i,4}$, should be taken for the analysis in the further stages.

3.2. Determining the Number of Impulses

To determine the number of impulses in the signal cleaned in the previous step, we calculate the moving maximum over a sliding window of length l_1 . We start from the first l_1 observations and move one sample to the right until there are enough elements to fill the window. In the next step, we compute the local peaks of the values taken by the moving maximum. As a local peak, we consider a data sample that is larger than its two neighboring samples. Then, the number of impulses in the signal is equal to the number of local peaks that are located above or on the level of a threshold chosen based on the calculated moving maximum values. For both considered simulated and real signals, the window

length is equal to $l_1 = 9000$ and the threshold value is chosen as a certain quantile of the values taken by the moving maximum (quantile of order 0.7 or 0.875 for simulated and real data, respectively).

3.3. Impulse Start Recognition

To locate where the impulse starts, we consider a part of the signal that contains samples both preceding and following the beginning of the impulse. For this part of the signal, we calculate the moving range over a sliding window of length l_2 similarly as in the previous stage of the analysis, see Section 3.2. Then, we examine the empirical probability density function (pdf) of the computed moving range values. If the considered part of the signal contains the beginning of the impulse, the distribution is bi-modal and the local minimum separating the modes can be treated as a threshold. Namely, the values taken by the moving range, which are greater than the set threshold, indicate the area of the impulse, and the first value meeting this condition determines the beginning of the impulse. We recall that to estimate the empirical probability density function from the observations $\mathbf{y} = (y_1, y_2, \dots, y_n)$ one can use the kernel density estimator of $f(y)$ defined as follows [47,48]

$$\hat{f}(y) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{y - y_i}{h}\right) \quad \text{for } y \in \mathbb{R}, \quad (4)$$

where $K(\cdot)$ is the non-negative kernel smoothing function, and h denotes the bandwidth. By choosing the kernel smoothing function, we determine the shape of the curve that is used to estimate the empirical pdf. For our purpose, we apply the Gaussian kernel of the following form

$$K(y) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{y^2}{2}\right\} \quad \text{for } y \in \mathbb{R}. \quad (5)$$

The bandwidth h is chosen according to the Silverman's rule of thumb to be optimal for estimating Gaussian densities, see [49]. The mentioned method is implemented in many mathematical packages (e.g., "ksdensity" in Matlab). Moreover, for both simulated and real signal, the window length for the moving range is chosen as $l_2 = 20$.

3.4. Impulse End Recognition

Since each impulse present in the signal is a damped vibration, its envelope can be approximated by an exponential function. As the end of the impulse, we can consider the moment when the fitted exponential function flatters. Let $\mathbf{r} = (r_1, \dots, r_N)$ indicate the signal cleaned at the first stage of the analysis, see Section 3.1, where N denotes its length. To extract the envelope of \mathbf{r} we use the Hilbert transform. Namely, let us consider the analytic signal composed as follows

$$\aleph(\mathbf{r}) = \mathbf{r} + i \tilde{\mathbf{r}}, \quad (6)$$

where $\tilde{\mathbf{r}} = (\tilde{r}_1, \dots, \tilde{r}_N)$ is the Hilbert transform of the original data, see [50]. Then, the envelope of the signal is given as the absolute value of the analytic signal given in Equation (6), namely $|\aleph(\mathbf{r})|$. In the next step, we smooth the envelope of the signal using the simple two-sided moving average method, see [10], and we fit an exponential function to the smoothed envelope of the impulse starting from the peak found at the previous stage of the analysis, see Section 3.2. As the end of the approximation interval, we choose the point where the envelope intersects the threshold determined based on the standard variation of the signal. In addition, consequently, as the end of the impulse, we consider the point where the fitted exponential function is flattered. In practice, we identify the moment when the difference in values taken by the fitted function for adjacent samples takes a value less than the selected ϵ . For real signal, the value of ϵ is chosen to be equal to 5.52×10^{-4} .

4. Simulated Data Analysis

In this section, we apply the proposed methodology to the simulated data presented in Figure 5. The analyzed signal is generated as a sequence of independent observations drawn from the standard normal distribution. It is simulated on the interval $[0, 5]$ with sampling frequency equal to 10,000 Hz. Additionally, starting at $t = 2$ we generate an exponentially decaying impulse of length 1000 that follows the formula

$$f(t) = a \exp(-bt) \sin(2\pi ft), \quad (7)$$

where $a = 25,000$, $b = 4$ and $f = 3000$, see Figure 5a for the generated impulse and Figure 5b for the simulated signal. In the first subsection, we demonstrate the performance of the method related to the removal of deterministic components proposed in Section 3.1. For this purpose, to the simulated signal we add additional deterministic components, namely two sine functions (with different amplitudes and frequencies) of the form

$$s_i(t) = A_i \sin(2\pi f_i t) \quad (8)$$

for $i = 1, 2$, where $A_1 = 1.4$, $A_2 = 1.2$, $f_1 = 15$, and $f_2 = 25$. The simulated signal with deterministic components is presented in Figure 5c. After verifying the performance of the method for removing determinism, in the subsequent subsections, we will focus on the remaining stages of the proposed procedure described in Sections 3.2–3.4, namely determining the number of impulses and their location in the signal.

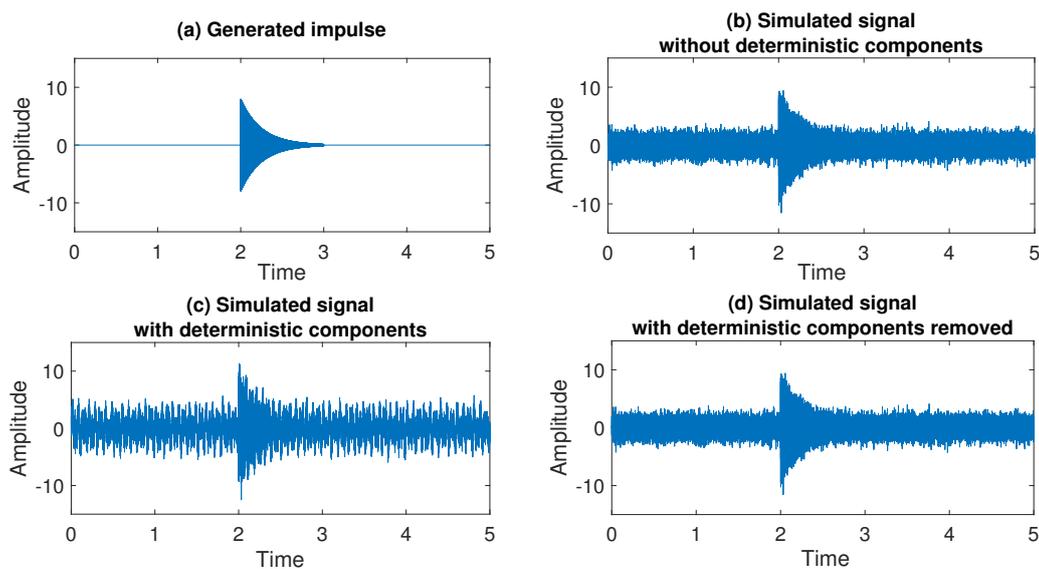


Figure 5. Generated impulse (a), simulated signal without and with deterministic components, (b) and (c) respectively, and simulated signal after iterative removal of two sine functions (d).

4.1. Simulation Study for Iterative Removal of Deterministic Components

Since in the simulated signal analyzed here, the parameters of $s_1(t)$ and $s_2(t)$ do not change over time, we proceed by omitting the step of dividing the signal into segments. Then, according to the proposed procedure, we iteratively fit the sinusoidal functions and determine the values of variance, interquartile range, and SACF^2 of the residual signals. Since we only have one segment in the signal, the statistic values after subtracting one, two, three, and four sine functions are presented in Table 1.

Table 1. The values of variance, interquartile range, and SACF² calculated for the signal with deterministic components and for the residual signals obtained after subtracting the first, second, third, and fourth sine function (denoted as r_1 , r_2 , r_3 and r_4 , respectively).

	Signal with Deterministic Components	r_1	r_2	r_3	r_4
variance	3.59	2.61	1.88	1.79	1.68
IQR	2.44	1.97	1.45	1.49	1.49
SACF ²	1003.66	723.86	140.07	131.44	103.60

As one can notice, the statistics values change most significantly after subtracting the first and second sine function. The third and the fourth sine fitted to the data do not change the statistics values so significantly. Additionally, the sine functions fitted in the first two iterative steps have a form that is very similar to the functions added to the original signal, namely

$$\tilde{s}_1(t) = 1.40 \sin(30\pi t + 0.02) \quad \text{and} \quad \tilde{s}_2(t) = 1.21 \sin(50\pi t + 0.01). \quad (9)$$

The signal reconstructed after subtracting two sine functions is presented in Figure 5d. We can see that visually it is very close to the one presented in Figure 5b. Additionally, the mean squared error of the original signal without deterministic components and the signal reconstructed by subtracting two sinusoidal functions is less than 0.0002 that confirms the efficiency of the method.

4.2. Simulation Study for Impulse Location Detection

To find the location of the generated impulse, first, we apply the method determining the number of impulses in the signal. According to the procedure presented in Section 3.2, we plot the values taken by the moving maximum, see Figure 6. Since there is one local peak located above a chosen threshold (a quantile of order 0.7), we conclude that there is only one impulse in the signal.

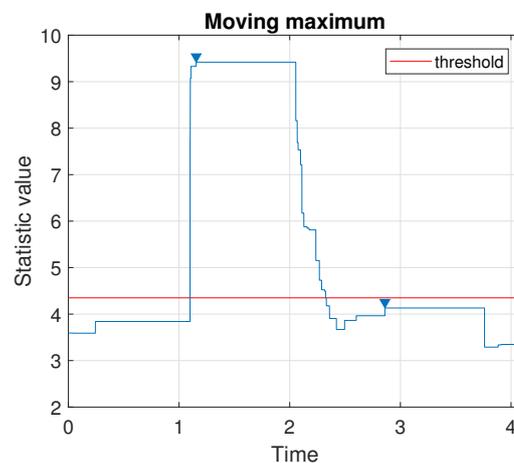


Figure 6. Values taken by the moving maximum calculated over a sliding window of length 9000. Triangle markers indicate the local peaks and the threshold is chosen as a quantile of order 0.7.

To identify the moment when the impulse starts, on the Figure 7a we plot a part of the signal around the sample determined by the moving maximum in the previous step. Following the procedure proposed in Section 3.3, we calculate the moving range (in Figure 7b) together with the corresponding empirical probability density function (in Figure 7c). As one can see, the distribution of the values taken by the moving range is bi-modal. By finding the local minimum separating the modes, we divide the values taken by statistic into two groups related to the area before impulse and the area of impulse, see in Figure 7d where the start of the impulse is marked with a red dashed line. In the next step, to find the end of the detected impulse we apply the method described in Section 3.4. The consecutive

steps of the procedure are presented in the Figure 8a–c. The threshold chosen to determine the end of the approximation interval, for the exponential function fitted to the envelope of the signal, is equal to 3 (three times the standard deviation of noise—according to the 68–95–99.7 rule). The end of the impulse is marked with a red dashed line in Figure 8c.

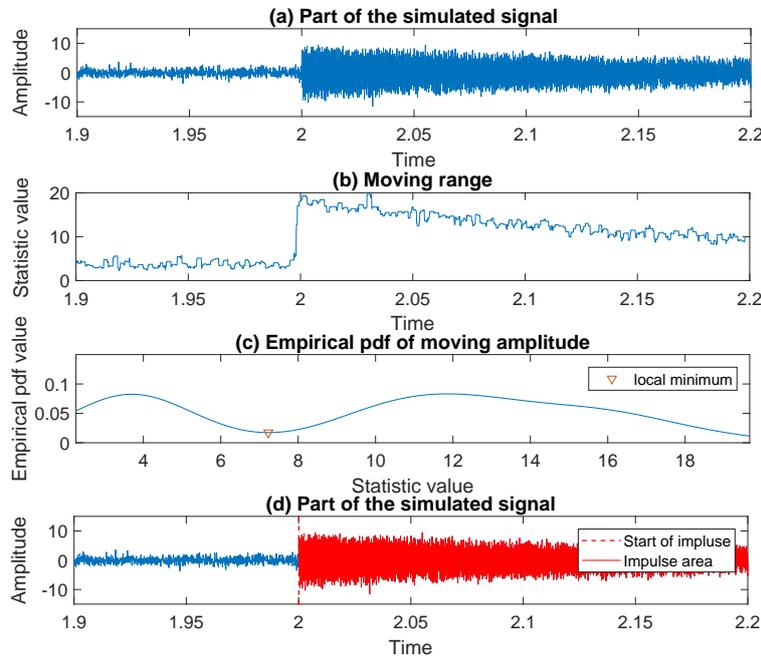


Figure 7. The consecutive steps of the procedure leading to the start recognition for an impulse in the simulated signal.

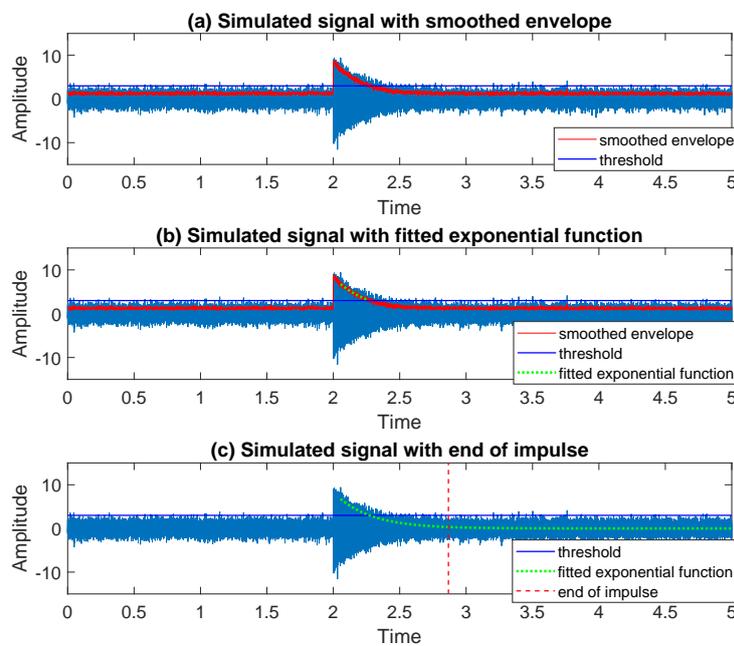


Figure 8. The consecutive steps of the procedure leading to the end recognition for an impulse in the simulated signal.

To summarize the results obtained in this section, in Figure 9 we present the simulated signal with the impulse marked in the original location—panel (a), and with the impulse marked in the location determined using the proposed methodology—panel (b).

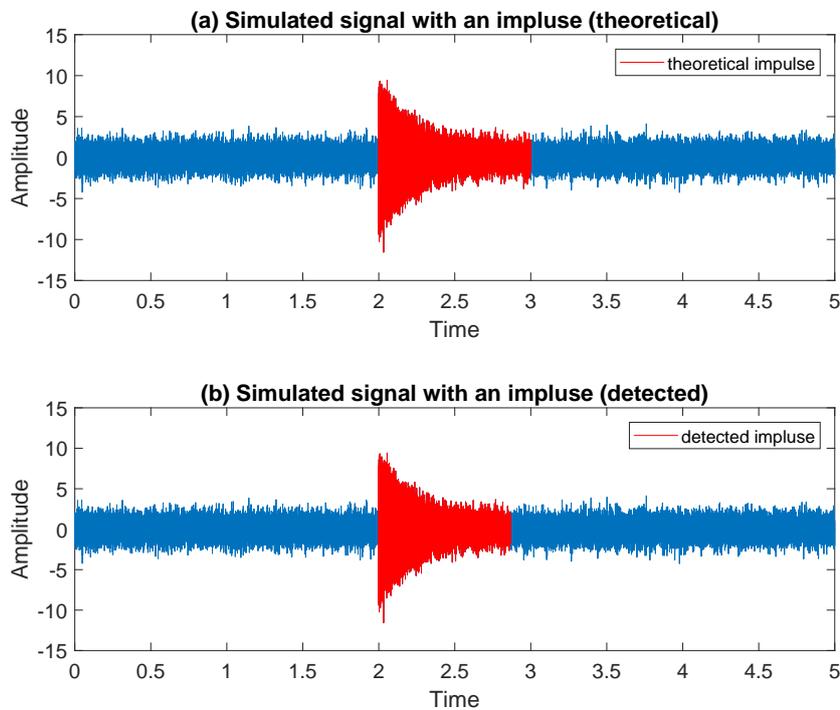


Figure 9. Simulated signal with an impulse marked in red: theoretical—panel (a), and detected using the proposed methodology—panel (b).

4.3. Simulation Study on the Influence of Impulse Amplitude on the Procedure Efficiency

To evaluate the influence of the parameter a in Equation (7), which is responsible for the amplitude of the generated impulse, on the correctness of the proposed procedure, we conduct additional simulation study. For each a changing from $a = 15,000$ to $a = 45,000$ with step equal to 1000, we generate $M = 100$ simulated signals as described above. For each signal, we localize the impulse using the method proposed in the paper. Then, we determine the absolute value of the differences between the theoretical values of its beginning and its end and those values that are determined by using our method. The results are presented in Figures 10 and 11.

In both figures, on the y -axis we plot the mean absolute differences between the theoretical and determined values. We can notice that in both cases, the method works better as the parameter a increases. The clearer the impulse, the easier it is to correctly determine its location. Moreover, the procedure for detecting the start of an impulse relatively quickly becomes error-free looking at the mean of the absolute difference.

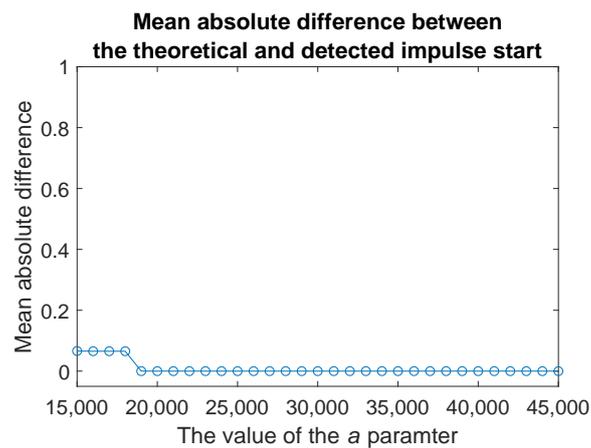


Figure 10. Mean absolute difference between the theoretical and detected impulse start calculated based on $M = 100$ simulated signals with exponentially decaying impulse as in Equation (7) where $b = 4$, $f = 3000$ and a is changing from 15,000 to 45,000.

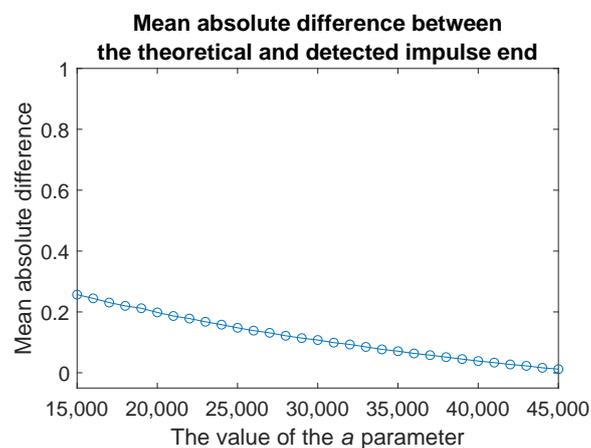


Figure 11. Mean absolute difference between the theoretical and detected impulse end calculated based on $M = 100$ simulated signals with exponentially decaying impulse as in Equation (7) where $b = 4$, $f = 3000$ and a is changing from 15,000 to 45,000.

5. Real Data Analysis

In this section, we analyze the real acoustic signal described in Section 2. To locate all impulses in the data, we proceed according to the presented methodology, and the following subsections correspond to the subsequent stages of the proposed procedure.

5.1. Iterative Removal of Deterministic Components

Following the methodology introduced in Section 3.1, to remove the deterministic components, we divide the signal into several segments. For each segment, we fit a sine function to the data. Then, we examine the frequencies corresponding to the subsequent sine functions (the number of sine functions is equal to the number of segments). In Table 2 for all considered segment lengths we present the sample variance of these frequencies. It can be seen that for the segments of length 40,000, the calculated sample variance takes the smallest value. For such a length, the frequencies of the fitted sines take very similar values in all segments.

Table 2. The length of one segment (the number of samples in one segment) and the corresponding sample variance calculated for the frequencies of the sine functions fitted to all segments.

Length	10,000	15,000	20,000	25,000	30,000	35,000	40,000	45,000	50,000	55,000
Variance	5505.8	8341.7	0.0117	14213	6.4279	23.5842	0.0003	15.236	10.336	23.805

Therefore, we divide the signal into 13 segments with 40,000 samples in each of them. The raw signal and the boxplots presenting the values in subsequent segments are presented in Figure 12. In the next step, we fit the sine function to each segment separately and compare the statistics determined for residual time series obtained after removing the matched functions. In Figure 13, one can see how the sample variance, IQR, and SACF² calculated for all segments change after subtracting consecutively first, second, third, and fourth sine function compared to the raw signal. It can be seen that all the statistics take much lower values, compared to the raw data, after removing the first and the second sine. Further operations do not make such a significant difference, therefore we decide to consider the signal cleaned by subtracting consecutively two sine functions. The signal obtained after removing the deterministic part is presented in Figure 14. One can notice that for the cleaned signal, the amplitude decreases and the impulses present in the data are more visible.

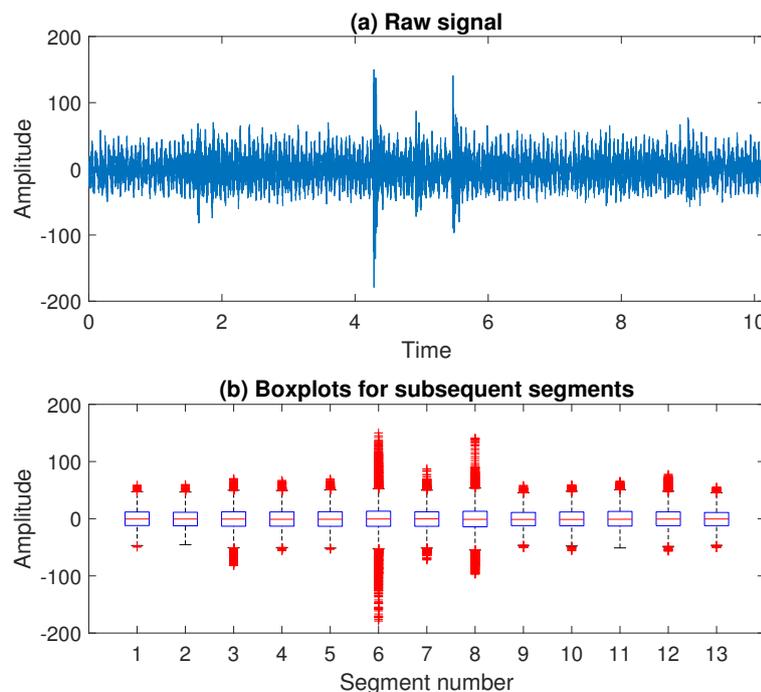


Figure 12. The raw signal and the boxplots presenting the values in the subsequent segments.

5.2. Detecting the Number of Impulses

Using the method presented in Section 3.2 we determine the number of impulses in the signal based on the values taken by the moving maximum, see Figure 15. As the threshold indicating the presence of an impulse, we choose the quantile of order 0.875 of the calculated statistics. One can observe that the method leads to detecting six impulses in the analyzed signal. In the following two subsections, we describe in detail how to use the proposed methodology to determine the start and the end for one of the impulses visible in the data. The signal with all six detected impulses marked in their locations will be presented at the end of the section.

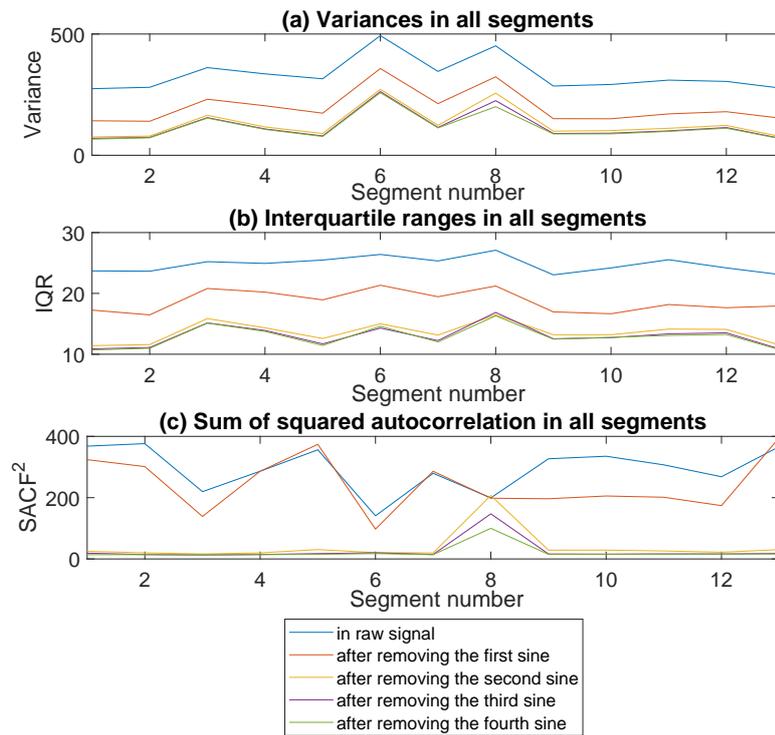


Figure 13. Sample variance, interquartile range and sum of squared auto-correlation calculated for the subsequent segments for the raw signal and the signals obtained after removing the sine functions.

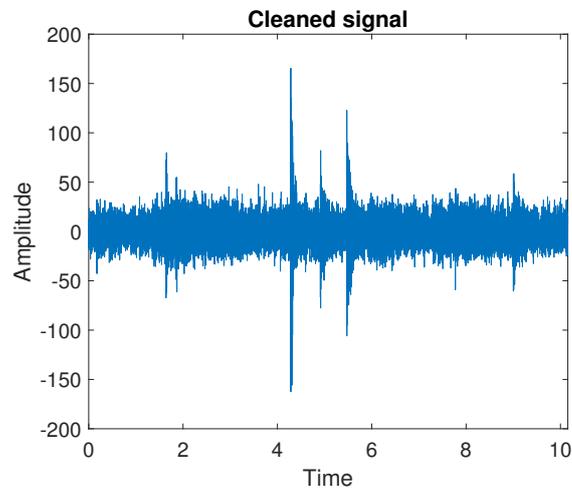


Figure 14. The signal obtained after removing the deterministic part.

5.3. Impulse Start Recognition for an Exemplary Impulse

Let us consider the third impulse detected in the signal on the previous stage of the analysis, namely, the one with the peak located between $t = 4$ and $t = 4.5$. To find its start we follow the procedure presented in Section 3.3. The consecutive steps are shown in Figure 16, where we present a part of the real signal—panel (a), the values taken by the moving range statistic—panel (b), and the corresponding empirical probability density function—panel (c). Finally, on panel (d) one can see the moment when the impulse starts marked with a red dashed line. As mentioned above, the same procedure is used to detect the beginning of all five remaining impulses present in the signal.

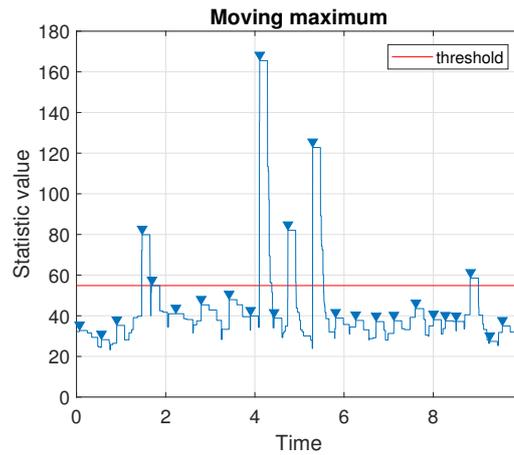


Figure 15. Values taken by the moving maximum calculated over a sliding window of length 9000. Triangle markers indicate the local peaks and the threshold is chosen as a quantile of order 0.875.

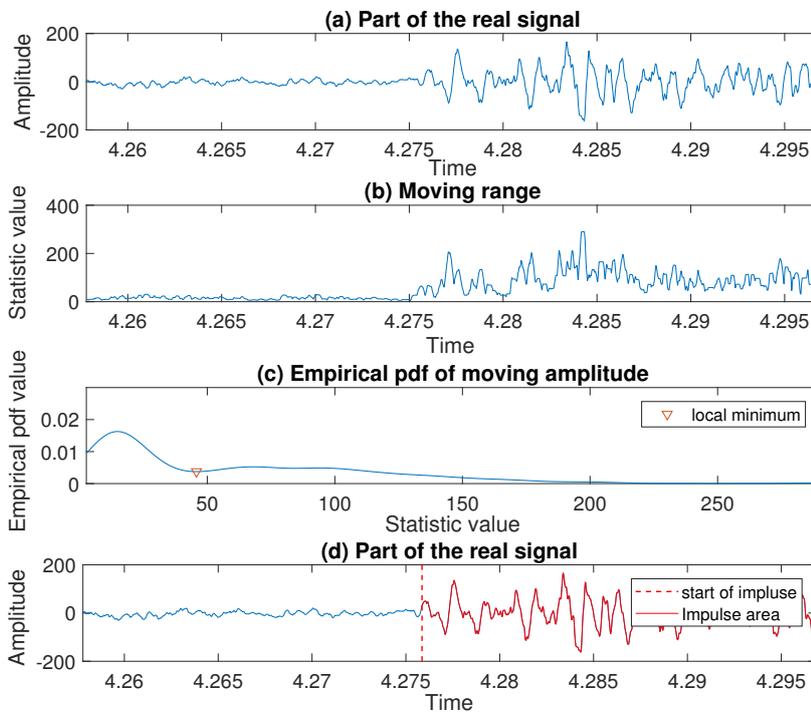


Figure 16. The consecutive steps of the procedure leading to the start recognition for an exemplary impulse in the real signal.

5.4. Impulse End Recognition for an Exemplary Impulse

For the impulse starting at the moment marked in Figure 16, we detect the end by using the methodology proposed in Section 3.4. All required steps are presented in Figure 17. Panel (a) shows the calculated smoothed envelope of the signal and the threshold chosen as the standard variation of part of the data where there are no impulses present (that is, the first 1.5 s of the signal). On panels (b) and (c), one can see the exponential function fitted to the signal envelope and its extension together with the end of the impulse marked with a red dashed line, respectively. The same methodology is applied to identify the end of all five remaining impulses present in the signal. The exact location of the third impulse considered here, taking into account its end and its beginning identified in the previous section, can be seen on the zoomed panel in Figure 18, which also shows the other five impulses detected in the signal.

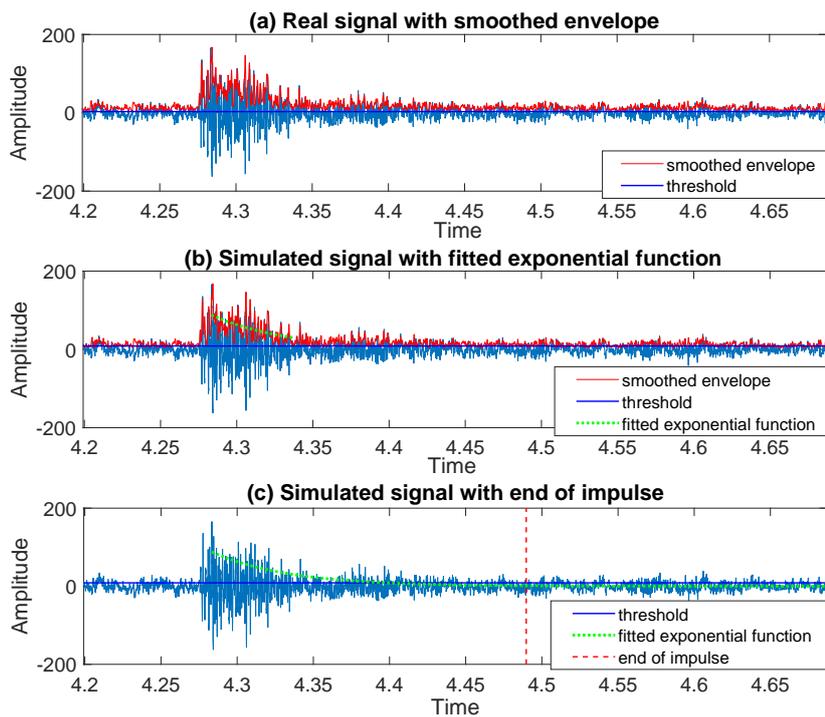


Figure 17. The consecutive steps of the procedure leading to the end recognition for an exemplary impulse in the real signal.

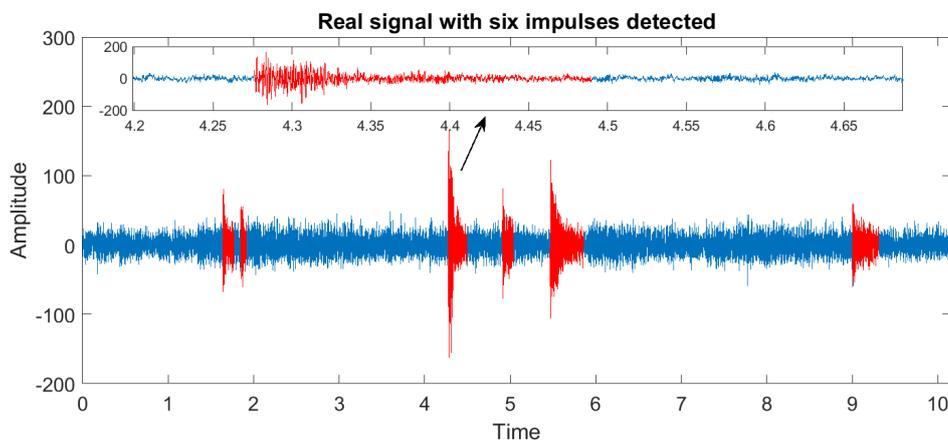


Figure 18. The location of all six impulses in the real signal.

6. Summary and Discussion

In this paper, a novel procedure for the identification, decomposition, and segmentation of impulsive vibration signals with complicated, time-varying structure has been proposed. The developed algorithm consists of a few steps that allow for effective deterministic component removal and impulses recognition.

The proposed methodology first is based on the identification and removal of several deterministic components (the sine wave functions). Due to AM/FM modulation (the amplitudes and frequencies of the sine wave function change over time), this step is based on segmentation and the identification of the deterministic components separately for each segment of the signal. That is a crucial step in the further analysis. Segment size, as well as the number of iterations, is derived experimentally, statistical-based criteria have been proposed here.

This step could be sensitive to large AM/FM modulations, which may happen for seriously varying speed (machine start/stop procedures). For such a very non-stationary regime, the size of the segment will be very short and it may result in poor approximation (fitting). However, it should be noted that from practice we know that the machine is in operation continuously, so the number of start/stop procedures is not significant.

After the deterministic component removal, we have proposed a new algorithm of impulse identification. Separately we detect the beginning of the impulse and its end as well as the peak amplitude of each impulse. The efficiency of this step will decrease for impulses with smaller amplitudes. In this algorithm, there is a need for several samples that protrude from the signal to build a model of envelope (to fit the exponential function). It is obvious that for weak impulses, even after deterministic component removal, several samples will be smaller so the precision of fitting will be worse. However, it should be noted that in the real environment we consider the signals with large amplitudes impulses (even 20–30 times larger than the averaged level of the noise). For small amplitudes impulses, there is no reason to use our procedure.

The proposed algorithm, which consists of a few important steps, is used for the very complicated signal coming from the sieving screen. Our idea was to propose a relatively intuitive method that can be used in real applications. Here we proposed a step-by-step procedure which effectiveness is demonstrated for the simulated signals. Finally, we applied this methodology to the signal from the machine working in a real environment.

Although the proposed algorithm seems to be sufficient in the considered case, there are additional limitations. The procedure is dedicated to the signals related to the machines with quasi-stationary work. In another case, it will be difficult to identify the optimal segments' length in the first step of the procedure and to fit the optimal sine wave functions. The second case that may cause the proposed method to be ineffective, is related to the impulses overlapping. When one impulse starts before the end of the previous one, the algorithm can return the wrong structure break point (i.e., the beginning of the new impulse and its end). Last but not least is the issue related to the number of impulses. It is obvious that a large number of impulses will extend the operation time of the entire procedure.

Author Contributions: Conceptualization, R.Z. and A.W.; methodology, R.Z., A.W., K.G., H.U. and A.G.; software, K.G., H.U. and A.G.; validation, R.Z.; investigation, K.G., H.U. and A.G.; writing—original draft preparation, A.G., R.Z. and A.W.; writing—review and editing, A.G., R.Z. and A.W.; visualization, A.G., R.Z. and A.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by European Institute of Innovation and Technology (EIT), a body of the European Union, under the Horizon 2020, the EU Framework Programme for Research and Innovation. This work is supported by EIT RawMaterials GmbH under Framework Partnership Agreement No. 18253 (OPMO—Operational Monitoring of Mineral Crushing Machinery).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Bartelmus, W.; Chaari, F.; Zimroz, R.; Haddar, M. Modelling of gearbox dynamics under time-varying nonstationary load for distributed fault detection and diagnosis. *Eur. J. Mech. A/Solids* **2010**, *29*, 637–646.
2. Bartelmus, W.; Zimroz, R. A new feature for monitoring the condition of gearboxes in non-stationary operating conditions. *Mech. Syst. Signal Process.* **2009**, *23*, 1528–1534.
3. Yang, Y.; Zhang, W.; Peng, Z.; Meng, G. Multicomponent Signal Analysis Based on Polynomial Chirplet Transform. *IEEE Trans. Ind. Electron.* **2013**, *60*, 3948–3956.
4. Wang, K.; Heyns, P. Application of computed order tracking, VoldKalman filtering and EMD in rotating machine vibration. *Mech. Syst. Signal Process.* **2011**, *25*, 416–430.
5. Feng, Z.; Qin, S.; Liang, M. Time-frequency analysis based on Vold-Kalman filter and higher order energy separation for fault diagnosis of wind turbine planetary gearbox under nonstationary conditions. *Renew. Energy* **2016**, *85*, 45–56.

6. Antoni, J.; Bonnardot, F.; Raad, A.; El Badaoui, M. Cyclostationary modelling of rotating machine vibration signals. *Mech. Syst. Signal Process.* **2004**, *18*, 1285–1314.
7. Randall, R.B.; Antoni, J. Rolling element bearing diagnostics—A tutorial. *Mech. Syst. Signal Process.* **2011**, *25*, 485–520.
8. Antoni, J. Cyclic spectral analysis of rolling-element bearing signals: Facts and fictions. *J. Sound Vib.* **2007**, *304*, 497–529. [[CrossRef](#)]
9. Randall, R.B.; Antoni, J.; Chobsaard, S. The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals. *Mech. Syst. Signal Process.* **2001**, *15*, 945–962.
10. Brockwell, P.J.; Davis, R.A. *Introduction to Time Series and Forecasting*; Springer: New York, NY, USA, 2002.
11. Cai, Z.; Xu, Y.; Duan, Z. An alternative demodulation method using envelope-derivative operator for bearing fault diagnosis of the vibrating screen. *J. Vib. Control* **2018**, *24*, 3249–3261.
12. Krot, P.; Zimroz, R.; Michalak, A.; Wodecki, J.; Ogonowski, S.; Drozda, M.; Jach, M. Development and Verification of the Diagnostic Model of the Sieving Screen. *Shock Vib.* **2020**, *2020*, 8015465.
13. Mauricio, A.; Qi, J.; Smith, W.; Sarazin, M.; Randall, R.; Janssens, K.; Gryllias, K. Bearing diagnostics under strong electromagnetic interference based on Integrated Spectral Coherence. *Mech. Syst. Signal Process.* **2020**, *140*, 106673. [[CrossRef](#)]
14. Cocconcelli, M.; Bassi, L.; Secchi, C.; Fantuzzi, C.; Rubini, R. An algorithm to diagnose ball bearing faults in servomotors running arbitrary motion profiles. *Mech. Syst. Signal Process.* **2012**, *27*, 667–682. [[CrossRef](#)]
15. Hebda-Sobkowicz, J.; Zimroz, R.; Wylomańska, A. Selection of the Informative Frequency Band in a Bearing Fault Diagnosis in the Presence of Non-Gaussian Noise—Comparison of Recently Developed Methods. *Appl. Sci.* **2020**, *10*, 2657. [[CrossRef](#)]
16. Hebda-Sobkowicz, J.; Zimroz, R.; Pitera, M.; Wylomanska, A. Informative frequency band selection in the presence of non-Gaussian noise—A novel approach based on the conditional variance statistic with application to bearing fault diagnosis. *Mech. Syst. Signal Process.* **2020**, *145*, 106971. [[CrossRef](#)]
17. Kruczek, P.; Zimroz, R.; Wylomanska, A. How to detect the cyclostationarity in heavy-tailed distributed signals. *Signal Process.* **2020**, *172*, 107514. [[CrossRef](#)]
18. Schmidt, S.; Zimroz, R.; Chaari, F.; Heyns, P.; Haddar, M. A simple condition monitoring method for gearboxes operating in impulsive environments. *Sensors* **2020**, *20*, 2115. [[CrossRef](#)]
19. Wodecki, J.; Michalak, A.; Zimroz, R.; Barszcz, T.; Wylomańska, A. Impulsive source separation using combination of Nonnegative Matrix Factorization of bi-frequency map, spatial denoising and Monte Carlo simulation. *Mech. Syst. Signal Process.* **2019**, *127*, 89–101. [[CrossRef](#)]
20. Szczerbowski, Z. Toward the Reliability of Geodetic Surveys in Study of Geodynamics—A Problem of Influence of Seasonal Variations. *Acta Geodyn. Geomater.* **2009**, *6*, 253–263.
21. Kruczek, P.; Wylomańska, A.; Teuerle, M.; Gajda, J. The modified Yule-Walker method for α -stable time series models. *Phys. A Stat. Mech. Its Appl.* **2017**, *469*, 588–603. [[CrossRef](#)]
22. Lee, M.; You, Y.; Kim, S.; Kim, K.; Kim, H. Decomposition of Water Level Time Series of a Tidal River into Tide, Wave and Rainfall-Runoff Components. *Water* **2018**, *10*, 1568. [[CrossRef](#)]
23. Tibshirani, R. Adaptive piecewise polynomial estimation via trend filtering. *Ann. Stat.* **2014**, *42*, 285–323. [[CrossRef](#)]
24. Jabloun, M.; Leonard, F.; Vieira, M.; Martin, N. A new flexible approach to estimate the IA and IF of nonstationary signals of long-time duration. *IEEE Trans. Signal Process.* **2007**, *55*, 3633–3644. [[CrossRef](#)]
25. Francos, A.; Porat, M. Analysis and synthesis of multicomponent signals using positive time-frequency distributions. *IEEE Trans. Signal Process.* **1999**, *47*, 493–504. [[CrossRef](#)]
26. Zhou, G. On polynomial phase signals with time-varying amplitudes. *IEEE Trans. Signal Process.* **1996**, *44*, 848–861. [[CrossRef](#)]
27. Li, Z.; Martin, N.; Vieira, M.; Gueguen, P. Parameter estimation of short-time multi-component signals using damped-amplitude & polynomial-frequency model. In Proceedings of the 2011 IEEE Statistical Signal Processing Workshop (SSP), Nice, France, 28–30 June 2011; pp. 545–548.
28. Gerber, T.; Martin, N.; Mailhes, C. Time-Frequency Tracking of Spectral Structures Estimated by a Data-Driven Method. *IEEE Trans. Ind. Electron.* **2015**, *62*, 6616–6626. [[CrossRef](#)]
29. Li, Z.; Martin, N. A Time-Frequency Based Method for the Detection and Tracking of Multiple Non-Linearly Modulated Components with Births and Deaths. *IEEE Trans. Signal Process.* **2016**, *64*, 1132–1146. [[CrossRef](#)]

30. Barkat, B.; Boashash, B. Instantaneous frequency estimation of polynomial FM signals using the peak of the PWVD: Statistical performance in the presence of additive Gaussian noise. *IEEE Trans. Signal Process.* **1999**, *47*, 2480–2490. [[CrossRef](#)]
31. Peng, Z.; Meng, G.; Chu, F.; Lang, Z.; Zhang, W.; Yang, Y. Polynomial chirplet transform with application to instantaneous frequency estimation. *IEEE Trans. Instrum. Meas.* **2011**, *60*, 3222–3229. [[CrossRef](#)]
32. Zimroz, R.; Urbanek, J.; Barszcz, T.; Bartelmus, W.; Millioz, F.; Martin, N. Measurement of instantaneous shaft speed by advanced vibration signal processing—Application to wind turbine gearbox. *Metrol. Meas. Syst.* **2011**, *18*, 701–712. [[CrossRef](#)]
33. Abboud, D.; Antoni, J.; Sieg-Zieba, S.; Eltabach, M. Envelope analysis of rotating machine vibrations in variable speed conditions: A comprehensive treatment. *Mech. Syst. Signal Process.* **2017**, *84*, 200–226.
34. Stack, J.; Harley, R.; Habetler, T. An amplitude modulation detector for fault diagnosis in rolling element bearings. *IEEE Trans. Ind. Electron.* **2004**, *51*, 1097–1102.
35. Bourdon, A.; Chesné, S.; André, H.; Rémond, D. Reconstruction of angular speed variations in the angular domain to diagnose and quantify taper roller bearing outer race fault. *Mech. Syst. Signal Process.* **2019**, *120*, 1–15.
36. Staszewski, W.; Tomlinson, G. Local tooth fault detection in gearboxes using a moving window procedure. *Mech. Syst. Signal Process.* **1997**, *11*, 331–350.
37. Lopatka, M.; Laplanche, C.; Adam, O.; Motsch, J.; Zarzycki, J. Non-Stationary Time-Series Segmentation Based on the Schur Prediction Error Analysis. In Proceedings of the IEEE/SP 13th Workshop on Statistical Signal Processing, Bordeaux, France, 17–20 July 2005; pp. 251–256.
38. Tóth, B.; Lillo, F.; Farmer, J. Segmentation algorithm for non-stationary compound Poisson processes: With an application to inventory time series of market members in a financial market. *Eur. Phys. J. B* **2010**, *78*, 235–243.
39. Hebda-Sobkowicz, J.; Gola, S.; Zimroz, R.; Wyłomańska, A. Identification and statistical analysis of impulse-like patterns of carbon monoxide variation in deep underground mines associated with the blasting procedure. *Sensors* **2019**, *19*, 2757.
40. Andreao, R.V.; Dorizzi, B.; Boudy, J. ECG signal analysis through hidden Markov models. *IEEE Trans. Biomed. Eng.* **2006**, *53*, 1541–1549. [[PubMed](#)]
41. Khanagha, V.; Daoudi, K.; Pont, O.; Yahia, H. Phonetic segmentation of speech signal using local singularity analysis. *Digit. Signal Process. A Rev. J.* **2014**, *35*, 86–94.
42. Sánchez-Pérez, L.; Sánchez-Fernández, L.; Suárez-Guerra, S.; Carbajal-Hernández, J. Aircraft class identification based on take-off noise signal segmentation in time. *Expert Syst. Appl.* **2013**, *40*, 5148–5159.
43. Popescu, T. Signal segmentation using changing regression models with application in seismic engineering. *Digit. Signal Process. A Rev. J.* **2014**, *24*, 14–26.
44. Wyłomańska, A.; Zimroz, R. Signal segmentation for operational regimes detection of heavy duty mining mobile machines—a statistical approach. *Diagnostyka* **2014**, *15*, 33–42.
45. Hossa, R.; Makowski, R.; Zimroz, R. Automatic segmentation of seismic signal with support of innovative filtering. *Int. J. Rock Mech. Min. Sci.* **2017**, *91*, 29–39. [[CrossRef](#)]
46. Gajda, J.; Sikora, G.; Wyłomańska, A. Regime Variance Testing—A Quantile Approach. *Acta Phys. Pol. B Proc. Suppl.* **2013**, *44*, 1015–1035. [[CrossRef](#)]
47. Hill, P.D. Kernel estimation of a distribution function. *Commun. Stat. Methods* **1985**, *14*, 605–620.
48. Bowman, A.W.; Azzalini, A. *Applied Smoothing Techniques for Data Analysis*; Oxford University Press Inc.: New York, NY, USA, 1997.
49. Silverman, B. *Density Estimation: For Statistics and Data Analysis*; Chapman & Hall: London, UK, 1986.
50. Oppenheim, A.V.; Schaffer, R.W.; Buck, J.R. *Discrete-Time Signal Processing*, 2nd ed.; Prentice Hall: Upper Saddle River, NJ, USA, 1999.

