Article

Application of UAV in Search and Rescue Actions in Underground Mine—A Specific Sound Detection in Noisy Acoustic Signal

Paweł Zimroz [1,*] , **Paweł Trybała** [1] , **Adam Wróblewski** [1] , **Mateusz Góralczyk** [1] , **Jarosław Szrek** [2] , **Agnieszka Wójcik** [1] and **Radosław Zimroz** [1]

[1] Department of Mining, Faculty of Geoengineering, Mining and Geology, Wrocław University of Science and Technology, 50-421 Wroclaw, Poland; pawel.trybala@pwr.edu.pl (P.T.); adam.wroblewski@pwr.edu.pl (A.W.); mateusz.goralczyk@pwr.edu.pl (M.G.); a.wojcik@pwr.edu.pl (A.W.); radoslaw.zimroz@pwr.edu.pl (R.Z.)

[2] Department of Fundamentals of Machine Design and Mechatronic Systems, Faculty of Mechanical Engineering, Wrocław University of Science and Technology, 50-371 Wroclaw, Poland; jaroslaw.szrek@pwr.edu.pl

* Correspondence: pawel.zimroz@pwr.edu.pl

Abstract: The possibility of the application of an unmanned aerial vehicle (UAV) in search and rescue activities in a deep underground mine has been investigated. In the presented case study, a UAV is searching for a lost or injured human who is able to call for help but is not able to move or use any communication device. A UAV capturing acoustic data while flying through underground corridors is used. The acoustic signal is very noisy since during the flight the UAV contributes high-energetic emission. The main goal of the paper is to present an automatic signal processing procedure for detection of a specific sound (supposed to contain voice activity) in presence of heavy, time-varying noise from UAV. The proposed acoustic signal processing technique is based on time-frequency representation and Euclidean distance measurement between reference spectrum (UAV noise only) and captured data. As both the UAV and "injured" person were equipped with synchronized microphones during the experiment, validation has been performed. Two experiments carried out in lab conditions, as well as one in an underground mine, provided very satisfactory results.

Keywords: UAV; acoustic signal processing; Euclidean distance; underground mine; search and rescue

1. Introduction

The biggest underground mines look like underground cities. They consist of many "streets" and "street crossings". Unfortunately, each mining void looks very similar to the other. Due to the complexity of the underground corridors network, localization of an injured person or any other incident, if not reported on an emergency line, is complicated and time-consuming. It imposes a serious danger in the cases, in which rapid contact with works supervisors or rescue units is crucial. The complex structure of an underground mine makes getting lost in it much more complicated to solve and serious than in the case of an accident on the ground level. Another important issue is related to extremely harsh environmental conditions. An example of an emergency situation caused by them can be given, like one of the co-workers in a pair or small group fainting. In such a case, the presence of a drone sent to the area in which worsened environmental parameters have been reported can significantly shorten the time needed to reach the victim and others at risk.

There are many natural hazards or unfavorable environmental factors, that can impose a significant risk to any humans performing work underground, such as spontaneous combustion of coal, rock bursts, methane ejections, water hazard, high dustiness [1–4]. According to the State Mining Authority, only in 2019 did Polish mine rescue units take part in 32 rescue operations in mining plants [5]. Most of them were related to roof/wall

collapse or fire occurrence. In case of such an extreme situation in the conditions of an underground mine, it is not difficult to imagine a situation in which it is necessary to search for workers. As a result of the risks present in the mining environment, such as lack of oxygen, presence of fire fumes or other harmful gases, leading to loss of consciousness, as well as other injuries, the miner may not be able to evacuate from the hazardous area. Furthermore, a person may get trapped due to a rock collapse or a fire occurring in the evacuation route. The scale of the problem can be seen in the example of the activity of Polish mining rescue units. Considering that mining rescue activities are not limited to Polish underground mines, it can be stated that on a global scale the issue of human location in emergency situations is a considerable challenge. As the area to search is large, consists of many corridors and time is a critical issue, there is a need for a quick method of finding victims of accidents.

In the paper, it is proposed to use an unmanned air vehicle (UAV) for audio signals acquisition for human voice detection. In the article, we propose a drone equipped with a microphone, capable of registering human voice and distinguishing it from the propellers' noise, to be used for detecting emergency situations in underground mines. A method using time-frequency decomposition of the signal is presented, together with the full procedure of separating the human voice and background noise. It is assumed that the person is not able to move, but can call for help by shouting. Practically, the mission may be performed using several drones searching corridor by corridor and using acoustic signals to find the location of the human.

From a signal processing perspective, the problem is related to voice detection in presence of time-varying noise with complex structure. Due to the spinning of the propellers, some deterministic and random components are present in the acoustic signal. Unfortunately, the energy of the voice signal captured by a microphone mounted below the drone is rather small and the high energy of the acoustic signal emitted by the drone makes the voice detection problem complicated. In the paper, a procedure for voice detection in presence of UAV noise is proposed. Our informative signal is in fact a human voice, but limited to specific words only like "help". So, we called it specific acoustic signal. We limit our research for "help" only, however we believe also other words might be detectable-but for sure not all. "Help" is the most intuitive in crisis situation. The experiments have been carried out in the lab as well as in the real mining corridor to test and validate the proposed procedure.

2. State of the Art

Voice activity detection (VAD), being a natural context for our sound detection method, is an inherent stage of any real-world application of speech processing. There have been numerous publications concerning various methods on this subject; however, formerly, emphasis was placed primarily on detecting voice messages between long periods of silence or low magnitude noise. If the background noise has lower energy than any material speech occurrence, then a simple thresholding criterion enables a recording or communication system to save energy, computational memory, and network bandwidth. Traditional VAD and noise estimation methods are used in VOIP (Voice over Internet Protocol), as well as various speech-based applications. Along with de-noising and feature extraction, it would be a stage of any speech recognition or voice control system [6].

More sophisticated methods were developed to deal with lower SNR (signal-to-noise ratio). Among more recent work on VAD methods, some were based on combining spectral analysis and statistical methods [7,8]. Closely related to them are the noise reduction techniques, which allow subtracting the noise to increase the intelligibility of the speech signal [9]. Extracting preliminary information from speech signals may include voiced/unvoiced speech classification [10]. Speech detection and de-noising were also considered in the context of the so-called cocktail party problem where audio sources (i.e., speech and noise) are separated through a binaural system consisting of two microphones [11].

The background noise has considerably high energy in the case of recording nearby UAV equipped with a multi-rotor propulsion system. It contains the sound of its engine and mechanical parts, such as the propellers, which cannot be damped in its natural operating mode. The methods of dealing with high-level background noise conditions are widely discussed in literature concerning acoustic measurement on machines with rotating elements such as turbochargers [12–14] among many other examples. The ego-noise of a small-size multi-rotor drone itself has been investigated by many researchers in recent years. It was thoroughly examined at static thrust [15]. Some recent research considers tracking change in noise characteristics under realistic flight conditions based on position and noise measurement of a moving drone using both camera and a multiple microphone system (which is called microphone array) [16].

There are some publications about de-noising methods concerning mainly a hovering drone. The azimuthal direction of sound arrival was analyzed (using both supervised and unsupervised methods) in [17,18] based on a microphone array mounted on a hovering drone. Though at static thrust drone was enabled to track moving sound sources. Analyzing the sound emitted by a UAV has recently acquired interest also in the detection of illegal drone activity [19], where the dynamics of a quadrotor drone was taken into account.

In our experiment, as the drone is moving from one point to another, the background noise is both high-energy and non-stationary. It originates from the variable mode of operating of a UAV and air response but also from air turbulences. We only use one microphone at a sound source, thus physical direction or distance cannot be tracked, but the distance is changing (in an approximately monotonous way) during our records.

When analyzing a non-stationary process, it is essential to perform some sort of time segmentation of the data [20]. Therefore, transformations preserving (to a certain extent) original time structure are very useful. One of the primary methods is representing a signal in the time-frequency domain via Short-Time Fourier Transform (STFT) [21].

To characterize the time intervals containing voice we adopt a method based on STFT which was discussed before in a context of cyclic impulse anomaly detection (local damage in driving units of a belt conveyor) based on calculating distance (metric) between spectral densities [22]. It was stated that informative frequency band selection may strongly affect the analysis based on the time-frequency map. To deal with it we perform both band-pass filtering and spectrogram normalization.

The application of robotics in mining industry and especially in missions related to rescue and search of the victims of an accident victims is well established and constantly developed since the beginning of the third-millennium [23–31]. The use of UAVs as a support for the search and rescue (SAR) operations has been presented among others in [32], where some significant considerations are included, such as energy limitations, the quality of sensors, and the exchange of information between UAVs and rescue teams. An example of a fixed-wing UAV-system capable of detecting accident victims employing real-time imaging, on-board processing, and assigning coordinates of the detected target can be found in [33]. Both-UAV's hardware and identification algorithm with a self-adjusting threshold were exhaustively described providing a valuable overview of the current technical tools valuable for SAR purposes in superficial conditions. A remote-controlled drone, presented by [34], is an example of how the technology of this type can not only prevent mining production obstructions but also increase safety on-site. The UAV used by the authors carried an analog thermal infrared video camera, digital video recorder, an RGB spectrum digital video camera, and a small battery power source. This solution based on commercially available devices allowed evaluation of the sub-surface temperatures in a flexible and low-cost way, nevertheless, the system was able to perform only short missions-lasting up to fifteen minutes due to the limitation of batteries weight. The character of sub-surface mining workplaces and accidents that take place there created a big demand for the introduction of urban SAR robotics in underground applications and their adaptation to solving similar problems in new-more demanding conditions. Together with maintenance-purpose robots, this technology will

allow the limitation of human involvement in the most hazardous areas, contributing to the transformation of the mining sector [27,35]. The vast majority of underground mining-related applications of aerial vehicles' proposed by researchers nowadays are aimed at mapping, localization of discontinuities, and general reconnaissance of sub-surface areas [36–40]. The aforementioned are undoubtedly valuable examples of how robotics can facilitate arduous and time-consuming processes, providing information about some available tools. However, recent development and increased accessibility of robotics in the raw materials sector is also aimed at increasing the safety of underground workers and providing quick and effective help in case of an accident. The detection of a human in an underground environment can be based on a video recording transmitted in real-time to an operator [41], automatic detection of humans with the use of image recognition algorithms [28], or, what we propose in this article—human voice detection. Three recent reviews of the usage of unmanned aerial vehicles [42–44] give an exhaustive insight into the spectrum of possible applications of them in both underground and open-cast mines, including rescue operations. Nonetheless, in most of the cases in which a UAV is considered to be applied in this area, the sensors in which it is equipped are the image-capturing ones. To the best authors' knowledge, the problem of voice detection and its extraction from the drone background noise have not been approached in the context described in this work (underground mining rescue action).

3. Theoretical Background of the Novel Procedure

We propose a procedure involving time-frequency decomposition of a signal. A time-frequency map (spectrogram) of a signal is generated using Short-Time Fourier Transform (STFT). The STFT of a vector of observations $x_0, \ldots, x_{N-1}$ is a function of discrete valued time and frequency. It is a DFT of a signal multiplied by a shifted window selecting the observations close to a given time point (equivalently in terms of filtering it is a discrete convolution of a frequency shifted signal with the window function) [45] which is given by the Equation

$$STFT(t, f) = \sum_{k=0}^{N-1} w_{t-k} x_k e^{-j2\pi fk/N}. \tag{1}$$

Both for further analysis and for plotting a spectrogram we will consider squared absolute value of the STFT and refer to it as a power spectral densities (PSD) matrix of size $T \times F$ containing values

$$PSD(t, f) = |STFT(t, f)|^2. \tag{2}$$

Its columns (time slices) we name power spectral densities and its rows (slices for given frequency) we name subsignals. There arises a problem of informative frequency band selection (see also [46–48]).

Hidden determinism associated with the motion of the drone propellers is present in the background noise. It was stated that such a character of data appears with high-energy low-frequency bands. To avoid the necessity of further investigating (see advanced filter design discussed among others by Wodecki [49]) and selecting the band of interest, we decided to use a data-driven method which proved very efficient in the context of cyclic anomaly detection. Reducing the impact of the hidden determinism involves sub-signals normalization. As a norm, we use the sum of the values of a sub-signal:

$$PSD_{\text{normalized}}(t, f) = PSD(t, f) / \sum_{u=0}^{T-1} PSD(u, f). \tag{3}$$

The main part of the presented method is calculating the measure of dissimilarity for every spectral density (time slice of the matrix) with a referential spectrum. Euclidean metric is used defined for power spectral densities $p(f)$ and $q(f)$ as follows:

$$D(p,q) = \sqrt{\sum_{f=0}^{F-1} \left(p(f) - q(f) \right)^2}. \tag{4}$$

The sound emitted by the drone is assumed to contain a notable noisy component. To increase SNR we do band-pass filtering in the frequency domain. According to the fact that the informative part of the human voice does not exceed 4 kHz, this value is used as a cutoff frequency. We also add a lower boundary at 0.2 kHz. Therefore, we modify the Equation (4) to obtain

$$D(p,q)_{\text{band-pass}} = \sqrt{\sum_{f=F_{\text{lower}}}^{F_{\text{upper}}} \left(p(f) - q(f) \right)^2}. \tag{5}$$

where $F_{\text{lower}} = 0.2$ kHz, $F_{\text{upper}} = 4$ kHz. For the sake of the visualization of results, we will also filter signals in the time domain.

As the referential, the mean power spectrum calculated from a reference interval (usually a few starting seconds depending on a particular sound signal) is assigned. It is assumed that in this time interval only the background noise containing the sound emitted by the drone is present. Thus for every time slice t on a spectrogram we obtain

$$D_t = D_{\text{band-pass}}\left(PSD_{\text{normalized}}(t,\cdot), PSD_{ref}(\cdot) \right). \tag{6}$$

where for a given reference interval I we have

$$PSD_{\text{ref}}(f) = \frac{1}{|I|} \sum_{t \in I} PSD_{\text{normalized}}(t,f). \tag{7}$$

To discriminate between the human voice and the drone noise we set a threshold on Euclidean distance. Beforehand, we apply moving mean smoothing (with parameter M) on the calculated metric:

$$\widetilde{D_t} = \frac{1}{M} \sum_{k=u-\lfloor \frac{M}{2} \rfloor}^{u+\lfloor \frac{M-1}{2} \rfloor} D_u. \tag{8}$$

As the threshold, the value of mode + standard error was chosen where the mode is read from the kernel smoothed empirical PDF (probability density function) as the argument of the maximum value. Values greater than the threshold are associated with the human voice therefore corresponding time interval is assigned to the voice-present cluster and marked with 1. Otherwise, a time interval is assigned to a voice-absent cluster and marked with 0. Final processing of the received output includes filling gaps (changing too short series of 0 to 1) and subsequent deleting immaterial voice (changing too short series of 1 to 0). It is accordingly necessary to beforehand define tolerance values Tol_0 and Tol_1 for both steps of this operation. Voice intervals merged by that means are marked with red color on the plots. At once, the detected intervals are exported to the output audio file. They are delimited with 1-second pause to distinguish between one another, that is, data that would be then sent by the drone to report emergency. The outline of our algorithm is as follows:

1. Calculating STFT of a signal and then its squared absolute value to obtain PSD(t,f) matrix;
2. Normalization of the subsignals (rows of the PSD matrix) in time domain and restriction of the frequency domain;
3. Defining of the referential spectrum;
4. Calculating Euclidean metric at every time interval of PSD matrix and using moving mean smoothing filter;
5. Setting a threshold on Euclidean metric;

6.   Assigning voice-present (1) and voice-absent clusters (0) to the original signal,
7.   Changing clusters in too short time intervals;
8.   Visualisation of the voice detection results on a band-pass filtered signal.

As a standard-setting, we use STFT with Kaiser window of order 5 and length 1024 which is also the number of FFT (Fast Fourier Transform) algorithm points. Other parameters which are needed to be set are the number of overlapping points in STFT, the start and end of the reference interval, the number of points in the moving mean filter, and tolerances (in seconds) for binary merging in final processing.

4. The Experiment

In the experiment, a manually controlled drone, DJI Mavic Mini was used. It is a small, commercially available quadrotor, weighing 249 g, using 4 pairs of light, 119 mm diameter propellers. This model was chosen for our case study as it represents well the wider class of micro aerial vehicles (MAV), which are, due to their small size and fairly good battery life, the most suitable type of unmanned vehicle to be used in an emergency situation in an underground mine. During a normal performance, its brushless motors are operating in the frequency range of 9000–15,000 RPM (150–250 Hz). Since the experiment was realized in indoor conditions, the propeller guard (provided by the UAV manufacturer) was mounted on the drone to avoid damage in case of the drone colliding with the walls.

To evaluate the possibility of the application of a microphone-equipped unmanned aerial vehicle in the detection of humans, two experiments have been performed. The first was a preliminary test performed in the corridors of a building (Figure 1), roughly approximating realistic conditions. Another took place in a slant of an inactive underground mine (Figure 2). In both experiments, a UAV with a microphone located underneath has been approaching a person laying in the side pit of the main corridor and calling for help. In the preliminary experiment, an additional microphone has been used, which was placed near the help-caller. The person was located behind the corner, loudly repeated the words 'HELP, HELP'. In the experiment performed in realistic conditions, a deep niche in the corridor has been selected as the 'victim's' location. The volume of help-calling, as well as the speed of the drone, were kept as constant as possible throughout the whole experiment given the technical and organizational capabilities. Some minor variations in both speed and volume of acoustic emission could be caused by a human factor, as well as variations in the speed of airflow in underground excavations.
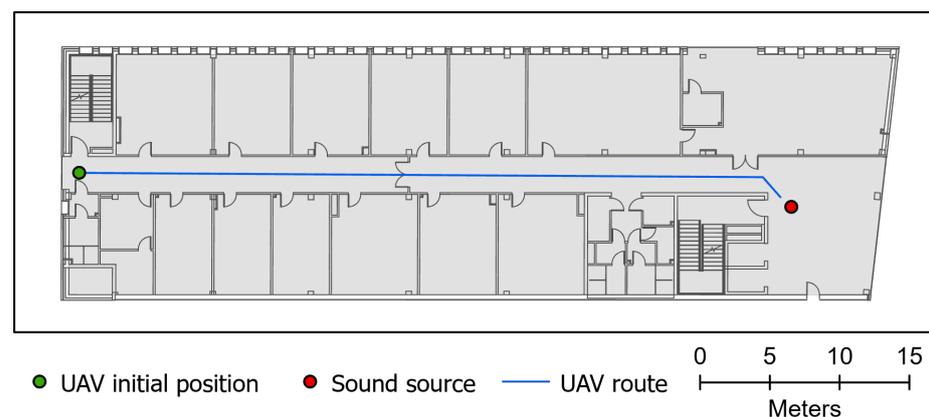

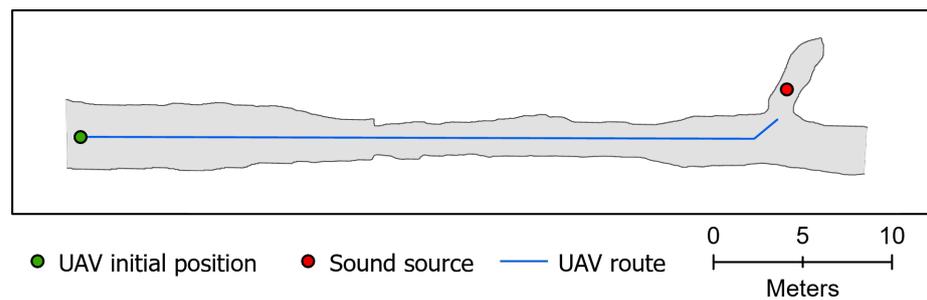**Figure 1.** Drone flight path during the first experiment.

Figure 2. Drone flight path during the second experiment.

5. Results

5.1. Overground Sounds

Data acquired during the first (overground) experiment consist of sounds recorded at the same time with two microphones located at different places: first, located nearby the UAV ('next-to-drone' record); second, nearby the person who was making voice calling for help ('next-to-victim' record). The length of the signals differ (they are 62 s and 66 s long respectively) and their real-time starting moments are shifted relative to each other. Both were recorded with a sampling frequency equal to 48 kHz. The procedure presented above was applied separately to each of the records. Time waveforms of both raw sounds are presented in Figure 3. As we can see, the first of them seems a noise-like signal while in the case of the second signal the time intervals containing human voice are clearly visible.

(a)	(b)
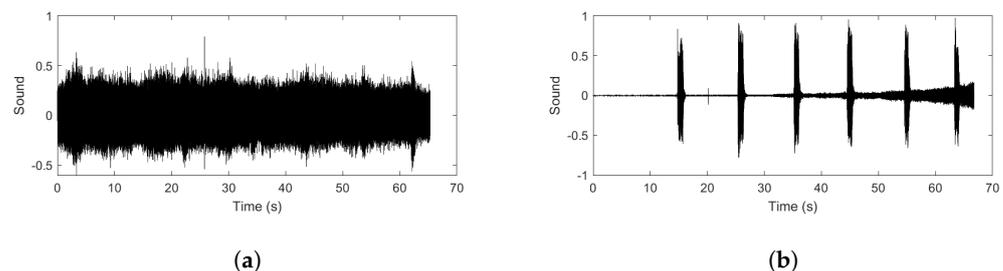
Figure 3. Raw signals of the sound recorded overground in the first realisation: (**a**) 'Next-to-drone' record, (**b**) 'Next-to-victim' record.

STFT was calculated with 80% overlapping. The resulting time-bin length is 4.3 milliseconds. Spectrograms of the raw signals which are presented in Figure 4 are records from both microphones. They illustrate the time-frequency decomposition of the sounds in terms of the log-scale PSD.

As it was stated the analysis of the data is based on the normalized spectrogram. This type of spectrogram is showed in Figure 5. It should be noted that frequencies in this case are restricted to the band 200–4000 Hz because, according to band-pass filtering of the data, we cut PSD to 0 outside of this band. In the case of the 'next-to-drone' record, we can distinguish particular moments in the evolution of the power spectrum only after normalization, the changes are visible mostly in a frequency bin located around 550–800 Hz. In the spectrograms based on the 'next-to-victim' recordings, on the other hand, the changes are by far more evident in numerous bands and valuable as the noise arises only in the ending part.

We visualize band-pass filtering in the time domain in Figure 6. In the case of the 'next-to-drone' record, this processing step makes the shape of the waveform more informative compared with the waveform of the raw signal presented before.
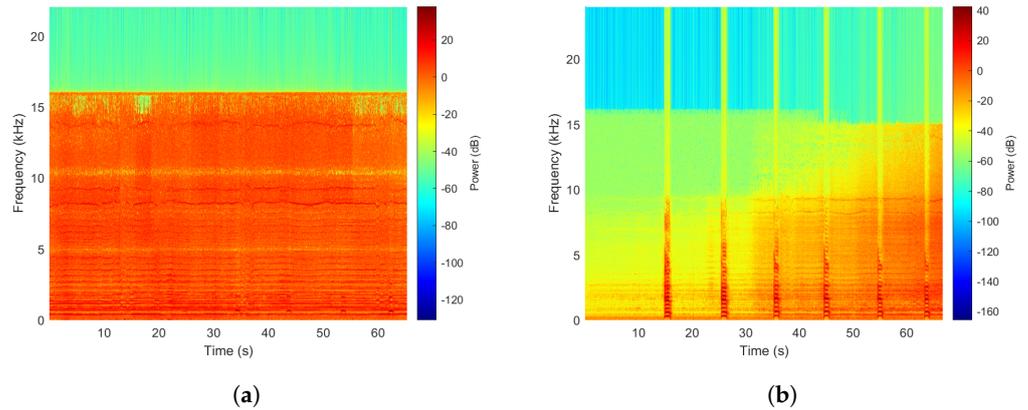
Figure 4. Spectrograms of the raw signal from Figure 3: (**a**) 'Next-to-drone' record, (**b**) 'Next-to-victim' record.
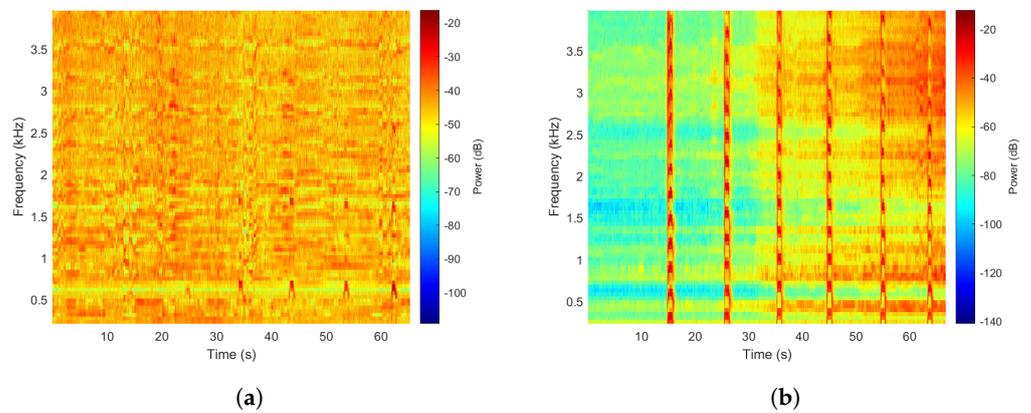

Figure 5. Normalized and band-pass filtered spectrograms of the signals from Figure 3: (**a**) 'Next-to-drone' record, (**b**) 'Next-to-victim' record.
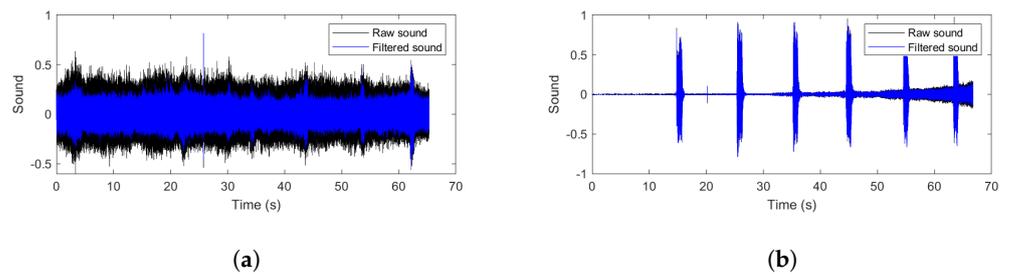

Figure 6. Band-pass filtering of the sounds from Figure 3: (**a**) 'Next-to-drone' record, (**b**) 'Next-to-victim' record.

The referential power spectral density was defined as a mean from 5 first seconds. Euclidean metric as a function of time is showed in Figure 7. The Euclidean metric smoothed with 20 samples moving mean is highlighted in green. The red line stands for the threshold which we obtain as mode + standard error value from the empirical distribution of our metric on the whole time domain. The empirical density is presented in Figure 8.

In Figure 9 the detected intervals of the found voice are highlighted as a part of the plot of the band-pass filtered signal. Tolerances Tol_0 and Tol_1 in final processing were both set to 0.25 s. The output comprises six voice-present regions selected from the 'next-to-drone' record which correspond to the analogous segments present in the 'next-to-victim' record. The last four of them can be heard in the original raw 'next-to-drone' record.

(**a**)

(**b**)
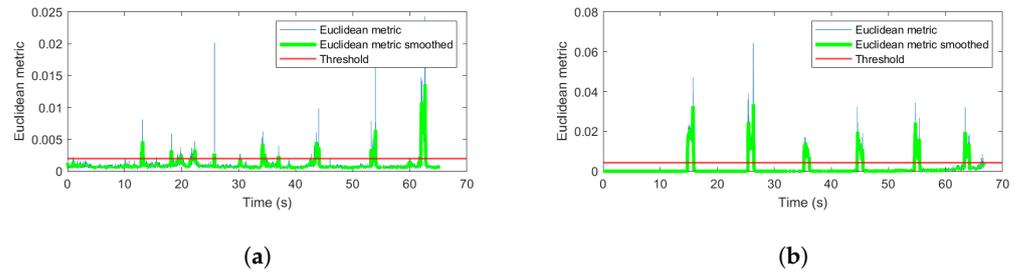
Figure 7. Smoothed Euclidean metric calculated from normalized spectrograms from Figure 5 as a function of time: (**a**) 'Next-to-drone' record, (**b**) 'Next-to-victim' record.
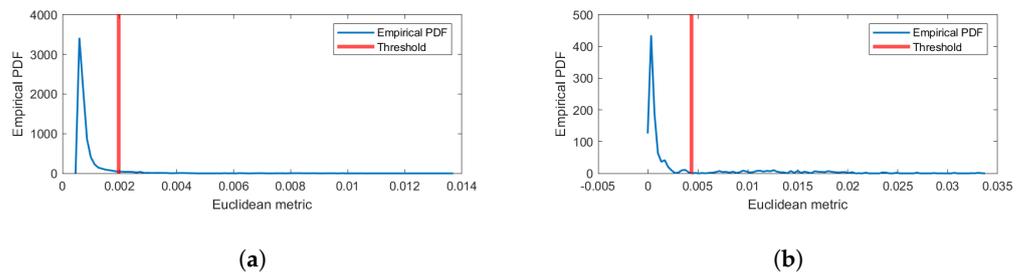

(**a**)

(**b**)

Figure 8. Empirical PDF of Smoothed Euclidean metric from Figure 7: (**a**) 'Next-to-drone' record, (**b**) 'Next-to-victim' record.

(**a**)

(**b**)
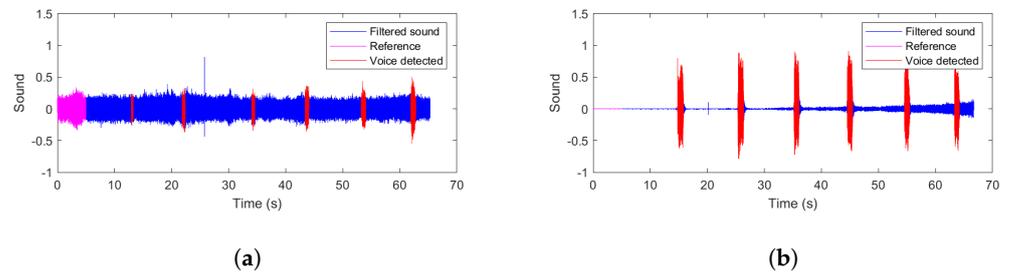
Figure 9. Detected voice time intervals extracted from filtered signals from Figure 6: (**a**) 'Next-to-drone' record, (**b**) 'Next-to-victim' record.

5.2. Underground Sounds

Data acquired during the second (underground) experiment include only sounds recorded nearby the drone with the sampling frequency equal to 25.6 kHz. From several sounds recorded, two were chosen for the analysis.

First, we analyze a short signal with a frequent and clearly audible human voice. The signal is 34 s long and we present it in Figure 10. Its spectrogram is presented in Figure 11. Spectral structure is dominated by the noise from the drone; showever, some voice related signature is detectable so it could be a basis for automatic signal processing.
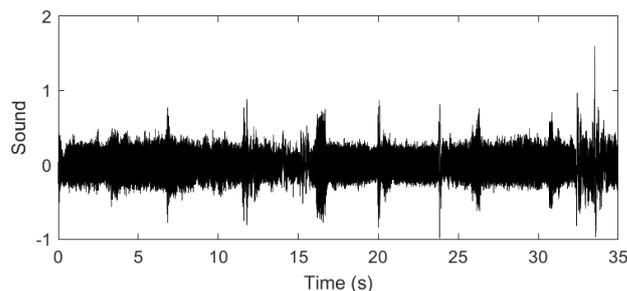

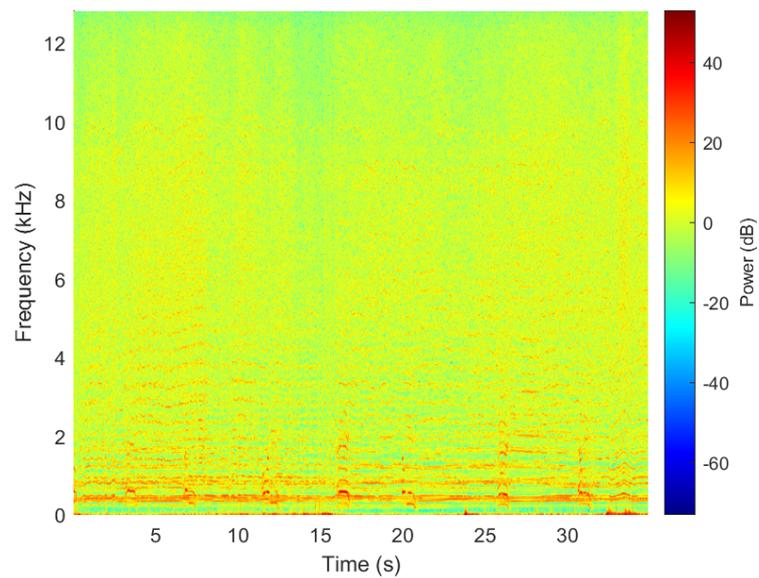**Figure 10.** Raw signal of the first sound recorded underground.

Figure 11. Spectrogram of the raw signal from Figure 10.

Since the voice regions are frequent in this realization of the experiment, this time we used lower overlapping equal to 50% so as not to blur the results. The time moments at which the voice was heard can be seen at the spectrogram, even more clearly after filtering and normalization (Figure 12).
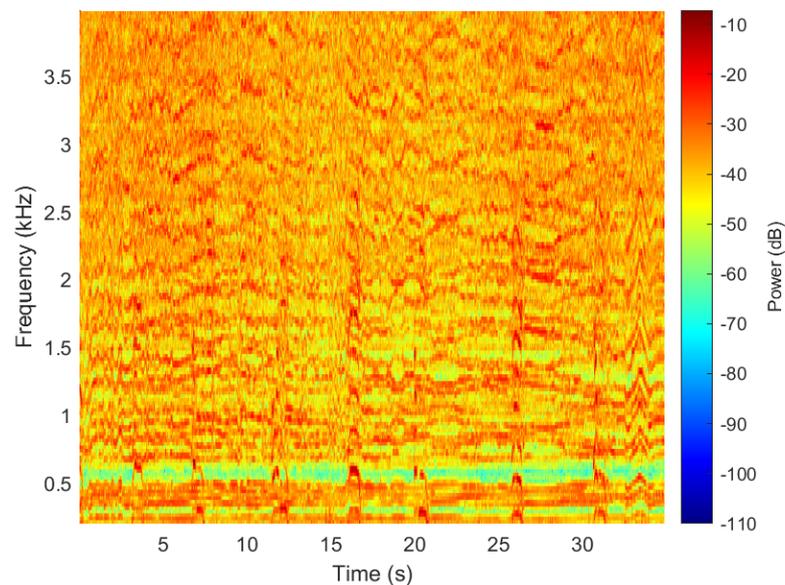

Figure 12. Normalized and band-pass filtered spectrogram of the signal from Figure 10.

A particularly nonstationary behavior of the drone alone in the ending time interval (which is also audible from the record) was successfully removed due to band-pass filtering as it can be noted at the plot visualizing band-pass filtering of the raw sound (Figure 13).

To determine the time intervals which correspond to the rescued person calling we calculate Euclidean metric as before. This length of the time bin is 20 ms resulting from the parameters of the STFT computation. The referential spectrum is calculated as a mean PSD from only first 2 seconds (so as to catch an interval without voice). Calculated metric and the determined threshold is presented in the time domain and on a kernel-smoothed density (Figures 14 and 15). Moving mean smoothing parameter was again set to 20. In both figures the chosen threshold equal to mode + standard error is shown. The detected

voice time intervals (Figure 16) are fully consistent with all seven calls for help, which are audible from the raw sound (using Tol_0 and Tol_1 equal to 0.5 s).
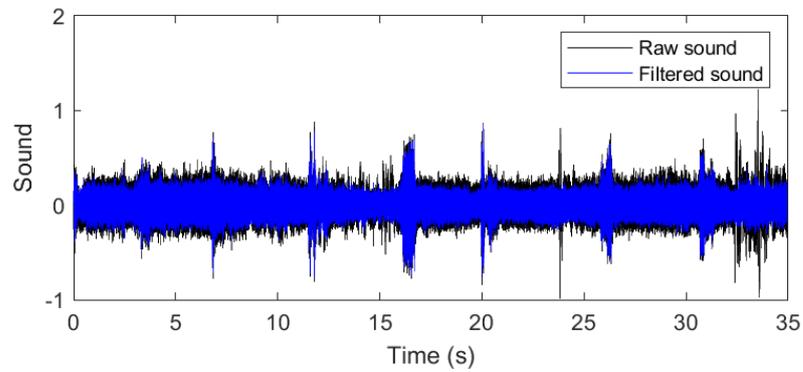

Figure 13. A band-pass filtering of the sound from Figure 10.
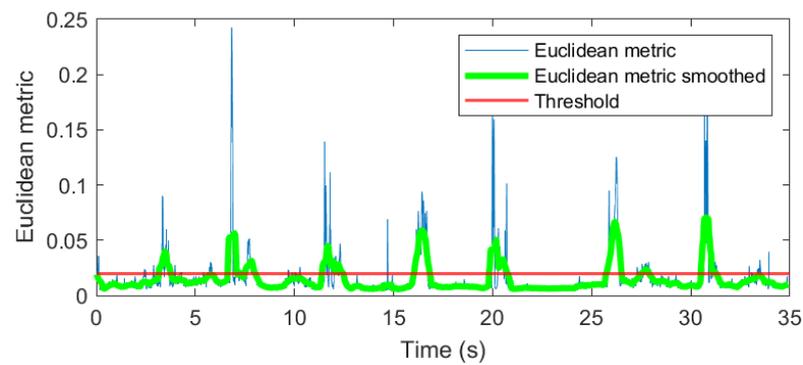

Figure 14. Smoothed Euclidean metric calculated from normalized spectrogram from Figure 12 as a function of time.
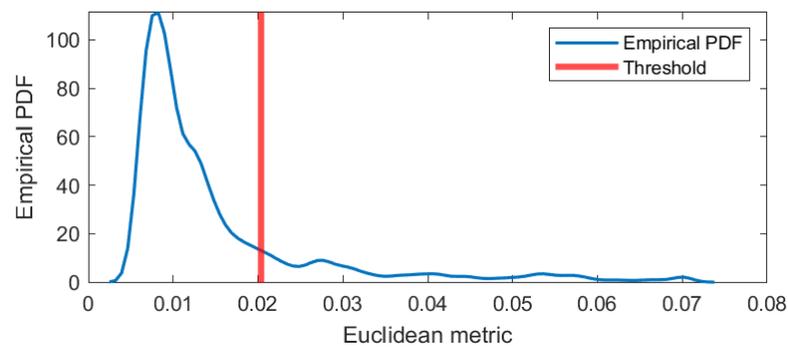

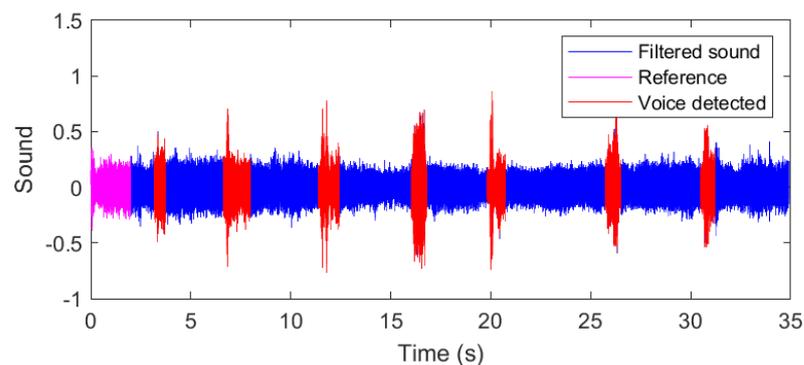**Figure 15.** Empirical PDF of Smoothed Euclidean metric from Figure 14.

Figure 16. Detected voice time intervals extracted from filtered signals from Figure 13.

The second sound signal is 90 s long and has a low signal-to-noise ratio (Figure 17). The way of approaching the rescued person was similar to that which was performed in the overground experiment (the background noise though differ due to the underground environment and other microphone used).
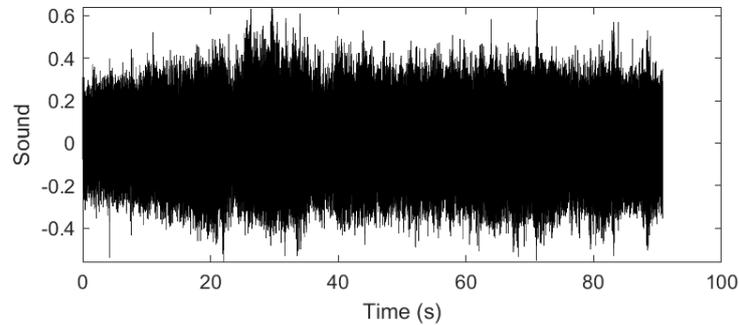

Figure 17. Raw signal of the second sound recorded underground.

STFT is computed with 80% overlapping. Its time-frequency characteristic is very volatile though harmonics, i.e., horizontal lines can be distinguished (Figures 18 and 19).
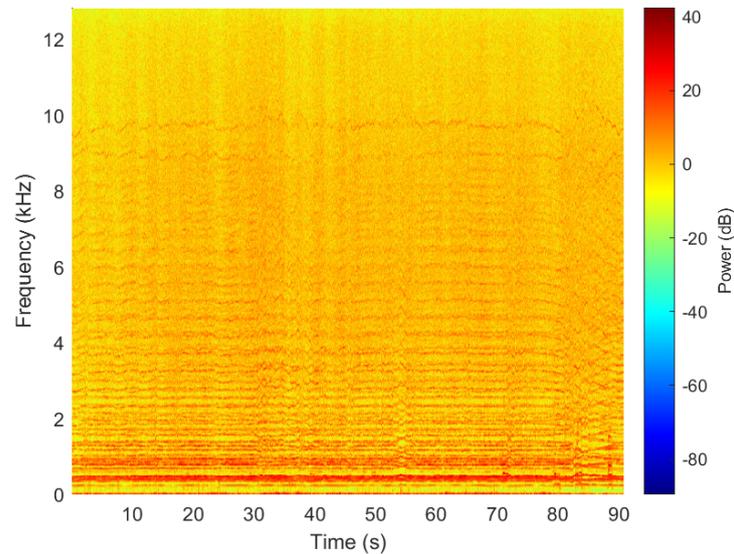

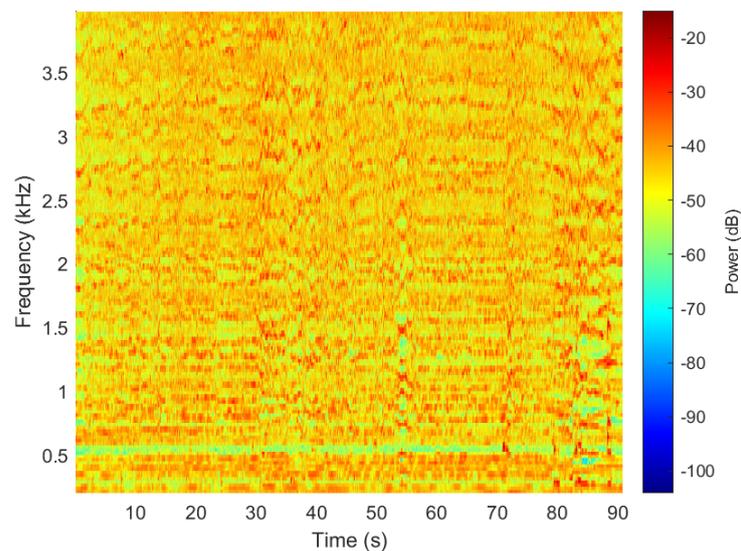**Figure 18.** Spectrogram of the raw signal from Figure 17.

Figure 19. Normalized and band-pass filtered spectrogram of the signal from Figure 17.

The signal even after filtering has still a very noise-like shape (Figure 20). Only some slow changes in mean magnitude (corresponding also to the power) are visible.

Time bins for calculation of spectral metric are 8 ms long. Smoothed (with 20 samples moving mean as before) Euclidean metric referring to a mean power spectrum from first 5 s is particularly high in the ending seconds (Figure 21). The threshold was in this case set using mean value instead of mode value. Kernel-smoothed PDF is presented in Figure 22.

In the set of the marked intervals no clear pattern can be seen (Figure 23). Nevertheless, audition of the detection output and comparing it with the input raw sound allows us to state that all audible voice is detected (included in 4 last intervals). Parameters of Tol_0 and Tol_1 were again fixed to 0.5 s.
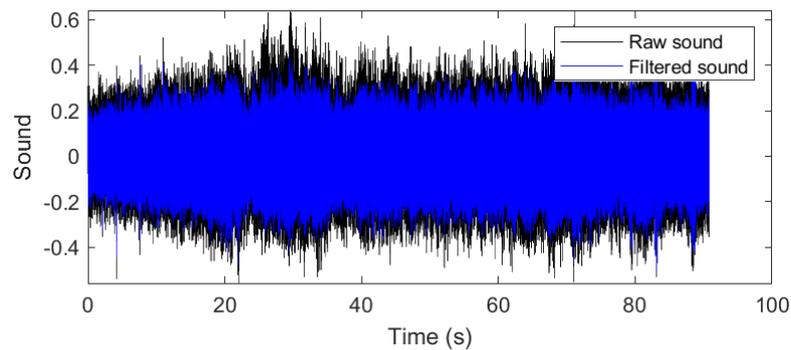

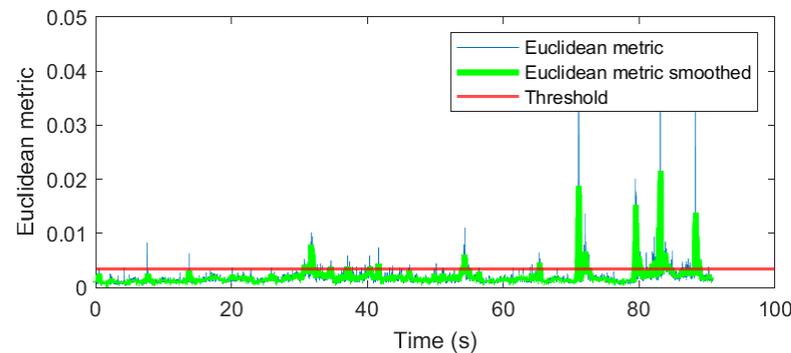**Figure 20.** A band-pass filtering of the sound from Figure 17.

Figure 21. Smoothed Euclidean metric calculated from normalized spectrogram from Figure 19 as a function of time.
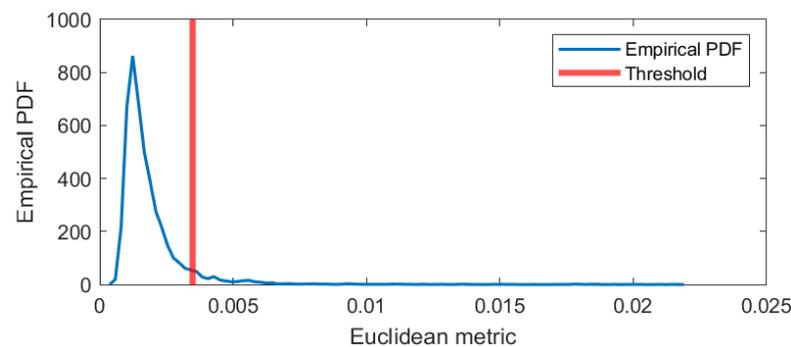

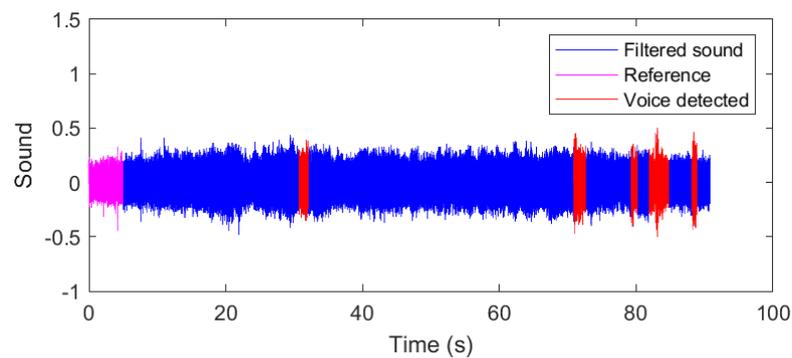**Figure 22.** Empirical PDF of Smoothed Euclidean metric from Figure 21.

Figure 23. Detected voice time intervals extracted from filtered signals from Figure 20.

6. Validation Experiments

To evaluate the efficiency of the presented method, several lab condition experiments have been performed. We investigated the impact of the drone's flight conditions on the performance of our algorithm. During the validation experiments, we have additionally tested two other drones of different sizes (see Figure 24) to evaluate how the noise generated by the rotors of different power they are equipped with, will mask informative components related to a human voice.
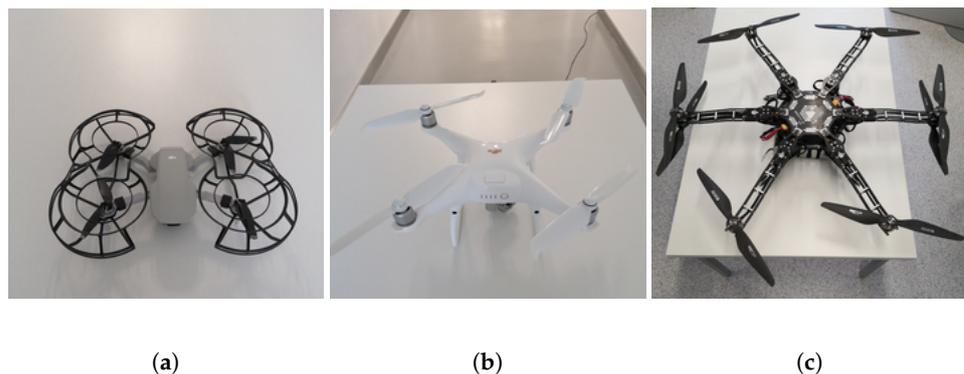

(**a**)　　　　　　　　　(**b**)　　　　　　　　　(**c**)

Figure 24. Drones used in the experiments: (**a**) small drone used both in mine and lab experiments, (**b**) medium size drone used in validation experiment, (**c**) large size drone used in validation experiment.

The second drone used for validation purposes is, similarly to the one used in the original experiment, a commercial quadcopter-DJI Phantom 4 Advanced, dimensions of which are 350×350 mm (the dimensions of Mavick Mini are 160×202 mm). The third one was a custom UAV, designed according to predefined specifications. The rotors' spacing is 900 mm. All sounds presented in this section were recorded using the same microphone with sampling frequency equal to 25.6 kHz.

First, almost stable flight conditions have been kept (see spectrogram in Figure 25). It means that the drone's ego-noise spectrum is more or less constant.

We intentionally introduced maneuvers during the flight to assess the variability of acoustic signal parameters. As we predicted, change of flight parameters will modify the spectrum significantly, including AM/FM modulations in the signal (see spectrogram in Figure 26).

The introduced changes in flight parameters cause that human voice activity detection is way more difficult to be achieved. When the drone's flight parameters are approximately constant, the effectiveness of the algorithm used is very high (see Figure 25). As can be seen in Figure 26 the proposed method faces segmentation problems while the drone is performing the additional maneuvers. First of all, not every voice activity has been detected. Secondly, the noise generated by the drone is incorrectly detected as a human voice.
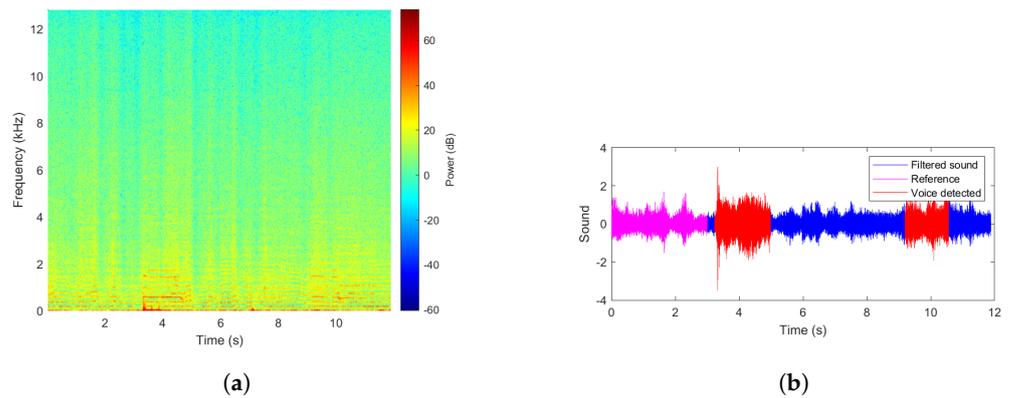
Figure 25. Example of a segment of the acoustic signal with almost constant flight conditions and neglectable spectral content fluctuation: (**a**) a raw spectrogram, (**b**) correct voice detection.
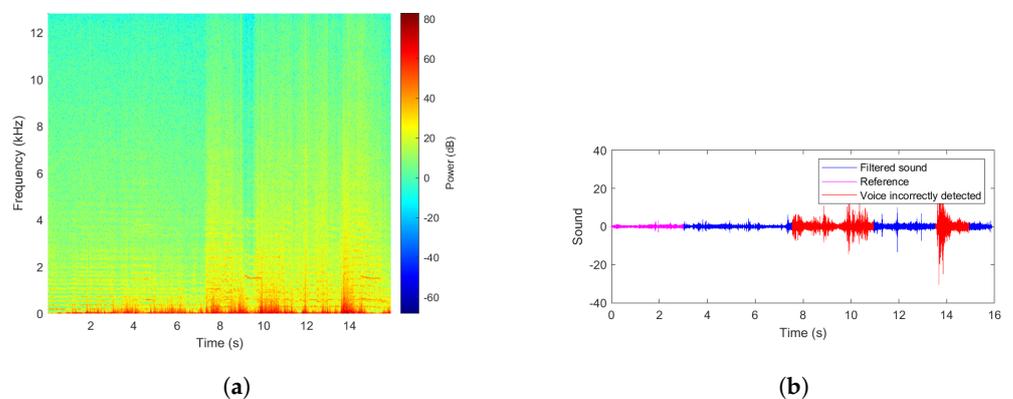

Figure 26. Example of a segment of the acoustic signal with highly variable flight conditions and significant spectral content fluctuation: (**a**) a raw spectrogram, (**b**) incorrect voice detection.

To simplify the problem, for two new drones (medium size and large size) we decided to analyze segments with stable flight conditions ignoring those with rapid maneuvers, i.e., strongly time-varying spectrum. The justification of this approach is the fact that in underground mines' conditions the dimensions of underground corridors are limited, hence the possibility of performing complex drone maneuvers during a flight is significantly limited. To reduce the risk of a situation in which the drone could be damaged as a result of a collision with excavation walls or other elements in the excavation area, the drone's flight speed must not be too high or variable, and the maneuvers should not be sudden. This translates into the type of recorded noise, the energy of which will not be high and the spectrum will not be highly variable in time. The above factors, although they have an impact on the effectiveness of the proposed method, will not occur often in real conditions in which the method is to be applied.

Results of additional experiments (focused on validation of the concept) are presented below in Figures 27 and 28. In Figure 27a, raw signal is presented. Only part of the signal plotted in black is analyzed using the proposed algorithm. Figure 27b presents the spectrogram of the analyzed signal, where a significant change in spectral content is clearly visible at ca. T = 5, 10, 15, 22, and 27 s. Figure 27c presents the Euclidean distance estimated for each spectrum. The threshold resulting from the PDF of such a measure is also plotted as a red line. Based on the proposed decision-making scheme, the final picture with detection results is presented in Figure 27d. All voice activities have been successfully detected.
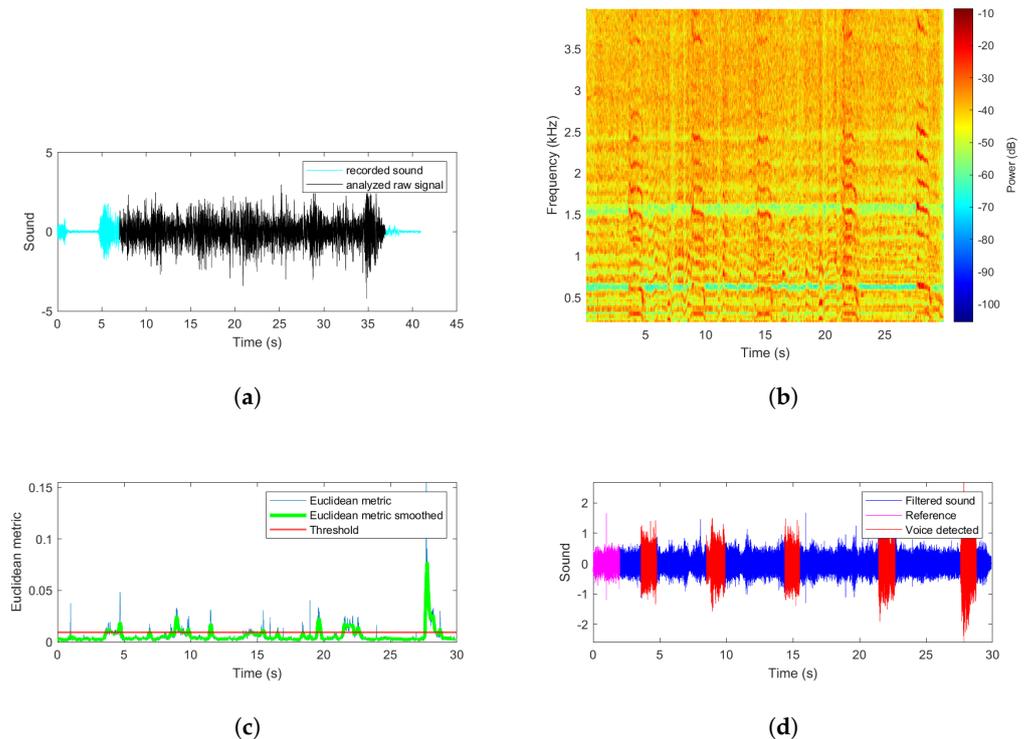
(**a**)

(**b**)

(**c**)

(**d**)

Figure 27. Results obtained with the medium size drone: (**a**) Original signal with segment selected for analysis, (**b**) normalized and band-pass filtered spectrogram selected for analysis, (**c**) estimated distance from reference spectrum with estimated threshold, (**d**) detection results.

By analogy, the same procedure has been applied in the third case (Figure 28). In Figure 28a a raw signal within the selected segment is presented. One may notice that for signals above T = 40 s its amplitudes are much higher; we did not select this part for further analysis. The max power of the drone and noise level were significantly higher resulting in lower SNR. We ignored this segment since in an underground mine, a drone ought to be steered very carefully and one should not risk flying with such parameters. The spectrogram (Figure 28b) for the 3rd case (selected part) is clear enough and allows detection of voice activities (Figure 28c). Final detection result visualization is presented in Figure 28d.

The results achieved in this case can raise concerns regarding the utility of the presented method when using bigger UAVs, nonetheless, taking into consideration the spatial restraints in underground mines, such drones are not recommended in this specific environment (bigger UAVs are much more cumbersome to be steered and can not reach as many locations as smaller ones).

One may conclude that based on the obtained results, the signal having its origin in a human voice is detectable in each case. Three different noise characteristics have been considered. During maneuvers (turning left/right, acceleration, etc.) the structure of drone ego-noise is time-varying and indeed it makes voice detection much more complicated. Given the aforementioned, we decided to skip the highly non-stationary segments during the detection procedure.
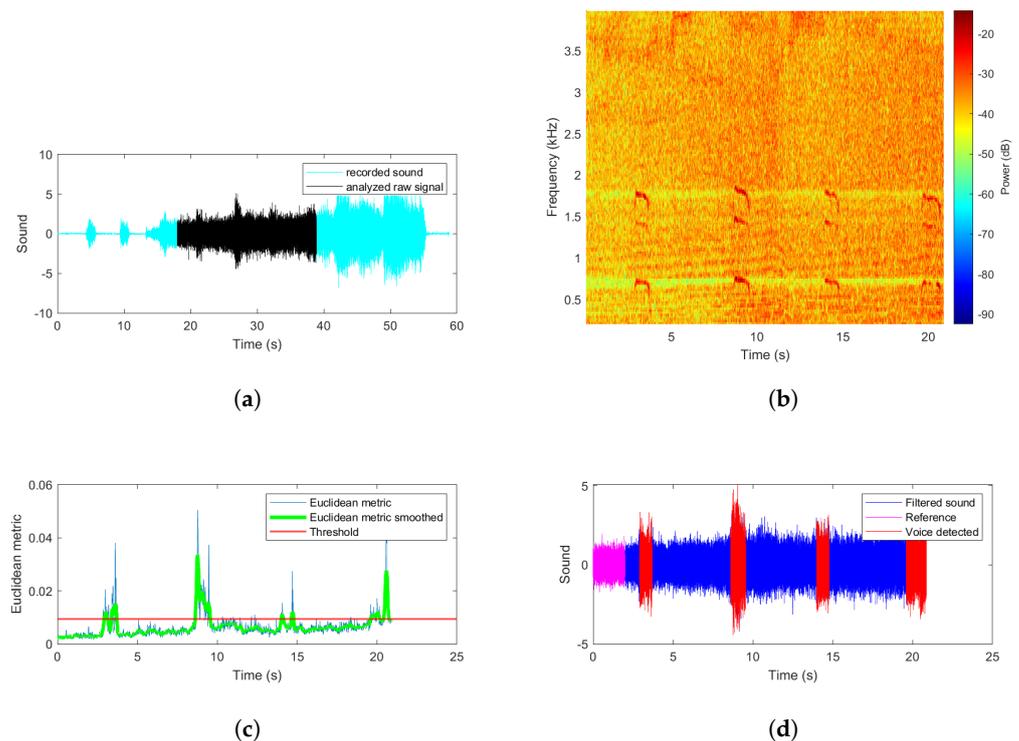
Figure 28. Results obtained with the large size drone: (**a**) raw sound recorded in the presence of a medium-size drone, (**b**) normalized and band-pass filtered spectrogram selected for analysis, (**c**) estimated distance from reference spectrum with estimated threshold, (**d**) detection results.

7. Discussion and Conclusions

A conception of drone usage for rescue operations in underground mines has been presented with the assumption of a situation in which an accident victim (being unable to rescue oneself from the accident's location) is capable of calling for help. Since the main problem to be solved in order to ensure human detection with the use of a microphone-equipped UAV was the distinction of a characteristic sound in the presence of heavy, time-varying noise, obtaining the automatic signal processing procedure is a valuable achievement, not only from the point of view of a rescue operation.

Our method used for voice detection tends to capture long phrases wholly, while a simple word /help/ is perceived mostly as /he/ or /e/. Hence high-energy voiced speech is better detected than low-energy noise-like unvoiced speech which is covered by the background noise. An improvement of the presented method can be sought in detecting voiced speech components having a harmonic structure. It should be noted that any source of the acoustic signal that is characteristic for a given situation and easily separable from drone-induced noise can be used to identify the event of interest. Taking this into consideration, future work will be conducted with the focus put on identifying other event-informative and distinguishable-from-noise sound sources, together with generalizing the signal processing technique presented. Such signals can be emitted by, for instance, electronic devices carried by underground workers. Characteristic spectral content obtained this way would be easier to detect than the human voice. The detection of human voice would then state a backup scenario in case of a lost or faulty sound emitting device.

The core of this research work—the procedure proposed in the article—is relatively simple, as it hinges on the usage of typical time-frequency representation of the registered signal and comparison of its instantaneous spectrum with a reference spectrum that is known to not contain the signal sought, but only drone's sound. It alludes directly to the novelty-detection approach, namely, if a spectrum emerges in the analyzed signal of an unknown or not expected structure, there is a high probability that it is related to an event

of interest. Taking into account that during a rescue action, the presence of various noise sources is limited (due to the stopped machinery activity for instance), the abovementioned assumption can be found as legitimate. It should be underlined that what is of the highest importance is not a detected and distinguished specific human voice or a phrase, but rather, a quick and reliable segmentation of useful data (containing more than just a drone's noise) from a big dataset created with the use of a team of microphone-equipped UAVs.

Several laboratory experiments and the one conducted in realistic underground mine conditions proved that the developed procedure gives satisfactory results, and is a promising basis for the development of a useful aid tool for rescue operations. Nonetheless, its usage can be also extended to other cases in which drone-aided inspection would be safer or more efficient, which is a common case in underground mining.

The presented study is a preliminary test of a concept to be further developed, and subjected to evaluation in various use-case scenarios in different conditions. The method was tested with the use of various drones having different engine powers emitting noise of various levels. One of the key issues is the identification of drone noise. In general, it is expected that noise from the flying drone will be a non-stationary signal with a time-varying structure, especially when flight parameters are changing. At this stage, to simplify the signal processing procedure, the flight condition was as stable as possible. No maneuvers, except take-off and landing, have been considered. During the flight, the flight speed was relatively slow (noise is smaller too) and stable as much as possible. Its ultimate autonomous flight could be stabilized by the algorithm so the variability of speed will be neglectable. It simplifies the structure of the instantaneous spectrum and makes detection easier. The speed has not been measured directly, but from the spectrogram, it can be seen that it is nearly stable (frequencies related to rotor speed are not changing too much).

Even though the method can be described as 'self-calibrating' (due to the automated calculation of the referential spectrum), repeating the experiment numerous times could be beneficial since it would allow examining the method's response and accuracy giving very significant information regarding the so-called 'false alarm ratio'. In this research, we were not able to estimate the relation between signal to noise ratio and probability of detection; hence, this is assumed to be the basis of future work. Implementation of the proposed method in a real-case scenario of rescue action conducted in an underground mine must be aware of possible detection of some signal something that is not expected voice-related source. According to the specific procedures concerning rescue actions, during such a situation, all activities in the mine are supposed to be stopped until the end of rescue action. There is nearly no chance to record other noise than from the drone's operation and potential human calling for help. In addition, any false detections in such a situation are of poor probability, since the proposed method is intended to support rescue teams, so every potential detection will be verified.

Another more complicated issue is the lack of detection due to poor SNR of the acquired signal. This could be minimized by decreasing the distance between the sound source and the drone. It has been shown that if the drone is closer to the source, the signal is more visible in the noise. The task here is not to detect all signatures, so classical measures in detection efficiency analysis are not relevant. The main target is to provide data by the drone before the rescue team will be able to realize the rescue action.

Employing an autonomous UAV should be considered to obtain the benefit of reduction of the human presence in the most dangerous places as much as it is possible. It should be also stated that detection ability is related to several factors: drone noise level, human voice source level, the distance between the voice source and the drone with a microphone, etc. One could expect precise information on how close a drone should be to detect a human voice based on acoustic signals. It can be determined experimentally on the basis of many experiments in the mine. It is important to say that voice propagation in the tunnel may be slightly different due to tunnel geometry and surface properties.

References

1. Tutak, M.; Brodny, J.; Szurgacz, D.; Sobik, L.; Zhironkin, S. The Impact of the Ventilation System on the Methane Release Hazard and Spontaneous Combustion of Coal in the Area of Exploitation—A Case Study. *Energies* **2020**, *13*, 4891. [CrossRef]
2. Szurgacz, D.; Sobik, L.; Brodny, J. Integrated method of reducing the threat of endogenous fires in hard coal mines. In Proceedings of the 14th International Innovative Mining Symposium, Kemerovo, Russian, 14–16 October 2019; Volume 105. [CrossRef]
3. Zhou, Q.; Herbert, J.H.; Hidalgo, A. Development of a stress-based approach for achieving the risk assessment of fault-related coal and gas outburst. *Int. J. Oil, Gas Coal Technol.* **2019**, *21*, 509. [CrossRef]
4. Ziętek, B.; Banasiewicz, A.; Zimroz, R.; Szrek, J.; Gola, S. A Portable Environmental Data-Monitoring System for Air Hazard Evaluation in Deep Underground Mines. *Energies* **2020**, *13*, 6331. [CrossRef]
5. State Mining Authority (Wyższy Urząd Górniczy), Assessment of Occupational Safety, Mine Rescue and General Safety in Relation to Mining and Geological Activities in 2019, 2020 (In Polish). Available online: https://wug.intracom.com.pl/download/WUG_Stan_bhp_19,7247.pdf (accessed on 11 June 2021).
6. Park, J.S.; Na, H.J. Front-End of Vehicle-Embedded Speech Recognition for Voice-Driven Multi-UAVs Control. *Appl. Sci.* **2020**, *10*, 6876. [CrossRef]
7. Cohen, I. Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging. *IEEE Trans. Speech Audio Process.* **2003**, *11*, 466–475. [CrossRef]
8. Sohn, J.; Kim, N.S.; Sung, W. A statistical model-based voice activity detection. *IEEE Signal Process. Lett.* **1999**, *6*, 1–3. [CrossRef]
9. Zheng, B.; Hu, J.; Zhang, G.; Wu, Y.; Deng, J. Analysis of Noise Reduction Techniques in Speech Recognition. In Proceedings of the 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chongqing, China, 12–14 June 2020. [CrossRef]
10. Bachu, R.; Kopparthi, S.; Adapa, B.; Barkana, B. Voiced/Unvoiced Decision for Speech Signals Based on Zero-Crossing Rate and Energy. In *Advanced Techniques in Computing Sciences and Software Engineering*; Springer: Dordrecht, The Netherlands, 2009; pp. 279–282. [CrossRef]
11. Roman, N.; Wang, D.; Brown, G.J. Speech segregation based on sound localization. *J. Acoust. Soc. Am.* **2003**, *114*, 2236–2252. [CrossRef]
12. Gagliardi, G.; Tedesco, F.; Casavola, A. An Adaptive Frequency-Locked-Loop Approach for the Turbocharger Rotational Speed Estimation via Acoustic Measurements. *IEEE Trans. Control. Syst. Technol.* **2020**, 1–13. [CrossRef]
13. Cavina, N.; Cesare, M.D.; Ravaglioli, V.; Ponti, F.; Covassin, F. Full Load Performance Optimization Based on Turbocharger Speed Evaluation via Acoustic Sensing. In *Instrumentation, Controls, and Hybrids*; Numerical Simulation; Engine Design and Mechanical Development; Keynote Papers; American Society of Mechanical Engineers: New York, NY, USA, 2014; Volume 2. [CrossRef]
14. Ravaglioli, V.; Cavina, N.; Cerofolini, A.; Corti, E.; Moro, D.; Ponti, F. Automotive Turbochargers Power Estimation Based on Speed Fluctuation Analysis. *Energy Procedia* **2015**, *82*, 103–110. [CrossRef]
15. Tinney, C.E.; Sirohi, J. Multirotor Drone Noise at Static Thrust. *AIAA J.* **2018**, *56*, 2816–2826. [CrossRef]
16. Zhou, T.; Jiang, H.; Sun, Y.; Fattah, R.J.; Zhang, X.; Huang, B.; Cheng, L. Acoustic characteristics of a quad-copter under realistic flight conditions. In Proceedings of the 25th AIAA/CEAS Aeroacoustics Conference. American Institute of Aeronautics and Astronautics, Delft, The Netherlands, 20–23 May 2019. [CrossRef]

17. Wang, L.; Sanchez-Matilla, R.; Cavallaro, A. Tracking a moving sound source from a multi-rotor drone. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018. [CrossRef]

18. Wang, L.; Cavallaro, A. Acoustic Sensing From a Multi-Rotor Drone. *IEEE Sens. J.* **2018**, *18*, 4570–4582. [CrossRef]

19. Djurek, I.; Petosic, A.; Grubesa, S.; Suhanek, M. Analysis of a Quadcopter's Acoustic Signature in Different Flight Regimes. *IEEE Access* **2020**, *8*, 10662–10670. [CrossRef]

20. Wyłomańska, A.; Zimroz, R. Signal segmentation for operational regimes detection of heavy duty mining mobile machines-A statistical approach. *Diagnostyka* **2014**, *15*, 33–42.

21. Wodecki, J.; Stefaniak, P.; Obuchowski, J.; Wylomanska, A.; Zimroz, R. Combination of principal component analysis and time-frequency representations of multichannel vibration data for gearbox fault detection. *J. Vibroeng.* **2016**, *18*, 2167–2175. [CrossRef]

22. Zak, G.; Wylomanska, A.; Zimroz, R. Local Damage Detection Method Based on Distribution Distances Applied to Time-Frequency Map of Vibration Signal. *IEEE Trans. Ind. Appl.* **2018**, *54*, 4091–4103. [CrossRef]

23. Liu, J. Current research, key performances and future development of search and rescue robot. *Chin. J. Mech. Eng.* **2006**, *42*, 1. [CrossRef]

24. Murphy, R.R.; Tadokoro, S.; Nardi, D.; Jacoff, A.; Fiorini, P.; Choset, H.; Erkmen, A.M. Search and Rescue Robotics. In *Springer Handbook of Robotics*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 1151–1173. [CrossRef]

25. Liu, Y.; Nejat, G. Robotic Urban Search and Rescue: A Survey from the Control Perspective. *J. Intell. Robot. Syst.* **2013**, *72*, 147–165. [CrossRef]

26. Said, K.O.; Onifade, M.; Githiria, J.M.; Abdulsalam, J.; Bodunrin, M.O.; Genc, B.; Johnson, O.; Akande, J.M. On the application of drones: A progress report in mining operations. *Int. J. Mining, Reclam. Environ.* **2020**, 1–33. [CrossRef]

27. Szrek, J.; Wodecki, J.; Błazej, R.; Zimroz, R. An inspection robot for belt conveyor maintenance in underground mine-infrared thermography for overheated idlers detection. *Appl. Sci.* **2020**, *10*, 4984. [CrossRef]

28. Szrek, J.; Zimroz, R.; Wodecki, J.; Michalak, A.; Góralczyk, M.; Worsa-Kozak, M. Application of the infrared thermography and unmanned ground vehicle for rescue action support in underground mine—the amicos project. *Remote Sens.* **2021**, *13*, 69. [CrossRef]

29. Zimroz, R.; Hutter, M.; Mistry, M.; Stefaniak, P.; Walas, K.; Wodecki, J. Why Should Inspection Robots be used in Deep Underground Mines? In *Proceedings of the 27th International Symposium on Mine Planning and Equipment Selection-MPES 2018*; Widzyk-Capehart, E., Hekmat, A., Singhal, R., Eds.; Springer International Publishing: Cham, Sitzerland, 2019; pp. 497–507.

30. Szrek, J.; Trybala, P.; Goralczyk, M.; Michalak, A.; Zietek, B.; Zimroz, R. Accuracy evaluation of selected mobile inspection robot localization techniques in a gnss-denied environment. *Sensors* **2021**, *21*, 141. [CrossRef] [PubMed]

31. Skoczylas, A.; Stefaniak, P.; Anufriiev, S.; Jachnik, B. Belt Conveyors Rollers Diagnostics Based on Acoustic Signal Collected Using Autonomous Legged Inspection Robot. *Appl. Sci.* **2021**, *11*, 2299. [CrossRef]

32. Waharte, S.; Trigoni, N. Supporting Search and Rescue Operations with UAVs. In *Proceedings of the* 2010 International Conference on Emerging Security Technologies, Canterbury, UK, 6–7 September 2010. [CrossRef]

33. Sun, J.; Li, B.; Jiang, Y.; Wen, C.Y. A Camera-Based Target Detection and Positioning UAV System for Search and Rescue (SAR) Purposes. *Sensors* **2016**, *16*, 1778. [CrossRef] [PubMed]

34. Malos, J.; Beamish, B.; Munday, L.; Reid, P.; James, C. Remote monitoring of subsurface heatings in opencut coal mines. In Proceedings of the 2013 Coal Operators' Conference. University of Wollongong, Wollongong, Australia, 14–15 February 2013; pp. 227–231.

35. Nanda, S.K.; Dash, A.K.; Acharya, S.; Moharana, A. Application of robotics in mining industry: A critical review. *Indian Min. Eng. J.* **2010**, *8*, 108–112.

36. Siebert, S.; Teizer, J. Mobile 3D mapping for surveying earthwork projects using an Unmanned Aerial Vehicle (UAV) system. *Autom. Constr.* **2014**, *41*, 1–14. [CrossRef]

37. Azhari, F.; Kiely, S.; Sennersten, C.; Lindley, C.; Matuszak, M.; Hogwood, S. A comparison of sensors for underground void mapping by unmanned aerial vehicles. In Proceedings of the First International Conference on Underground Mining Technology. Australian Centre for Geomechanics, Perth, Austria, 11–13 October 2017. [CrossRef]

38. Ren, H.; Zhao, Y.; Xiao, W.; Hu, Z. A review of UAV monitoring in mining areas: Current status and future perspectives. *Int. J. Coal Sci. Technol.* **2019**, *6*, 320–333. [CrossRef]

39. Li, H.; Savkin, A.V.; Vucetic, B. Autonomous Area Exploration and Mapping in Underground Mine Environments by Unmanned Aerial Vehicles. *Robotica* **2019**, *38*, 442–456. [CrossRef]

40. Turner, R.M.; MacLaughlin, M.M.; Iverson, S.R. Identifying and mapping potentially adverse discontinuities in underground excavations using thermal and multispectral UAV imagery. *Eng. Geol.* **2020**, *266*, 105470. [CrossRef]

41. Reddy, A.H.; Kalyan, B.; Murthy, C.S. Mine Rescue Robot System–A Review. *Procedia Earth Planet. Sci.* **2015**, *11*, 457–462. [CrossRef]

42. Lee, S.; Choi, Y. Reviews of unmanned aerial vehicle (drone) technology trends and its applications in the mining industry. *Geosyst. Eng.* **2016**, *19*, 197–204. [CrossRef]

43. Park, S.; Choi, Y. Applications of Unmanned Aerial Vehicles in Mining from Exploration to Reclamation: A Review. *Minerals* **2020**, *10*, 663. [CrossRef]

44. Shahmoradi, J.; Talebi, E.; Roghanchi, P.; Hassanalian, M. A Comprehensive Review of Applications of Drone Technology in the Mining Industry. *Drones* **2020**, *4*, 34. [CrossRef]

45. Allen, J. Short term spectral analysis, synthesis, and modification by discrete Fourier transform. *IEEE Trans. Acoust. Speech Signal Process.* **1977**, *25*, 235–238. [CrossRef]

46. Wyłomańska, A.; Żak, G.; Kruczek, P.; Zimroz, R. Application of tempered stable distribution for selection of optimal frequency band in gearbox local damage detection. *Appl. Acoust.* **2017**, *128*, 14–22. [CrossRef]

47. Hebda-Sobkowicz, J.; Zimroz, R.; Wyłomanska, 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]

48. Hebda-Sobkowicz, J.; Zimroz, R.; Pitera, M.; Wyłomańska, 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*. [CrossRef]

49. Wodecki, J.; Michalak, A.; Zimroz, R. Optimal filter design with progressive genetic algorithm for local damage detection in rolling bearings. *Mech. Syst. Signal Process.* **2018**, *102*, 102–116. [CrossRef]