Detection of a predefined acoustic pattern by a measurement system on a drone and its application to search for a missing man in an underground mine

Paweł Zimroz 1,* , Adam Wróblewski 1 , Paweł Trybala 1

1 Faculty of Geoengineering, Mining and Geology, Wrocław University of Science and Technology, Na Grobli 15, 50-421 Wrocław, Poland

Corresponding author: pawel.zimroz@pwr.edu.pl

Abstract.

Due to the very difficult working conditions and a certain number of hazards that do not occur in other industries, the operation of rescue units in underground mines is necessary. The area of exploitation is usually very large, thus determining the location of a person, which may be in need due to the accident, is not an easy task. As the time for reaching such a person is crucial, there is a strong need for a solution that would provide a quick establishment of the victims’ location. Moreover, conducting a rescue mission is always associated with risk exposure for rescuers’ life and health. Thus, in this paper, we propose a solution based on an unmanned aerial vehicle (UAV) for a predefined acoustic pattern detection to support rescue units in human location assessment in the underground mine. The presented method is based on measuring the dissimilarity between the subsequent short-time power spectra and the referential spectrum characterizing the UAV’s ego-noise. This relatively general and data-driven approach is applied both to generated narrow-band harmonic patterns and to the human voice. As the analyzed signals of interest are of specific frequency content they can be selected from the background noise with the use of band-pass filtering.

1. Introduction

The dynamic development of new technologies creates opportunities to improve work safety and streamline production in the mining industry. The utilization of a number of possibilities offered by robotic solutions for these purposes seems to be of particular interest. The use of Unmanned Ground Vehicles (UGV) and Unmanned Aerial Vehicles (UAV) in difficult conditions of underground mines has been proven right in many publications. According to [1] robotic solutions implementation may be divided into several main groups: robotics for main and auxiliary technological processes, which are presented in [2, 3, 4, 5, 6, 7], robotics for inspection applications, shown in [8, 9, 10], as well as robotics for rescue missions, described in [11, 12, 13, 14, 15, 16].

UAVs, commonly known as drones, are widely used in the industry. One of the benefits of exploiting small drones, particularly important from the underground mining point of view, is that they can penetrate hard-to-reach places and take the risk which otherwise a human would need to take. An important issue is to provide the UAV with tools and algorithms that would allow it to replace humans in performing particular actions.
One of the potential applications for UAV is predictive maintenance. Detection of cyclic anomalies such as local damage in machines is a subject widely studied recently for real application to monitoring of the mining machinery condition [17, 18, 19]. Latest efforts have been given for using sound signals for diagnostics issues (see for example [20, 21]). We want to show how similar methods may be applied for distress call detection ensuring safety in mines. As a part of the AMICOS project [22] the miners would be equipped with a pocket device able to emit particular sounds. During a rescue operation, one or more people would be tracked by a drone patrolling the tunnels and recording the sound continuously. When the drone detects a sound of interest being a novelty in its record, it could report this information to the rescue crew bundled with its current location (which could be determined using one of the techniques shown in [7]).

As each audio signal recorded in the presence of a drone contains a quadrotor drone’s ego-noise, a problem of signal segmentation arises. UAV noise has been widely investigated in the last years [23, 24, 25]. It has many harmonic components and comprises high-energy low-frequency bands. Besides, numerous publications can be found on the detection of a sound of interest (especially speech) in noisy conditions [26, 27]. The methodology presented here exploits metrics between short-time spectra. It corresponds to a technique of anomaly detection used in [28] for local damage detection in the driving units of belt conveyors. Our method was presented recently in the context of human voice detection (some experiments were performed underground) [29]. However, it should be noted that the signal of interest having a wide spectrum (in terms of Fourier Transform) interlaces with the spectrum of the drone’s noise causing informative frequency band selection to be a significant issue. For this experiment, we have tested the sound emitted by a rescued person in the form of a repeated harmonic test sound at a given frequency. Nonetheless, more extensive studies should be conducted on the optimal emergency signal frequency selection strategies for the specific UAV planned to be used in a search and rescue mission.

Voice activity detection has been also recently studied in the context of rescue operations [30]. The method presented here can be used for detecting harmonic patterns as well as repeated sounds emitted by a rescued person (on the condition that he or she is able to make voice).

In this article, we present an experiment that concerns the detection of a predefined acoustic pattern. Its aim is the future application of such a system in a real scenario of an underground search and rescue operation.

## 2. Methodology

The proposed procedure is based on the calculation of the STFT (Short-Time Fourier Transform) [31] on discrete set of time and frequency values $T \times F$. As a standard-setting we use STFT with Kaiser window of order 5 and length 1024, which is also the number of FFT points and 80% overlapping.

Both for further analysis and for plotting a spectrogram we will consider the squared absolute value of the STFT:

$$PSD(t, f) = |STFT(t, f)|^2.$$  

Columns of such spectrogram matrix are power spectral densities and its rows are called subsignals (each for a given a frequency).

To reduce the impact of the hidden determinism of the drone’s noise, the sub-signals are normalized (the sum of the values of a sub-signal is used as a norm):

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

In this way, the overall level of all frequencies is aligned. The next step is band pass filtering. As it was mentioned, the signals of interest are by default single harmonic tones. In this case,
the filters used are of narrow-band type with limits $f_1 = f_0 - \varepsilon$ and $f_2 = f_0 + \varepsilon$ where $f_0$ is the frequency of the detected tone.

To make a robust procedure for distinguishing spectral components present in the recorded sound, we calculate the metrics between the power spectra. To characterize the spectrum of the noise produced by a drone (the reference spectrum), the spectra are averaged on some time interval $I$ at the start of the experiment when the drone is hovering and no other significant sound is recorded:

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

(3)

As a measure of dissimilarity with a referential spectrum, the Euclidean metric changing in time (defined as below) is used:

$$D_t = D \left( PSD_{\text{normalized}}(t, \cdot), PSD_{\text{ref}}(\cdot) \right) = \sqrt{\sum_{f=f_1}^{f_2} \left( PSD(t, f) - PSD_{\text{ref}}(f) \right)^2},$$

(4)

Note that the calculation is restricted to the frequency band passed by the filter. The obtained series is smoothed with a moving mean filter (with parameter $M$):

$$\tilde{D}_t = \frac{1}{M} \sum_{k=u-\lceil \frac{M-1}{2} \rceil}^{u+\lceil \frac{M-1}{2} \rceil} D_u$$

(5)

For discriminating between the searched sound component and the drone noise, a threshold is set on Euclidean distance. As the threshold, the value of mode (read from the kernel smoothed empirical PDF) + standard error is set. As a final result, the sound is filtered and segmented to the regions with and without detection of a predefined spectral pattern (i.e. a short harmonic signal). Segments are merged so that a region of each type is at least half-second long.

3. Experiment and results
The experiment was performed using a quadrotor drone (DJI Mavic Mini) depicted in Fig. 1. Weighting 249 g, it is a good Micro Aerial Vehicle (MAV) class representative. The drone rotors are operating in the frequency range of approximately 150 Hz to 250 Hz.

As it was discussed in [29] in target real scenario a UAV (with a built-in microphone on its deck) may be alternately hover during sound recording and fly on patrolling the area. To make the experiment easier to perform a source of the sound of interest was moved instead of the drone. It doesn’t affect the possibility of validating the method for different distances from the rescued person.

The acoustic measurement was carried out using built-in microphone of a cell phone located under the hovering drone (kept very close to it as it had been limited by organizational abilities and safety issues).

The quadrotor was steered manually to hover at the height of about 1 m above the floor. At the same time, a person was playing sounds (generated for the experiment). Starting from the distance of 3 meters after each pair of test sounds, the person was quietly walking away from the UAV for about 3 meters else. The time interval between the consecutively played sounds is about 10 s.
A repeated sound at the pitch of 4.5 kHz was generated. Its power short-time spectrum is presented at Fig. 2. A record containing drone noise and a periodically generated test beacon (Fig. 3) was analysed. Fig. 4 shows further processing of the data, i.e. subsignals normalization and band-pass filtering with $f_0 = 4500$ Hz and $\varepsilon = 50$ Hz. The Euclidean metric (Fig. 5a) is calculated in reference to a 2 s long estimation of the UAV noise. The referential segment of the record and the regions labeled by the algorithm as 4.5 kHz sound detection are denoted by the waveform of the filtered signal (Fig. 5b).

The test sounds have been fully detected at all distances from the drone ending at 15 meters.
Figure 3: A record containing UAV noise and high frequency patterns.

(a) Raw signal.  
(b) Spectrogram of the raw signal.

Figure 4: Further processing of the signal from Fig. 3a

(a) Spectrogram with normalized subsignals.  
(b) Narrow band-pass filtering.

Figure 5: Results for the analysis of signal from Fig. 3a

(a) Euclidean metric changing in time.  
(b) Resulting segmentation of the filtered signal.
In the next step, we repeated the experiment with the addition of periodical occurrence of another specific sound which was the human voice. A distress call, the phrase 'help!', was made by a person alternately with 4.5 kHz beacons. Another change was made in the pattern of the test sound (Fig. 6). Instead of one long sound, we generated three subsequent short sounds. Spectrograms (raw and normalized) of this sound record are presented on Fig. 8. In total there were generated 8 harmonic beacons and the phrase 'help!' was uttered 4 times.

Two different band-pass filters were applied to this signal depending on the objective. First was the narrow-band filter used as before to select the 4.5 kHz proximity. Second filter was designed to catch the informative band of the human voice. In this case, we set \( f_1 = 0.2 \, \text{kHz} \) and \( f_2 = 3.5 \, \text{kHz} \). Different bands selected resulted in different values of Euclidean metric (in reference to the drone noise estimated from the first 2 s of the recording). The detection of both signals of interest in the filtered signals along with the Euclidean metrics and their empirical PDF are depicted on Fig. 9 and Fig 10 for 4.5 kHz pattern detection and 'help!' call detection respectively.

It can be noted in the detection results noted that two harmonic patterns were not detected and in the case of human voice detection we obtained one false alarm. The problems with detection occur only with the distance from the source of the generated sound/voice bigger than 9 meters.

![Figure 6: 4.5 kHz test beacon (three sounds pattern).](image)

![Figure 7: A record containing UAV noise, human voice and high frequency patterns.](image)
Figure 8: Spectrograms of the signal from Fig. 7.

(a) Spectrogram of the raw signal.  
(b) Spectrogram with normalized subsignals.

Figure 9: Detection of the high frequency tone patterns in the record from Fig. 7.

(a) Raw and filtered signal.  
(b) Euclidean metric changing in time.

(c) Empirical PDF of Euclidean metric.  
(d) Resulting segmentation of the filtered signal.
4. Conclusions
Experiments provided satisfactory results. 4.5 kHz patterns were successfully detected. Better performance was achieved for a longer test sound (100 % accuracy in 15 m range) than for a pattern consisting of three short sounds. However, in the latter case it was still possible to identify the sought signals (100 % accuracy in 9 m range). Achieved ranges of successful signal detection are sufficient for being used in a search and rescue mission utilizing a drone and would allow it to detect human presence during the mine exploration.

As selecting a band of interest allows us to detect either high frequency patterns or the human voice, the algorithm proved to be useful for general purposes. In the real scenario of a rescue operation, the two approaches could be therefore combined. If a person is calling for help and is unable to use the emergency sound emitter, their voice will be detected. Detecting a harmonic sounds may be more steady (as its band is narrower) and it is still possible if the person is for some reason unable to make voice. Besides, we confirmed that the presence of the human voice does not disturb the detection of the generated harmonic patterns.

Different frequencies of the emergency signal have been discussed. High pitch of the test sound is favorable as the informative band of speech can be unambiguously separated from it. The higher the frequency, the more it becomes unpleasant to the ear, so the level of 4.5 kHz was chosen as the acceptable level. Nonetheless, it would be also technically possible to use lower frequencies (e.g. from 1 to 3 kHz).

Acknowledgments
This activity has received funding from 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. 19018 (Autonomous Monitoring and Control System for Mining Plants - AMICOS).
References

[1] Zimroz R, Hutter M, Mistry M, Stefaniak P, Walas K and Wodecki J 2019 Proceedings of the 27th International Symposium on Mine Planning and Equipment Selection-MPES 2018 (Springer) pp 497–507

[2] Stefaniak P, J W, Jakubiak J and Zimroz R 2017 IOP Conference Series: Earth and Environmental Science vol 95 (IOP Publishing) p 042028

[3] Stefaniak P, Wodecki J, Jakubiak J and Zimroz R 2017 IOP Conference Series: Earth and Environmental Science vol 95 (IOP Publishing) p 042027

[4] Szrek J, Wodecki J, Blażej R and Zimroz R 2020 Applied Sciences 10 4984

[5] Nüchter A, Elseberg J and Borrmann D 2013 IFAC Proceedings Volumes 46 119–124

[6] Maity A, Majumder S and Ray D N 2013 International Conference on Robotics, Biomimetics, Intelligent Computational Systems (IEEE) pp 242–246

[7] Szrek J, Trybala P, Góralczyk M, Michalak A, Zietek B and Zimroz R 2021 Sensors 21 141 URL https://www.mdpi.com/1424-8220/21/1/141

[8] Green J, Bossepa P A, Candy L, Hlopile K, Coetzee S and Brink S 2010 25th International Conference on CAD/CAM, Robotics and Factories of the Future (CARsFOF) (CSIR International Convention Centre)

[9] Jiang B, Sample A P, Wistort R M and Mamishev A 2005 ICAR ‘05. Proceedings., 12th International Conference on Advanced Robotics, 2005. 673–679

[10] Yinka-Banjo C, Bagula A and Osunmakinde I O 2012 Ubiquitous Computing and Communication Journal 7 1316

[11] Green J 2013 2013 6th Robotics and Mechatronics Conference (RobMech) (IEEE) pp 111–116

[12] Murphy R R, Kravitz J, Stover S L and Shoureshi S 2009 IEEE Robotics & Automation Magazine 16 91–103

[13] Alkhadi R, Inkis M and Alkhadi A 2016

[14] Ma G, Tong Z, Tong M and Tang S 2018 2018 Eighth International Conference on Instrumentation & Measurement, Computer, Communication and Control (IMCCC) (IEEE) pp 259–263

[15] Atoaibi E T, Algeafari S S and Koubaa A 2019 IEEE Access 7 55817–55832

[16] Dang T, Mascarich F, Khattak S, Nguyen H, Hirsh S, Reinhart R, Papachristos C and Alexis K 2020 2020 IEEE Aerospace Conference (IEEE) pp 1–8

[17] Wodecki J, Michalak A and Zimroz R 2021 Measurement 169 108400 ISSN 0263-2241 URL https://www.sciencedirect.com/science/article/pii/S0263224120309349

[18] Hebda-Sobkowicz J, Zimroz R and Wyłomańska A 2020 Applied Sciences 10 2657 URL https://doi.org/10.3390/app10082657

[19] Wodecki J, Kruczek P, Bartkowiak A, Zimroz R and Wyłomańska A 2019 Mechanical Systems and Signal Processing 130 585–596 ISSN 0888-3270 URL https://www.sciencedirect.com/science/article/pii/S0888327019303279

[20] Gagliardi G, Tedesco F and Casavola A 2020 IEEE Transactions on Control Systems Technology 1–13 URL https://doi.org/10.1109/tcst.2020.3007075

[21] Cavina N, Cesare M D, Ravaglioli V, Ponti F and Covassin F 2014 Volume 2: Instrumentation, Controls, and Hybrids; Numerical Simulation; Engine Design and Mechanical Development; Keynote Papers (American Society of Mechanical Engineers) URL https://doi.org/10.1115/icef2014-5677

[22] Zietek B, Banasiewicz A, Zimroz R, Szrek J and Gola S 2020 Energies 13 6331 URL https://doi.org/10.3390/en13236331

[23] Djurek I, Petosic A, Grubesa S and Suhaneck M 2020 IEEE Access 8 10662–10670

[24] Zhou T, Jiang H, Sun Y, Fattah R J, Zhang X, Huang B and Cheng L 2019 2015 25th AIAA/CEAS Aeroacoustics Conference (American Institute of Aeronautics and Astronautics) URL https://doi.org/10.2514/6.2019–2587

[25] Tinney C E and Siriohi J 2018 AIAA Journal 56 2816–2826 URL https://doi.org/10.2514/1.j056827

[26] Cohen I 2003 IEEE Transactions on Speech and Audio Processing 11 406–475 URL https://doi.org/10.1109/tasas.2003.811544

[27] Sihn K, Kim N S and Sung W 1999 IEEE Signal Processing Letters 6 1–3 URL https://doi.org/10.1109/97.736233

[28] Zak G, Wyłomańska A and Zimroz R 2018 IEEE Transactions on Industry Applications 54 4091–4103

[29] Zimroz P, Trybala P, Wróblewski A, Góralczyk M, Szrek J, Wójcik A and Zimroz R 2021 Energies 14 ISSN 1996-1073 URL https://www.mdpi.com/1996-1073/14/3/3725

[30] Yamazaki Y, Tamaki M, Premachandra C, Perera C J, Sumathipala S and Sudantha B H 2019 2019 Third IEEE International Conference on Robotic Computing (IRC) pp 555–559

[31] Allen J 1977 IEEE Transactions on Acoustics, Speech, and Signal Processing 25 235–238