



This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits use, distribution, and reproduction in any medium, provided the original publication is properly cited. No use, distribution or reproduction is permitted which does not comply with these terms.

SAFETY IN CITIES AND TRANSPORT THROUGH SOUND MONITORING

Dalibor Smažinka, Martin Hrinko *

College CEVRO Institute, Prague, Czech Republic

*E-mail of corresponding author: martin.hrinko@seznam.cz

Resume

This paper deals with the use of modern technology for security purposes, focusing on the situation in cities (urban noise and shooting) and in transport (passenger rail). The authors aimed to highlight the possibilities and capabilities of installation of technology in the field of applied physical acoustics (sensors, evaluation system with the possibility of setting sensitive sensing filters, end output) and the possibilities of monitoring, including the response of security forces or security measures to the detected danger. In the paper, the authors also include graphical representations of the energy-normalized spectra of the different sound backgrounds and classification classes of acoustic events and types of acoustic events to clarify the differences of the selected monitored events.

Article info

Received 15 February 2022

Accepted 16 August 2022

Online 14 September 2022

Keywords:

urban safety
traffic safety
sound events
sound classification
video surveillance
audio surveillance
safety system response

Available online: <https://doi.org/10.26552/com.C.2022.4.F72-F81>

ISSN 1335-4205 (print version)

ISSN 2585-7878 (online version)

1 Introduction:

For years, the world's security forces have been concerned with the use of modern technology to strengthen their response to dangers that may arise, either intentionally or negligently, from the weakest link in the security chain, the human factor. The authors of this paper present a description of one such modern and now available technology, the essence of which is the use of applied physical acoustics in the installation of systems in the streets and squares of cities, with the ability to evaluate and compare the specified criteria as correct or incorrect. All this according to the sounds that a modern sensitive acoustic sensor, through set filters of monitored sounds (acoustic events), placed at target locations in the city, evaluates and transmits the information either to the operator or autonomously evaluates as a danger and triggers protective measures. Some examples, among many that can give the readers of this paper a better overview, include the recognition of the sound of gunshots on a busy street, in a square, in a means of transport etc. Further, for example, detecting the sound of a railway train passing through a location, whereby the correct time of passage can be compared at a time when the barriers are to be lowered, or evaluating the sound of a railway train passing through a section at a time when no passage through that

location is expected. Reaction and the so-called checking or comparison checking with the described technology will thus help to prevent collision of rail trains directed to one track, check the lowered barriers, alert security forces to firearms firing at a particular location etc. All these measures are aimed at saving the life and health of the public and the efficiency of security forces and measures. Nowadays, one is increasingly encountering the so-called "smart cities", but behind this label, in the vast majority of cases, one only find a system that counts parking spaces in car parks in parts of cities, near hospitals or shopping centres. That is why the authors were looking at a new way of strengthening the security.

2 Goal

The aim of this paper was to discuss, describe and investigate the evolution of sound detection and its application for security, starting from the original knowledge about the origin and physical characteristics of sound, to the current multifunctional use of modern technologies and systems. Based on several studies from the transportation field, the authors evaluated the current potential of using the sound event detection (SED) systems specifically for the purpose of improving

the transportation safety in practice and present their conclusions to the readers.

3 The use of applied physical acoustics in the traffic safety and smart cities

Scientific experiments with sound date back to ancient Greece, when scientists and philosophers were concerned with locating the sources of sound. However, the modern study of this topic did not begin until the late 19th century, motivated by the desire to scientifically describe the principle by which the human brain is able to locate the direction of sound. The first experiments described were conducted by Lord Rayleigh in his garden, trying to locate the position of talking people. The result of his investigations was the conclusion that the phenomenon that makes this possible is binaural hearing, the fact that humans have two organs of hearing. Simply put, the human brain has the ability to evaluate the delay of individual sound signals in both ears to locate the source [1]. This type of experiment began the modern investigation of acoustic cues used to source and localize sound.

In the first half of the 20th century, other psycho-acousticians such as Jeffress, Mills, Newman, Rosenzweig, Wallach and many others joined in and came up with further research that underpins today's understanding of sound source localization [2].

The results of research in acoustics were used as early as World War I for sound detection and location systems to identify and locate potential firing incidents and to set up an appropriate response. Those systems grew from simple microphone configurations used for position estimation to complex arrays that relatively accurately located gunfire at long range [3].

In later years, methods for audio pattern recognition and audio signal analysis, data pre-processing, feature extraction and classification algorithms were added. There has been a shift from audio analysis of recorded audio to real-time audio analysis.

Such advanced systems can now be used as an effective means of audio monitoring, i.e. detecting, locating and classifying audio events for security purposes. They can significantly increase the efficiency and effectiveness of security systems and contribute to prevention, especially when integrated with other security technologies, such as video surveillance systems, alarm systems or active voice loudspeakers.

Modern audio detection systems can reliably detect sound, identify its source from a selected list, locate its position and transmit this data in real time either directly to security guards or trigger a subsequent automated process through integration with downstream communication or security systems. Category lists of detected sounds for security purposes most commonly include gunshots, explosions, human screams, glass breaking and various alarm sounds such as car or

security system alarms.

The application of security audio-detection systems is suitable in public areas of cities and towns, in the premises of transport hubs and public transport vehicles, in buildings such as schools, hospitals, banks, office centres, shopping centres etc. Complementing the traditional security technologies, used in public areas outside or inside buildings with sound detection significantly improves their effectiveness. It brings a much greater situational awareness, brings a proactive approach to these systems and thus contributes to preventative security. It also contributes to more efficient communication and work of security forces and emergency services personnel. By providing additional situational awareness, sound detectors are also an important factor in eliminating false alarms when integrated with other security technologies.

The integration of security technologies into cooperating units has the great economic benefits for the operator, as well. Seemingly higher input costs bring savings in the long run, because some damages do not occur at all due to prevention. This is true when it comes to protecting assets. When it comes to protecting health or human life, the benefit is obvious.

4 Benefits of sound detection for security purposes

The vast majority of security monitoring systems are based on video surveillance using cameras. This has the great advantage of providing visual information, which is the most conclusive material (we trust what we see). However, the use of audio detection and monitoring systems has some undeniable advantages over video monitoring.

The biggest advantage is that while standard cameras have a limited angular field of view, microphones can be omnidirectional, i.e., with a spherical monitoring field. (Eyes can only see in one direction, ears can hear in all directions)

The second important factor that speaks in favour of use of the audio monitoring systems is the fact that some audio events, important for surveillance, such as screams or gunshots, have little or no visual counterpart. These are events that have little or no visual manifestation.

Other advantages are the independence of sound from lighting conditions and visibility, for example in adverse weather conditions or in the dark. Although the video technology has undergone great developments in recent years, using technologies such as thermal imaging and highly sensitive image sensors to advance the use of video technology even in very poor visibility, it is still limited in this respect. Another advantage of audio monitoring systems is their low costs of transmission over a data network. The bitrate of the processed audio is smaller than the video bitrate. This can be an advantage for larger installations of audio

detectors, for example in urban environments where the complicated conditions for data signal transmission are expected.

It could also be noted here that from a psychological point of view, audio surveillance is perceived as less invasive than video surveillance and can be a valid substitute in all the situations where privacy is a concern. To this end, it is important to note that audio surveillance does not usually involve automatic speech recognition [4].

5 Forms of audio detection from a data network architecture perspective

Historically, the digital audio monitoring systems have evolved architecturally by processing data streams from audio sensors on separate dedicated servers with software developed for this purpose. Such an architecture with a central server brings the advantage of high computing power for audio processing and analysis. However, it also has many disadvantages. The sound detected by the sensor has to be transmitted to this server over the data network. However, there may not be a good data link between the monitoring site and the server itself and transmission with minimal loss is expected for reliable sound detection and analysis. In recent years, with development of the small computers and the rise of edge computing, a new variant of network architecture is also coming to the fore for audio monitoring systems. The audio is analysed in a distributed manner directly at the point of detection and only the metadata in the form of extracted information and alarms is sent to a central server. This architecture saves the computing power of the server and is also significantly less demanding on the capacity of the transmission network between the detector and the server. However, this option requires each sensor to be equipped with hardware with computing power, power and data connectivity.

Considering the facts stated in the previous two options, a third option is logically offered, which uses other security system equipment already installed at the location of the intended audio monitoring of events. The use of surveillance cameras is offered as the best solution for several reasons. Firstly, they are already electrically powered devices, equipped with data connectivity and in some cases also an application platform with computing power ready to run additional applications directly on the camera. Currently, some manufacturers of security surveillance cameras offer such an open environment [5].

Another significant advantage is the possibility of logical integration of events detected by audio analysis with video verification directly at the level of the camera and transferring this information to the monitoring center or security control room comprehensively. This aspect is further discussed in the following section.

6 Technical aspects of the use of sound detection and classification

6.1 Noise and signal level of detected sounds

For an initial assessment for possibility of deploying and successfully using the sound detection and classification for security purposes, it is important to first map the acoustic environment of the site where the sound detection is planned to be used. This is particularly the case for common ambient sounds, as noisy environments make sound detection difficult. Considerably higher success rates can be expected in indoor environments where the amount of background sounds (noise) is limited or known. In outdoor environments, especially in urban environments, detection accuracy will be limited by the amount and intensity of background sounds. In addition, in outdoor environments, sounds caused by weather conditions must be considered. Many factors influence the quality and thus the success rate of sound event detection. Sound pressure levels in decibels (dB) typical in different acoustic environments and for some specific sound sources (measured at a distance of 1m) are visualized in Figure 1.

The most important is the so-called signal to noise ratio (SNR). This is a measure that compares the level of the desired signal with the level of background noise. The SNR is defined as the ratio of signal power to noise power expressed in decibels. The SNR is affected by several factors. The most important ones are the background noise level, the signal power of the event being monitored and the distance of the microphone from the sound source. All these factors make the problem of background sound subtraction a challenging task [4].

6.2 Measuring the accuracy of sound detection and classification

The detection, classification and localization of acoustic events for security purposes are the focus of many researchers. Experiments focus on the accuracy of detections in different acoustic environments while using different audio analysis algorithms. For accuracy tests, four different well-known audio detectors, based on different sound processing algorithms, are most commonly used:

Impulse detector - detects sound based on short-term signal levels - used mainly to detect sudden, loud impulse sounds.

Speech detector - detects sound based on the harmonicity of the signal - mainly used to detect speech and human screams.

Variance detector - detects sound based on changes in signal characteristics over time - particularly suitable for detecting sudden narrowband changes in the signal being analysed

Histogram detector - detects sound based on the differences in the energy distributions for the signals

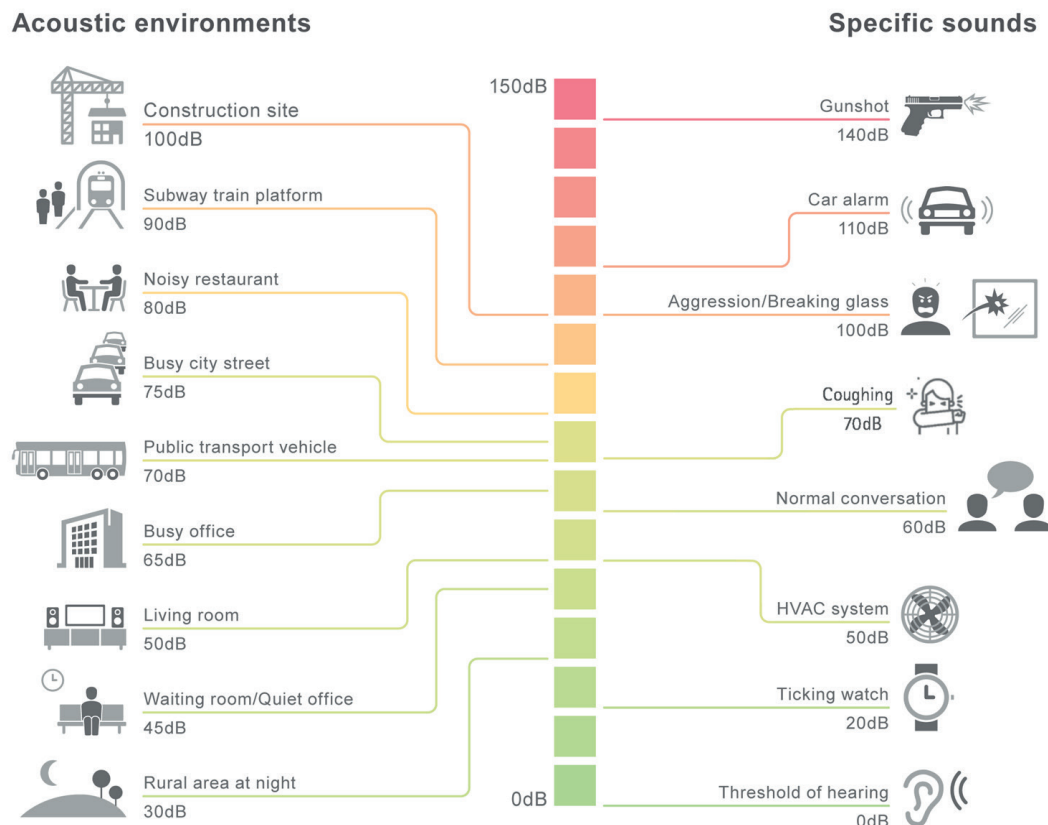


Figure 1 Sound pressure levels in decibels (dB) typical in different acoustic environments and for some specific sound sources (measured at a distance of 1 m), [6]

overall difference between the event and background spectra - particularly suitable for detecting any abnormal sounds (uses a histogram in 1/3 octave frequency bands to model the spectrum of the acoustic background).

For example, Polish researchers from the Technical University of Gdansk in their experiment investigated the accuracy of detection, classification and localization of sound events [7]. To test the detection and classification accuracy, they used four classification classes of threatening events C1 to C4 and one classification class of non-threatening events C5 (other sounds).

For the background sound, the test used four simulated environments with typical background sounds, two outdoor and two indoor:

- outdoor scene with traffic noise recorded on a busy street (traffic)
- outdoor scene with railway traffic noise recorded at the station (railway)
- indoor scene with noise from cocktail parties recorded in the university dining hall (cocktail-party)
- indoor scene with noise recorded in the main auditorium of the university (indoor).

Each sound background has its own characteristics. To compare them, it is necessary to normalize their sound spectra energetically. As can be seen in Figure 2,

used are clearly visible. The sound background typical of the indoor scene has energy concentrated in the middle part of the spectrum (200 Hz-2000 Hz). The very high level of tonal components for the railway noise was due to the train braking.

They used recordings of typical sounds of hazardous situations - gunshot, explosion, broken glass and human scream - as the types of sound events they investigated [7]. They used the ratio of True Positive (TP) and False Positive (FP) detections to determine the detection accuracy rate. The detection rate is then equal to the number of correctly detected events (TP) that match the events in the Ground Truth (GT) reference list divided by the total number of events in this list. A correct detection (TP) is then understood if the difference between the detection time and the event time in the GT reference list is less than 1 second. An incorrect detection (FP) is understood when an event is detected that does not match an event in the GT reference list and is classified as one of the four event types of the reference list (C1 to C4) [7].

Figure 3 shows the correct detection rates for different classification classes and types of acoustic events. It can be seen that, on average, the detectors perform best in the presence of cocktail-party noise compared to other types of disturbance signals. The

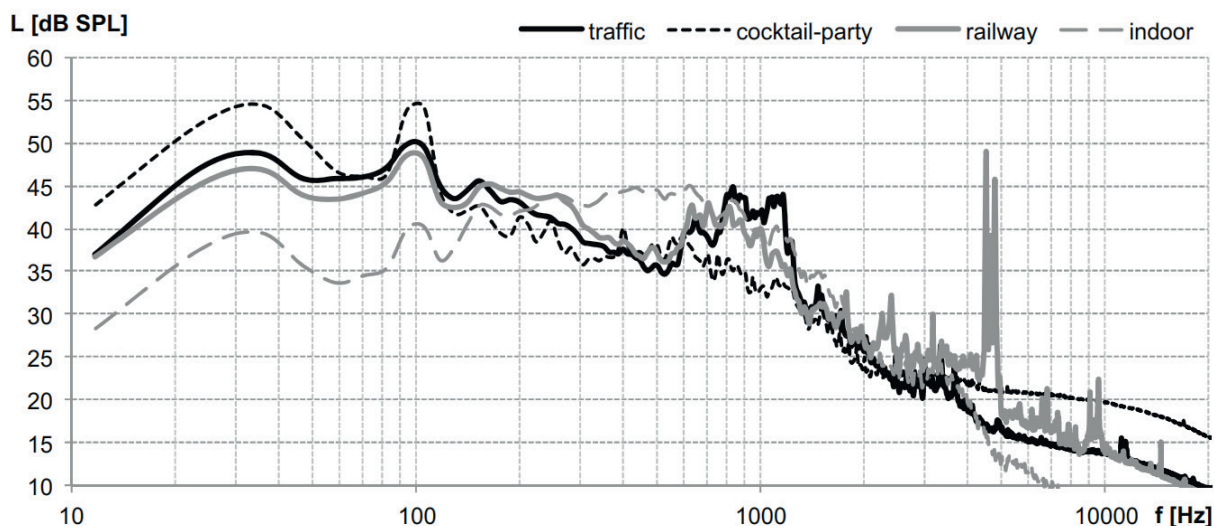


Figure 2 Energy-normalized spectra of individual background sounds, [7]

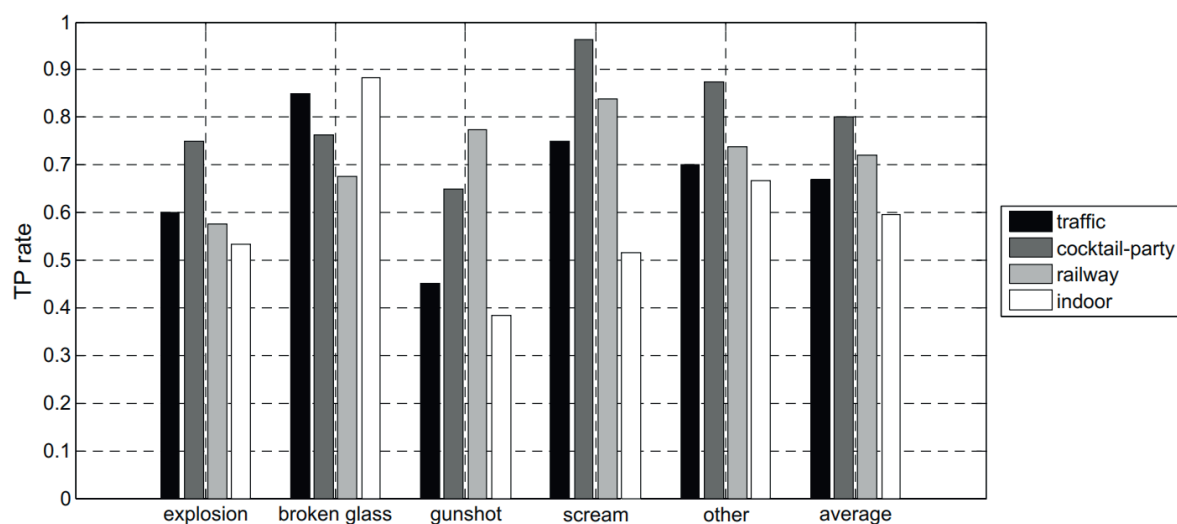


Figure 3 Correct detection (TP) rates for different acoustic event classification classes and types of acoustic events. [7]

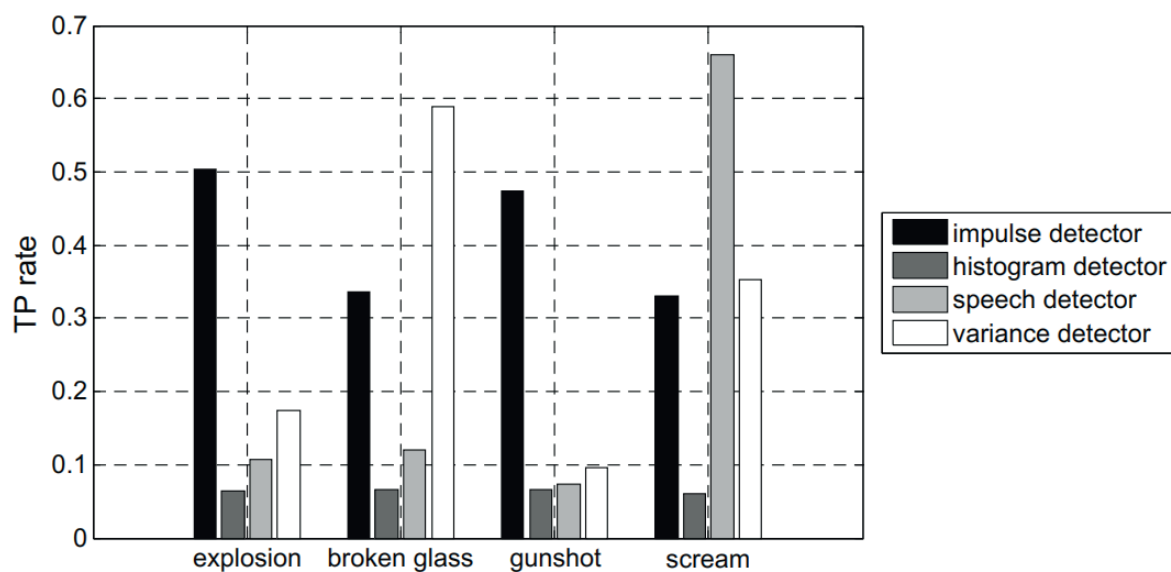


Figure 4 Correct detection (TP) rates for different classification classes and types of acoustic events, [7]

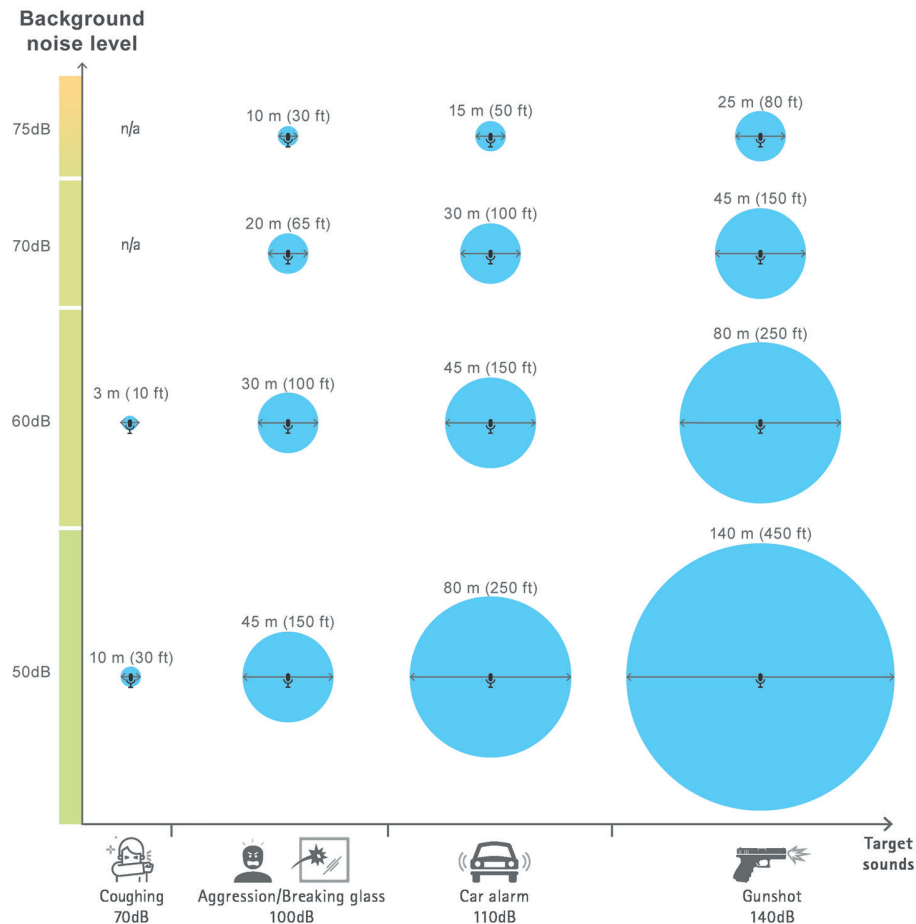


Figure 5 Typical sizes of areas covered by effective sound detection for security purposes for different cases based on experience, [6]

worst detection rates are achieved in simulated noise typical of indoor environments. (Indoor) It can also be observed that some classes of acoustic events are strongly masked by specific noise types. For example, gunshots have a TP level of 0.45 in the presence of the traffic noise (traffic) and 0.74 in the presence of the railway noise (railway).

Figure 4 shows how different detection algorithms performed in the test in recognizing different event classification classes. The results in the graph indicate the average precision detection rates (TP) for all SNR values. The presented dependencies show that the different detection algorithms used are complementary and suitable for recognizing the specific types of events. For example, the speech detector responds to the tonality present in screams, while the variance detector responds to sudden changes in sound features associated with a glass breaking event [7].

6.3 Spatial coverage by sound detection

The size of the microphone coverage area depends to a large extent on the signal-to-noise ratio (SNR). The diameter of the circle around the microphone is

used to measure the detection coverage, i.e. the area where the microphone can detect sounds. When the background noise (noise) increases, detection coverage decreases. Sounds that are desirable to detect, such as aggressive voices or gunfire, must be 10-30 dB higher than the background noise level. The information below is intended as an informative guide and is based on many years of experience of implementing the sound detection solutions, (Figure 5). The X-axis shows the most commonly detected sound sources for security purposes and the Y-axis indicates increasing ambient noise levels in the detected area.

7 Integration with surveillance system

Improving the safety and security of the public transport system is a top priority for transport companies, which are installing camera systems for this purpose. However, the number of these surveillance cameras is constantly increasing and their monitoring requires too much workload for operators to maintain a high level of attention and short reaction times. Over the last decades, many software developers and researchers have been involved in developing image processing tools that

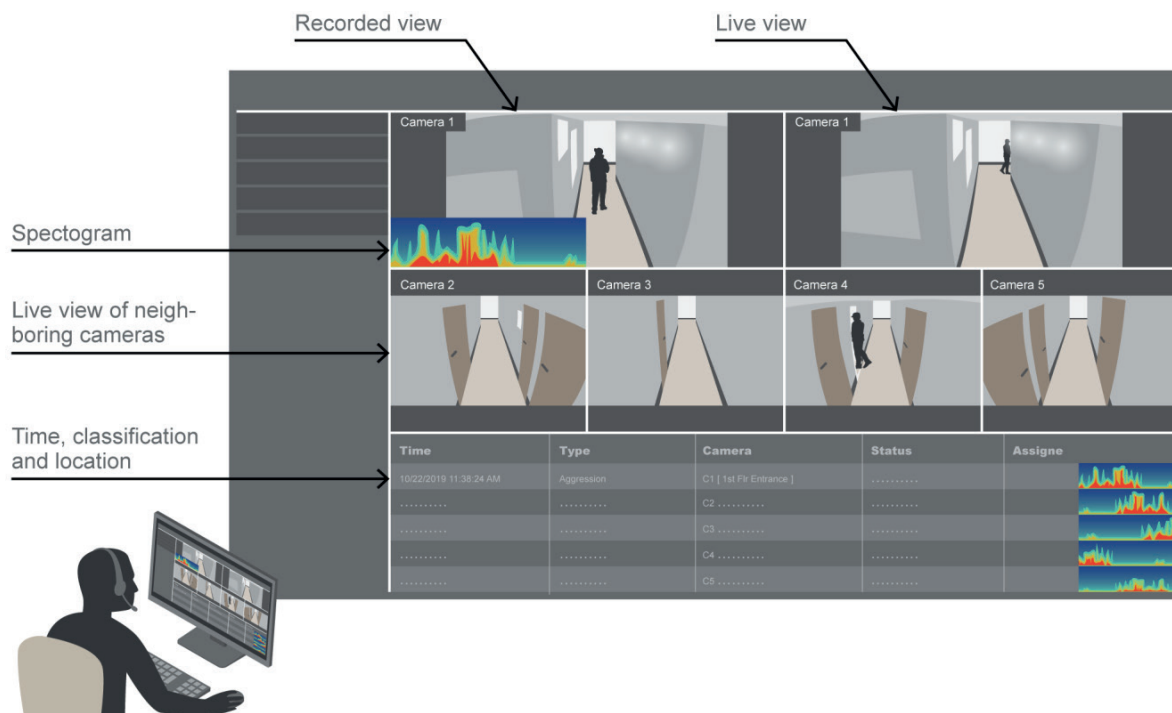


Figure 6 Example of integration of audio monitoring with a video surveillance system in a surveillance center, [6]

help the operator to automatically detect suspicious and dangerous situations. Many video analytics software applications, specifically designed for the transportation environment, have been created and have been validated in the real-world situations [8]. Analyzing audio events in a very effective way increases the efficiency of these surveillance systems.

As already mentioned, the combination and integration of safety technologies significantly increases their efficiency, reliability and preventive effect. The integration of audio monitoring with a video surveillance system brings the greatest synergy. In general, during the security incident, the situation can evolve rapidly and is usually not static. The security operator should be able to visually verify the situation quickly and easily. To achieve this, it is recommended to automatically link alarms triggered by the sound detection with live and recorded video provided by the CCTV system. If only live footage is displayed during detection, it may be that the perpetrator has already moved out of the field of view of the camera. Therefore, it is important to display the recorded video starting a few seconds before detection (video buffer) and a live view of surrounding cameras that cover the locations where the perpetrator may have moved. This is shown in Figure 6, where one can see the surveillance wall in the security operations center and the recorded view and live view windows at the top of the wall when an alarm triggered by a sound detector is detected. Below them, one can see automatically displayed windows with live view of surrounding cameras to get a situational overview. Below them, the recorded alarm data time, sound

source classification and location are displayed for the quick response and future possible forensic analysis. As additional information, a spectrogram is also displayed on the screen and recorded for a visual representation of sound levels at different frequencies over time.

8 Existing studies on acoustic event detection (SED) for the traffic safety purposes

Historically, there have been several studies that focus on combining audio and video analytics to enhance security. Some of them even directly in the traffic environment. This section examines these in order to evaluate the suitability of using audio and video analysis to enhance safety in traffic and to describe the current limitations.

In 2007, a team of three researchers from the University of Verona addressed the question of whether the recognition and detection of audio and visual patterns in a scene and the subsequent integration and synchronization of this audio and visual data are possible and which method is effective. Experimental results on real sequences showed promising results [9].

Separate audio and video signals were processed using the two different adaptive modules aimed at accounting for audio and visual information in a unique way, using only one camera and one microphone. Then, the two models were integrated using the concept of synchronization to achieve audio-video event recognition. This fusion was implemented using the AVC matrix, a feature that allows both to detect and

Table 1 Accuracy of audio event classification for four different scenarios [9]

| Scenario | Audio | Video | Audio-Video |
|----------|---------|--------|-------------|
| A | 100.00% | 83.35% | 100.00% |
| B | 60.87% | 95.65% | 95.65% |
| C | 95.24% | 85.71% | 95.24% |
| D | 62.12% | 66.67% | 89.39% |

segment AV events as well as to discriminate between them. Experimental results on real sequences showed that detection of this type is applicable in practice. From today's perspective of using neural networks and deep learning, these are early experiments, but they were already applicable. Table 1 shows the classification accuracy of audio and video analysis in four scenarios with increasing complexity. The results show a relatively high level of accuracy.

A team of researchers from France has investigated the effectiveness of a combined audio-video surveillance system for an automatic monitoring in public transport vehicles. This was in 2006, as a part of the SAMSIT (Système d'Analyse de Médias pour une Sécurité Intelligente dans les Transports publics) project, which aimed to design a solution for automatic surveillance in public transport vehicles (e.g. trains and metro cars) by analysing human behaviour based on interpretation of the audio-video stream. Its aim was to take into account the specific transport environment for the design of effective security surveillance systems.

The system consisted of six modules: face detection and tracking, audio event detection and audio-video scenario recognition. The audio event detection module detected abnormal audio events, which were precursors for detecting security incident scenarios that were predefined by end users. The audio-video scenario detection module performed the high-level interpretation of observed objects by combining audio and video events based on spatio-temporal reasoning. System performance was evaluated for a series of predefined audio, video and combined events.

Their study concluded that despite the challenging visual conditions, the system was able to successfully recognize several scenarios. They further stated that much work is still needed to develop video and audio algorithms to obtain a reliable system. They further stated that the future direction lies in improving the ability of the video algorithms in particular to cope with the complex lighting conditions in moving vehicles [9].

An interesting study in this regard was presented by a group of researchers from Ankara Science University in 2018 at the 6th International Conference on Control Engineering and Information Technology (CEIT). For their research, authors developed deep neural network (DNN) models for scream and traffic accident recognition. Tests of their models showed that they can be reliably used in real-world applications and in transportation [10].

However, they also pointed out that although there are already a number of papers that show the possibility of using audio as a source for video surveillance security, there are still few studies that address cost, scalability and practicality. In general, recent work tends to perform such tasks on supercomputers equipped with powerful GPUs and using complex neural network architectures. Thus, for wider use in practice, algorithms and, more importantly, their computational power requirements will need to be streamlined. A suitable solution is to distribute these algorithms from the central servers directly to sensors (cameras, microphones). The most modern terminal devices already allow this with their performance and chipset architecture.

In the last few years, Intelligent Traffic Monitoring Systems (ITS) have also been developed, which, among other things, aim to improve the road safety by ensuring a timely response to events such as traffic accidents and congestion.

The research on the effectiveness of different algorithms for audio detection in traffic, specifically within the framework of Intelligent Traffic Monitoring Systems (ITS), is the subject of a 2020 paper by a team of researchers from the University of Skopje, who conducted a comparative study of different audio algorithms suitable for practical use for these purposes. Their goal was to design a robust system capable of detecting traffic audio events in a real environment. At the core of this system is a deep learning model capable of detecting anomalous events and classifying them based on their acoustic characteristics. The results showed that the proposed model can be successfully applied in practice and that algorithms using the convolutional neural networks (CNN) are the most suitable for the detection of acoustic events [11].

9 Conclusion and future work

The aim of this paper was to discuss, describe and investigate the evolution of sound detection and its application for security, starting from the original knowledge about the origin and physical characteristics of sound to the current multifunctional use of modern technologies and systems. Based on several studies from the transportation field, the authors evaluated the current potential of using sound event detection (SED) systems specifically for the purpose of improving the transportation safety in practice and present its

conclusions to the readers. The idea of combining video and sound to make security surveillance and warning systems more effective has been the subject of several research in history. Some of them even directly in the transport environment. The aim was to evaluate their suitability for use and to describe their current limitations.

Over decades of development and research in the field of physical acoustics, a number of algorithms have been developed that can be effectively applied in different cases and environments. The outcome of most of the research has been the realization that improvements in the performance of the algorithms are necessary for an effective use. For audio analytics algorithms, this is primarily reliability in background noise environments, i.e., an increase in the signal-to-noise ratio (SNR). For video analytics algorithms, this is primarily their ability to cope with complex lighting conditions in the image, which is typical in moving vehicles, for example [12].

In this respect, technology has advanced significantly in the last few years. This is mainly due to the multiplication of computational power in surveillance cameras due to more powerful graphics cards chipsets (GPU) and thus improved image parameters for video analysis. Development of video and audio analysis algorithms has also advanced significantly, mainly due to the use of artificial intelligence principles such as deep learning and neural networks. Thus, the use and effectiveness of the combination of audio and video detection and recognition for security purposes, not only in transport, has moved forward significantly.

In the last few years, Intelligent Traffic Monitoring Systems (ITS) have also been developed to, among other things, improve the road safety by ensuring a timely response to events such as traffic accidents and congestion. This field is also beginning to make significant use of a combination of audio and video

detection for the automated acquisition of reliable traffic data.

Sound events detection systems are proving to be a very effective technology for improving traffic and city safety with a strong preventive effect. Especially in conjunction with video surveillance systems, they theoretically provide an exceptionally effective safety tool. Over decades of development and research in the field of physical acoustics, several algorithms have been developed that can be effectively applied in a variety of cases and environments. It is surprising that acoustic detection, classification and localization has not been used widely and massively to a similar extent as the video surveillance systems have been over the decades. According to information on urban and traffic security systems, the use of audio detection systems is very poor. For example, according to a survey of 110 small and medium-sized municipalities in the Czech Republic in 2020 and 2021, not one of them is considering deploying an audio detection system for their city and traffic safety, even though all of them either already have or are planning to build a video surveillance system [13].

This paper dealt with the problem of audio event detection and its combination with video surveillance systems purely from a technological point of view. Legislation and processes applicable to the transport sector will certainly be another important factor in expanding their use.

Further research is suggested here in the form of finding out the reasons for such a low deployment of audio detection systems in transport and cities, as well as research into the percentage increase in safety in specific use cases in transport and urban public spaces when audio detection technology is deployed in conjunction with video surveillance systems or other technologies.

References

- [1] BORING, E. G. *Sensation and perception in the history of experimental psychology*. New York: Appleton-Century-Crofts, 1942.
- [2] BLAUERT, J. *Spatial hearing: the psychophysics of human sound localization*. Cambridge: MIT press, 1997. ISBN 0-262-02413-6.
- [3] GRAVES, J. R. Audio gunshot detection and localization systems: history, basic design and future possibilities. ProQuest Dissertations and Theses. Denver: University of Colorado at Denver, 2012.
- [4] CROCCO, M., CRISTANI, M., TRUCCO, A., MURINO, V. Audio surveillance: a systematic review. *ACM Computing Surveys* [online]. 2016, **48**(4), 52. ISSN 0360-0300, eISSN 1557-7341. Available from: <https://doi.org/10.1145/2871183>
- [5] AXIS camera application platform (ACAP) - Axis Communications [online]. White paper. 2021. Available from: https://www.axis.com/files/tech_notes/wp_ACAP_en_2103.pdf.
- [6] Audio analytics for professionals - Sound Intelligence Design Guide [online]. 2021. Available from: <https://www.soundintel.com/>
- [7] LOPATKA, K., KORTUS, J., CZYZEWSK, A. Detection, classification and localization of acoustic events in the presence of background noise for acoustic surveillance of hazardous situations. *Multimedia Tools*

- and Applications* [online]. 2016, **75**, p. 10407-10439. ISSN 1380-7501, eISSN 1573-7721. Available from: <https://doi.org/10.1007/s11042-015-3105-4>
- [8] ROUAS, J. L., LOURADOR, J., AMBELLOUIS, S. Audio events detection in public transport vehicle. In: IEEE Conference on Intelligent Transportation Systems ITSC: proceedings [online]. IEEE. 2006. p. 733-738. Available from: <https://doi.org/10.1109/ITSC.2006.1706829>
- [9] CRISTANI, M., BICEGO, M., MURINO, V. Audio-visual event recognition in surveillance video sequences. *IEEE Transactions on Multimedia* [online]. 2007, **9**(2), p. 257-267. ISSN 1520-9210, eISSN 1941-0077. Available from: <https://doi.org/10.1109/TMM.2006.886263>
- [10] ARSLAN, Y., CANBOLAT, H. Performance of deep neural networks in audio surveillance. In: 6th International Conference on Control Engineering and Information Technology CEIT: proceedings, 2018. p. 1-5.
- [11] CHAVDAR, M., GERAZOV, B., IVANOVSKI, Z., KARTALOV, T. Towards a system for automatic traffic sound event detection. In: 28th Telecommunications Forum TELFOR: proceedings [online]. IEEE. 2020. Available from: <https://doi.org/10.13140/RG.2.2.19496.29448>
- [12] VU, V.-T., BREMOND, F., DAVINI, G., THONNAT, M., PHAM, Q.-C., ALLEZARD, N., SAYD, P., ROUAS, J.-L., AMBELLOUIS, S., FLANCQUART, A. Audio-video event recognition system for public transport security. In: International Conference on Crime Detection and Prevention ICDP'2006: proceedings. 2006. p. 414-419.
- [13] Help protect your town – Blind Spots / Pomozte chránit vasi obec - Slepa Mista (in Czech) [online]. 2021. Available from: <https://www.slepamista.cz/>