MINISTERUL EDUCAȚIEI

UNIVERSITATEA TEHNICĂ
DIN CLUJ-NAPOCA

Electronics Engineering, Telecommunications and Information Technologies

DOCTORAL THESIS

– SUMMARY –

Acoustic sensors for geographically isolated area surveillance

PhD student:
Ing. Olimpiu Pop

PhD supervisor
Prof.dr.ing. Corneliu RUSU

- Cluj-Napoca -
2024

1. Introduction

The forest is one of the oldest living ecosystems on Earth. Approximately 20% of the total volume of carbon emissions is due to uncontrolled deforestation, the criminal phenomenon being omnipresent in this field. In 2018, the control authorities managed to identify only 1% of the total volume of illegal forest cutting in Romania, i.e. approximately 200,000 cubic meters of the 20 million cubic meters of wood that disappeared every year without legal documentation from the country's forests. By protecting forests and wooded areas, using modern technologies, we can make a significant contribution to the fight against climate change and biodiversity conservation.

Due to the specific conditions that characterize isolated geographical areas such as electricity supply, poor coverage of 3G, 4G data networks and weather conditions, implementing an efficient and reliable technical solution is a real challenge. Compared to image sensors that can be severely limited and unpredictable due to specific environmental conditions (light, day/night, fog, dense vegetation), plus the large amount of data generated, audio data can be provided robustly using a small volume of information and low-capacity communication channels.

The acoustic sensor network can also be used to study and monitor various wildlife species. Audio detection involves the use of sensors sensitive enough for the intended purpose, at a reasonable price. Recognizing the types of audio signal sources is relatively difficult to achieve because in such an environment the signal-to-noise ratio of the sound signal usually has low values. Also, the monitored acoustic environment may contain a large number of other sound sources, which are not necessarily of interest. The methods using passive acoustic recording have a great potential for monitoring different species of living things. A major challenge in this case is the very large variety of background sounds and noises depending on the geographical area and the specific vegetation where the monitoring is carried out.

There is increasing interest in the use of dedicated, high-energy-efficiency sensors in various large-scale and long-term monitoring applications of natural resources and criminal phenomena with environmental impact (e.g. poaching, illegal logging). Research in this field involves the creation of energy-independent sensors that use techniques to transform renewable sources through the use of solar panels or wind turbines. The amount of electricity these charging techniques can produce is limited by the surrounding environment. This aspect can
greatly influence the process of generating the electrical energy required for the operation of the sensor

Currently, acoustics is used in most research projects in the field of ecology and environmental conservation. An example of equipment available in the market is the Acoustic Song Meter. Due to the exceptional quality of the microphones, this device has a high price ($800-1000), which is why it cannot be considered an affordable product, its use being restricted to research projects with a considerably high budget. For a series of applications used in research projects or for surveillance, with a low budget, there is no need for a high precision of the device, this being an aspect that does not influence signs
another disadvantage of these commercial devices is the fact that they are made with closed program sources, which is why they are difficult or even impossible to adapt to the particular requirements of the hardware or software specific to another application. For the reasons listed there is a growing trend among research communities to design and use custom acoustic monitoring equipment

Due to the complexity of the construction of these devices, average researchers without considerable electronics skills could not make them. The situation changed with the advent of modular microcomputers, which can provide the necessary framework for customizing and optimizing the requirements of an application.

Alternatively, there are acoustic monitoring devices on the market that solve these problems. They are low-cost open source devices that meet the requirements for such an open and flexible platform for acoustic monitoring applications. One such product is AudioMoth, with an open source hardware and firmware with permissive licenses, which allow it to be updated and expanded for any application. With the help of this device, monitoring of a geographical area can be achieved by long-term recording of certain sounds filtered with a sound source identification algorithm (Deploying Acoustic Detection Algorithms on Low-Cost Open-Source Acoustic Sensors). The disadvantage of this device is the fact that it only operates a single omnidirectional microphone which reduces its ability and performance to identify the sound source and excludes the possibility of determining the direction of arrival of the sound signal.

2. Previous results of the research group

For several years, the effort of the research team I belong to (the Signal Processing Group of the Technical University of Cluj-Napoca) has focused on creating a monitoring system for isolated natural habitats through sensor networks.

- People have an important role in ensuring the integrity of forests and wild places. Although these regions are protected by law, they are often targeted by criminals and poachers for logging and hunting. Even the disturbance of wildlife by some people could damage endangered species.
- Not only forested land reserves are the target of these illegalities, but also protected lakes or coastal regions (such as the Danube Delta) are places for illegal fishing or hunting of bird species strictly protected by international laws.

In general, wilderness monitoring and intruder detection systems are currently needed and would likely ensure better conservation of these protected areas.

For several years, the research group's work focused on the design of an acoustic monitoring system. Its use against other monitoring systems such as video surveillance has several advantages:

- simplicity of implementation;
- less information to process;
- does not depend on ambient light;
- represents a much cheaper solution.

Much of the effort of the research group focused on the development of processing procedures for audio signals acquired in the natural environment. The results were encouraging, concrete and in several doctoral theses supported, however, the problem of acquiring the audio signal from the natural environment in an efficient way remained open. The present doctoral thesis is mainly addressed to this subject.

Many of the problems of processing the acquired audio signal were the subject of other works by members of the Signal Processing Group of the Technical University of Cluj-Napoca. For this reason, we will not focus on these aspects, mentioning them to the extent that they are important for our objectives.

3. Objectives

Applications based on acoustic sensors collect information by capturing the sound signal that is later used by the processing algorithms. By adapting the acoustic and electrical parameters of the sensor, respectively a set of sensors, to certain types of sound signals, the efficiency of the system can be increased by possible to integrate an in situ data processing. By filtering and processing the primary information, the main parameters are extracted from the recorded acoustic signals that will be used in the subsequent processing. In this way, the volume of data is reduced, making it possible to send them to a collection point for further processing. The transmission of an optimized data volume means a lower consumption of energy and radio bandwidth, a very important aspect for the development of a network of smart, autonomous sensors, sustainable from the point of view of electricity supply

To begin with, we propose to create an experimental model for performing measurements in the laboratory and in the field. With the help of this test model we will be able to determine the characteristics of some microphone sets in different configurations and environmental conditions, such as directivity, frequency response, signal/noise ratio. Another proposed goal of the thesis is to create an equipment capable of recording in audio file format several signals (in our case there will be four in the end, coming from the four microphones of the proposed quadraphonic sensor) and transmit this file through a data transmission network (internet over IP) in the first phase upon request , we are going to develop the application to be able to do this automatically, when a sound source appears in the supervised area. The operation of the experimental model at the current stage of the research requires the existence of a data network in the area supervised geographical area, this requirement quite seriously limiting the efficiency of the application. Further, other solutions for data transmission such as a wireless network with a central communication point that has access to a data network can be sought.

In the last part, we will analyze the results of measurements carried out in the laboratory and in the field in real conditions in order to optimize the experimental model (types of microphones, sensor geometry, filtering and sampling of signals, optimization of the transfer rate and reduction of the volume of transmitted data.)

4. Requirements and constraints

The choice of the optimal number of microphones for the given application must be made taking into account several criteria that, as a rule, cannot be fulfilled simultaneously. Depending on the intended purpose, a compromise must be made in choosing a set of criteria that will satisfy the requirements of the application.

A main characteristic of an acoustic sensor is the directivity. There is a great variety of microphones depending on their constructive directivity diagram: omnidirectional, cardioid, double cardioid, hypercardioid, etc.

Another important parameter is the sensitivity of the acoustic sensor. The combination of these two characteristics determines another parameter, perhaps the most important of the microphone: the signal/noise ratio (SNR - Signal to Noise Ratio) for the received signal expressed on a logarithmic scale, in decibels.

The solution to increase the signal-to-noise ratio for the received signal is to use a larger number of microphones with good directivity to limit the summation of ambient noise with the useful signal. It is practically feasible to attenuate the energy of the wave fronts coming in certain directions and implicitly also of the noise than to amplify them by passive processes those coming in a direction of interest.

5. Structure of the thesis

Starting from the previously mentioned hypotheses, in this doctoral thesis, sensor configurations capable of meeting the specific operational requirements were studied: efficiency, reliability, robustness and, last but not least, a convenient price, compared to other products on the market.

The acoustic sensor arranged in a certain configuration is the basic cell of the surveillance network, the element on which the network structure is built. A series of methods for determining the direction of arrival of the sound signal and identifying the sound source are cataloged in the specialized literature . A part is presented in this doctoral thesis of these methods that I have experimented with with the help of various microphone configurations. The constructive parameters have been optimized so that this sensor can be used effectively in most methods of determining the direction of arrival, always considering efficiency, reliability, robustness, as well as low consumption (energy and material)

Next, the thesis is structured in two large parts: a first part in which certain results from the specialized literature are presented with a predominantly theoretical character, and a second part in which the achievements of the doctoral student are detailed. Chapter 2 is dedicated to the methods of locating sound sources, and chapter 3 on the modeling of sound sources in space. The methods of locating acoustic sources using acoustic sensor systems are described in chapter 4.

The acoustic sensor with four microphones is the subject of chapter 5. It starts with a first implementation of the microphone array and the acquisition board, discusses the application based on the support provided by the Raspberry Pi, and then moves on to the second implementation of the acoustic sensor, i.e. to the mannequin head with 4 microphones. The chapter concludes with a set of results and comments.

Chapter 6 is dedicated to the modeling and practical implementation of the orthogonally arranged acoustic sensor. To begin with, the general principles for determining the transfer function of the human head are reviewed, the analysis of captured signals using the autocorrelation function is presented, as well as the identification of sound sources and the identification of frequencies of interest. The application of the Goertzel algorithm for the identification of sound sources concludes this chapter.

In chapter 7 the costs and perspectives of the proposed implementation are presented, and in the last chapter the personal contributions are summarized and the dissemination of the results is detailed.

6. Passive localization methods

The main passive localization methods analyzed in this work were: ILD-Interaural Level Difference, TDE-Time Delay Estimation, ITD-Interaural Time Difference, TDOA-Time Difference Of Arrival, HTRF-Head Related Transfer Function.

In the methods based on TDE and ILD, the calculation algorithm contains two stages:

1. estimating the delay time or intensity level differences;
2. calculating the position of the sound source.

Correlation is the most frequently used calculation method for estimating the delay time. With this method, the most important requirement is the approximation with the highest possible precision of the delay time of the sound wave between the pair of microphones. Also, with the method based on ILD, the most important requirement is the most accurate approximation of the sound intensity differences between the two microphones.

A major obstacle for calculating the position of the sound source is the complexity of the equations and the relatively high time of information processing. There are numerous methods and techniques that involve a large amount of calculation, which is why they cannot be applied in real situations, especially when we are talking about isolated geographical areas where there are no electricity and data networks**.**

7. Spatial correlation of microphone signals

If we focus on the noise component of the captured sound we will distinguish two components

1. the uncorrelated internal noise of the microphones which is a limiting factor for the spatial selectivity of a microphone array. A typical electret microphone produces a signal to noise ratio SNR of 60 dB with a reference level of 96 dB SPL (Sound Pressure Level).
   Considering the normal level of the human voice of 60-65 dB SPL and taking into account the noise introduced by cables, preamplifiers and converters we can say that the uncorrelated noise is at a level of 30 dB below the vocal sound level;
2. the external (ambient) noise captured by the microphones. The main difference compared to the previous noise is that this noise is the same for all microphones in the area, possibly slightly offset and delayed. In practice this is not true, especially in the upper part of the frequency band.

In conclusion, noise is a combination of correlated and uncorrelated components. The ratio between them is frequency dependent and varies depending on the geometry of the microphone area and the microphones used. The noise level varies between 5-20 dB. In general, the ambient noise is 10-25 dB above the internal noise level. Because of multiple reflections of the sound signal that arrives attenuated and delayed, the reverberation will obviously influence the degree of correlation.

8. Methods of locating acoustic sources using microphone systems.

The localization of acoustic sources is carried out by using several methods that are based on the calculation of some parameters of the signal received from the distributed sensors:

- calculating the differential amplitudes of the received signals;
- calculating the time difference for signals arriving at different sensors with known position (TDOA - Time Difference Of Arrivals);
- **determination and comparison of directions of arrival (DOA - Direction Of Arrivals);**

• **calculation of acoustic energy.**

Depending on the environmental conditions and the conformation of the ground which can produce reflections, reverberations, which can introduce attenuation, each of the methods mentioned above can be more or less effective. In the following, we will review each of the these four methods and at the end of the chapter we will discuss the influence of the sampling rate value in the localization of sound sources using acoustic sensor networks.

The localization method, based on the comparison of the amplitudes of the received signals, is strongly influenced by the conditions in the surrounding environment (for example, vegetation) and by the conformation of the land that behaves like anisotropic attenuators. Without the implementation of a complex model of signal attenuation, the value of the received signal amplitude is insufficient to determine the distance.

The second calculation method, based on TDOA, requires a precise acquisition of the phase of the signals arriving at the nodes of the network of acoustic sensors. This method involves processing all the signals received from the i nodes of the network in the central node to determine the position of the sound source. The method is based on the calculation of the arrival time difference, in the central node, of the signals from the i nodes. This arrival time determination is mostly done in the frequency domain using GCC-PHAT (Generalized Cross-Correlation POwer PHAse Transform). The method is quite precise, but it requires the transmission of the entire audio signal to the central processing node, reducing in this way the operating time of the node due to energy consumption. An important requirement of this method is that these signals must be precisely timed (differences below 1 μs). This desired is difficult to achieve if each node has an independent time generator. The non-synchronization between the nodes seriously affects the accuracy of this method.

The increase in the number of nodes and the processing power can be limited by doing a preprocessing of the data in the acquisition node, followed by the transmission of a reduced volume of data. In this case, the processing load of the central node decreases significantly.

The third method is based on the determination of the direction of arrival using in the nodes of the network the areas of microphones that are each capable of determining a direction of arrival independently of the rest of the nodes in the network. By combining these results, the position of the sound source can be estimated.

Extracting phase features with sufficiently good accuracy from a narrowband signal can become difficult if the environment is noisy. In addition, the technique based on coherent processing of signals from different network nodes (correlation method) is limited by acoustic coherence properties of the environment and does not work well if the nodes are more than 10m apart. In open, reverberant spaces over large areas, this method of localization can become ineffective.

The fourth method is based on energy calculation, more precisely on monitoring the acoustic energy around the node. In practice this approach is not very accurate and can also be affected by noise

9. The influence of the value of the sampling rate in the localization of sound sources using networks of acoustic sensors

Traditionally the sampling frequency used for sound sources is 44.1 kHz. This sampling rate, which respects Nyquist's law for the audio frequency range (20 Hz-20 kHz), allows the original signal to be re-converted into a high-fidelity acoustic signal. This sampling rate requirement is necessary to be able to make, for example, high-fidelity music recordings, where the frequency range of the acoustic signal is between 20 Hz and 20 kHz.

In the case of a network of acoustic sensors that is built to locate acoustic sources, the band of the captured signal can be even smaller. In the specialized literature and in the international standards for data transmission in wireless networks (IEEE 802.15.4), the maximum sampling rate is 17.7 kHz. This value can vary depending on the number of nodes in the network. Fulfilling the Nyquist condition in this situation generates a large volume of data, but not necessarily information, the occupied bandwidth being unjustified.

For example, at a typical sample rate of 44,100 Hz, with samples represented by 16 bits, the transfer rate is 705.6 kbps. This value, as I stated above, is much higher, almost three times, than the maximum allowed in the standard, of 250 kbps.

Several authors whose research object is wireless microphone networks use lower sampling rate values, without arguing the chosen value in any way. Depending on the mathematical algorithm chosen for locating the acoustic source, the optimal frequency of the sampling rate may differ.

As a conclusion of this section, we can say that the sampling rate influences the accuracy of the calculation of the time difference between the captured distributed audio signals, and the localization of the sound source can be achieved by different algorithms in the time or frequency domain.

The obtained experimental results show us that the algorithms that work in the time domain and use subsampled signals generate unstable results and lose information due to the low resolution on the time axis.

Algorithms that operate in the frequency domain (for example, algorithms that use the envelope of signal spectra), can use low sampling rates without significantly affecting the accuracy of the results. This is possible because these methods depend on the spectral content of the signal and not on the amplitude value of the signal.

After analyzing several algorithms, we can consider that a lower sampling frequency than the usual one for 44.1 kHz audio signals can be used depending on the algorithm used. This conclusion is very important and significantly contributes to establishing the criteria that must be taken into account when we build a network of wireless acoustic sensors.

10. First deployment of microphone array and acquisition board

To begin with, when making the set of microphones, I used a number of 4 ECM-670 shotgun microphones placed on a flat surface in the shape of a square, arranged diagonally at a 90 ∘ angle between them. The distance between the microphone capsules can be changed in a beach of 10-20 cm. This possibility of changing the distance between the capsules of the set of microphones is necessary when studying the phase differences between the acoustic signals captured for a certain bandwidth (maximum detectable frequency without phase ambiguity).

To eliminate confusion during measurements in the laboratory and in the field, I associated the microphone number with a cardinal point as follows:

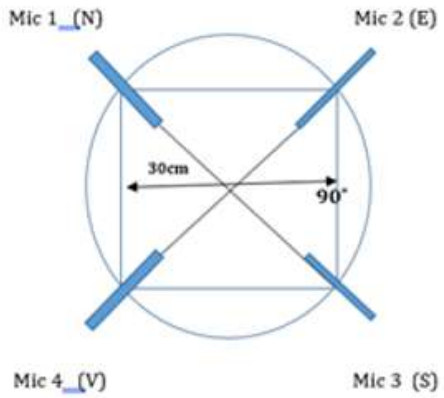

Figure 1. Geometry of the microphone set

The set of microphones connected to an audio acquisition board that has 4 balanced XLR inputs with phantom power (48V), necessary to power the four condenser microphones. Also, each channel has an amplifier with adjustable level and frequency filters. When making the measurements, we used a constant amplification level and did not introduce bandpass filtering in order to be able to evaluate the entire present spectrum. Also, in order to establish the signal/noise ratio in each situation, we did not use frequency band filtering. The sampling frequency for all recorded signals is 44 100 Hz.
Using the serial port of the acquisition board, the connection was made with the Raspberry Pi application, a versatile minicomputer that has, among other things, an Ethernet port for connection to a data network. Next the signal from the four audio channels were transferred to one
audio signal processing program Audacity, an open source program widely used for applications that require the capture and processing of audio signals. A simplified schematic diagram is shown in Figure 2.

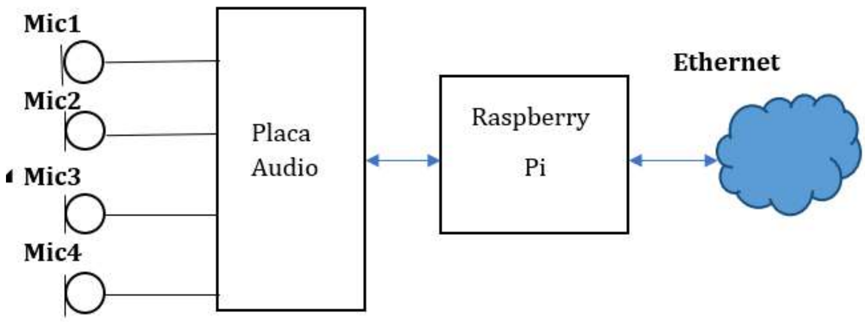

Figure 2. The acquisition system

11. The quadraphonic sensor

The realization of these microphone arrays must take into account several considerations:

• If this network is an ad-hoc distributed one, the microphones in the network nodes may not be calibrated and also their position may not be well established or even known. In this case the information related to the geometry of the network can no longer be used.

- In many applications the position of the microphones and their number may vary. In most cases there are very serious limitations regarding the bandwidth of the signals to be transmitted between the network nodes. Most of the time, the bandwidth available for transmission is insufficient, as there are dozens or hundreds of microphones whose signal must reach from one node to another or to a central point. For this reason it is very important that this bandwidth available for transmission is used as efficiently as possible;
- In the case of these distributed networks, it must be taken into account that due to the constraints related to the transmission band, the primary processing of the data in the network nodes is needed as efficiently as possible without affecting their content as much as possible, in order to reduce to the maximum the amount of information that must be transmitted, thus saving transmission bandwidth. In this way, the energy consumption, often limited, of the central processing node can be reduced.
- Another problem is the synchronization of the transmitted data. Since each sensor in the network nodes has its own system clock, differences will appear between the sampling rates of the signals from the microphones in the network. This problem will seriously affect post-processing using algorithms based on signal coherence.

These special requirements lead to the need for a different approach in terms of signal processing, their modeling and the algorithms used to implement microphone networks.

12. Second implementation of the quadraphonic sensor

Following the analysis of the acoustic source localization methods listed above and taking into account the parameters to be determined and calculated, we experimented with several types of mechanical constructions (platforms, reflectors, supports) and microphone models. The proposed goal was to obtain parameters of directivity, signal/noise ratio and linearity în band close to those of the microphones used in the initial experiments, but at a much lower cost. In this sense, I chose a condenser lavalier microphone, AKG C 417.

13. Dummy head with four receivers

The requirements that are imposed for the realization of the sensor cell are related to directivity, linearity in the band of interest, reliability, protection against bad weather. Equally important is that the signals captured by the sensors provide useful data that can be used in methods of determining the direction of arrival or locating the sound source.

Numerous methods of locating sound sources are based on a model built on the structure of the human body, namely the cranial box and the human auditory organs (auricle, auditory canal, eardrum, interaural distance, etc.).

One of the previously listed methods is based on the human head transfer function (HRTF). HRTF is the Fourier transform of HRIR and thus represents the filter characteristics of the ensemble formed by the head, ears, back and shoulders, due to the diffraction and reflexive phenomena of sound waves.

In our case, this filter had to be built and optimized to meet the requirements and the proposed purpose. More precisely, sound "targets", due to the fact that they are at a relatively large distance (of the order of hundreds of meters) and at a short distance from the ground, can be considered as being in a plane, that is, in a two-dimensional space. For this reason the directivity diagram of the sensor must be limited relative to this plane. Also, the information coming in other directions must be attenuated as much as possible, because they only contribute to the decrease of the signal/noise ratio. The opening angle of the directivity diagram in the plane must be limited to an angle smaller than 90 ∘ , for a system composed of 4 microphones. In this way, a network can be built with sensors in the corners of a quadrilateral (for example, a square). The accuracy of the localization is not critical, it can be done with errors of the order of tens of meters without affecting its result. Our goal is above all the most accurate identification and localization with a sufficiently good precision in a limited area.

Taking into account the above I built the following configuration composed of 4 microphones inserted in a mechanical structure. The 4 microphones are positioned at an angle of 90 ∘ to each other inside a feed horn type channel, at a distance of approx. equal to the distance between the eardrums (8-10 cm). On the two directions N-S and E-W there are a connecting channel so that the sound received in one direction is also picked up by the microphone located in the opposite direction, of course at another level and with a delay. This channel, whose size and shape will be established experimentally, has the role of attenuating the spectral composition of the signal captured by the microphone located on the opposite direction to the sound source (Figure 3). Their size must be chosen in such a way as to introduce an attenuation in the spectrum in the area that is not of interest from the point of view of the positioning of the fundamental harmonics of the sound sources that want to be identified and located. It will be established experimentally that this spectral zone is located in the range 1kHz-3kHz.

Figure 3: Dummy head with 4 receivers horizontal section and execution

In our experiment we made an orthogonal arrangement of the four sensors. We also implemented a system of orthogonal tubing inside the mannequin head through which we interconnected the four sensors (Figure 4). The role of these tubes is to eliminate unwanted effects such as reflections and reverberations that would occur if the cavities where the microphones are placed were closed. At the same time, the tubes in which the microphones are placed also act as a filter for the acoustic signal. The size of the tubes allows us to place the capsules of the 4 microphones at variable distances. In this way we can optimize the directivity diagram of the microphones in order to minimize the secondary lobes that appear depending on this distance between the microphones and that can generate erroneous results when determining the direction of arrival of the sound signal.

Figure 4: Dummy head with 4 receivers, practical realization

14. Results obtained

To model the parameters of the mannequin head with 4 receivers, several series of measurements were performed in different environments and using different sound sources. The recordings were imported into a previously mentioned audio signal processing and analysis application (Audacity). This program allows simultaneous recording of 4 or more audio channels.

There is also the possibility of processing this information oscillogram, spectral analysis, etc.). With the help of this program we generated the graphs that attest the results of the measurements.

The first set of measurements was carried out in an acoustically equipped room (concert hall with a noise level below 30dB) for which an amplified sinusoidal signal generator was used. A series of measurements were made using several frequency values in the range 50 Hz-5kHz and the frequency response characteristic of the dummy head with 4 receivers was determined under these conditions.

In all the measurements carried out I assumed that the sound sources are at a sufficiently large distance to be able to consider them coplanar with the plane determined by the set of microphones, a plane parallel to the terrestrial horizontal plane. We can consider this hypothesis correct in most cases, since the types of sound sources that we want to authenticate and determine their direction are at ground level (chain saws, machines of transport, weapons, animals of hunting interest, etc.). The spatial position of the sound sources will be determined in a plane and will be given by the angle of the arrival direction vector with respect to the axes of a Cartesian coordinate system.

Another hypothesis that we have to take into account is that the speed of the sound source is much lower in relation to the speed of sound, so that the precision of determining the direction of arrival is within the established error limits. This hypothesis is otherwise correct if we refer to the types of sound sources mentioned above.

In order to determine the frequency characteristic of the set of microphones, a multiburst signal was generated in the range of 50 Hz - 5 kHz, which we amplified and broadcast in the mentioned acoustic enclosure where the set of microphones was also located, at a distance of 10 meters , positioned in a plane parallel to the floor. This type of acoustic enclosure was chosen to eliminate the effects produced by the reverberation of the sound emitted by the sound source, more precisely the overlapping of the reflected wavefronts over the direct wave, a phenomenon that would produce phase differences and random amplitude differences between the 4 captured signals by the 4 acoustic sensors, thus affecting the

precision of the measurements.Analyzing the obtained data, we observe a linear frequency characteristic with signal level variations of 3-5 dB in the range 50 Hz-1400 Hz. We notice that at signal frequencies between 1400 Hz and 1600 Hz, the set of microphones behaves like a resonant system. Because this frequency range is not in the area where we expect to
we find sound sources that we want to monitor (as we will see later), the resonance will not affect the accuracy of determining the direction of arrival by generating harmonics that stand out in the spectrum and that could overlap a useful signal.

As a first conclusion, we note that the level of signals received on the 4 directions is different, with differences of 3-4 dB. The North direction was chosen as the direction of arrival for the test signal. We notice that in the opposite direction we have a level difference of 9-10 dB, and on the East-West axis of 3-4 dB. In ideal conditions and with this working hypothesis, we should have the same signal level on East and West sensors. The difference in level is due to the calibration error of the amplifier chain for the two signals. In order to have good accuracy even at low signal levels, especially, these amplifier channel calibrations need to be performed as precisely as possible.

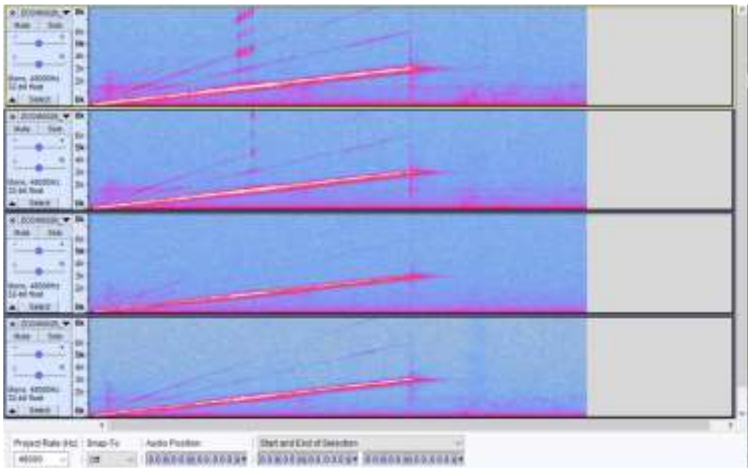

Figure 5: Multiburst signal captured with the set of 4 microphones, from the North direction

In the outdoor space, the noise will be composed of the correlated one, usually generated by the sound source itself, and an uncorrelated component, the ambient noise (wind, rustling leaves, insects, etc.). The energy level of the correlated noise, usually generated by the sound source itself, can be assessed by calculating the autocorrelation function of the received signal and can help determine the direction of arrival. Uncorrelated noise, however, depending on its level, can affect or even compromise the determination of the direction of arrival for signal sources located at greater distances. Through a mechanical construction adapted to reduce uncorrelated noise, its level can be reduced (for example, the use of sponge windbreaks specially dedicated for this purpose).

Next, we recorded sound signals in the range 50 Hz - 4kHz to be able to analyze even more precisely the frequency characteristics of the quadraphonic sensor, emphasizing the frequency band in which we will have sound signals of interest. An example is presented in Figure 6.

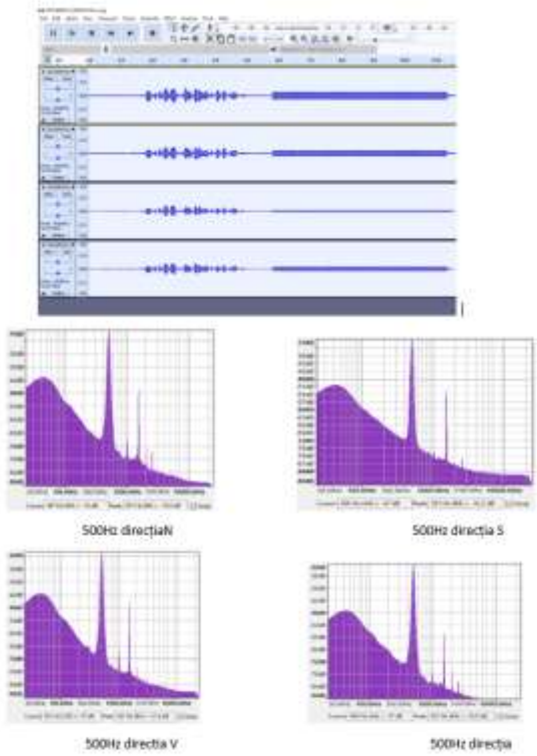

Figure 6. Frequency characteristics for a sound source of 500 Hz..

After analyzing the experimental results and calculation methods for determining the direction of arrival of the acoustic signal, we built an optimized model of the quadraphonic sensor. It started from the human head model for which we added an additional number of two orthogonally arranged sensors. The directive microphones (expensive) were replaced with omnidirectional microphones (cheap) and their directivity diagram was modeled through the mechanical construction adapted for the frequency range in which the sound sources of interest are located in the case of surveillance of isolated geographical areas. Also, during its construction, the elimination of uncorrelated noise was taken into account as effectively as possible.

15. Analysis of captured signals using the autocorrelation function

To identify the sound source and determine the direction of arrival of the sound signal, the autocorrelation function was applied for the four signals from the four microphones arranged orthogonally in the ducts in the mannequin head. In order to reproduce the conditions as close as possible to the Reality in the field, a signal composed of a spectrum signal was used wide, white noise-like, and a 500 Hz sinusoidal signal. To avoid ambiguity, at the beginning only one sound source was used that generates this signal.

By applying the autocorrelation function, the following results were obtained which we represented graphically in Figure 7

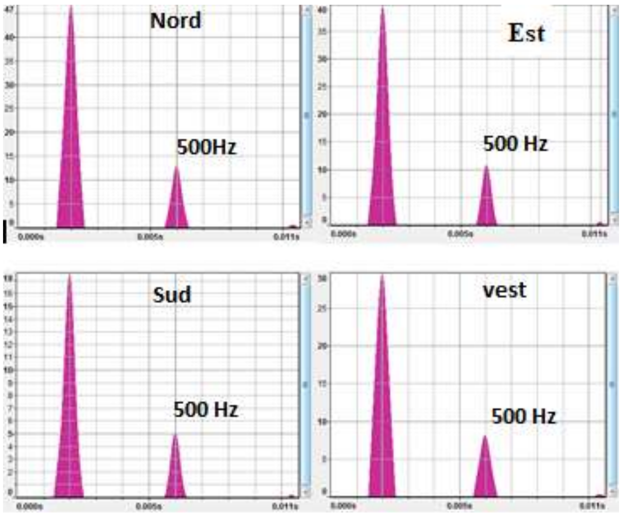

Figure 7: Autocorrelation function plot of 500 Hz sinusoidal signal with white noise

As I stated in the previous chapter, the analysis of the signal by applying the autocorrelation function highlights two important aspects: the value and the direction of the correlated noise. The value of the autocorrelation function close to the origin highlights the level of correlated noise as can also be seen from the plot of the function. We can also observe the energy value 500 Hz sinusoidal signal. The values read from the graph are presented in Table 1.

Direction	N	S	E	V
Correlated noise	47	19	40	30
Sinusoidal signal	13	5	10	7

Figure 8: The value of the energy of the 500 Hz sinusoidal signal captured in the 4 directions

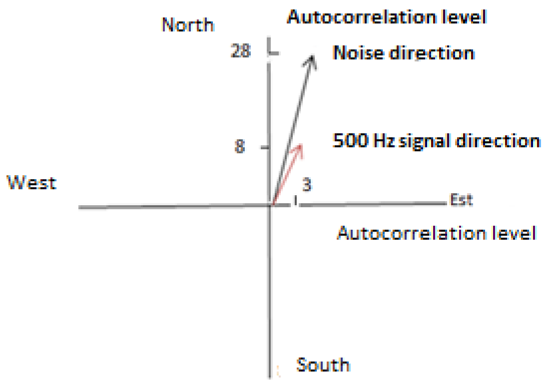

Figure 9: The direction of arrival vector

If there are several signal sources in the same perimeter, we will have an ambiguous situation. In the following example, there are three signal/noise sources in the monitored perimeter: two chainsaws located in different directions and an electric lawnmower. The read values of the signal levels are listed in Table 2

Direction	N	S	E	V
155 Hz	2	2	2	3
134 Hz	1	1,5	1	1
101 Hz	1,5	1	1,5	1,2
Noise	14	17	15	5

Table 2: Value of signal energy from two chainsaws and an electric lawnmower captured in the 4 directions

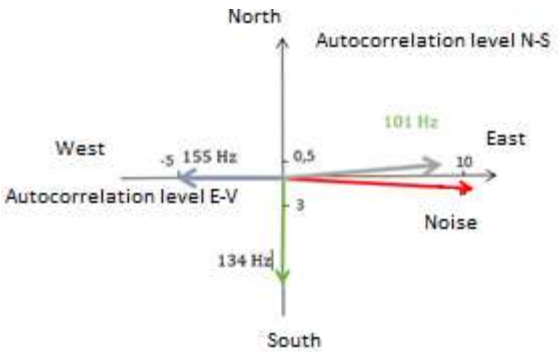

Figure 10: Direction vectors for three sound sources and noise

Other results of the measurements at different distances for a chainsaw-type audio source, in the N N-E direction, at an estimated angle between 80-90 ◦ are presented in Tables 3 and 4

Direcție \ Distanță		50 m	100 m	300 m	500 m	1000 m
N	Nivel intensitate [dB]	-25,5	-20,9	-30,4	-20,8	-19,76
	Nivel autocorelatie	15,83	100,12	58,94	181,95	251,27
S	Nivel intensitate [dB]	-45,6	-43,4	-54,92	-34,2	-32,56
	Nivel autocorelatie	1,43	7,10	3,5	38,78	23,12
E	Nivel intensitate [dB]	-40,1	-28,4	-40,95	-30,1	-40,88
	Nivel autocorelatie	6,62	42,27	19,66	62,31	58,02
V	Nivel intensitate [dB]	-30,1	-27,9	-38,14	-26,5	27,43
	Nivel autocorelatie	7,07	44,77	23,79	93,69	104

Direction \ Distance		50m	100m	300m	500m	1000m
N	Intensity level [dB]	-25,5	-20,9	-30,4	-20,8	-19,76
	Autocorrelation level	15,83	100,12	58,94	181,95	251,27
S	Intensity level [dB]	-45,6	-43,4	-54,92	-34,2	-32,56
	Autocorrelation level	1,43	7,10	3,5	38,78	23,12
E	Intensity level [dB]	-40,1	-28,4	-40,95	-30,1	-40,88
	Autocorrelation level	6,62	42,27	19,66	62,31	58,02
V	Intensity level [dB]	-30,1	-27,9	-38,14	-26,5	27,43
	Autocorrelation level	7,07	44,77	23,79	93,69	1004

Table 3: Measurement results at different distances for a chainsaw-type audio source – option2.

Angle of arrival \ Distance	50m	100m	300m	500m	1000m
ILD	87°	88,72°	83,45°	74,95°	43°
Autocorrelation	88°	88,44°	85,7°	83,5°	79°

Table 4: Angle of arrival for different chainsaw distances - option 2.

16. Evaluation of the proposed solution

As we saw in the previous chapters, in this doctoral thesis I used a structure of audio sensors adapted to the proposed purpose starting from the shape, dimensions and functioning of the human head. The processing of the data recorded by the network of sensors was carried out by implementing an open source software Audacity. This audio signal processing program was installed on a Raspberry Pi 5 minicomputer. For data transmission, a LoRa Module RFM96, 868MHZ, ultra long range, 3.3V for Raspberry Pi, with a frequency band of 868 MHz (for Europe) was used) and a directional Yagi antenna. I used a solar kit with a photovoltaic panel to supply the device with electricity of 100W and a battery with a capacity of 17 Ah. In conclusion, for the entire equipment the cost reaches somewhere around several hundred euros. In this calculation, we did not take into account the additional costs due to human resources and those specific to any product in the development, production and sales phase, and for the beginning we limited ourselves only to material expenses.

By transferring the received data and by not using an equipment to process them in situ, the energy consumption decreases, making it possible to use smaller electrical energy capture and storage elements or to ensure the continuous operation of the sensor cell. The cost for a single cell to use the mobile phone data network, depending on the provider, is approximately 5 euros/month.

For the surveillance of an area of 1 km2, using sensor cells at a distance of 500 meters, arranged in the corners of a square, 9 cells would be needed. The cost of data transfer in this case would be 45 euros/km2.

In the situation where the area to be monitored is not covered by a data network and the transmission of a large volume of information can be a difficult, even impossible mission, it is necessary to minimize it. The solution is to process the data in situ using a Raspberry Pi-type minicomputer on location. In this case, a point-to-point transmission technique is used using one of the wireless communication protocols (LoRa). With the help of this communication protocol, data transfers of the order of tens of kbits/s can be carried out, at a distance of 5-10 km, using a direct radiating system. In this case, the data transfer rate would have costs close to zero, instead it would significantly increase the cost of a sensor cell by adding the minicomputer to the structure and increasing the energy capacity required to maintain the sensor's operation, under conditions that would significantly increase consumption (approximately 5 times).

Following the measurements made in real conditions, in open fields, hilly areas, forested areas, we determined that the average effective distance of the sensor is approximately 250 m. In conclusion, 9 sensor cells would be needed to cover an area of 1 km2. If the monitored area increases, the number of sensors needed per km2 will decrease. Thus for 2km2, we would need 15 sensors, for 4km2 25 sensors are needed.

In conclusion, to cover an area of 4 km2, the number of sensors can vary between 13 and 25. The cost of a surveillance system can vary between 3600 Euro and 7500 Euro. The average price per km2 would be 800 euro-1870 euro, depending on the number of cells used

The materials used in the construction of this sensor have good resistance in conditions of high humidity and extreme temperatures. Placing the microphones inside the structure, in PVC tubes at a distance of 4 cm from the external surface, from which they are separated by an anti-wind sponge, gives them increased protection against rain or dust. The outer surface a of the sensor is treated with a rubberized silicone paint, equally resistant to the weather. The electronic equipment is housed in a sealed housing with a rubber gasket to prevent moisture and dust from entering. A wintertime problem can be snow covering the solar panel, which is why using a solar panel would be helpful
with heating system.

17. Final Assessment

The 4 acoustic sensors, in our case, are (cheap) omnidirectional electret microphones, but whose directivity diagram was modeled by the mechanical construction of the sensor and by the acoustic, sound-absorbing characteristics of the materials used. The proposed goal was to obtain a relatively inexpensive sensor with very good directivity characteristics over a frequency range and a high signal-to-noise ratio. All these features determine a good quality of the recorded sound signal without the need for sophisticated noise reduction methods and other types of interference that require complex signal processing. In this way, an algorithm that does not require a lot of computing power can be used, this one
being a very big advantage in terms of the hardware used and energy consumption, both of which are limited for such a sensor.

By using this type of sensor and the autocorrelation method for determining the signal maximum in the 4 directions and establishing the angle of arrival of the signal, we obtained a direction estimation accuracy between 5∘ and 15∘. The accuracy of determining the direction of arrival is influenced by the distance at which the sound source is located and the signal/noise.

The results obtained by the autocorrelation method were more accurate than those obtained by the other methods (amplitude calculation), knowing that this method has a high immunity to noise, which is indicated for filtering signals from highly noisy **environments.**

The maximum distance for which correct evaluations can be made, within the tolerance allowed depending on the application, was between 300-500m, of course this distance varies greatly depending on the nature of the sound source and its intensity.

Following the experiments carried out, the accuracy of estimating the direction of arrival as a function of distance is graphically represented in Figure 11. According to measurements carried out in real conditions (ambient noise, wind, etc.), the accuracy of determining the direction of arrival, with both measurement methods (amplitude and autocorrelation) is good up to a distance of 250 m, after which the autocorrelation method becomes superior to the other method. This fact is due to the fact that the autocorrelation method has an increasing immunity to noise, the results obtained even in conditions of low signal/noise ratio being better than those obtained by other methods.

18. Conclusions

The key element of the surveillance technique is the set of microphones arranged orthogonally. In the specialized literature there are a number of methods used to determine the direction of arrival of the sound signal and to identify the acoustic source. In this work, these methods were studied and most of the algorithms were experimented with the help of the constructed quadraphonic sensor. Following these experiments, it was possible to optimize the constructive parameters of the quadraphonic sensor so that it can be used with maximum efficiency in most methods of determining the direction of arrival and identifying the sound source

The starting point for the construction of the sensor was the human head. The constructive and dimensional characteristics of the human head were studied and adapted, which were used in the mechanical construction of the sensor. By studying the transfer characteristic of the human head, the constructive dimensions of the quadraphonic sensor were optimized, the mechanical positioning of sensors, the composition of the materials used in such a way that it can be used for most methods of determining the direction of arrival and identifying the sound source.

The accuracy of the captures of the sound sources with the 4 microphones that make up the quadraphonic sensor allowed the use of calculation methods of medium to low complexity, which do not require high processing power, with very good results in terms of determination precision.

By using this mechanical construction, a directivity with an optimal F/B Ratio (Front to Back Ratio) was obtained, using cheap microphones (omnidirectional condenser or electret type). In this way, the use of methods that require complex processing or that require the use of synchronized signals at reception was avoided, a wish that is difficult to fulfill and that requires the use of equipment with a high degree of complexity. Also by adapting the mechanical constructive parameters, the unwanted influences of reflections and reverberations were considerably reduced, two important factors that can alter or even compromise the results of the determination.

A minicomputer (Raspberry Pi) using an open source operating system (Linux) and an audio signal processing application, implemented on the Audacity platform (open source) was used to process the sound signals captured by the quadraphonic sensor in situ. The amount of data required to be transmitted for direction determination is 8-10 bytes/sample.

The sound signal identification function is performed with a simplified algorithm (Goertzel algorithm). Using this algorithm is accurate enough for the intended purpose without generating unnecessary data volume for the application. The sound source identification vector generated on the analyzed sample is one byte in size. Using these **determination methods, the volume of data to be transmitted is very small, which is why narrowband data transmission methods and techniques can be used over long distances, on the order of kilometers, for example LoRa.**

The sensor cell was powered by converting a renewable source (10W solar panel) with storage in a 12V, 7Ah battery. The average energy accumulated during a day with average illumination of 6 hours per day was sufficient for the uninterrupted operation of the cell. A possible optimization would be the attachment of a small wind turbine, capable of producing energy also in periods of low illumination, for example cloudy weather or night.

In short, the main contributions of this doctoral thesis can be summarized as follows:

1. the set of microphones arranged orthogonally in a housing similar to a human head;
2. optimization of the constructive parameters of the quadraphonic sensor;
3. obtaining audio signals with high accuracy by capturing sound sources with the 4 microphones that make up the quadraphonic sensor;
4. obtaining a localization with a directivity and with an optimal F/B ratio, using cheap microphones;
5. considerable reduction of unwanted influences of reflections and reverberations through the mechanical construction of the proposed sensor.