AS 026
iOwlT: Sound Geolocalization System

Team members：Davi Almeida, Gabriel Firmo, Matheus Farias

Organization: Universidade Federal de Pernambuco

Instructors: Daniel Filgueiras, Edna Barros

I. High-level Project Description

Acoustic systems of location and event identification have several applications in the everyday world, being present in security systems, earthquake recognition, sonar and various types of man-machine iteration.

Shooting sound mapping techniques began to be implemented in the last decades, even though it has been a problem of interest since the mid-First World War. In addition to military practices and environmental protection (e.g. detection of hunters in forbidden areas), this mechanism can be used in urban areas, providing instantaneous data to the local police or collecting data for further study of violence in certain areas.

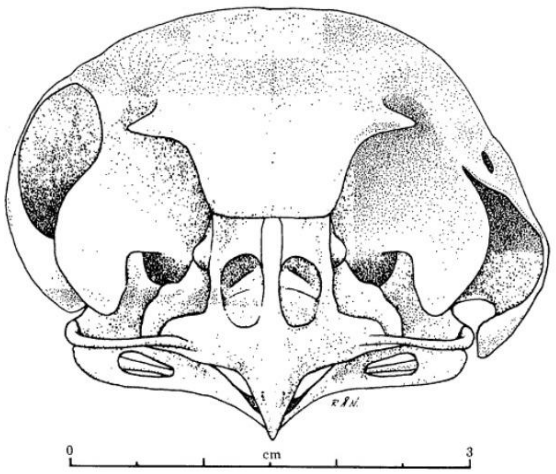
Aiming to recognize and map specific types of sounds, an idea of an intelligent and self-adaptive system was developed based on the functioning and learning of the auditory system of species of owls.

Owls are animals that possess a powerful hunting ability during the night, and to accomplish such a feat, as at night the sight is naturally more overshadowed by the absence of light, the owl has to use other benefits of evolution to improve accuracy of predicting the location of your dinner, one of them is the sound.

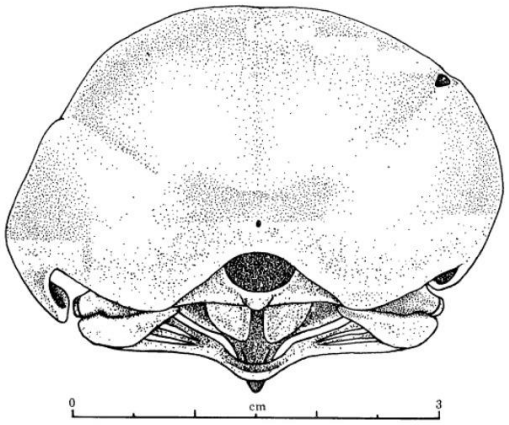
Experiments conducted by neurobiologists Eric I. Knudsen and Masakazu Konishi[1] have been able to prove, using barn owls as the species of study, that this species of owl is able to locate a prey being immersed in a totally dark room, only using the sound emitted by its prey.

A barn owl

The great evolutionary advantage present in this species is related to the considerable asymmetry that exists between their ears, it is known that the left ear is positioned around centimeters below the right ear, and with this difference of height, the owls can receive the information of the emitter with phase shift. From this difference, it becomes possible to accurately measure the location of its target, much like the triangulation positioning process, widely used in telecommunications engineering with telephone networks, or even in satellites. A very interesting video produced by the BBC[2] demonstrates the whole hunting process of this species.



Front vision of barn owl skull



Back vision of barn owl skull

Owls that locate their prey using a sharp hearing aid are not born[3] with this technique already well developed, thus necessitating an apprenticeship to adapt to their own physical characteristics (skull diameter, height difference between the ears, etc.) that can vary significantly in the same species, beyond that, the owls have on the side of the head, channels of rigid feathers that can regulate the passage of sound. Thus, these animals have a very efficient adaptive control, allowing that the accuracy in the location prediction maintains high even when dealing with different environmental conditions or physiological differences inherent to the species.

The technique of finding the coordinates of an unknown source from delays in reception of the signal in receivers distributed in a known manner in space is part of a technique called **multilateration**, which has no trivial solution. It is possible to show with algebra that in an N dimensional space N+1 receivers are needed, with known positions, to uniquely determine the coordinates of an unknown source.

Taking a case of easier visualization, there are 3 known receptors $\mathbf{R_1}$, $\mathbf{R_2}$ and $\mathbf{R_3}$ and a target **T** with unknown location in an x-y plane.

$$R_1 : (x_1, y_1),\ R_2 : (x_2, y_2),\ R_3 : (x_3, y_3),\ T : (x, y)$$

When **T** emits a sound, the receivers detect the signal at different time. Without loss of generality, consider that $\mathbf{R_1}$ will receive the information first in a time **t**, $\mathbf{R_2}$ in $\mathbf{t+dT_1}$ and $\mathbf{R_3}$ in $\mathbf{t+dT_2}$.

To calculate the distance between **T** and the i-receptors we have:

$$d_i = v \cdot t_i$$

Where **v** is the sound velocity and $\mathbf{t_i}$ is the time of arrival of the signal from **T** to the i-th receptor

$$d_1 = v \cdot t$$

$$d_2 = v \cdot (t + \Delta t_1) = d_1 + v \cdot \Delta t_1 = d_1 + \Delta s_1$$

$$d_3 = v \cdot (t + \Delta t_2) = d_1 + v \cdot \Delta t_2 = d_1 + \Delta s_2$$

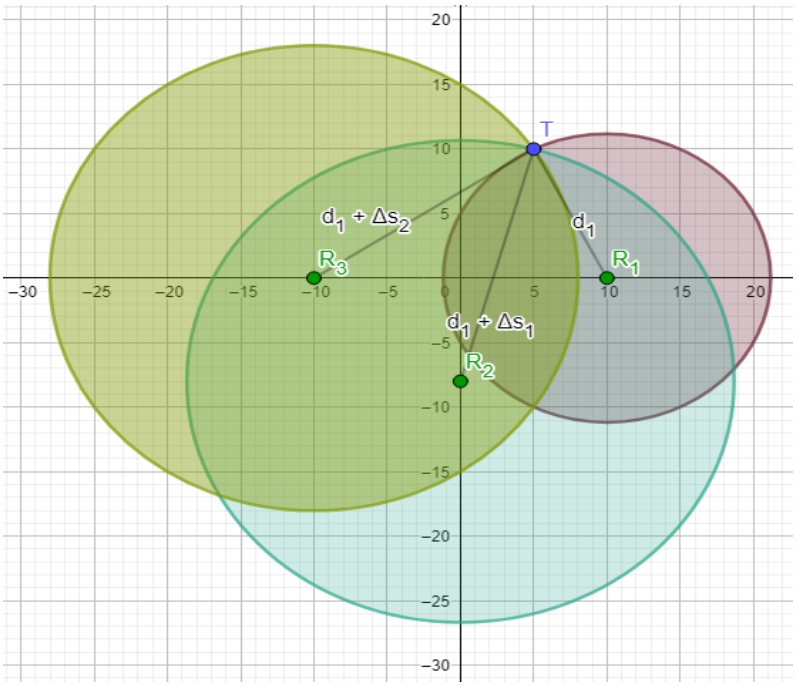
Centered at each of these receptors one can draw the circles $\mathbf{C_1}$, $\mathbf{C_2}$ and $\mathbf{C_3}$:

$$C_1 : (x - x_1)^2 + (y - y_1)^2 = d_1^2$$

$$C_2 : (x - x_2)^2 + (y - y_2)^2 = (d_1 + \Delta s_1)^2$$

$$C_3 : (x - x_3)^2 + (y - y_3)^2 = (d_1 + \Delta s_2)^2$$

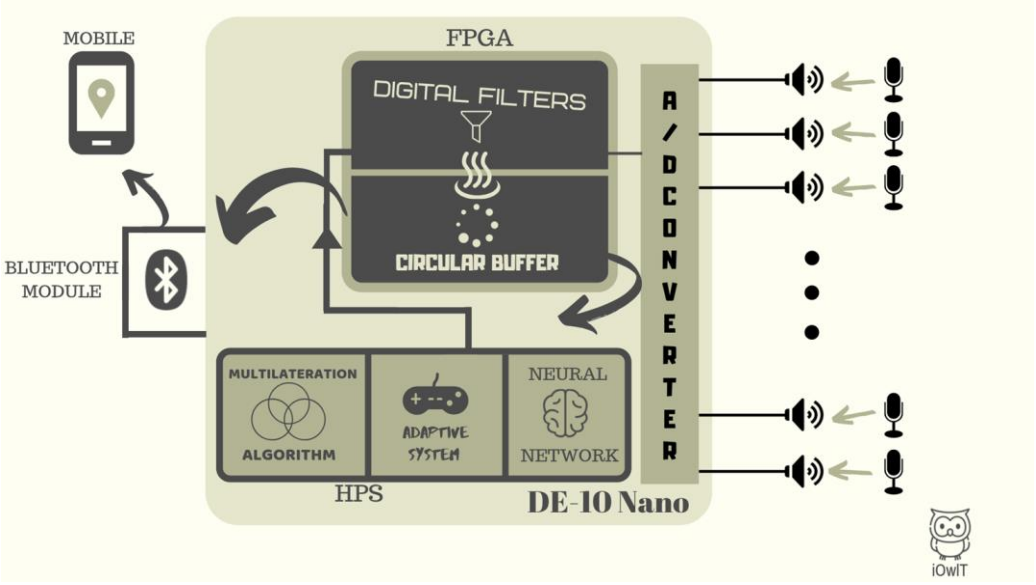
Drawing the circles in an x-y plane, we have:



The only unknown variables to this system of equations are **x**, **y**, and $\mathbf{d_1}$. For purposes, it is possible to solve this system by applying direct minimization techniques, otherwise, in the real case, with noises and inaccuracies, these circles do not have an intersection and we need to define cost functions with numerical algorithms (e.g. gradient descent) that minimizes the error and find and approximate value for T.

II. Block Diagram

Figure bellow is the system block diagram which can be divided into three parts: The Acquisition Circuit, FPGA and HPS (ARM).



2.1 Acquisition Circuit

This is an amplifier circuit that was designed to obtain the signal received by the microphones in voltage form varying in the range allowed by the *DE-10 Nano* A/D Converter, which receives the signal.

The *DE-10 Nano* board has an analog-to-digital converter of only one output, so the signals picked up by the sound detectors pass through a 8:1 multiplexer, observing the datasheet, it is noted that this mux takes a total of 3µs to switch, and therefore the largest possible delay in the acquisition of the signals, i.e. considering 8 sound detectors, is 3x7 = 21µs, as the audible frequency is in the range of 20Hz to 20kHz, using the Nyquist theorem, the sampling rate for the set of observed signals is 40kHz, and this results in 25µs, so the analysis of the signals is practically simultaneous, leading to an increase of considerable performance as well.

2.2 FPGA

The FPGA will have 4 essential modules, the A/D converter controller, that do the communication with the A/D converter and control the sampling rate of the signal, the Digital Filters module, that will process the sound digitalized by the A/D converter and remove low frequency noise, using precalculated parameters

adapted to the type of noise and type of impulsive sound that the system will recognize, the Circular Buffer module, that stores the sound signals in circular buffers to after send to the HPS, and the Bluetooth communication, that simply guarantee the communication between the cellphone and the equipment.

2.3 HPS (ARM)

The HPS will have 3 modules, the Adaptive System module, that will change some threshold parameter to adapt the solution to the environment, the Multilateration Algorithm, which is the correlation operation combined with the

effective measurement of where is the sound emitter using the multilateration technique and the Neural Network, which is responsible to determine if the sound is the desired sound or not.

III. Intel FPGA Virtues in Your Project

3.1 Adapt to changes

Processing of sound signals made in the FPGA is supported by an adaptive threshold system, that will vary depending on the distance of the sound source, type of sound (being more general than gun shots) and ambient noise. This will result in a change in the threshold variable which will be adapted to help in the sound recognition. Therefore, the iOwlT system is adaptable to this feature.

The present project can be used not only embedded on a police car. Depending on the use of the technology, the iOwlT system may well be positioned in static strategic positions, such as on traffic lights, an interesting application would be to identify a possible earthquake imminence, since the onset of a seismic shake is determined much earlier by sound signals of high intensity but with very low frequency, being audible to animals like horses but not to humans. Such sonorous signals could be identified by the iOwlT system, and therefore there would be a longer preparation time for the coming earthquake.

3.2 Boost Perfomance

The iOwlT system, using FPGA technology, can performs the multilateration algorithm with an outstanding precision, using the idea of a circular buffer to be

the data structure that stores the sound received by each microphone. As the microphones receives the signal with a phase difference, it's possible to see very clear the phase difference by counting the pivot index difference with correlation. As discussed in FPGA section of Block Diagram, the almost simultaneity of signal analysis contributes to increase considerable performance as well.

The sampling rate control system and real-time audio capture is of great importance for sound recognition and phase difference calculation between microphones. Having this circular buffered system implemented in hardware is guaranteed that the system will function properly. This same implementation would be very difficult to perform on a normal microprocessor system due to severe time constraints.

3.3 Expands I/O

The analog inputs of the *DE-10 Nano* board will mostly be occupied by sound detectors, although 4+1 detectors would solve the problem of precisely determining the target (1 to be the reference), as a form of security, adding extra microphones does not increase the cost considerably and ensures the reliability of the signal that will be further processed.

The output of the multilateration will result in the location of the sound event, such output will be sent by the Bluetooth module to the connected mobile phone, so that the location of the event can be shown in the application to easily see in a cellphone.

IV. Functional Description

Following the natural flow of the project, in order, there is the acquisition of sound data by the A/D Converter, the Digital Filters, Neural Network and in the last the Multilateration Algorithm.

4.1 Digital Filters

At this point, the received data goes through digital filters, whose purpose is to clear the received signal from possible low frequency noises. To perform such cleaning, a high pass digital filter is applied using the expression below that characterizes a moving average filter. Since the amplitude of the A/D converter signal has an offset (DC component), this filter also removes this component, centering the signal in zero, something that facilitates the after processing.

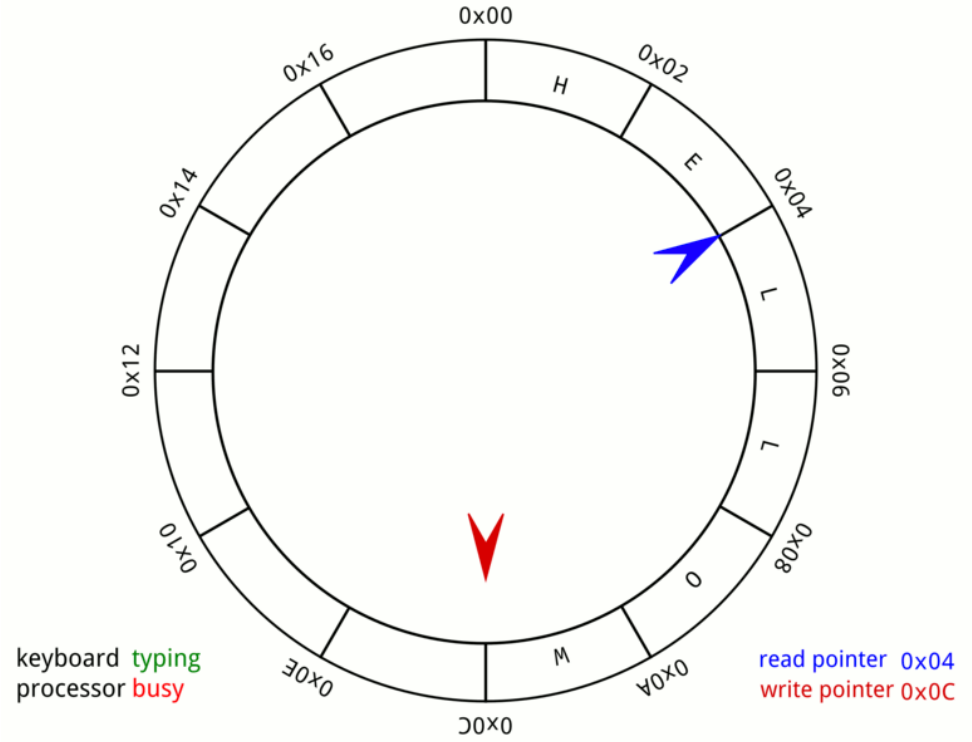
$$y[k] = \alpha(x[k] - x[k-1]) + \alpha y[k-1]$$

In the above expression, **y** represents the output vector of the filter, **x** the input vector and **α** the filter parameter, precalculated based on the sampling rate of the system and the desired cutoff frequency.

Once filtered, the data is sent to circular buffers, where there is one circular buffer for each microphone.

4.2 Circular Buffer

Circular buffers are data structures that are defined by a pivot, where the first data that was placed on the structure is located, and its tail, which is the last data, as shown in the figure below.

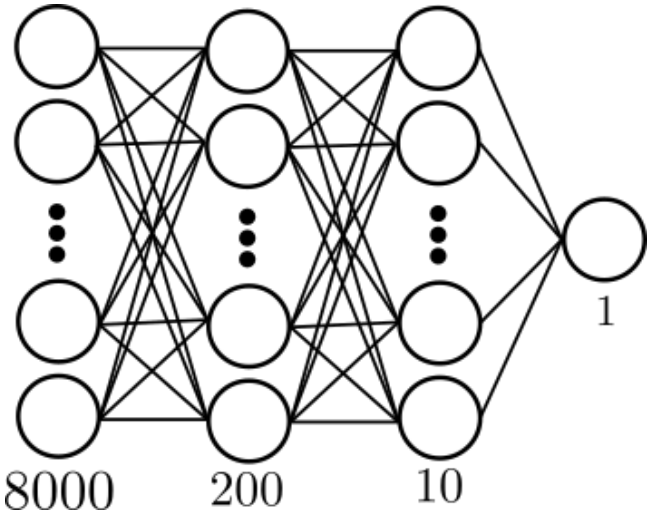


If some signal that is stored in the circular buffer pass the threshold barrier, this one is sent to the HPS, and the sound recognition is done with the help of a Neural Network.

4.3 Neural Network

Before determining the cross-correlation between the signals to obtain the phase difference between then, one must first know if these signals are really a desired sound, and for this, neural networks are used. Firstly, so that the system does not keep processing the neural network all time, a threshold based on the impulsivity of the signal is used. Since the received signal is considered impulsive, the signal is processed by the neural network and is determined whether the signal is the desired sound or not.

The neural network architecture used in the project is 4-layer MLP (input + 2 hidden layers + output), with neurons of each layer in the order: input, 200, 10, 1. The input layer quantity neurons depends on both sampling rate, feature extraction techniques and the average time that defines the signal (window time). An example of an MLP architecture for the system with 16 kHz sampling rate, window time of 0.5 s and no feature extraction is illustrated in the figure below.

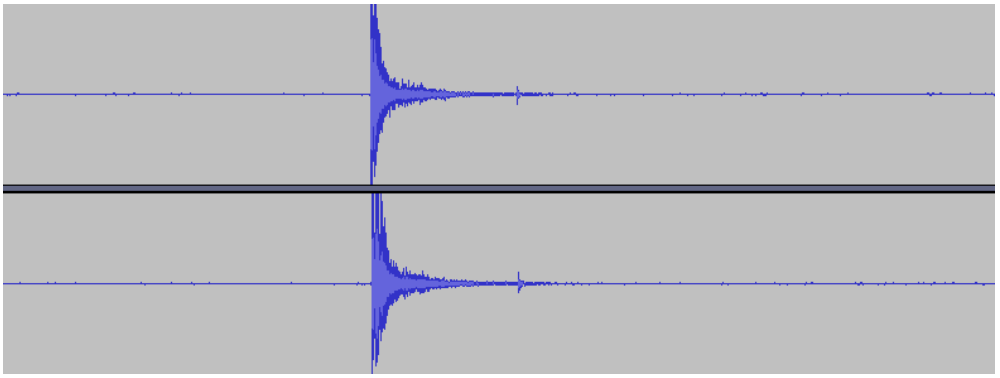


It is also important to highlight that due to the flexibility of the neural network, the system can be trained to identify other sound events (if trained correctly).

If the Neural Network identify the candidate signal as a desired sound, the system executes the multilateration algorithm, to locate the possible source of the sound.

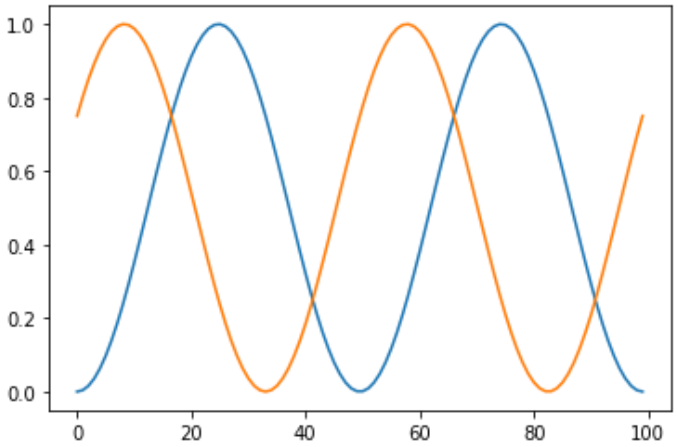
4.4 Multilateration Algorithm

In iOwlT, since the microphones are arranged at a well-defined distance, the signal is received by the microphones at different time. Two sound waves of a gunshot sound between two system microphones were recorded, and Audacity[4] software was used to show then. Is virtually imperceptible the lag between the sounds, as shown in the figure below.

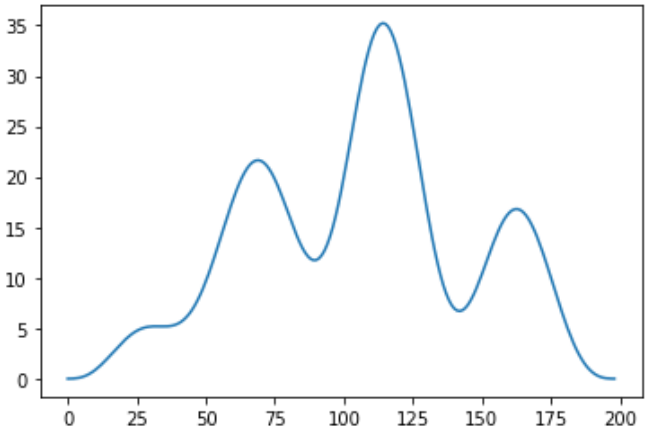


Two gunshot waves recorded with two system microphones

However, the idea behind the circular buffer is that two microphones will have their pivots shifted from a given number of samples that can be found through signal similarity analysis methods, the use of the FPGA at this time is crucial because one slight acquisition delay can lead to a considerable calculation error, since the speed of sound is high. In this case, the method used is the cross correlation.



Example of two sine signals with phase shift



The cross correlation of the sine functions

It is observed that by applying the method, a peak is obtained in the operation, and the distance from that peak to the center represents the *N* amount of samples shifted between the analyzed sine waves, and it is now possible to defined the lag time **τ** using the sampling rate SR. In iOwlT system, the SR is 16 kHz. In the example above, the phase shift is observed seeing that the peak is not in the center.

With the delays between the microphones calculated using cross correlation, it is now possible to execute the true multilateration algorithm, which consists in the solution of the following system.

$$\begin{bmatrix} \frac{2x_2}{v\tau_2} - \frac{2x_1}{v\tau_1} & \frac{2y_2}{v\tau_2} - \frac{2y_1}{v\tau_1} & \frac{2z_2}{v\tau_2} - \frac{2z_1}{v\tau_1} \\ \frac{2x_3}{v\tau_3} - \frac{2x_1}{v\tau_1} & \frac{2y_3}{v\tau_3} - \frac{2y_1}{v\tau_1} & \frac{2z_3}{v\tau_3} - \frac{2z_1}{v\tau_1} \\ \frac{2x_4}{v\tau_4} - \frac{2x_1}{v\tau_1} & \frac{2y_4}{v\tau_4} - \frac{2y_1}{v\tau_1} & \frac{2z_4}{v\tau_4} - \frac{2z_1}{v\tau_1} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} v\tau_1 - v\tau_2 + \frac{x_2^2+y_2^2+z_2^2}{v\tau_2} - \frac{x_1^2+y_1^2+z_1^2}{v\tau_1} \\ v\tau_1 - v\tau_3 + \frac{x_3^2+y_3^2+z_3^2}{v\tau_3} - \frac{x_1^2+y_1^2+z_1^2}{v\tau_1} \\ v\tau_1 - v\tau_4 + \frac{x_4^2+y_4^2+z_4^2}{v\tau_4} - \frac{x_1^2+y_1^2+z_1^2}{v\tau_1} \end{bmatrix}$$

The multilateration approximate linear system

Where τ_i is the time difference of the i-th microphone related to a microphone set as reference, **v** is the sound velocity and x_i, y_i and z_i are coordinates of each microphone related to a microphone set as reference.

V. Performance metrics / goals

In the iOwlT system, there are three important performance parameters that are analyzed:

5.1 Threshold

The first important parameter to be analyzed is the threshold, as it determines whether the neural network should judge whether the detected impulsive sound is a gunshot sound or not. Therefore, tests were made with impulsive sounds such as firework sounds, plastic bags and shooting itself.

5.2 Neural Network

The second important parameter to be analyzed is the neural network. To observe the behave of gunshot and impulsive sounds, a partnership was made with BOPE (brazilian SWAT), where it was possible to create a considerable

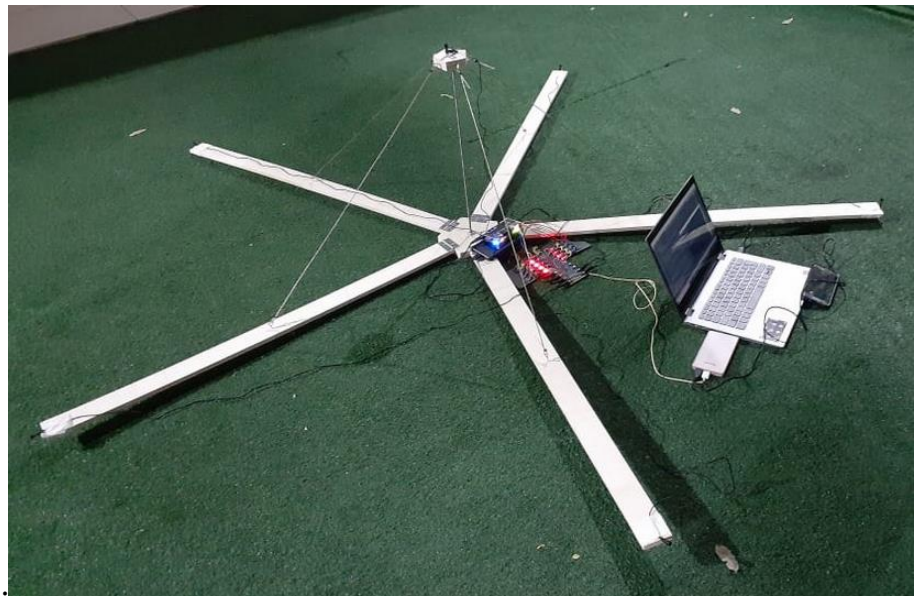
dataset of gunshots. The sounds were recorded in an open environment with a pistol.

Using Holdout validation technique, the neural network took an average perfomance of 91.38%, with the average confusion matrix shown below:

Real/Prediction	Not Shot	Shot
Not Shot	105	12
Shot	3	54

5.3 Multilateration Algorithm

The third important parameter to analyze is the multilateration algorithm. To this, an arrangment of 5-legs umbrella-shaped microphones was constructed, where each leg has a microphone in the end that will be used for the calculation of multilateration, and in the center has a microphone that has the purpose of the threshold and neural network process.



The iOwlT system

The coordinate system origin was defined at the center of the pentagon, and the y-axis as one of the legs. Thus, using a measuring tape, the actual distance values of a sound emitted were compared with the values found by the multilateration algorithm. The main idea was to analyze the system error (both distance and direction) for the four quadrants and varying distances (4 times for each point and took the average), as shown in the table:

	5m	10m	15m	20m
1^{st} Quadrant	1.63%	2.33%	3.21%	4.45%
2^{nd} Quadrant	0.79%	1.68%	2.54%	3.02%
3^{rd} Quadrant	1.01%	3.12%	3.88%	4.74%
4^{th} Quadrant	1.52%	2.16%	3.44%	5.12%

Table of angle direction system error

	5m	10m	15m	20m
1^{st} Quadrant	8.63%	10.11%	13.57%	15.12%
2^{nd} Quadrant	5.63%	7.77%	8.81%	12.24%
3^{rd} Quadrant	7.09%	9.21%	11.95%	17.47%
4^{th} Quadrant	8.00%	12.34%	17.37%	21.54%

Table of distance system error

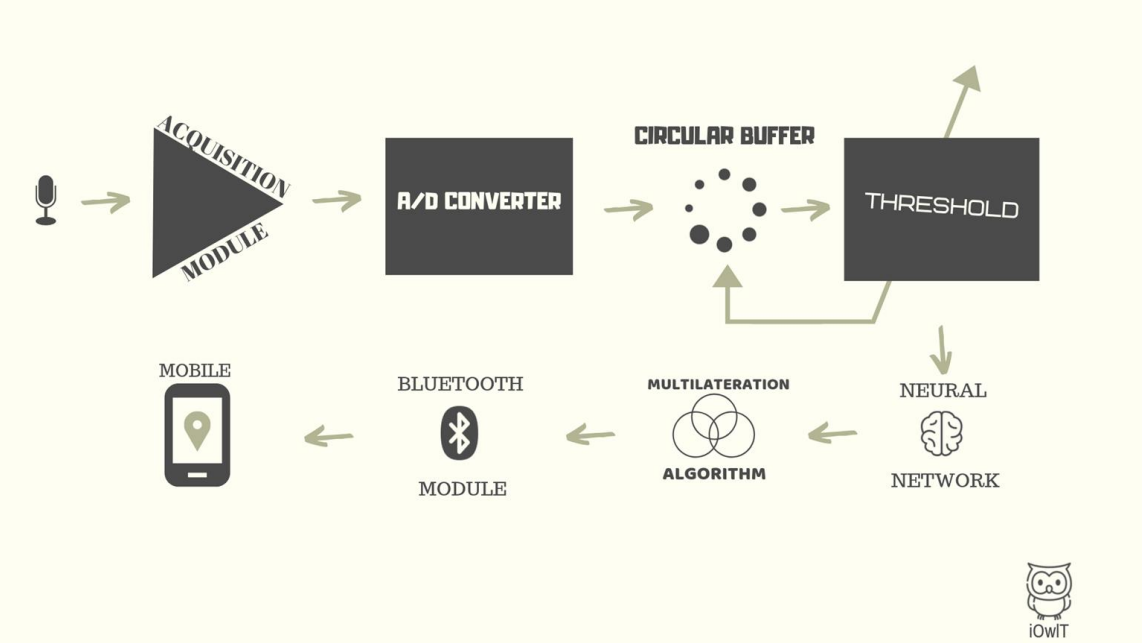
5.4 The echo problem

It is curious to observe that the iOwlT system was very good at determining the direction of the sound detected, with a 5.12% error precision in the worst case, as seen in the table. Otherwise, the error precision to the measurements of distance was bigger, this can be explained with the echo phenomenon.

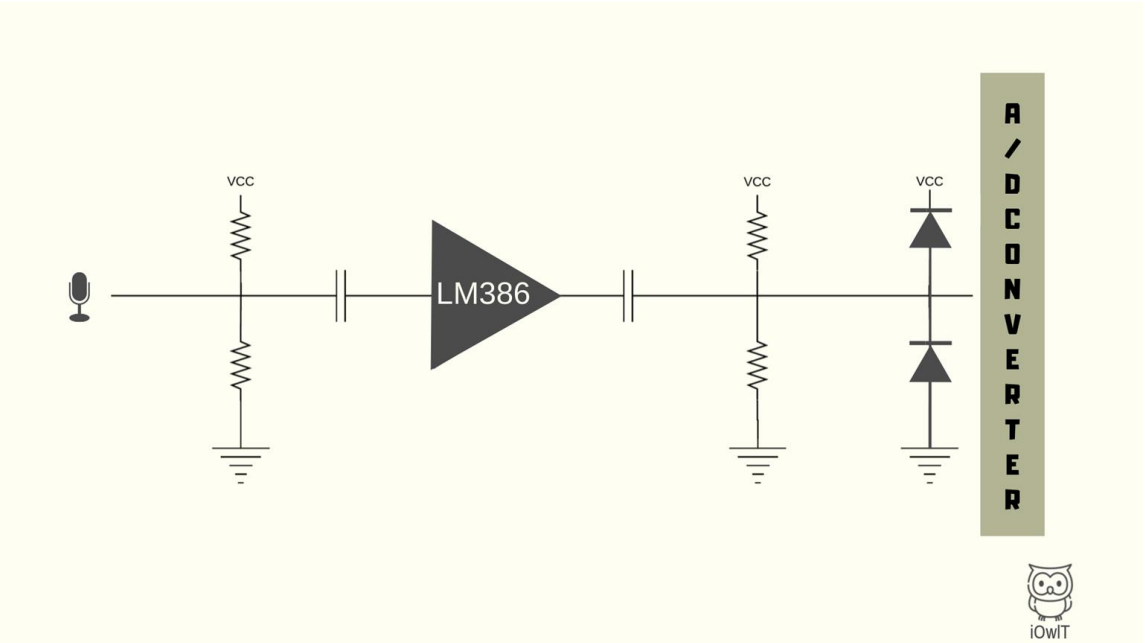
The great problem is that the echo mimetizes the sound emitted by a source in another position, creating a pseudo-source emitting the same sound, which can impose a confusion to the iOwlT system, as it uses only the phase difference of the sound to doing calculations.

One possibility to amenize the error distance due to echo problem is to treat the signal with a filter. In more open areas, the echo is not a problem, so the addition of those filters is a decision that can be done based on the application.

VI. Design Method

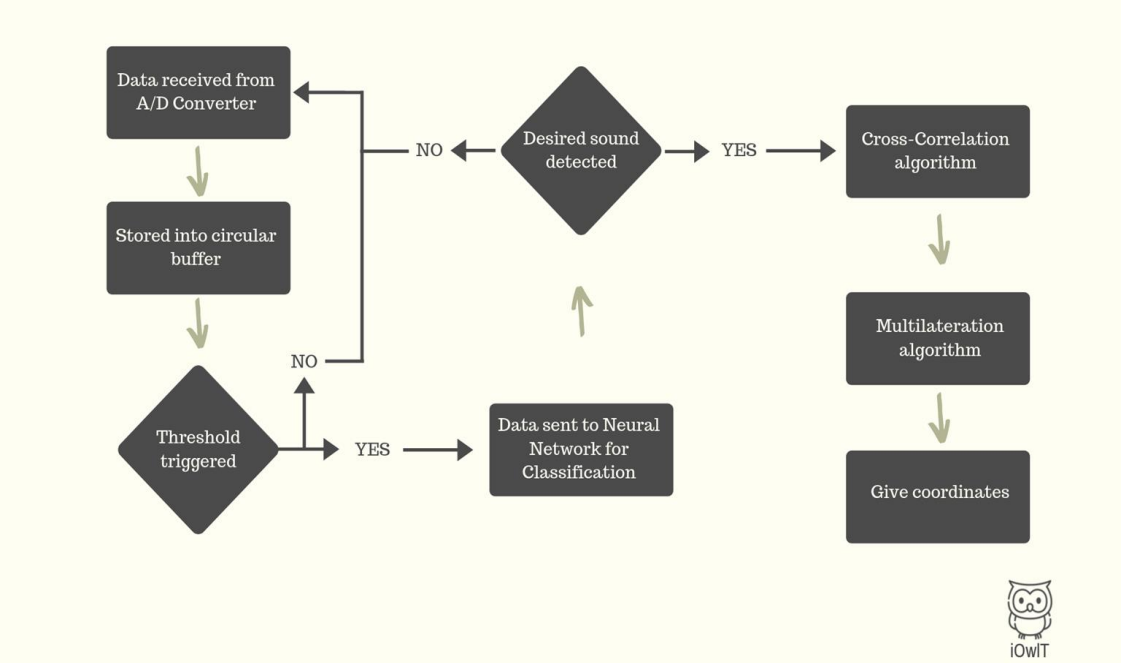


iOwlT system design scheme



Hardware circuit for every microphone

Each of the microphones used required an external FPGA circuit before they could be connected to the A/D converter. This circuit aims to polarize the microphones to allow them to operate, amplify the signal, and offset the signal to the A/D converter conversion range. In addition, there are two protection diodes that prevent a very high voltage from being delivered to the FPGA.



Software flow

VII. Conclusion

The system as a whole achieved good results, the main goal of the project is to demonstrate the technical possibility of locating sound events with sound treatment alone, and certainly iOwlT proved the technical feasibility with good accuracy.

It is important to highlight the fundamental use of FPGA to develop the system. As the system is based on the multilateration algorithm, it is extremely necessary that the phase difference measurement of the microphones be very accurate, since the speed of sound is high and therefore any error in the measurement of time leads to a considerable deviation in the final result of event position. As the phase difference between the microphones is very small, the level of accuracy in time recording of microphone signals would not be possible using a circular buffer structure implemented in software. At this point, hardware implementation was critical to good system performance, the parallelism and synchronization of hardware implemented circular buffers underlines the importance of the system.

Of course, the system can get even better, both in hardware, i.e. hardware that enables a faster A/D conversion could result in a smaller prototype as its size is directly related to time measurement accuracy, or even more logical elements,

for a more robust implementation with feature extraction and FPGA neural networking, and even software enhancements, with more BOPE visits for a wider shot dataset creating a better neural network training set.

VIII. References

[1] Knudsen, E.I. & Konishi, M. J. Comp. Physiol. (1979) 133: 13. https://doi.org/10.1007/BF00663106

[2] How Does An Owl's Hearing Work? | Super Powered Owls | BBC https://www.youtube.com/watch?v=8SI73-Ka51E

[3] Knudsen, E. Instructed learning in the auditory localization pathway of the barn owl. *Nature* 417**,** 322–328 (2002) doi:10.1038/417322a

[4] Audacity. https://www.audacityteam.org/

[5] J. Pak and J. W. Shin, "Sound Localization Based on Phase Difference Enhancement Using Deep Neural Networks"

[6] Renda, William & Zhang, Charlie. (2019). Comparative Analysis of Firearm Discharge Recorded by Gunshot Detection Technology and Calls for Service in Louisville.

[7] Mandal, Atri & Lopes, C.V. & Givargis, T & Haghighat, A & Jurdak, Raja & Baldi, Pierre. (2005). Beep: 3D indoor positioning using audible sound. 2005.