



Final Project

Microphone array processing

in Partial Fulfillment of the Requirements for the Degree
Telecommunication Technical Engineering. Sound and Image Speciality

Presented to the Faculty of Technical School for Industrial Engineering and
Telecommunications
of the Public University of Navarre (Spain) by

Héctor Ángel Navarro Contreras

in September 2010

Supervisor: Christian Debes
Co-Supervisor: Carlos del Rio Bocio

Declaration

Hereby I declare that I wrote this thesis myself with the help of no more than the mentioned literature and auxiliary means.

Up to now, this thesis was not published or presented to another examinations office in the same or similar shape.

Pamplona, 2010-09-14

place and date

signature (Héctor Ángel Navarro Contreras)

Abstract

Microphone arrays consist of multiple microphones functioning as a single directional input device: essentially, an acoustic antenna. Using sound propagation principles, the principal sound sources in an environment can be spatially located. Distinguishing sounds based on the spatial location of their source is achieved by filtering and combining the individual microphone signals. The location of the principal sounds sources may be determined dynamically by analyzing peaks in the correlation function between different microphone channels.

As an acquisition device for speech, microphone arrays have a distinct advantage as they place few constraints on the user, freeing them from the need to ‘wear’ a microphone (as in lapel, headset, or mobile devices) or be near to and speaking towards the microphone (as in web cams and lecturn microphones, for example). Unfortunately, arrays also some disadvantages compared to individual microphone, including increased hardware and processing requirements due to the multiple channels, and acoustic losses due to the distance from the speaker. Research into microphone array processing is thus required to minimize these issues, enabling the advantages of microphone arrays to be exploited in practical systems.

The more accurate the system, better results can be obtained. It is therefore important to choose and place the sensors in the most effectively. This is demonstrated in this paper, implementing an algorithm to estimate the direction of arrival and position of the sound source, in the most possible efficient way.

Acknowledgments

Acknowledgments

I would like to thank Christian Debes for his diligence and attention during my work as well as how he welcomed me.

Also Rachel Fandos and Stefan Leier for helping me, thank you very much.

Finally, thank on the patience who has had Carlos del Río, serving me very well every time I have gone to his office.

Table of contents

Table of contents

▶	Abstract	iii
▶	Acknowledgments	v
▶	Table of contents	vii
▶	1 Introduction	1
▶	▶ 1.1 Project Object.....	1
▶	▶ 1.2 Motivation	1
▶	2 Microphone Recording	2
▶	▶ 2.1 Hardware Implementation.....	2
▶	▶ 2.1.1 Condenser Microphones.....	2
▶	▶ 2.1.2 Electret Microphones:.....	4
▶	3 Microphone Array Processing	17
▶	▶ 3.1 Time-Delay estimation.....	17
▶	▶ 3.2 Direction-of-arrival estimation (DOA).....	17
▶	▶ 3.2.1 Methods of estimation.....	19
▶	▶ 3.2.2 Experiment description.....	20
▶	▶ 3.2.3 Development work.....	24
▶	▶ 3.2.3.1 Stages and Ejecution times	24
▶	▶ 3.3 Speaker location estimation.....	27
▶	4 Experimental results	30
▶	▶ 4.1 DOA results.....	30
▶	▶ 4.2 Speaker location estimate results.....	32
▶	5 Conclusions	34
▶	6 APPENDIX	35
▶	▶ 6.1 Matlab code for micrphones testing:.....	35
▶	▶ 6.1.1 Functions 'filtrar2' and 'espectro' used:.....	37
▶	▶ 6.2 Matlab code of the algorithm to estimate the direction of arrival:..	39
▶	▶ 6.3 Matlab code to determine the position of the sound source, using three pairs of microphones:.....	41
▶	7 Bibliography	45

Table of contents

List of Figures

Figure 1: Condenser Microphone.....	3
Figure 2: Internal circuit of an electret microphone.....	4
Figure 3: Circuit of the first solution proposed.....	5
Figure 4: Implementation of the circuit of the first solution.....	6
Figure 5: Schematic of the second solution (specifications manual).....	6
Figure 6: Implementation of the circuit of the second solution.....	6
Figure 7: Schematic of the third solution (specifications manual).....	7
Figure 8: Implementation of the circuit of the third solution.....	7
Figure 9: Schematic of the fourth solution (specifications manual).....	8
Figure 10: Implementation of the circuit of the fourth solution.....	8
Figure 11: Original signal and its spectrum.....	9
Figure 12: Original signal (green), recorded signal (red) and recorded signal increased 10 times (blue).....	10
Figure 13: Filtered signal and its spectrum.....	10
Figure 14: Original signal (green), recorded signal (red) and recorded signal increased 10 times (blue).....	11
Figure 15: Filtered signal and its spectrum.....	11
Figure 16: Original signal (green), recorded signal (red) and recorded signal increased 10 times (blue).....	13
Figure 17: Filtered signal and its spectrum.....	13
Figure 18: Original signal (green), recorded signal (red) and recorded signal increased 10 times (blue).....	14
Figure 19: Filtered signal and its spectrum.....	14
Figure 20: Theoretical representation of the delay between two sensors.....	16
Figure 21: Relative phase difference between sensors depending on the angle of arrival.....	17
Figure 22: Seating Plan of the room (plan view and side view).....	19

Table of contents

Figure 23: Laboratory conducting the experiment.....	20
Figure 24: Loudspeaker.....	21
Figure 25: A/D converter (ADA8000).....	21
Figure 26: Condenser microphone B-5 (Behringer) with cardioid capsule.....	22
Figure 27: An example of the cross-correlation between two similar signals (one displaced in time over the other), and a zoom in of the central point (maximum). It shows the delay N in number of samples	24
Figure 28: An example of the detection of the position of the sound source, obtained with the intersection of the angles of incidence of two pairs of microphones.	26
Figure 29: Position of the microphones in the room. Each pair and axis represented with one different color.....	27
Figure 30: Position of the sound source. Cutoff of the three lines. (Ideal situation).....	28
Figure 31: An example of the estimation of the incidence angle of the sound signal.	29
Figure 32: Another example of the estimation of the incidence angle of the sound signal.	30
Figure 33: An example of the estimation of the incidence angle of the sound signal, with the speaker in front of the microphone pair.	30
Figure 34: An example of the estimation of the position of the sound source. .	31

1 Introduction

1.1 Project Object

Different methods for direction-of-arrival (DOA) estimation and location are considered and applied to experimental data recorded in a typical office environment.

The purpose of this project is to determine the origin of human voice recorded with microphone array, estimating the direction of arrival (angle) and the location of the speaker in a room (coordinates).

For this will use microphones to record the voice signal and then processing the signal to obtain the information necessary to make such estimates.

1.2 Motivation

Digital microphone array has been proposed for sonar, audio, teleconferencing, multimedia, and hearing aid applications. A microphone array can enhance the SNR in a noisy environment, by forming a focused beam toward the desired speech source, attenuating background noises and rejecting discrete spatial interferences.

There are many applications:

- Systems for extracting voice input from ambient noise (notably telephones, speech recognition systems, hearing aids).
- Surround sound and related technologies.
- Locating objects by sound: acoustic source localization, e.g. military use to locate the sources of artillery fire. Aircraft location and tracking.
- High fidelity original recordings.
- Speech enhancement in a reverberant space.
- Blind Speech Separation (for example, to separate the speech of two talkers simultaneously speaking).

These are some examples that may be applied to the digital microphone array.

2 Microphone Recording

2.1 Hardware Implementation

It is intended to locate where it comes from the voice. For this, we use microphones to capture the signal, and then we process all the signals received to estimate the angles of incidence. Once we get this, is identified the position of the sound source, corresponding to the cutoff point of the lines that start from the axes of the microphones to infinity and whose inclination angle is calculated above.

Possible solutions are based on the type of microphone used. In this case, we consider two appropriate solutions:

2.1.1 Condenser Microphones

The condenser microphone is also called a capacitor microphone or electrostatic microphone. They have a microphonic capsule which consists of two capacitor plates: one fixed and one movable, separated by an insulating material. The diaphragm acts as one plate of a capacitor, and the vibrations produce changes in the distance between the plates, and this causes a variation in the voltage stored (electrons are gained or lost in the insulating material placed between two plates).

The capacitor plates need electric current to operate, hence, these microphones are not independent but require power that can be supplied either by a internal battery or external, which in the field of microphone is called phantom power. The external power arrives to the microphone from the mixer (48 volts usually). In addition to providing energy to the plates, phantom power also provides the power required to operate the preamplifier circuit (pre-amp) that the condenser microphones need, since its output signal is weak.

2 Microphone Recording



Figure 1: Condenser Microphone

Features

Much of condenser microphones are variable addressability. That is, have a switch (in this case, interchangeable capsules) that lets you choose the different pickup patterns (cardioid, omni-directional or more appropriate to a given audio jack).

The resonant frequency of condenser microphone is in the area of acute (12-20 kHz), however, as the diaphragm is not very heavy, do not get very high peaks.

Condenser microphones are commonly used by professionals because it is the type of microphone that provides greater frequency response: 20 Hz to 18,000 Hz

The great advantage of the condenser microphone is that the size of it's diaphragm is not limited by having to fit in a particular magnetic field.

The main disadvantage of condenser microphones is that, by its high sensitivity, if the sound source is too high, can cause overload distortion.

Another drawback is that they have a very high output impedance, so the cable length should be short so that no losses. In addition, they also have other great inconveniences: are affected by humidity and temperature, are very fragile and have a high cost. [AIP]

2.1.2 Electret Microphones:

Electret microphones (also called electret condenser microphone), is a variant of condenser microphone that uses an electrode (fluorocarbonato) laminal of plastic which eliminates the need for a polarizing power supply by using a permanently-charged material, however, it required a power supply to provide power to the preamplifier. Plates are polarized means they are loaded in perpetuity from the time of manufacture (are polarized once and can last many years).

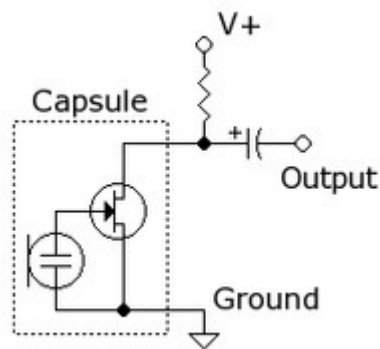


Figure 2: Internal circuit of an electret microphone

The diaphragm weighs less (it has less mass), so the frequency response of electret microphone is closer to the answer that provides a dynamic microphone, than that offered by a conventional condenser microphone. It is usual to use a 1.5V battery, although you can use the phantom power, is not appropriate, since microphone supercharge constantly shortens its life.

In terms of directivity, can be omnidirectional or directional.

A very important characteristic of these microphones is that they are robust (flexible), so they can be handled easily. And also, as they are small, they are used in applications that exploit this advantage:

- As a Lapel Microphone: Most of the lapel microphones are used in television are electret type. Moreover, mass production allows its cost is affordable.
- As a microphone of a small portable recorders that use external professionals (for radio statements, etc.).
- As microphones to be attached to specific instruments, percussion, brass, acoustic pianos, strings, etc.
- As mobile phones microphones.

2 Microphone Recording

Electret microphones have a fairly good frequency response (50 to 15,000 Hz), although distant from that of the condenser microphones that are much more sensitive in the area of the treble. It is also little flat.

The main drawback of electret microphones is that they are very sensitive to changes in humidity and temperature, which along with the dust, their performance deteriorates with use. An electret microphone starts to indicate that it should be removed (which he ended his life) when it begins to produce an unexplained hum (noise). [AIP]

The motive of using these microphones is to take advantage of they are small, flexible and cheap to put in place that we want. In fact, it can be placed very close together, and in a very stable way, and also we can obtain an absolute fixation because they're easy to paste them. The fact that they are cheap, makes possible to use lots of them, sparing no expense. This will allow greater opportunities for the design of the array. The closer the sensors are placed (in this case, the microphones), better results can be obtained in terms of measurement uncertainty. You get more precise estimates if the microphones are placed considerably closer.

The main problem is to connect the electret microphones to the Analogic/Digital converter, because they have different impedances. It has attempted to address this problem through different ways:

The first solution was to implement the follow circuit:

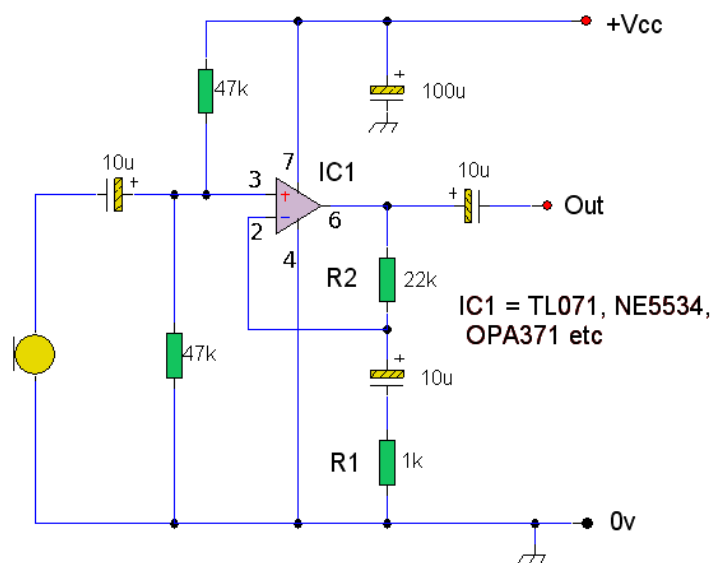


Figure 3: Circuit of the first solution proposed

2 Microphone Recording

who tried to match the impedances of the electret microphone and the PA.

We used the components necessary for its successful implementation in a mounting plate.

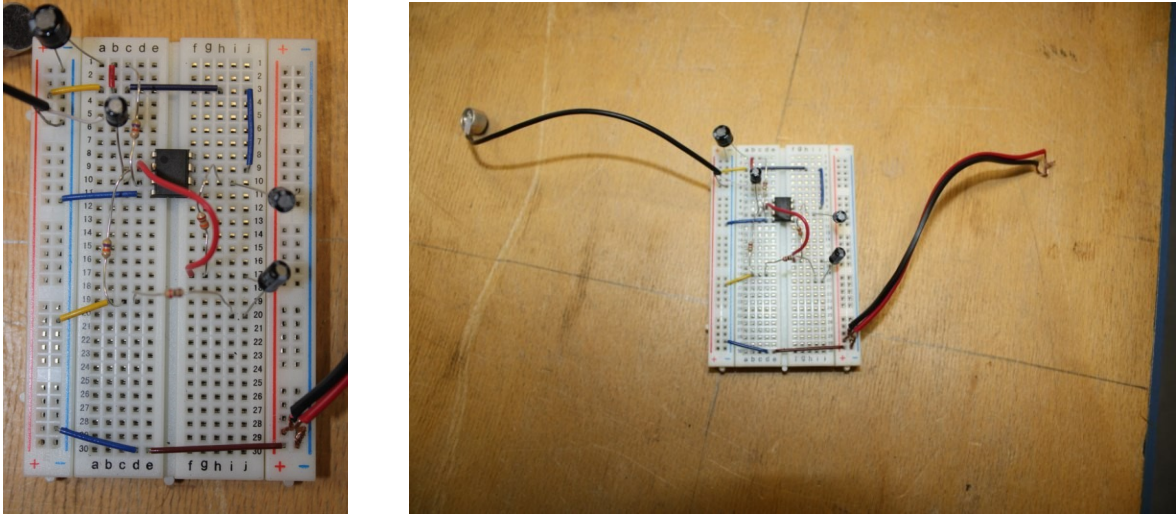


Figure 4: Implementation of the circuit of the first solution

Once tested this circuit, it was observed that the signal quality was not very good, so that it was raised another solution:

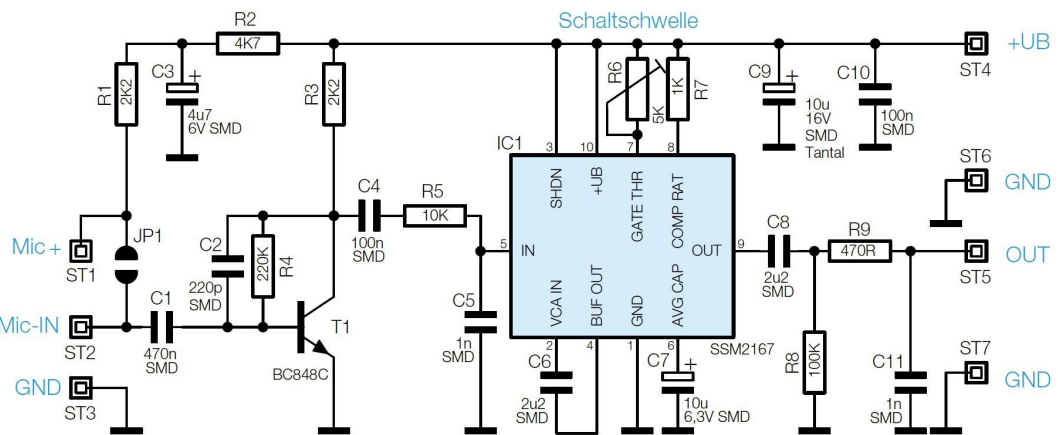
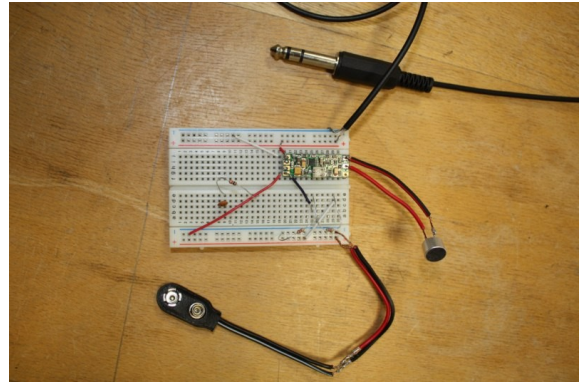
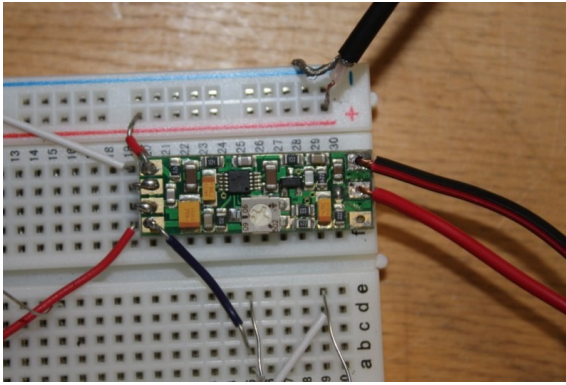


Figure 5: Schematic of the second solution (specifications manual)

This solution is an integrated circuit. Special features include an automatic gain control (ALC), avoiding saturation, and an adjustable squelch (noise gate).

Figure 6: Implementation of the circuit of the second solution

2 Microphone Recording



This model has the great advantage that the circuit is rather small in size, and this can be leveraged to build an array of microphones with a very small distance between them.

Although this model worked, before deciding whether to use it or not, two more circuits were tested:

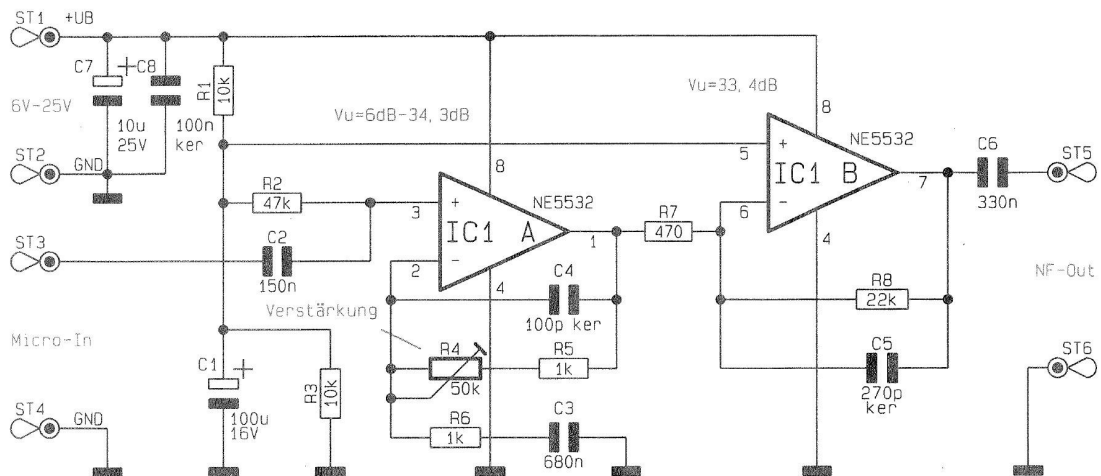


Figure 7: Schematic of the third solution (specifications manual)

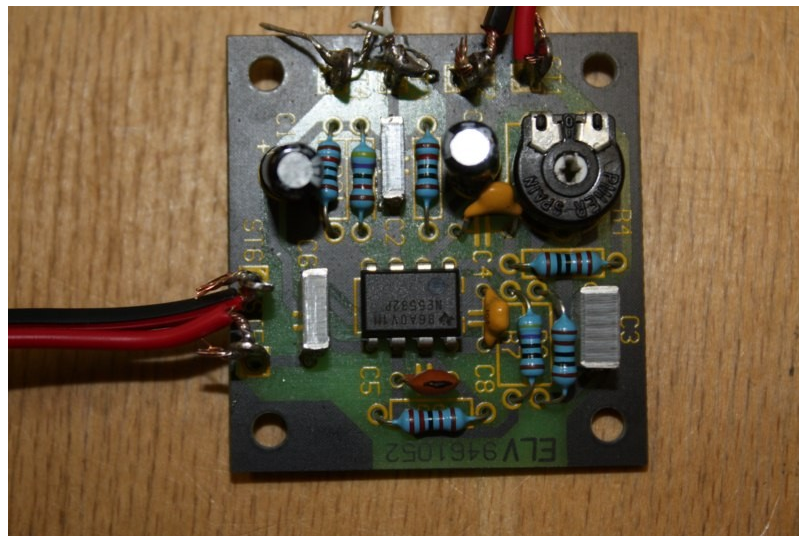


Figure 8: Implementation of the circuit of the third solution

2 Microphone Recording

And the last tested solution with electret microphone:

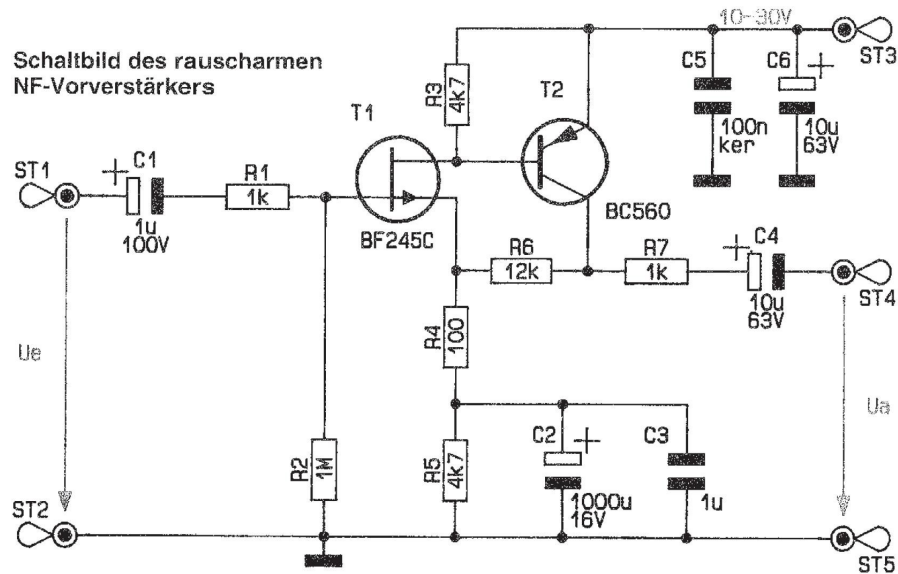


Figure 9: Schematic of the fourth solution (specifications manual)

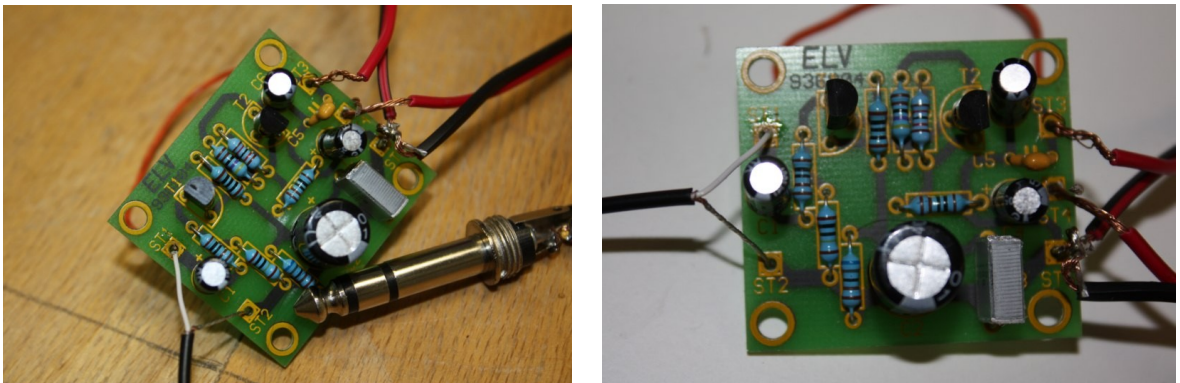


Figure 10: Implementation of the circuit of the fourth solution

All circuits were tested and it came to a conclusion that both offered a similar response, although the latter gave slightly better results than the rest. That's why testing was done with the fourth solution.

To see if it's possible to use these microphones with preamplifiers circuits, a test was performed:

- The microphone is placed at one point and a loudspeaker in front of him at a distance.
- A signal is reproduced from the speakers while is recorded with the microphone
- That signal is stored in the computer for analysis.
- It is filtered with a low pass filter to eliminate unwanted noise.

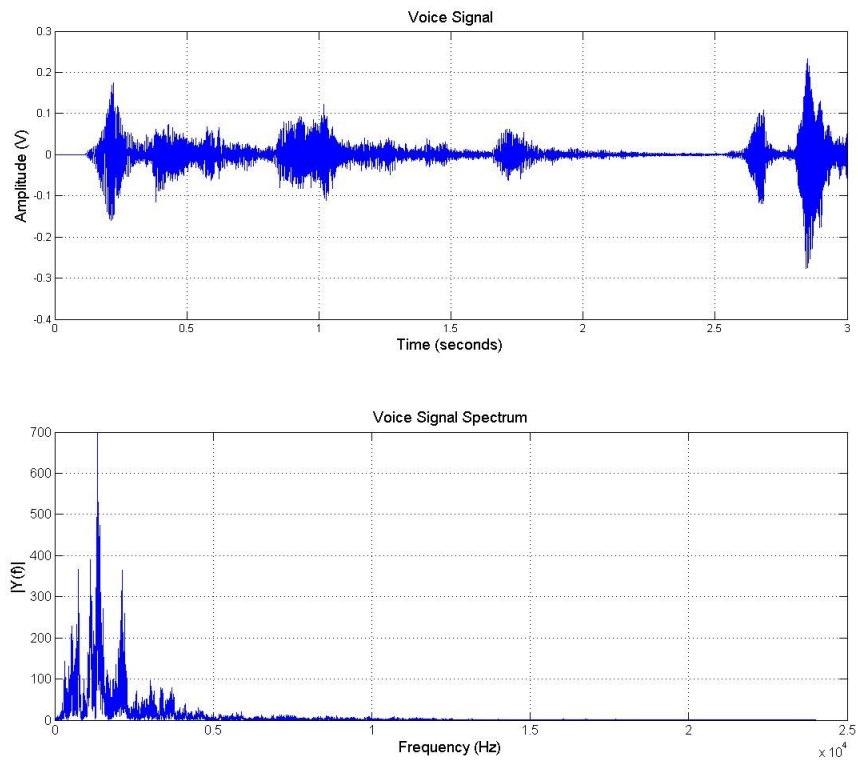
2 Microphone Recording

- The original signal, the recorded signal and filtered signal, are shown and compared.

The results are shown in the following pictures. These figures show the waveform and spectrum of the original signal to be reproduced by the speaker and later will be picked up by the microphone. So we will can compare then with the signal obtained by the microphone.

2 Microphone Recording

Figure 11: Original signal and its spectrum



Below are the results. First place the sound source at a distance of one meter, and then to three meters. It is represented the waveform of the signal captured compared with itself amplified 10 times, and with the original, to see the amplitude. After, it is depicted the waveform of the filtered signal and its corresponding spectrum.

2 Microphone Recording

(Distance between the microphone and the loudspeaker = 1 meter)

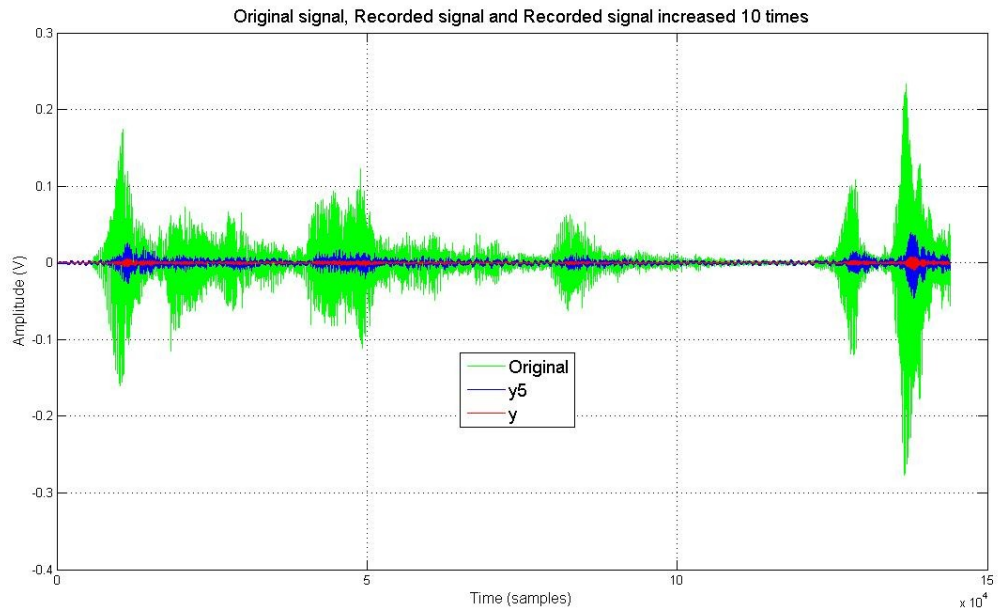


Figure 12: Original signal (green), recorded signal (red) and recorded signal increased 10 times (blue)

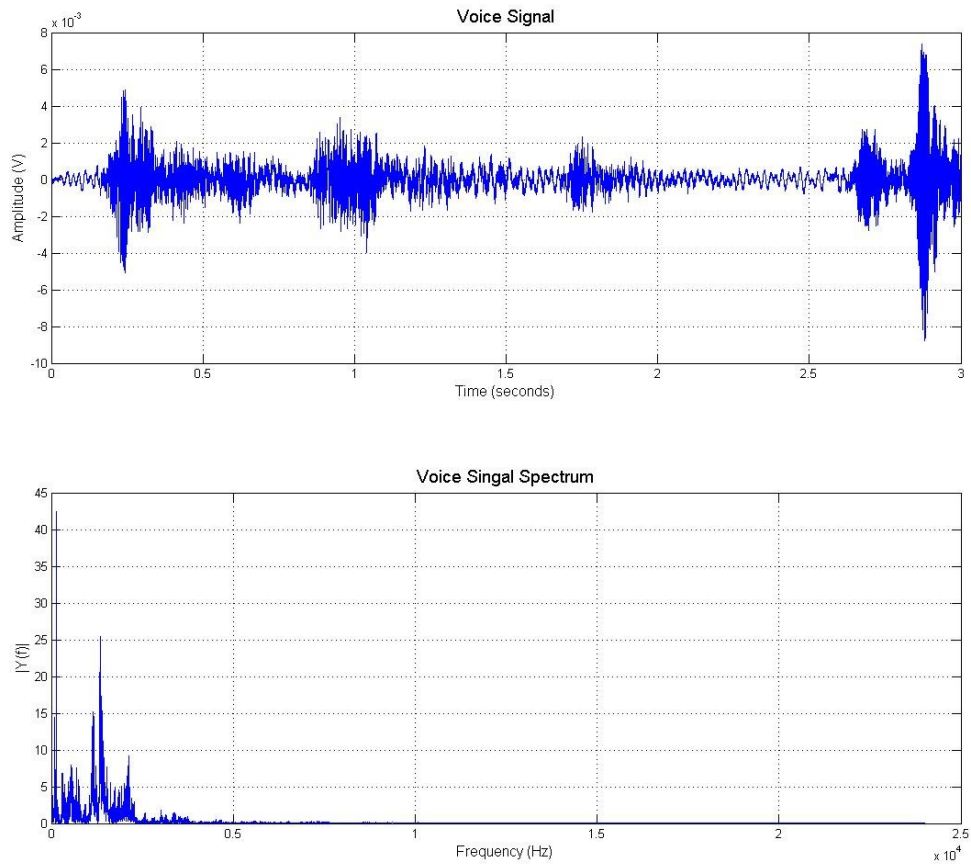


Figure 13: Filtered signal and its spectrum

2 Microphone Recording

(Distance between the microphone and the loudspeaker = 3 meters)

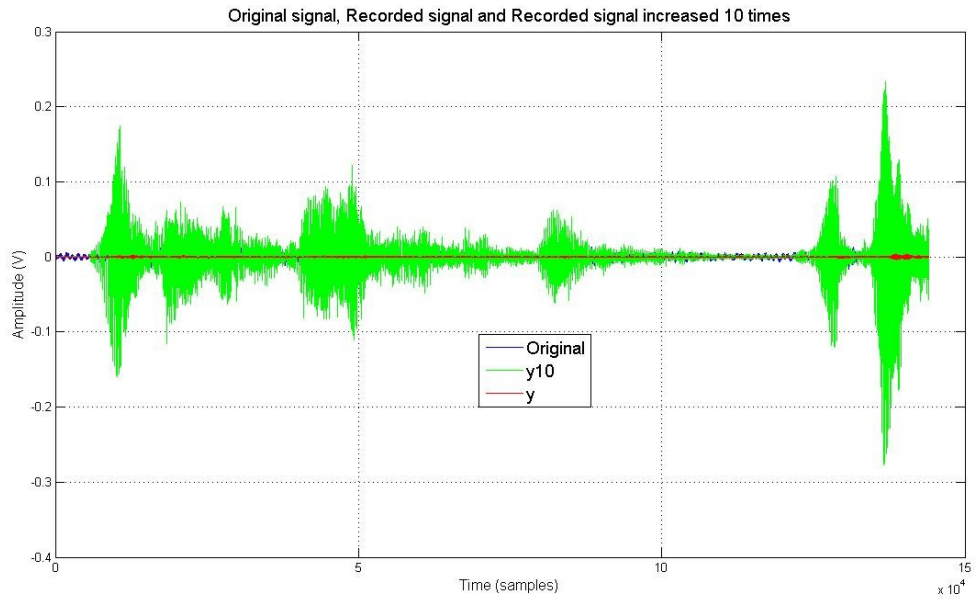


Figure 14: Original signal (green), recorded signal (red) and recorded signal increased 10 times (blue)

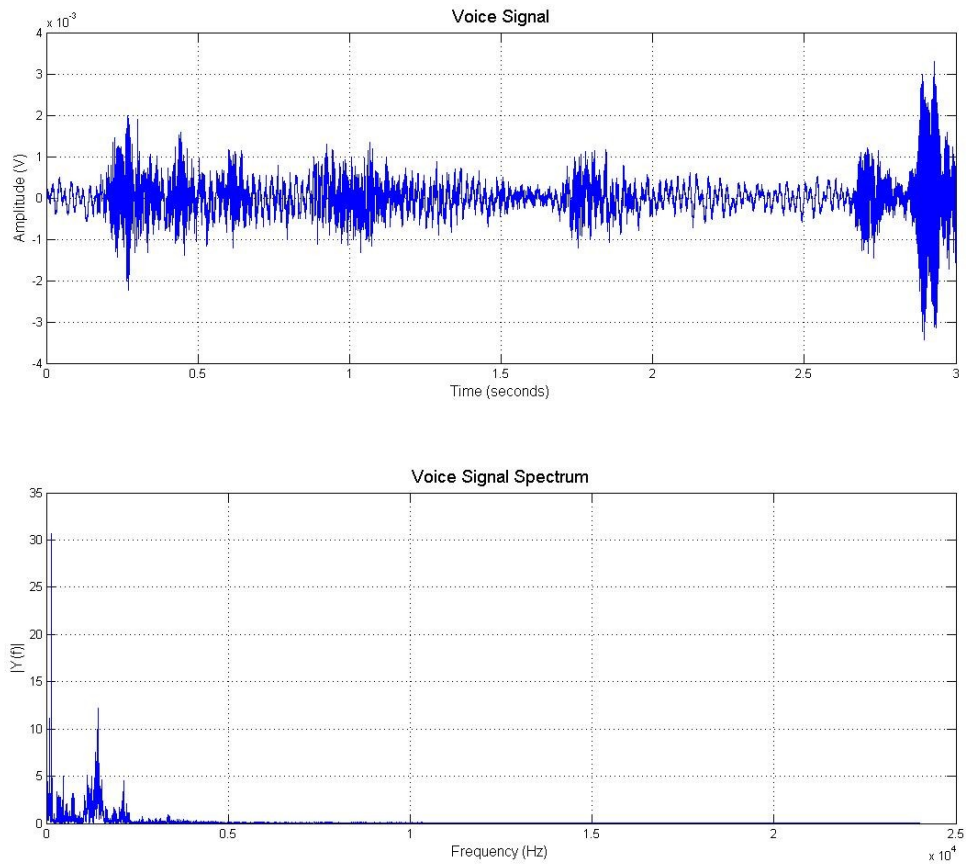


Figure 15: Filtered signal and its spectrum

2 Microphone Recording

As we can see, the recorded signal has a really low amplitude. Also, if the distance increases, the results are worse. The signal obtained from 3 meters is too low. It's impossible to process it with precision.

By way of comparison, we proceed to perform the same test but this time with a condenser microphone.

We put the microphone in a microphone stand, at the same high that the loudspeaker, and we do two different measures: one at one meter of distance and other at three meters of distance between de loudspeaker and the microphone.

Let's see the graphics of the results:

2 Microphone Recording

(Distance between the microphone and the loudspeaker = 1 meter)

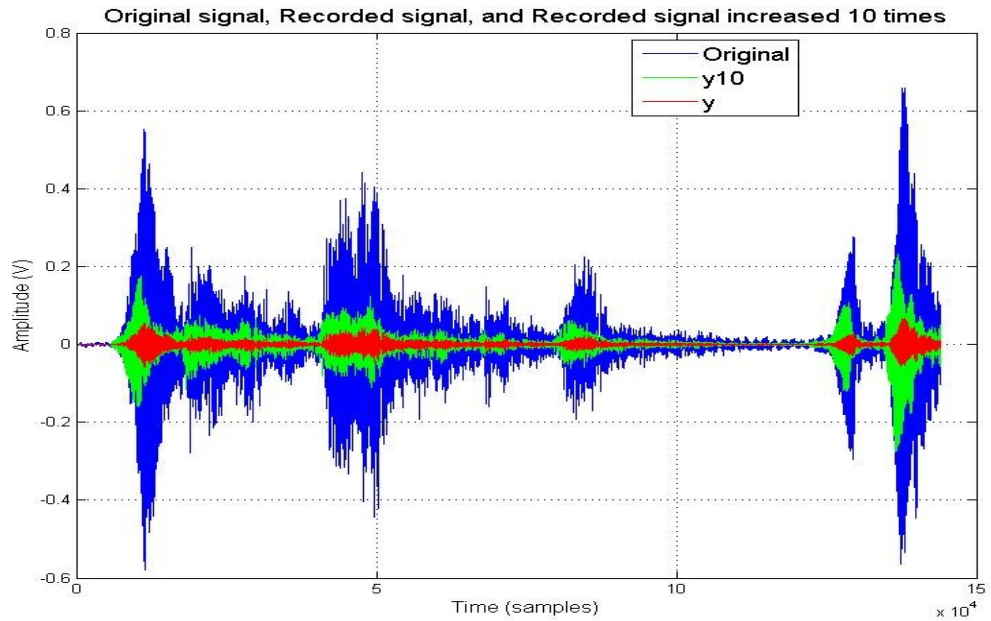


Figure 16: Original signal (green), recorded signal (red) and recorded signal increased 10 times (blue)

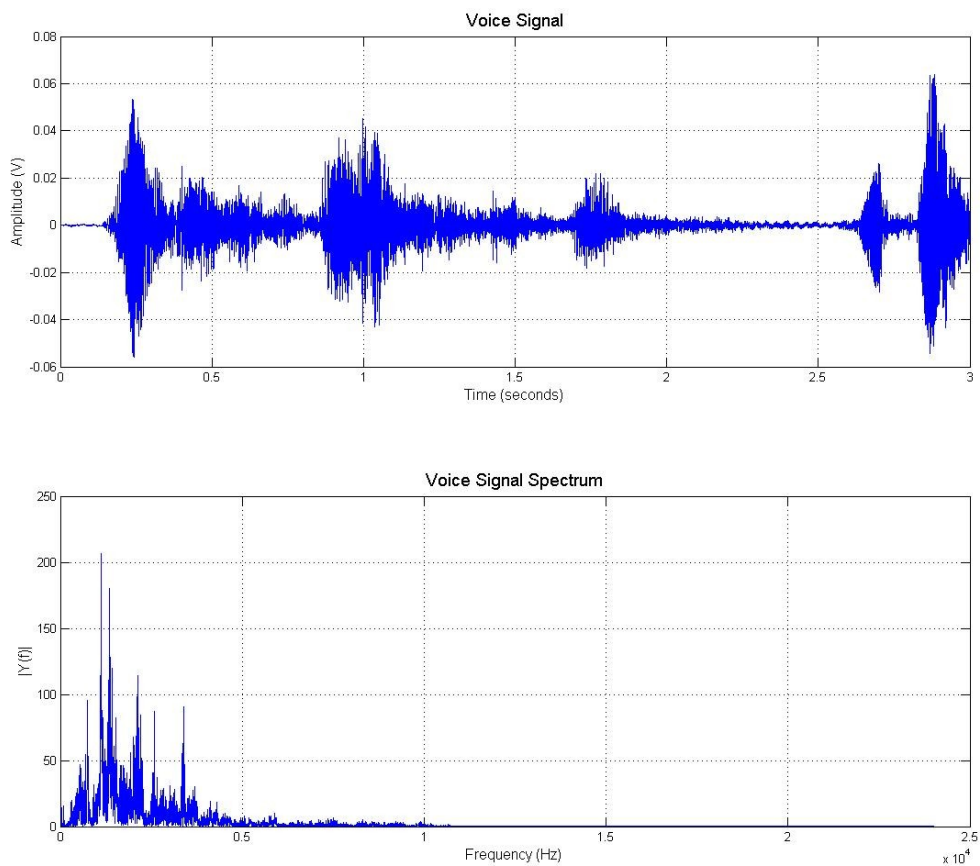


Figure 17: Filtered signal and its spectrum

2 Microphone Recording

(Distance between the microphone and the loudspeaker = 3 meters)

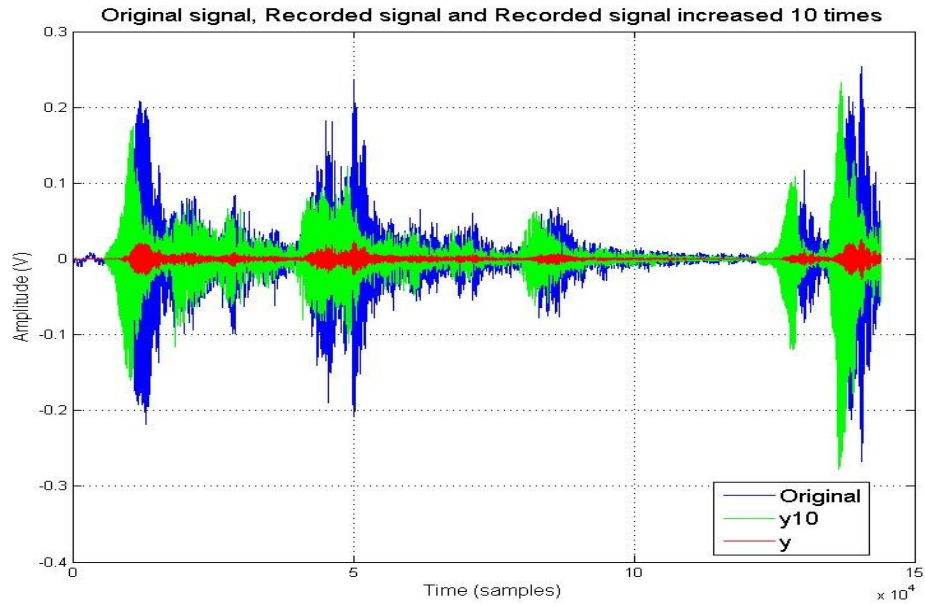


Figure 18: Original signal (green), recorded signal (red) and recorded signal increased 10 times (blue)

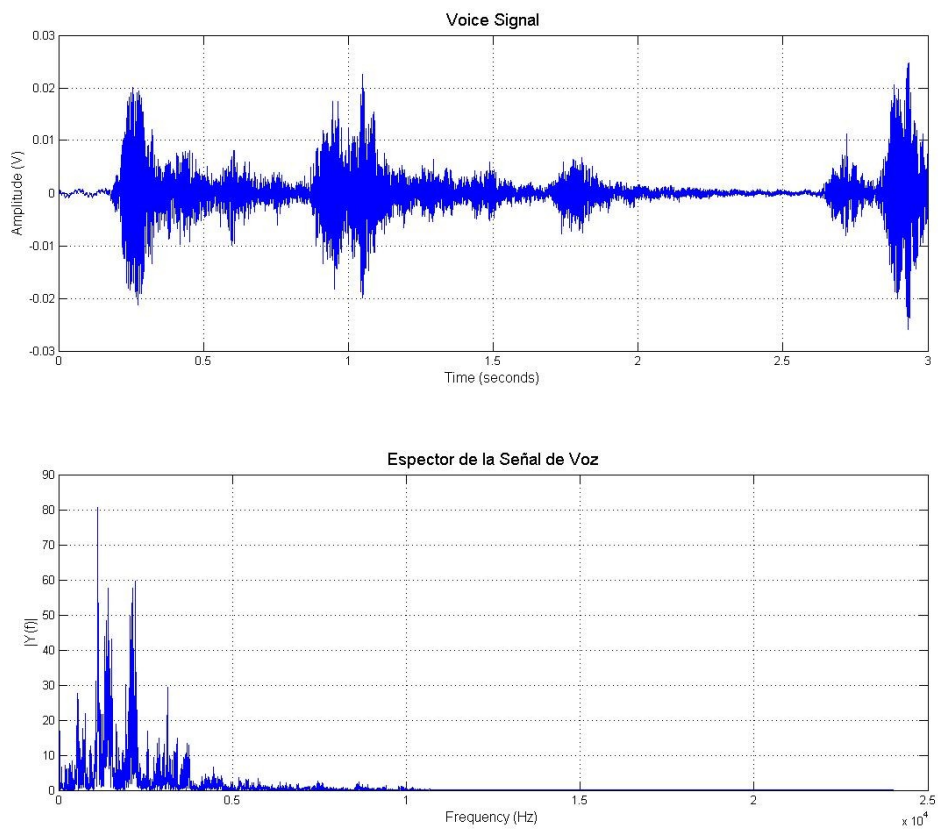


Figure 19: Filtered signal and its spectrum

2 Microphone Recording

The blue signal is the original one. The red one is the signal captured by the microphone. Obviously, it is lower than the original one, but it is very higher than the captured by the electret microphone.

We conclude from these results that the electret microphones are undesirable for this kind of experiment, because we need a good response from the microphones, and the sound source will be significantly far away (around 4 meters). Therefore, is decided to use condenser microphones. Although we can not finally profit from the flexibility of electret microphones, it is considered more important the quality of the recorded signal, since the final results will depend directly from it.

Obtaining good signals through microphones allow us to achieve better results in accuracy. This is not possible with electret microphones, from which we obtain signals with lot of unwanted noise.

3 Microphone Array Processing

3.1 Time-Delay estimation

They estimate the delay of the wavefront between sensors. At the peak of the cross-correlation, we have the time delay between the signals of two sensors.

$$r_{ij}(\tau) = \int_{-\infty}^{\infty} y_i(t) \cdot y_j(t - \tau) \cdot d\tau \quad (1)$$

$$D_{ij} = \max(r_{ij}(\tau))$$

Therefore, developing the equality we obtain that the gap is represented as a delta centered in the delay between signals.

$$\delta(t - D_{ij}) = TF^{-1} \left\{ \frac{TF\{y_i(t)\} \cdot TF\{y_j(t)\}^H}{\text{Normalization factor}} \right\} \quad (2)$$

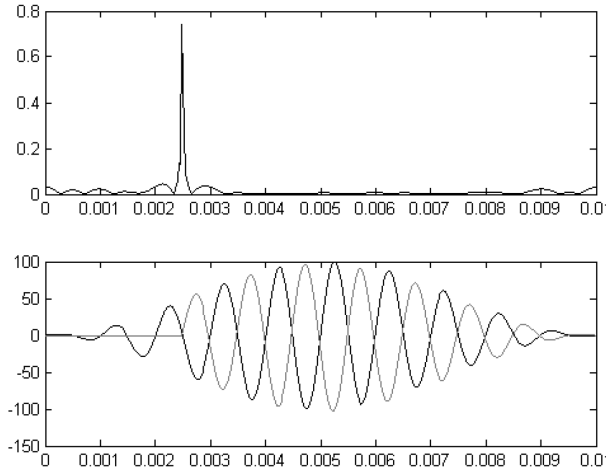


Figure 20: Theoretical representation of the delay between two sensors

So, to estimate the angle of arrival:

$$\theta = \arccos\left(\frac{c \cdot D_{ij}}{d}\right) \quad (3)$$

[OSP]

3.2 Direction-of-arrival estimation (DOA)

Below are various DOA (Direction of Arrival) methods. They are based on microphones arrays in real scenarios. A comparison is performed between various methods based on different theories:

- Cross Power Spectrum Phase, which study the delay between two sensors.

3 Microphone Array Processing

- Spectral estimation methods, analyzing energy in space axis
- Subspace methods, based on eigenvalues and eigenvectors to separate signal and noise subspaces.

Results of the speech and low signals tones show advantages and disadvantages of the methods simulated in real situations.

The aim of the work is to detect the angle of arrival in a room with the data captured by an array of microphones. It will be used a set of sensors located in different locations in space to measure the propagation of audio signals. From processing of the multichannel output we obtain information of the angle from which the source emits sound.

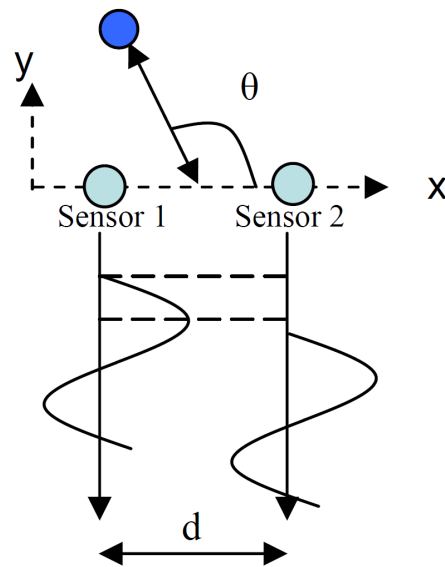


Figure 21: Relative phase difference between sensors depending on the angle of arrival.

The information of the direction of arrival is implicit in the information received by each sensor in the form of gap on adjacent sensors. For sensor m the time delay corresponds to:

$$\tau_m = m \frac{d \cdot \sin(\theta)}{c} \quad (4)$$

[MA]

3.2.1 Methods of estimation

To carry out the detection of DOA it has to be raised a signal model consistent with the scenario in which sources are found. There are different methods of DOA estimation:

3.2.1.1 Methods for estimating the delay between sensors

This method is based on obtaining the delay of the signal between two sensors doing the cross-correlation between signals received by both microphones. Making the correct calculations (explained above), we obtain the delay of the signals, and knowing the distance between the microphones and the propagation speed of sound, it can be obtained the value of the angle of incidence of the signal. [TDE]

3.2.1.2 Methods for estimating the power spectral density

The model defines a vector approach that it is formed as from the basis of information obtained from the sensors.

If is supposed we work in the far field and therefore the wave fronts are flat. The immediate consequence is that the only difference between two signals captured by each sensor is a delay dependent only on the position.

The methods are based on narrowband signals. It will therefore be necessary to treat different subbands in the case of voice. In dealing with speech signals, we do not have the imaginary part of the information. For this reason, when we detect a signal, the conjugate is in the position of angles of opposite sign. [TDE]

3.2.1.3 Methods of decomposition in principal components or estimate of the angular position.

It is based on implementing the features of vector spaces. The covariance matrix is decomposed into two matrices of eigenvectors and eigenvalues.

$$\hat{R} = \underline{\underline{E}} \cdot \underline{\underline{D}} \cdot \underline{\underline{E}}^H \quad (5)$$

where E is the matrix that contains in its NS columns eigenvectors and D is the diagonal matrix containing NS eigenvalues sorted in descending.

It is a method with high resolution in non-directional noise scenarios and distributed sources. [TDE]

3.2.2 Experiment description

Having chosen the method that we use and the types of microphones, we proceed to the description of the room where the experiment is performed, and a description of the equipment used for it.

3.2.2.1 Room

The room where the experiment is performed is the Signal Processing Laboratory of the TU Darmstadt. Its dimensions are: 5.46m x 5.43m x 2.78m (width x length x height). Inside there are tables, cabinets, chairs, two computers, three windows and various objects.

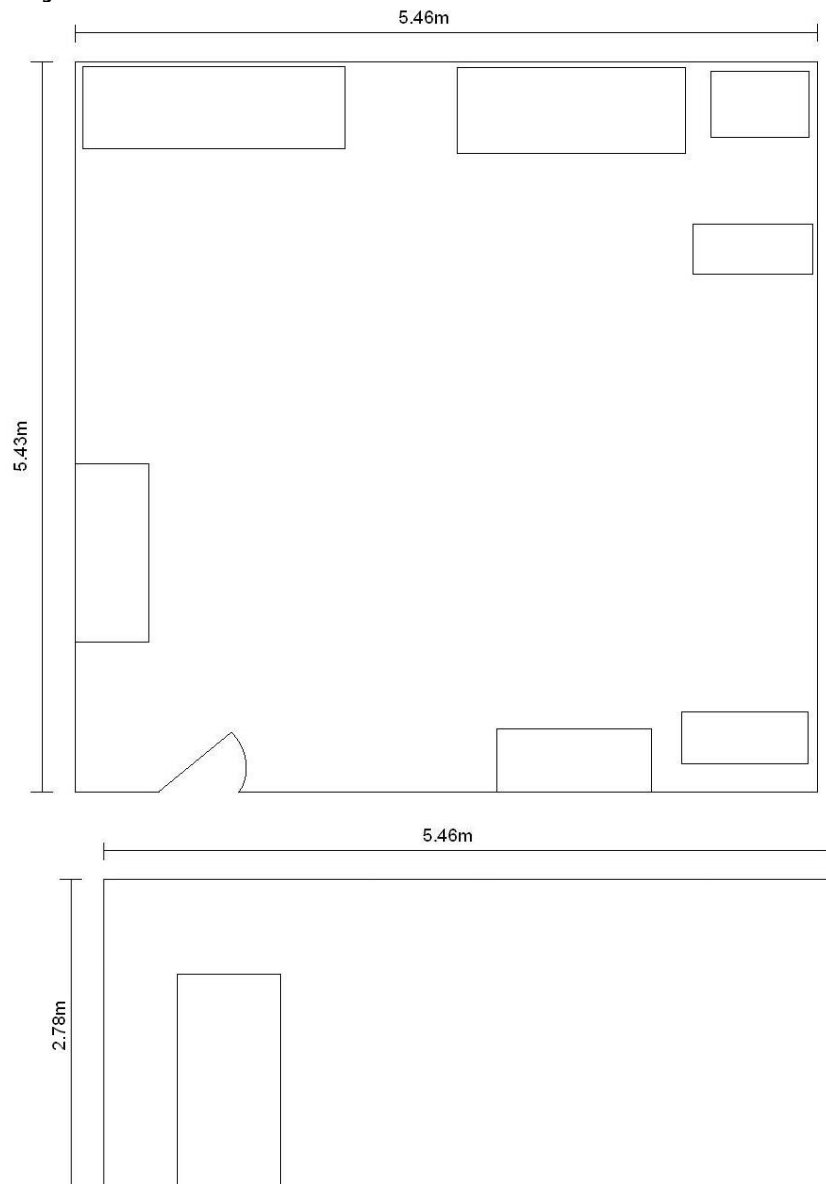


Figure 22: Seating Plan of the room (plan view and side view)

3 Microphone Array Processing

Can see that the room has parallel walls and is full of objects:



Figure 23: Laboratory conducting the experiment

3.2.2.2 Loudspeaker

The loudspeaker is used as a sound source rather than the voice of a person so that all tests are under the same condition (volume, position, height, directivity), and also for always use the same signal to obtain more reliable results (a human is unable to perform two equal tests).

The loudspeaker which is used is the model HS50M Yamaha, which has a white polypropylene cone of 5", a $\frac{3}{4}$ "-dome tweeter, it's 70-watt biamplified power and also has room control and frequency response switches. To place it at the right height, is used a height-adjustable stand. It is connected to the A/D converter via its XLR output with an XLR cable.

3 Microphone Array Processing



Figure 24: Loudspeaker

3.2.2.3 A/D Converter

It is used an analog / digital converter to recording up to 8 signals simultaneously in the PC. In addition, it allows to play back a signal from the speaker while is recording. Specifically, the model used is ADA8000 (Behringer)



Figure 25: A/D converter (ADA8000)

It has 8 different input channels, with the choice between XLR connector and Line In. The output of the converter is connected directly to the PC sound card, so it is enough for connecting the microphones to the PC.

3 Microphone Array Processing

3.2.2.4 Microphones

The microphones are responsible for receiving the acoustic signals to send them to the computer and process them later.

There are used six microphones with cardioid capsule to avoid capturing sounds from behind them (as well as reflections on walls)



Figure 26: Condenser microphone B-5 (Behringer) with cardioid capsule

3.2.3 Development work

We have described the method, location and components we use. This section explains how has been developed the work to achieve the ultimate goal and the steps that it have been followed.

3.2.3.1 Stages and Ejecution times

The first goal is to achieve that, by using a pair of microphones, the computer give us back the angle of incidence of the human voice on the axis of the microphones.

To do this, we first need to put a pair of microphones on an axis, which will be perpendicular to the imaginary line that connects them and will pass through its midpoint. We also place the speaker anywhere in the room. We develop an algorithm in matlab which is responsible for recording the signal with the two microphones at the same time, using two different channels. These two signals will be stored on the PC for later perform the cross correlation between them. In signal processing, cross-correlation is a measure of similarity of two waveforms as a function of a time-lag applied to one of them. This is also known as a sliding dot product or inner-product.

If you perform the cross correlation between the signals recorded by both microphones,

$$R_{x_1x_2}(\zeta) = E x_1(n+N)x_2(n+\zeta) = R_{x_1x_2}(N-\zeta) \rightarrow \max @ N \quad (6)$$

will produce a signal whose maximum is displaced from the center in N samples. The two recorded signals are almost the same, but one of them will be displaced in time over the other, because the sound wave arrives before to one microphone that to the other one. Therefore, the cross correlation indicates the gap between the two signals in number of samples.

3 Microphone Array Processing

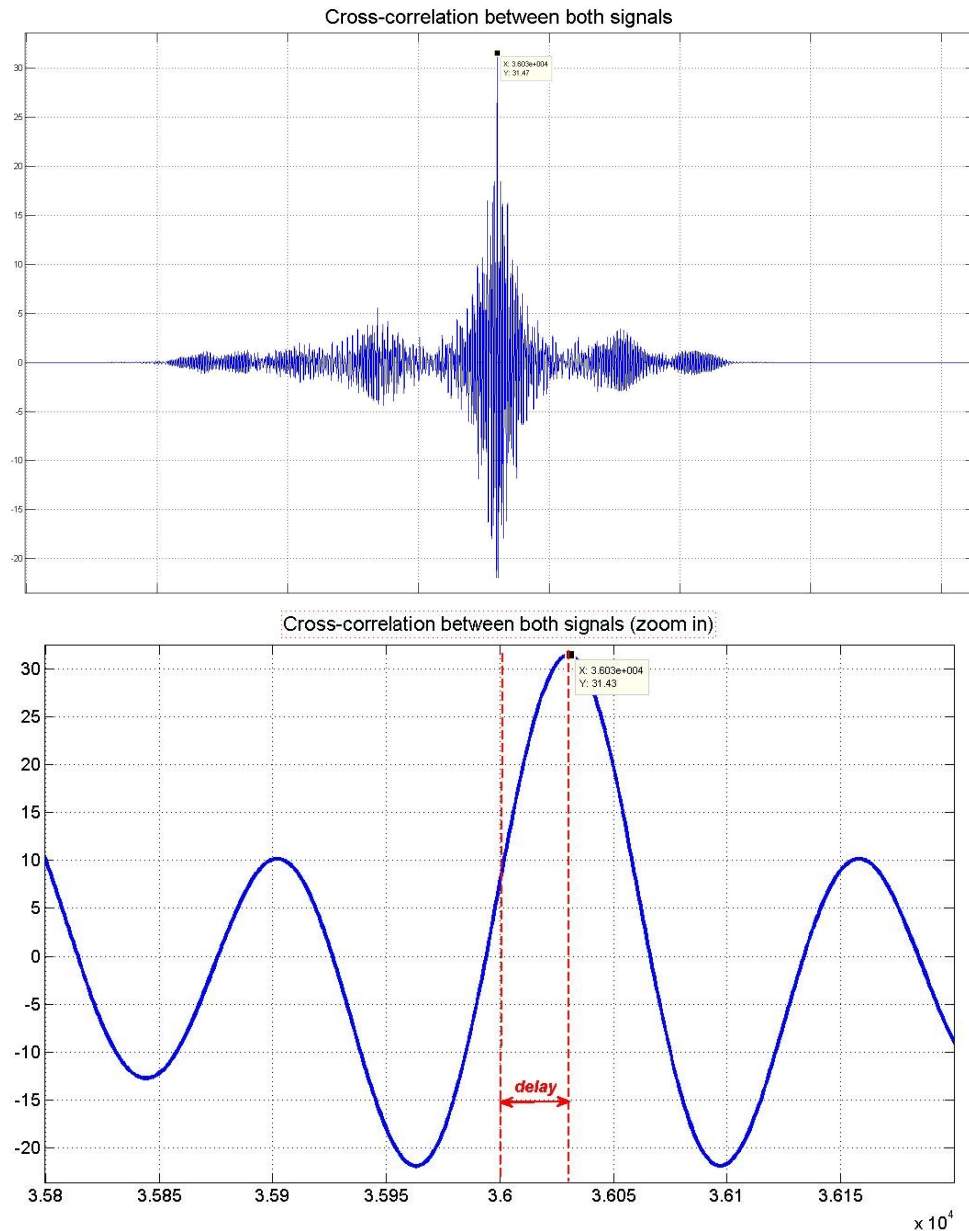


Figure 27: An example of the cross-correlation between two similar signals (one displaced in time over the other), and a zoom in of the central point (maximum). It shows the delay N in number of samples

Once obtained the delay of both signals in number of samples N , is transformed into time t (in seconds) by dividing the number of samples displaced by the sampling frequency f_s :

$$t = \frac{N}{f_s} \quad (7)$$

3 Microphone Array Processing

For obtain the angle of incidence *theta*, simply calculating the inverse cosine of the result of multiplying the time delay *t* by the speed of propagation of sound *v* (344 m / s) and divided by the distance between the two microphones *d*:

$$\cos(\vartheta) = \left(\frac{v \times t}{d}\right) \quad (8)$$

$$\vartheta = \arccos\left(\frac{v \times t}{d}\right) \quad (9)$$

Finally, we convert the angle from radians to degrees, by multiplying the previous result by 360 and dividing by pi:

$$\varphi(\text{degrees}) = \frac{\vartheta \times 360}{\pi} \quad (10)$$

3.3 Speaker location estimation

The goal now is to locate the position where is the speaker

The next step is to do the same thing just done, but with a new pair of microphones placed in another different position.

So it will be get two different angles, which cut in space in the approximate location of the sound source.

The same procedure it's followed, except that now we must take into account that we have two pairs of microphones, and therefore, two pairs of signals. The two cross-correlations are determined, and then, the angles of incidence are estimated.

Once this is done, is represented on the plane to see the position of the cutoff point, which should coincide with the position (coordinates) of the sound source.

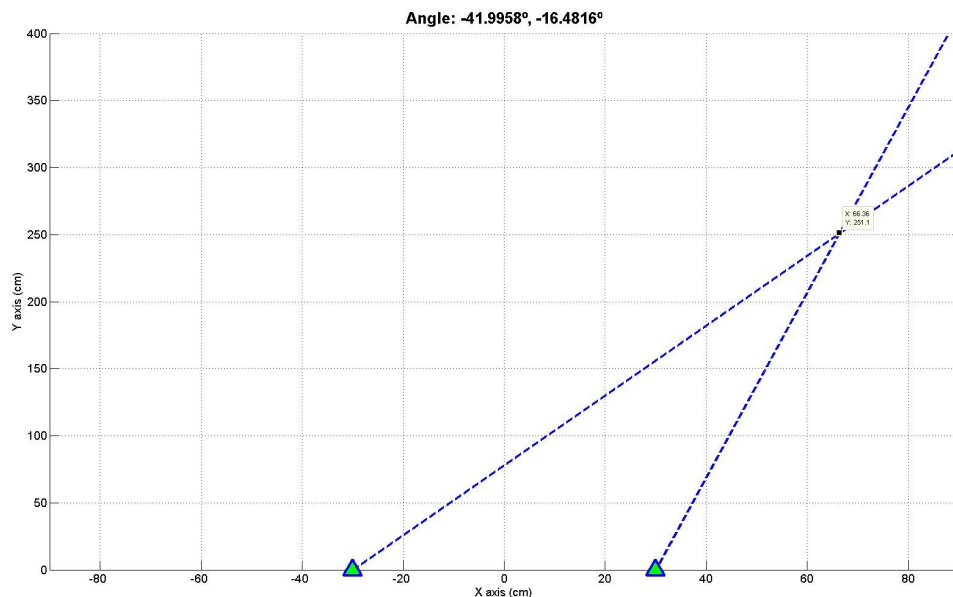


Figure 28: An example of the detection of the position of the sound source, obtained with the intersection of the angles of incidence of two pairs of microphones.

To make this more precise still, we design a new model with three pairs of microphones. In this way, can be considered more reliable delivery of the position of the sound source.

The process followed is the same as in the previous two cases.

3 Microphone Array Processing

Now, we must define a space in which pairs of microphones are placed on the same coordinate system, but each pair with its own axis. It is calculated the angle of each pair separately, and when the representation in the same coordinate systems will take into account their orientation, to obtain the correct position of the sound source.

Ideally, the three lines formed by the angles of incidence should cut at the same point, as shown in the figure below.

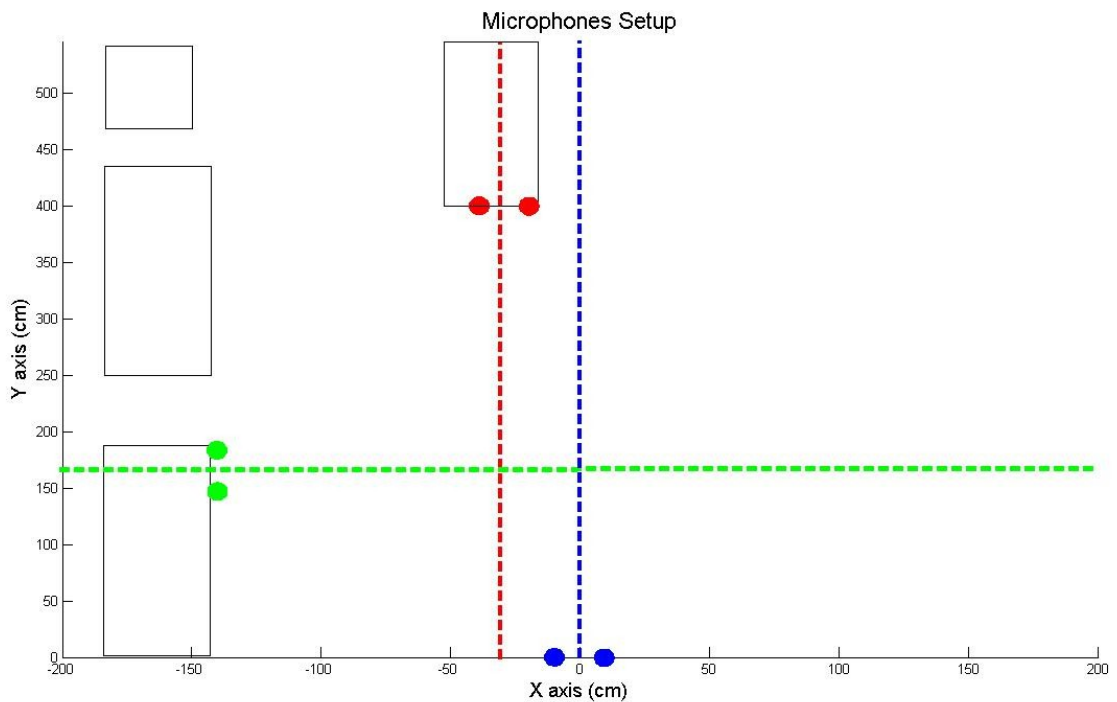


Figure 29: Position of the microphones in the room. Each pair and axis represented with one different color.

3 Microphone Array Processing

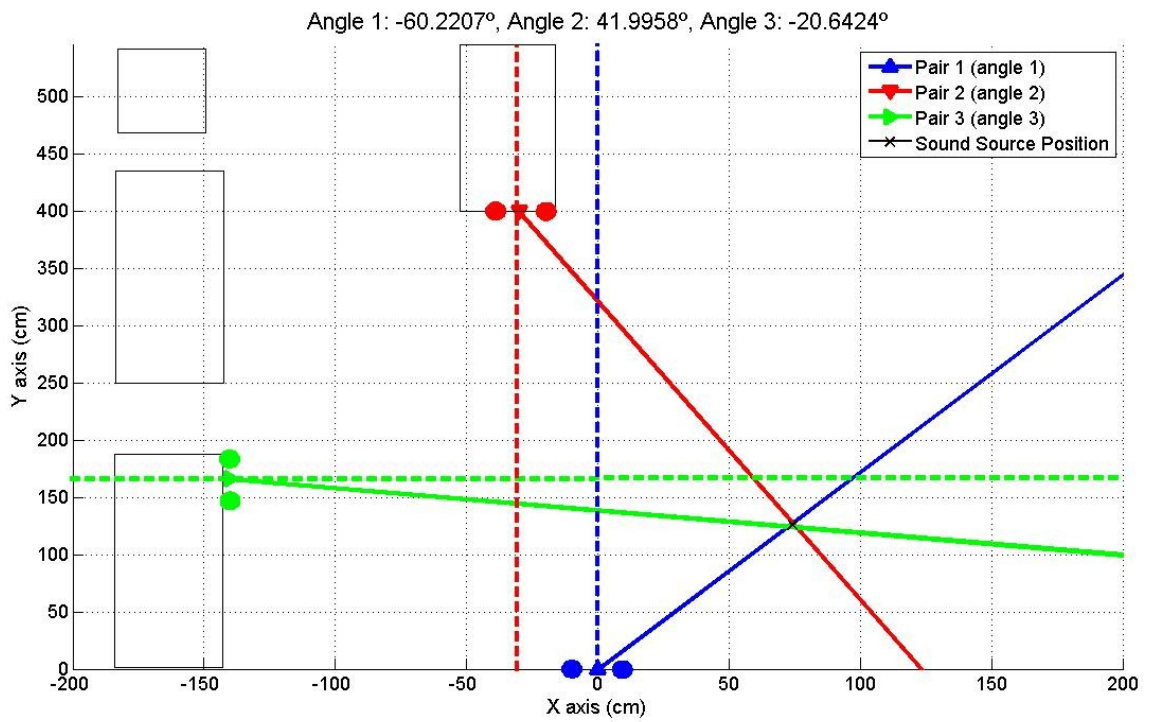


Figure 30: Position of the sound source. Cutoff of the three lines. (Ideal situation).

4 Experimental results

Below is the result after the execution of the experiment.

4.1 DOA results

For the first experiment, we obtain a graph that shows the point where the axis of the microphones are located (coordinate 0,0), and an infinite dashed line forming an angle with the perpendicular to the axis between the two microphones. That angle is the obtained of the algorithm, and corresponds to the direction of arrival of the sound source.

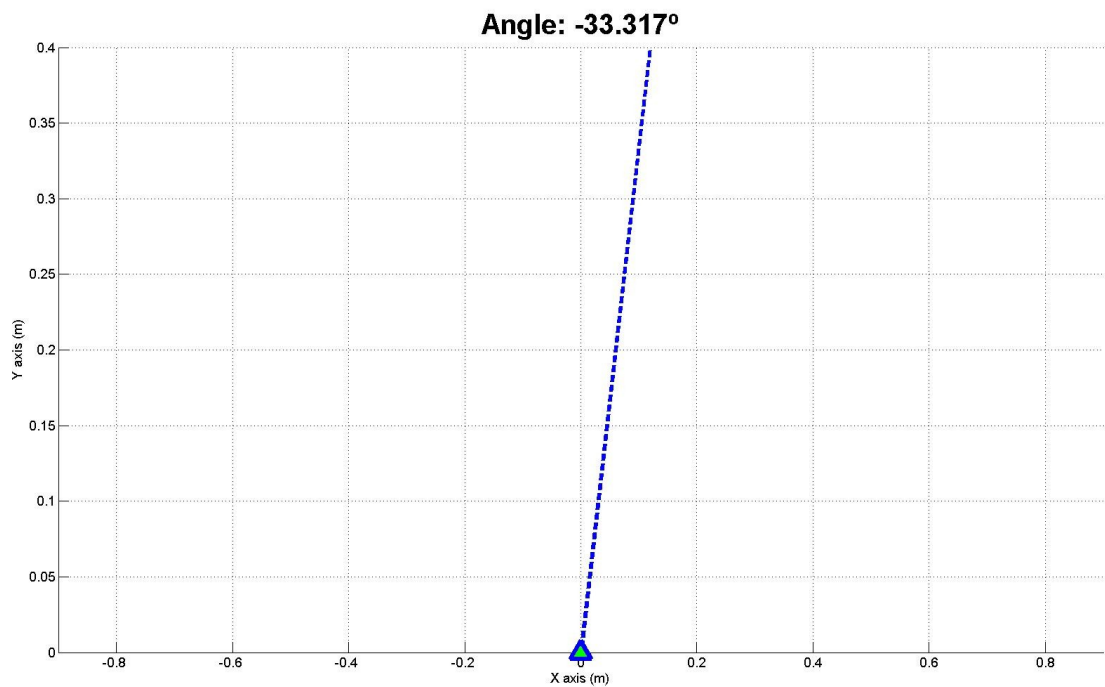


Figure 31: An example of the estimation of the incidence angle of the sound signal.

A full frontal impact is considered with 0° : Incidence from the left side (as seen from above) corresponds with 90° , and an angle of -90° incidence on the right side. The remaining angles follow this criterion.

4 Experimental results

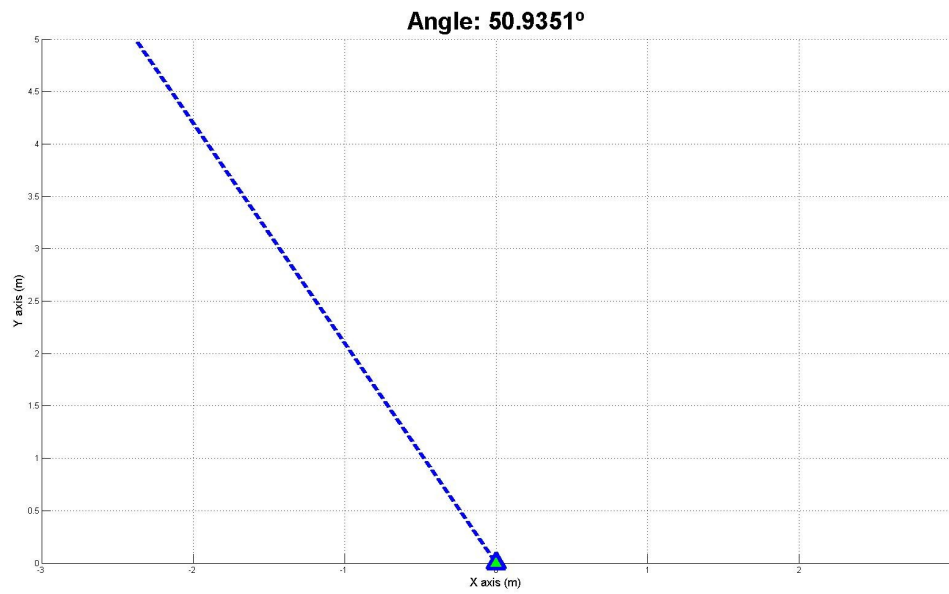


Figure 32: Another example of the estimation of the incidence angle of the sound signal.

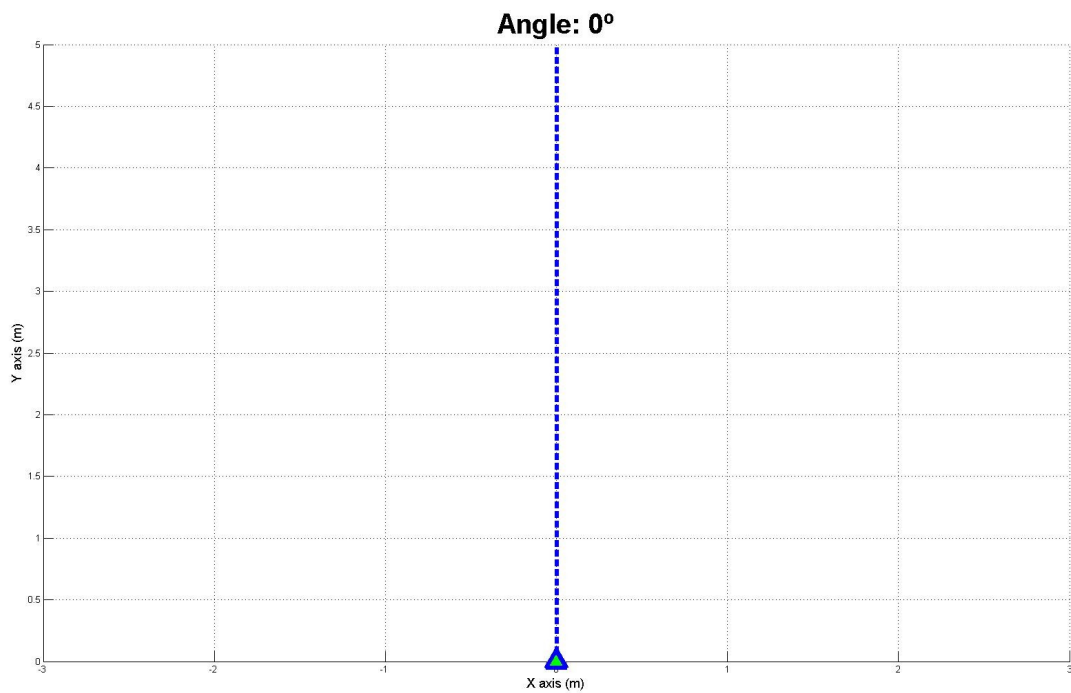


Figure 33: An example of the estimation of the incidence angle of the sound signal, with the speaker in front of the microphone pair.

4.2 Speaker location estimate results

As mentioned before, if we estimate the incidence angles in each pair of microphones, the three lines formed by the angles of incidence should be cut at the same point. But this is only an ideal case. In practice, what we get is a triangle formed by the cutoff points of the lines in pairs. It can be seen in the figure below:

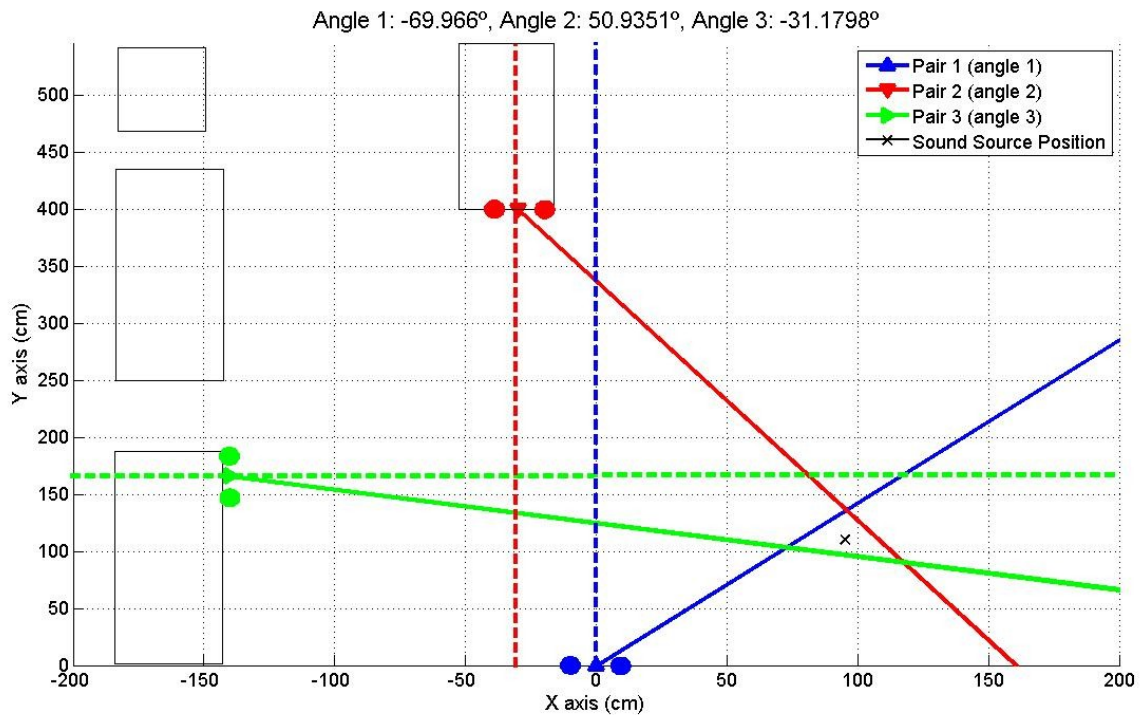


Figure 34: An example of the estimation of the position of the sound source.

Therefore, we calculate the midpoint of the resulting triangle, that will give us an approximation of the estimated position of the sound source.

This is done across the linear equation:

$$y=mx+n \tag{11}$$

Knowing the slope of each line (obtained from two points, by dividing the difference of their coordinates),

$$m=\left(\frac{By-Ay}{Bx-Ax}\right) \tag{12}$$

we can get the origin ordinate isolating n of the previous formula and using any point on the line. Then: $n=y-mx$.

4 Experimental results

Once obtained the equations of lines, we find the cutoff points between both lines.

$$\begin{aligned}X1 &= \frac{(n1 - n2)}{m2 - m1} \\ Y1 &= (m1 \times X1) + n1\end{aligned}\tag{13}$$

And finally, once obtained the three points of the triangle, we find the midpoint, which is to be considered as the coordinates where is the sound source.

5 Conclusions

First, we wanted to work with electret microphones to take advantage of their flexibility and small size to enable a better calibration and greater accuracy in the estimation, since, by placing sensors nearly, it increases the range detectable angle.

The quality of the signal obtained after being captured by an electret microphone was not good enough for this kind of processing because of the noise captured, but perhaps using a large number of microphones can eliminate this noise by correlation, and obtain interesting part of the signal to detect the gap between them, and thus estimate the angle of incidence.

Clearly, with professional condenser microphones can achieve optimum results to process the signals, but the fact that it is too difficult to place them close together, it extends the uncertainty of estimating the direction of arrival, and therefore, uncertainty estimation of the position of the sound source.

We have implemented an algorithm to estimate the direction of arrival of a sound source by using a pair of microphones through the Technical "Delay Between Sensors"

It has also been implemented another algorithm to estimate the location of a sound source based on the same technique but this time using three pairs of microphones.

The results obtained are not entirely reliable because the experiments were not performed in an acoustically suitable room. The lab features make that the measurement conditions are not good. The noise from the road and street in general was quite substantial, since the sound insulation of windows that fill the entire wall bordering to the street is not good. In addition, the geometry of the room and the materials of the walls, floor and ceiling did not help to avoid the reflections causing a disturbance in the signal captured by microphones, and that influences the delay estimated by cross correlation. Still, despite not getting the 'ideal' results, the estimate of the position of the speaker that was achieved was quite good.

6 APPENDIX

```
filtrada=filtrar2(y);           % It Filters the recorded signal
filtrada10=filtrada*10;       % Increases 10 times the filtered signal

% % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % %

pa_wavplay(y,48000,0,'asio');   % Reproduced recorded signal
pa_wavplay(filtrada,48000,0,'asio'); % Reproduced filtered signal
pa_wavplay(y10,48000,0,'asio'); % Reproduced filtered signal x10
pa_wavplay(filtrada10,48000,0,'asio'); % Reproduced filtered signal x10

% % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % %

% % % % ORIGINAL, FILTERED AND FILTERED x10 SIGNAL REPRESENTATION % % %
figure
plot(x,'g');
hold on;
plot(filtrada10);
plot(filtrada,'r');
h = legend('Original','filtrada10','filtrada',0);
set(h,'Interpreter','none');
title('Original signal, Filtered signal and filtered signal increased 10
times'),...
xlabel('Time (samples)'), ylabel('Amplitude (V)'), grid on

% % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % %

% REPRESENTATION OF THE ESPECTRUM OF THE ORIGINAL, RECORDED AND FILTERED
% SIGNALS

espectro(x);
espectro(y);
espectro(filtrada);
```

6.1.1 Functions 'filtrar2' and 'espectro' used:

```
function filtrada = filtrar2(y);
%LOWPASS_FILTER_3500 Returns a discrete-time filter object.

% M-File generated by MATLAB(R) 7.10 and the Signal Processing Toolbox
6.13.

% Generated on: 28-Jul-2010 14:14:36
% Equiripple Lowpass filter designed using the FIRPM function.

% All frequency values are in Hz.
Fs = 48000; % Sampling Frequency

Fpass = 3500; % Passband Frequency
Fstop = 3700; % Stopband Frequency
Dpass = 0.057501127785; % Passband Ripple
Dstop = 0.01; % Stopband Attenuation
dens = 20; % Density Factor

% Calculate the order from the parameters using FIRPMORD.
[N, Fo, Ao, W] = firpmord([Fpass, Fstop]/(Fs/2), [1 0], [Dpass, Dstop]);

% Calculate the coefficients using the FIRPM function.
b = firpm(N, Fo, Ao, W, {dens});
Hd = dfilt.dffir(b);

% [EOF]
B=Hd.numerator;

filtrada=filter(B,1,y);
```

6 APPENDIX

```
% % % % % % % % % % % % % % % SPECTRUM % % % % % % % % % % % % % % % % % % % % % % %
function [e] = espectro(y)
y_t=y.'; % Organizes the data in row vector.
Fs=48000; % Sampling frequency.
T=1/Fs; % Sampling time.
L=length(y_t); % Signal length.
t=(0:L-1)*T; % Time vector.
NFFT=2^nextpow2(L); % Next power of 2 of the length of 'y_t'.
Y=fft(y_t); % Apply Fourier Transform.
f=Fs/2*linspace(0,1,48000/2); % Frequencies vector.

% Displays the voice signal and the signal spectrum.
figure
subplot(2,1,1), plot(t,y_t), title('Voice Signal'),...
xlabel('Time (seconds)'), grid on
subplot(2,1,2), plot(f,2*abs(Y(1:48000/2))),...
title('Voice Signal Spectrum'),...
xlabel('Frequency (Hz)'), ylabel('|Y(f)|'), grid on
```

6.2 Matlab code of the algorithm to estimate the direction of arrival:

```

% % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % %
% % % % % % % % ESTIMATION OF SIGNAL INCIDENCE ANGLE % % % % % % % % % % % % %
% % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % %
fs=48000; % sample frequency
N=48000; % no. of samples
t=1; % length of the signal, in seconds
N=t*N; % length of the signal, in samples
n=0:N-1;

x=wavread('voice.wav');
x=x(1:N,1); % Cut the signal (only t seconds)
longitud=length(x);

% simultaneous playback and recording (CHANNELS 7 & 8)

y=pa_wavplayrecord(x,0,fs,N,3,8,0,'asio');

y1=y(:,1); % Signal microphone 1
y2=y(:,2); % Signal microphone 2

c2=xcorr(y2,y1); % Cross correlation between both signals

[maximo2,indice2]=max(abs(c2)); % Maximum of the cross-correlation
signal. Value and index.

[theta,linea_x,linea_y]=angulo(indice2,fs,longitud); % Calls "angulo"
function to calculate the angle

(['The signal comes with a ',num2str(theta),'° angle with the axis of the
microphones'])

figure
grid on
axis ([-3 3 0 5]);
line(linea_x,linea_y,'LineWidth',5);
title(['Angle: ', num2str(theta),'°'])

```

6 APPENDIX

```
function [angle,linea_x,linea_y] = angulo(indice,fs,longitud)
indice=indice-longitud;

v=344;                % Sound propagation velocity (m/s)
d=0.05;              % Distance between the microphones (m)
% tiempo=indice/fs;

angle_rad=acos(v*indice/(d*fs)); % Incidence angle (RADIANS)
angle=(angle_rad*360/pi)-180;    % Incidence angle (DEGREES)

% % % % % % % % % % % REPRESENTATION % % % % % % % % % % % % % % % % %
m=15;
A_x=0;
A_y=0;

B_x=((cos(angle_rad))*m);
B_y=(sin(angle_rad))*m;

linea_x=[B_x A_x];
linea_y=[B_y A_y];
```

6.3 Matlab code to determine the position of the sound source, using three pairs of microphones:

```
% % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % %
% % % % % % % % ESTIMATION OF POSITION OF SOUND SOURCE % % % % % % % %
% % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % %

fs=48000; % sample frequency
N=48000;  % no. of samples
t=3;     % length of the signal, in seconds
N=t*N;   % length of the signal, in samples

% play tones
n=0:N-1;

x=wavread('voice.wav');
x=x(1:N,1); % Cut the signal (only t seconds)
longitud=length(x);

% simultaneous playback and recording (CHANNELS 7 & 8)
y=pa_wavplayrecord(x,0,fs,N,3,8,0,'asio');

y1=y(:,1); % Signal microphone 1
y2=y(:,2); % Signal microphone 2
y3=y(:,3); % Signal microphone 3
y4=y(:,4); % Signal microphone 4
y5=y(:,5); % Signal microphone 5
y6=y(:,6); % Signal microphone 6

c1=xcorr(y4,y3);
[maximo1,indice1]=max(abs(c1));

c2=xcorr(y6,y5);
[maximo2,indice2]=max(abs(c2));

c3=xcorr(y2,y1);
[maximo3,indice3]=max(abs(c3));

angulo2(indice1,indice2,indice3,fs,longitud); % Calls "angulo" function
to calculate the angle
```


6 APPENDIX

Function that determines the position of the sound source, from the delays (in number of samples) for each pair of microphones:

```
function [angle1, linea_x, linea_y] =
angulo2(indice1, indice2, indice3, fs, longitud)
indice1=indice1-longitud;
indice2=indice2-longitud;
indice3=indice3-longitud;

v=344; % Sound propagation velocity (m/s)
d=0.10; % Distance between the microphones (m)
d2=0.08;

angle_rad1=acos(v*indice1/(d*fs)); % Incidence angle (RADIANS)
angle1=(angle_rad1*360/pi)-180; % Incidence angle (DEGREES)

angle_rad2=acos(v*indice2/(d*fs)); % Incidence angle (RADIANS)
angle2=(angle_rad2*360/pi)-180; % Incidence angle (DEGREES)

angle_rad3=acos(v*indice3/(d2*fs)); % Incidence angle (RADIANS)
angle3=(angle_rad3*360/pi)-180; % Incidence angle (DEGREES)

% % % % % % % % % % % REPRESENTATION % % % % % % % % % % % % % % % %
m=2500;
A_x=0;
A_y=0;
B_x=((cos(angle_rad1))*m);
B_y=(sin(angle_rad1))*m;
linea_x1=[B_x A_x];
linea_y1=[B_y A_y];

C_x=-30;
C_y=400;
D_x=((cos(angle_rad2-pi))*m)-30;
D_y=((sin(angle_rad2-pi))*m)+400;
linea_x2=[D_x C_x];
linea_y2=[D_y C_y];

E_x=-140;
E_y=166;
F_x=((sin(angle_rad3))*m)-166;
F_y=(-cos(angle_rad3))*m+140;
linea_x3=[F_x E_x];
```

6 APPENDIX

```
linea_y3=[F_y E_y];

figure
grid on
axis ([-300 300 0 500]);
line(linea_x1,linea_y1,'MarkerFaceColor',[0 0 1],'MarkerSize',10,...
     'Marker','^',...
     'LineWidth',2,...
     'Color',[0 0 1]);
title(['Angle 1: ', num2str(angle1),'°, Angle 2: ', num2str(angle2),...
     '°, Angle 3: ', num2str(angle3),'°'],'FontSize',16)

line(linea_x2,linea_y2,'MarkerFaceColor',[1 0 0],'MarkerSize',10,...
     'Marker','v',...
     'LineWidth',2,...
     'Color',[1 0 0]);
line(linea_x3,linea_y3,'MarkerFaceColor',[0 1 0],'MarkerSize',10,...
     'Marker','>',...
     'LineWidth',2,...
     'Color',[0 1 0]);
% Create xlabel
xlabel('X axis (cm)','FontSize',14);

% Create ylabel
ylabel('Y axis (cm)','FontSize',14);

% % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % %

m1=(B_y-A_y)/(B_x-A_x);
m2=(D_y-C_y)/(D_x-C_x);
m3=(F_y-E_y)/(F_x-E_x);

n1=0;
n2=400+30*m2;
n3=166+140*m3;

X1=(n1-n2)/(m2-m1);
Y1=m1*X1+n1;
P1=[X1 Y1];

X2=(n1-n3)/(m3-m1);
Y2=m1*X2+n1;
P2=[X2 Y2];
```

6 APPENDIX

```
X3=(n3-n2)/(m2-m3);
Y3=m3*X3+n3;
P3=[X3 Y3];

XM=[X1 X2 X3];
XM_M=mean(XM);
YM=[Y1 Y2 Y3];
YM_M=mean(YM);
P=[XM_M YM_M]
hold on
plot(XM_M, YM_M, 'MarkerFaceColor', [0 0 0], 'MarkerSize', 10, 'Marker', 'x', ...
      'LineWidth', 2, ...
      'Color', [0 0 0]);
```

7 Bibliography

AIP: , Theory, calibration, and measurements, 1995

MA: , Microphone Arrays, 2001

OSP: Sophocles J. Orfanidis., Optimum Signal Processing: An Introduction., 1985

TDE: Michael S. Brandstein, John E. Adcock, HarveyF. Silverman., A practical Time-Delay Estimator for Localizing Speech Sources with a Microphone Array.,

Don H. Johnson and Dan E. Dundgeon. *Array Signal Processing: Concepts and Techniques*. Prentice Hall, 1993.