POLITECNICO DI MILANO

Master of science In Music and Acoustic Engineering

Implementation of a low-cost acoustic camera using arrays of MEMS microphones

Author: Dario De Lucia

Student ID: 927373
Advisor Fabio Antonacci
Academic Year 2020-21

…to Marta

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor Prof. Fabio Antonacci for his guidance throughout this thesis. I really appreciate his professionalism and his friendliness. I Would like to thank the Ing. Daniele Ponteggia for his valuable advices and for being a point of reference.
I would like to offer my special thanks to Enrico Ricciardi for his constant assistance, for having any tools that exists in the word that I needed and for being a wise and precious guide.
I would like to thank my brothers and my parents who have encouraged and support me throughout my study and life. I hope to repay your confidence.
I would also like to thank my friends Adriano and Tommaso for the projects shared together and for always being present.
The final and greatest thanks goes to my woman Valeria, for her infinite patience and the immense support shown throughout my studies. Thank you for understanding the particular moments I went through. Love you!

CONTENTS

LIST OF FIGURES

ABSTRACT

In the last few years viewing an audio source has become a widely used tool not only in teleconferencing, speech enhancement and recognition, video games etc, but also in the acoustic field, especially regarding noise localization and sound insulation.
Although its potentiality, an acoustic image is difficult to achieve in environments with a large amount of noise and reverberation. An effective approach to obtaining a clean recording of desired acoustic signal comes from Beamforming Theory coupled with Microphone Array.
The latter together with a camera is usually referred to as acoustic camera, a device used to locate sound sources and to characterize them.
In this thesis we have designed an acoustic camera using a microphone array mini-DSP UMA 16 and MATLAB software in order to determine sound power estimation performance and sound source separation ability. We have implemented it in two steps: a static setup in which the audio signals have been acquired by a microphone array and have been processed at later time, while in a second step we have extended it in a real time application.
Alter an investigation about the techniques utilized for an acoustic camera application, we have explored the uses of the array and the beamformer in order to obtain a sound intensity map and reconstruct the acoustic scene.
The obtained results presented in this thesis show that reliable application of beamforming techniques can generate a sound intensity map with a fair accuracy both in static setup and in real time mode.

1 INTRODUCTION

The term acoustic camera has been widely used during the 20th century to designate various types of acoustic devices, such as underwater localization systems or active systems used in medicine.
Nowadays it designates for every transducer array coupled with a camera which are enabled to localize a sound source and visualize it [11][17].
The advantage to "see" the sound and characterize it, has stimulated many companies to produce their own acoustic camera. But the signal processing required is very intensive and the process needs powerful hardware. So, cameras that do perform signal processing in real time tend to be large and very expensive.
Starting from several studies in this field [25][26][27][28], this thesis investigates different approaches designing a low-cost acoustic camera using micro electro-mechanical microphones (MEMS) collected into a plane array [18]. MEMS are micro-scale devices that provide high fidelity acoustic sensing and are small enough to be included in a tightly integrated electronic product [23]. They are used to measure the propagating wavefield and transferring the filed energy to electrical energy. Since the wavefield is assumed to have information about the signal source, it is sampled into a data set to be processed in order to extract as much as possible information from it. This process is a branch of array signal processing, and it involves several algorithms which are known as beamforming techniques. The aim of these techniques is to detect and to estimate many properties and parameters of the signal such as power level and direction of arrival (DOA) [32].
The estimation problem is a key research area in the array signal processing and many engineering applications [15], such as wireless communications, radar, sonar, other devices that need to be supported by direction of arrival estimation. For example, the job of radar systems is to detect and locate objects [20]. A signal emitted at pre specified frequency hits the objects. One of the reflected rays comes back in the direction of the radar to be detected by a receiving antenna which delivers to the receiver information about the distance between the radar and the object location. Same

principle is involved in sonar systems. The difference between radar and sonar is that sonar is used underwater, and it uses acoustic waves instead electromagnetic waves which are used in radar systems.
Moreover, the highly growing demands for mobile communications services have increased the needs of more efficient beamforming techniques [21].
However, it is possible to distinguish different methods of beamforming such as Conventional Beamforming, Statistically Optimum Beamforming or Adaptive Beamforming. Conventional beamforming is also referred to as the Delay and Sum method (DAS). The idea is to scan across the angular region of interest, and whichever direction produces the largest output power is the estimate of the desired signal's direction. The problem associated to DAS algorithm is the poor resolution which is a significant weakness of the method. For time varying signal environments, such as wireless cellular communications systems, statistics change with time as the target and interferers move around. Furthermore, for the time-varying signal propagation environment, a recursive update of some beamforming parameters is needed to track a moving interference so that the spatial filtering beam will adaptively steer to the target time-varying DOA, thus resulting in optimal transmission/reception of the desired signal. Algorithms that do this job are called Adaptive Beamforming and due to their high resolution, they are widely used in different applications.

Although the array signal process is a fundamental aspect in DOA estimation problem, building a low-cost acoustic camera involves a suitable hardware system. Most previous studies have shown how the accuracy of the device depends on many factors. One of the main factors is the number and the type of the array elements [1]. Generally, the greater number of array elements the better will be the estimation performance. At the same time parameters as directivity or SNR can modify the frequency response of the system as well as the amplitude and phase of microphone sensor or their position on the array.

In order to implement a low-cost acoustic camera, we have employed the mini-DSP UMA 16 microphone array together with a CMOS C5014 usb camera which perfect fit the right compromise between price and resolution and we have applied two different beamforming techniques.

The aim of this thesis is to develop an acoustic camera that is cheaper than the commercial one. With this in mind, we have divided the work in two steps about a static acquisition and a real-time implementation respectively. For the static acquisition, the idea was to explore different beamforming techniques and evaluate in terms of accuracy, resolution, and computational time which technique could be more suitable for our system in real time application. In particular, we have implemented a classical beamforming (DAS) and subspace method called Multiple Signals Classification (MUSIC). Regarding the former, depending on the position of the sound source, we have noticed that it can be influenced by the reflected sound that should be seen by the system as another source. In terms of accuracy, it varies in both Azimuth and Elevation angles from 4 to 20 degrees while the resolution depends on the scanning point of the steering vector. Obviously, the greater the number of scanning point the greater will be the computational time. Different from DAS, MUSIC algorithm does not present the problem of the reflected sound and it also shows an improvement of the accuracy which decreases from 20 to 5 degrees for both Azimuth and Elevation, while the resolution depends on the scan angles too.

Due to its relative localization error and a faster time response, we have decided to implement MUSIC method in the real time section. For easy to use we have developed an application using App Designer tool of Matlab in order to create a simple graphic user interface (GUI).

As a result, we have designed and implemented an acoustic camera with a frequency range between 200 and 4000 Hz that is also able to detect and locate sound sources in the reverberating environments.

2 STATE OF ART AND BACKGROUND

The goal of the acoustic camera is to localize a sound source drawing on its position a power map that corresponds to its sound intensity. Today all commercial acoustic cameras use a microphone array and a camera on its center in order to collect the incoming signal and visualize the sound scene. This signal is analyzed through a beamforming technique, and after a localization, the system displays a power map on the source position that represents the sound intensity. This section explains the basic principles of microphone array and beamforming techniques. We will review some previous studies about the suitable number of microphones in the array and its shape and we will analyze some of widely used beamforming techniques. Finally, we will present the device and software used to implement our low-cost acoustic camera

2.1 MICROPHONE ARRAY

A microphone array is a composition of spatially distributed microphones [22][24]. This technology is used in many fields, most notably to operate with electromagnetic waves (radar, radio astronomy, tomography) or ultrasound waves like a sonar.

Over the past two decades, microphone arrays have been increasingly utilized for sound source separation and amplification.

Moreover, as the array technologies have become cheaper and accessible, there has been growing interest to use such setups for capturing contextualized audio events for building context-aware applications.

More recently, rapid advances in processor technologies have propelled small-scale microphone arrays into many consumer electronics products such as smartphones and personal gaming devices. In fact, most smartphones have at least two microphones for noise cancellations or even they have three microphones, which will enable more microphone array-based applications such as automatic voice tracking.

However, the commonest widely applications of the microphone array are designed for beamforming, a technique used to amplify sound from one direction and suppress sound coming from other directions [2][8].

2.2 ARRAY GEOMETRY

In an acoustic enclosure the microphone array features the capability of obtaining the actual three-dimensional position of sound sources by estimating several direction-of-arrival given geometrical considerations.

The way in which the microphones are arranged on the array defines the array's shape. There have been developments of different microphone arrays such as linear array, planar array and spherical array [35]. In a linear microphone array the sensors are linearly distributed with equal inter-element distances. In the same way planar arrays have the microphones equally distributed on a plane. Rectangular (or Square) and Circular arrays are the widely used array of this type. The same goes for spherical arrays where all the microphones lie on the area of a sphere.

Basically, all shapes are imaginable, but in practice shapes with specific symmetrical properties show certain acoustic capabilities.

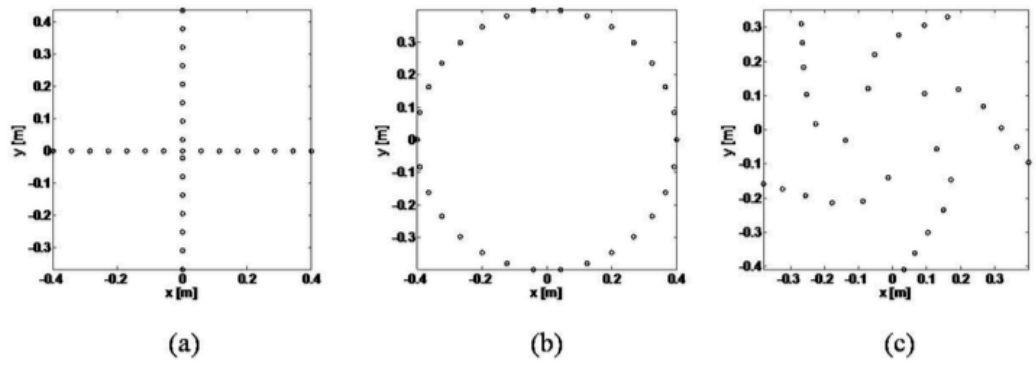

Figure 1 *The array shapes of interest consisted of 30 microphones. (a) a cross type, (b) a circular type, (c) a modified spiral type array. Image taken from [9].*

Several simulations were performed considering square and circular arrays composed by MEMS microphones with varying number of microphones and varying spacing between them [3]. The obtained results at broadband frequency range demonstrated that the best results are obtained using a square microphone array with 16 microphones.

The same study proceeded investigating the possibility to design an acoustic camera with better performance for a broadband frequency range. It was decided to simulate three new square arrays with 12, 24 and 48 microphones respectively at the following frequencies 1 kHz, 2 kHz and 4 kHz. The results demonstrated that all the attained arrays have a respectable main lobe Gain and significant attenuation of side lobes but the best design for the acoustic camera is the square array of 24 microphones with a microphones distance equal to 0.1 m.

In addition, it can be noticed that the gain in the desired direction as well as the number of side lobes and their gain is very frequency dependent since at 2 kHz and 4 kHz all arrays have a better attenuation of side lobes.

2.3 Directivity Pattern

The response of a receiving aperture is inherently directional in nature, because the amount of signal seen by the aperture varies with the direction of arrival. The aperture response as a function of frequency and direction of arrival is known as the aperture directivity pattern or beam pattern [33][34].

The ability for an acoustic camera to determine the direction of arrival from impinging sound waves is limited by the performance of the microphone array used.
Those performance can be quantified through three main performance parameters:

- 3 dB beamwidth
- Relative side-lobe level
- Peak-to-zero distance

All the performance parameters are functions not only of the array geometry, but also of the number of microphones and of the frequency of the impinging signal.

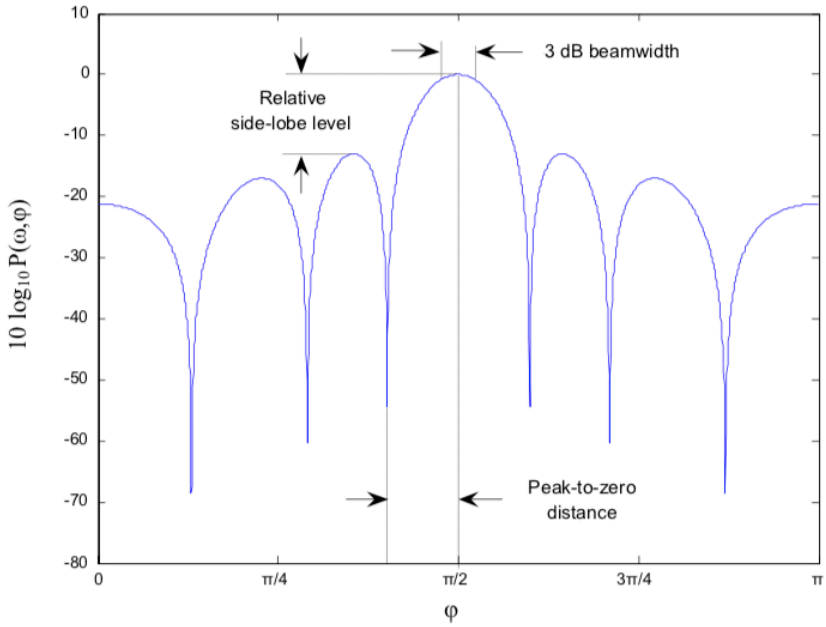

Figure 2 *Microphone Array parameters at 1 KHz. Image taken from [6]*

The 3 dB beamwidth, also called half power beamwidth, is the region where the main lobe has not decreased by more the 3 dB. The relative side lobe level expresses the relative sensitivity of the first side lobe compared with the main lobe. Peak-to-zero distance measures the angle from the main lobe maximum to the first minimum.

In previous research [3], it was shown that increasing the number of microphones produces a larger and narrowed main lobe, which means a higher gain of the signal in the desired direction. Furthermore, it was noticed that increasing the spacing between adjacent microphones will result in an increase of the side lobe.

Therefore, it was concluded that the main parameters of the microphone array are the number of microphone and the spacing between adjacent microphones.

2.4 OPERATING PRINCIPLE

Consider a sound wave impinging on a Uniform Linear Array as shown in figure 3. The spatial sound is recorded by the array and a separate signal is generated for each microphone. Due to the difference in the distance the sound must travel to reach each microphone in the array and time delay occurs between the signals that are being recorded by the microphones. This type of delay is called *Time Difference of Arrival TDOA*. Knowing the geometry of the array and the distance between microphones *d*, through beamforming techniques the *Direction of Arrival DOA* of a sound source can be detected from the *TDOA*.

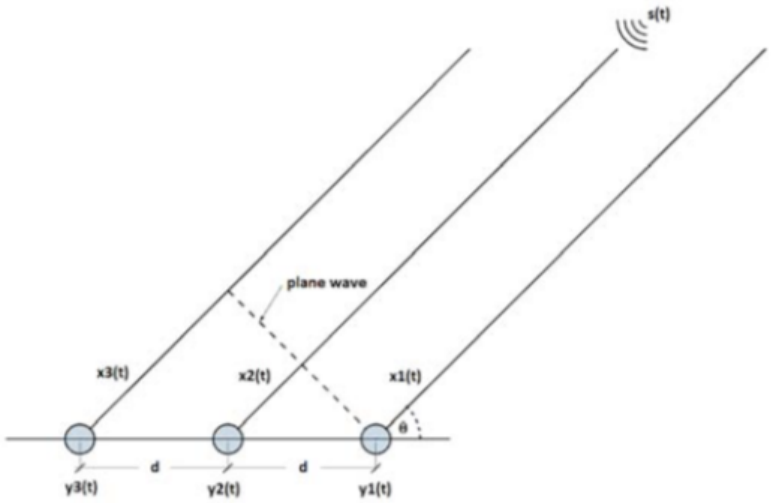

Figure 3 *A liner microphone array in a far-field. Image taken from [5]*

The calculation of the *TDOA* can be reduced by making a fundamental assumption:

the wave front of a sound source is curved, and it introduces much more complexity in the actual position calculation using the TDOA. However, this wave front can be approximated as flat, but it introduces an error in the calculated position. The relative size of this error compared to other error sources depends on the distance *r* between the sound source and the microphone array and it depends on the size of the array. If the distance between the sound source and the array is large compared to the dimensions of the array, the approximation does not introduce much additional error in the calculated position. This is referred to as a *Far-Field* case. If the distance between the sound source and the array is small compared to the dimension of the array this approximation cannot be made and it is referred to as a *Near-Field* case.
In most beamforming applications the assumption of Far-Field case simplifies the analysis. It means that the signal source is located far enough away from the array so that the wave fronts impinging on it can be seen as a plane wave. The equation which verifies the far-field condition is

$$r \gg 2\lambda \qquad (2.1)$$

Where λ is the wavelength of the incoming signal.
When the microphone array is in the far field, the sound from the source is radiated as plane wave meaning that the incident angle θ, with respect to the array axis, will be equal for every microphone in the array. The signal from the source is represented by *s(t)*. The additional distance that the sound has to travel to reach the *i-th* microphone m_i, with respect to the reference microphone m_1, is $d_i \cdot cos\ \theta$.

In the case of linear array, the microphones in the array are positioned equidistantly, i.e., $d_i=d$. For such linear array it is possible to define the time delay τ_1 for each microphone m_i. That time delay represents the *TDOA* for uniform linear array in the far filed and it can be calculated as

$$\tau = \frac{d \cdot cos\theta}{c} \qquad (2.2)$$

Where *c* is the speed of sound.

2.5 SPATIAL ALIASING

Consider the linear uniform array in figure 2.2.
The output of the *ith* microphone can be modeled as [4]

$$y_k(n) = h_k(n) \cdot s(n - \tau_k) + e_k(n) \qquad (2.3)$$

Where:

- *s(n)* is the source signal measured at reference point
- $h_k(n)$ is the impulse response of the microphone
- τ_k is the propagation delay between the reference point and the microphone
- $e_k(n)$ is the additive noise at the microphone

Using the same notation, the STFT of the microphone signals can be written as

$$y_k(t) = H_k(\omega_c) \cdot s(t) \cdot e^{-j\omega_c \tau_k} + e_k(t) \qquad (2.4)$$

Where $H_k(\omega_c)$ is the frequency response of the *kth* microphone evaluated at $\omega = \omega_c$.

Adopting a vector representation, it is possible to derive the complete array model for a single source

$$y(t) = a(\theta) \cdot s(t) + e(t) \qquad (2.5)$$

Where

- $y(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_M(t) \end{bmatrix}$ is the *Array Vector*

- $a(\theta) = \begin{bmatrix} H_1(\omega_c)e^{-j\omega_c\tau_1} \\ H_2(\omega_c)e^{-j\omega_c\tau_2} \\ \vdots \\ H_M(\omega_c)e^{-j\omega_cM} \end{bmatrix}$ is the *Propagation Vector* or *Steering Vector*

- $e(t) = \begin{bmatrix} e_1(t) \\ e_2(t) \\ \vdots \\ e_M(t) \end{bmatrix}$ is the *Noise Vector*

Now the explicit dependence of the time delay as a function of the *DOA* θ can be derived as

$$\tau_k = (k-1) \cdot \frac{d \cdot sin(\theta)}{c} \quad \text{for } \theta \in [-90^\circ, 90^\circ] \qquad (2.6)$$

Usually, the microphones in the array are identical and omnidirectional. It means that the impulse responses of the microphones are the same.

$$H_1(\omega_c) = H_2(\omega_c) = \cdots = H_M(\omega_c) \qquad (2.7)$$

Using (2) and (7) the *Propagation Vector* becomes

$$a(\theta) = \begin{bmatrix} e^{-j\omega_c \frac{dsin\theta}{c}} \\ e^{-j\omega_c \frac{dsin\theta}{c}} \\ \vdots \\ e^{-j(M-1)\omega_c \frac{dsin\theta}{c}} \end{bmatrix} \quad (2.8)$$

The terms of the Propagation Vector correspond to the samples of complex sinusoid

$$e^{-j\left[\omega_c dsin\frac{\theta}{c}\right]k},\ k = 0, 1, \ldots,\ M-1 \quad (2.9)$$

And the frequency of this sinusoid is the so-called *Spatial Frequency*

$$\omega_s \triangleq \omega_c \frac{dsin\theta}{c} \quad (2.10)$$

Moreover, the complex sinusoid in the Propagation Vector $a(\theta)$ is sampled with a unitary sample period $k = 0, 1, \ldots, M-1$

$$a(\theta) = \left[1 \ \ e^{-j\omega_s} \ \ \ldots \ \ e^{-j(M-1)\omega_s}\right]^T \quad (2.11)$$

Thus, for the sampling theorem must satisfy

$$|\omega_s| \leq \pi \quad (2.12)$$

The anti-aliasing condition can be expressed as a function of the microphone spacing *d*

- Wavelength*:* $\lambda = \frac{c}{f_c} = \frac{c}{(\omega_c/2\pi)} = \frac{2\pi c}{\omega_c}$

- Spatial Frequency: $\omega_s = \omega_c \frac{dsin\theta}{c} = 2\pi \frac{dsin\theta}{\lambda}$

$$\left|2\pi \frac{dsin\theta}{\lambda}\right| \leq \pi \xrightarrow{yields} d|sin\theta| \leq \frac{\lambda}{2} \xrightarrow{yields} d \leq \frac{\lambda}{2} \qquad (2.13)$$

While the increase of the number of microphones increases the gain and improves the separation of sources, the distance between adjacent microphones can generate spatial aliasing and consequently the wave arrival direction can be ambiguous.
In addition, the anti-aliasing condition introduces an important weakness of the array.
Knowing that

$$\lambda = \frac{c}{f} \qquad (2.14)$$

And substituting (2.14) into (2.13)

$$d < \frac{c}{2f} \qquad (2.15)$$

the rearranged equation (2.15) is

$$\mathrm{f} < \frac{c}{2d} \qquad (2.16)$$

The equation (2.16) highlights a fundamental aspect of the application. It sets the upper limit of the frequency range of the incoming signals.

2.6 BEAMFORMING

Beamforming is a signal processing technique. The objective is to estimate the signal arriving from a desired direction in the presence of noise and interfering signals [5].

If the desired signal and the interferers occupy the same temporal frequency band, then the temporal filtering cannot be used to separate the signal from the interferers. However, the desired and the interfering signals generally originate from different spatial locations. This spatial separation can be exploited to separate the signals from the interference using an array of sensor in a particular configuration.
Typically, a beamformer linearly combines the spatially sampled waveform from each sensor in the same way a FIR filter linearly combines temporally sampled data.

Beamformers are classified as either *data independent* or *statistically optimum*, depending on how the weights are chosen. The weights in a data independent beamformer do not depend on the array data and are chosen to present a specified response for all signal and interference scenarios. On the other hand, the weights in a statistically optimum beamformer are chosen based on the statistics of the array data to optimize the array response.

2.7 DELAY AND SUM BEAMFORMING

One of the most common and simplest beamforming data independent algorithms is the Delay and Sum (DAS) algorithm. The main idea upon which the algorithm is based is that the signals of individual microphones can be

delayed by the calculated TDOA in order to obtain a constructive summation of the signals coming from the desired direction. The signals coming from other directions are out of phase and are therefore summed up destructively. Thus, the amplification of the signal from the desired direction, as well as attenuation of the signal from other directions, is achieved.

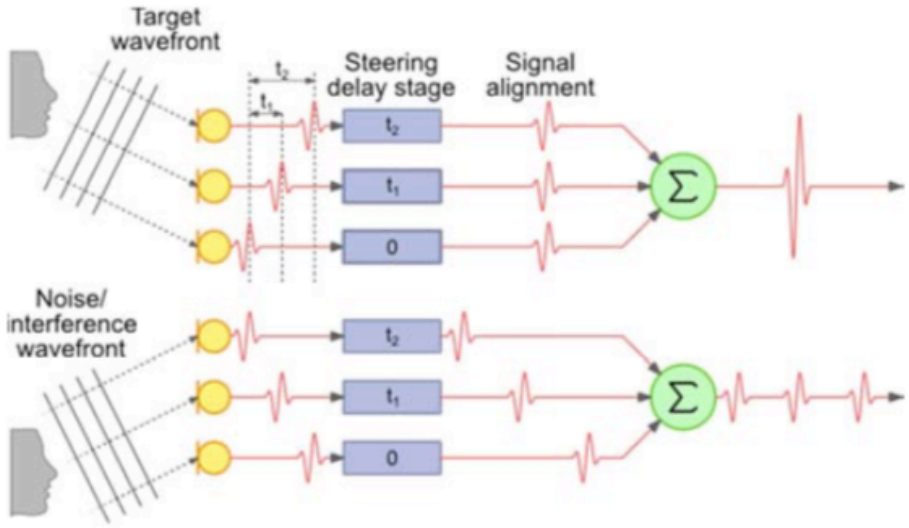

Figure 4 *DAS algorithm operating principle. Image taken from [5]*

The Delay and Sum beamforming algorithm can be performed in *time* and in *frequency domain*.

2.7.1 Time Domain DAS

The functionality of the delay and sum beamforming in the time domain can be resume as following: the sound source reaches the microphone on different paths. The signals captured by the microphones are similar in wave form but show different delays and phases. Both are proportional to the cover distances. The delays can be determined from the speed of sound, the distance between microphone and the sound source [29].
Then the delays of the signals can be calculated with respect to the reference microphone to ensure that the part of the desired signal has the

same phase on all channels. From the signal model [5], the delayed signal of the *i-th* microphone is:

$$y_i(t) = a_i \cdot s[t - \tau_i(\theta)] + e_i(t) = x_i(t) + e_i(t)c \quad (2.17)$$

The output of the beamformer is the summation of signals on all channels normalized respect to the number of the microphones in the array:

$$y(t) = \frac{1}{M}\sum_{1}^{M}[x_i(t) + e_i(t)] \quad (2.18)$$

In other words, the *DAS* algorithm in the time domain separates the delay into an integer multiple of the sampling period and a non-integer part. The integer part is obtained by delaying the signal with a certain amount of samples, whereas a *Finite Impulse Response (FIR)* filter is used to add the non-integer part of the delay. The algorithm cannot process negative delay values and therefore all the delays need to be additionally shifted so that the delay with the smallest, i.e. most negative value, becomes the reference value.

2.7.2 Frequency Domain DAS

The Delay and Sum beamformer in a frequency domain is based on a similar principle as in the time domain. The idea is that linearly combining the microphone signals results into enhance the signal coming from a desired direction and attenuate the signal coming from all other directions [4].

Using the weights

$$h = [h_1, h_2, \dots, h_M]^H \quad (2.19)$$

the beamformer output can be express as:

$$y_F(t) = \sum_{k=1}^{M} h_k \cdot y_k(t) = h^H y(t) \qquad (2.20)$$

Assuming that the filter h is band-pass, centered at θ in the spatial domain, the poer of the filtered signal should give a good indication of the energy coming from the direction θ. This power of the filtered signal can be calculated as:

$$E\{|y_F(t)|^2\} = h^H R h \qquad (2.21)$$

Where R is the *Covariance Matrix* of the array data

$$R = \{y(t)y^H(t)\} \qquad (2.22)$$

It means that the *DOA* of the incoming signal can be estimated evaluating the equation (21) known as *pseudo-spectrum* function.

Since the Covariance Matrix is only related to the incoming signal, the problem is reduced to design a spatial band-pass filter $h(\theta)$ such that it passes undistorted the signal with a target *DOA* $\overline{\theta}$ and it attenuates all the other *DOAs* different from $\overline{\theta}$ as much as possible.

The first condition is met when the spatial response at $\overline{\theta}$ is unitary, i.e. when

$$h^H(\overline{\theta})a(\overline{\theta}) = 1 \qquad (2.23)$$

Then assuming that the signal is spatially white, minimizing the power from all the direction but $\overline{\theta}$ means solving the following problem:

$$h(\overline{\theta}) = \arg\min h^H h \quad \text{subject to} \quad h^H a(\overline{\theta}) = 1 \qquad (2.24)$$

A suitable spatial filter which is a solution of the problem is:

$$h(\overline{\theta}) = \frac{a(\overline{\theta})}{a^H(\overline{\theta})a(\overline{\theta})} = \frac{a(\overline{\theta})}{M} \qquad (2.25)$$

Thus, the resulting pseudo-spectrum is:

$$p(\theta)E = \{|y_F(t)|^2\} = h^H(\overline{\theta})Rh(\overline{\theta}) = \frac{a^H(\overline{\theta})Ra(\overline{\theta})}{M^2} \qquad (2.25)$$

Practically, the resulting filter weights in $h(\overline{\theta})$ correspond to pure delays. The effect of the filter is that of re-phasing the microphone signals accordingly to the propagation delays there will be constructive interference for the direction of interest and destructive interference for all other directions.

2.8 MULTIPLE SIGNAL CLASSIFICATION (MUSIC)

The *Multiple Signal Classification* algorithm is a high-resolution beamforming technique. Based on an eigen subspace decomposition method, it divides the observation space into signal and noise subspaces by decomposing the array correlation matrix into its eigen-structure form [16][37][38].

Let's consider the complete data matrix *X* that contains the signals at each microphone in the case of multiple signal sources:

$$\vec{X} = \vec{A}\vec{S} + \vec{V} \qquad (2.26)$$

Where A is the steering vector, S is the complex impinging waveforms at each sensor and *V* is the noise at each element of the array.

Under the hypothesis that:

- The number of sources *N* is exactly known in advance, and they are smaller than the number of the microphone of the array:

$$N < M \qquad (2.27)$$

- The source signals $\vec{S}$ are such that their covariance matrix is non-singular:

$$rank(R_s) = N, \qquad R_s = E\{s(t)s^H(t)\} \qquad (2.28)$$

- The sensor noise is spatially white, with independent and identically distributed components having identical variance:

$$E\{e(t)e^H(t)\} = \sigma^2 I_M \qquad (2.29)$$

- Sensor noise is uncorrelated to source signals
- All the *DOSs* are different, thus they lead to different spatial frequencies

Under these assumptions, the input covariance matrix can be decomposed as the sum of two components:

$$R = AR_sA^H + \sigma^2 I_M = U\Lambda U^H \qquad (2.29)$$

Where U represent the unitary matrix and Λ is a diagonal matrix of real eigenvalues ordered in a descending order

$$\Lambda = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_M\} \qquad (2.30)$$

Any vector orthogonal to *A* is an eigenvector of R with value σ^2 and there are *M-N* such vectors. The remaining eigenvalues are larger than σ^2, which enables to separate two distinct eigenvectors-eigenvalues pairs, the signal pairs and the noise pairs.

The signal pairs are governed by the signal eigenvalues-eigenvectors pairs corresponding to the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_N \geq \sigma^2$, while the noise

pairs are governed by the noise eigenvalues-eigenvectors pairs corresponding to the eigenvalues $\lambda_{N+1} = \cdots = \lambda_M = \sigma^2$.
The covariance matrix can be express as:

$$R = U_s \Lambda_s {U_s}^H + U_n \Lambda_n {U_n}^H \quad (2.31)$$

Where U_s and U_n are the signal and the noise subspace unitary matrix.
The key issue in estimating the direction of arrival consists in observing that all the noise eigenvectors are orthogonal to *A,* the columns of U_s span the range space of *A* and the columns of U_n span the orthogonal complement of *A* which is the nullspace of A^H. By definition, the projection operators onto the noise and signal subspace are:

$$P_s = U_s {U_s}^H = A(A^H A)^{-1} A^H \quad (2.32)$$

And

$$P_n = U_n {U_n}^H = I - A(A^H A)^{-1} A^H \quad (2.33)$$

Since the signals are linearly independent, $AR_s A^H$ is a full rank and since the eigenvectors in ${U_n}^H$ are orthogonal to A, it leads to:

$${U_n}^H A = 0, \theta \in \{\theta_1, \dots \theta_N\} \quad (2.34)$$

In other words, the estimated signal covariance matrix will produce an estimated orthogonal projection onto noise subspace $P_n = U_n {U_n}^H$.

Finally, the *DOAs* can be retrieved from the *N* highest peaks of the *Music* spatial “pseudo-spectrum” function defined as:

$$P_{MUSIC}(\theta) = \frac{1}{a^H(\theta)P_n a(\theta)} \qquad (2.35)$$

Basically, the *music* algorithm estimates the distance between the signal and the noise subspaces in a direction where a signal is present and, since the two subspaces are orthogonal to each other, the distance between them at that every angle will be zero or near zero. Similarly, if no signal is present at a particular direction the subspaces are not orthogonal and the result will be zero.

2.9 MINI DSP UMA 16

Classical microphone arrays usually consist of condenser set of microphones mounted on a single device. But in order to collect signals from each microphone they also need an analog to digital convert ADC that may introduce latency which should create some problems in real-time application.

The choice of the Mini DSP UMA 16 is due to its flexibility and its high performance.

The UMA-16 is a sixteen channels microphone array with plug&play USB audio connectivity. Its system architecture consists of two core elements:

- The microphone array PCB which has 16 x SPH1 668LM4H MEMS Knowles laid out in a Uniform Rectangular Array.
- A nano-Sharc kit. A 400MHz HARC ADSP21489 + 500MHz multicore CPU

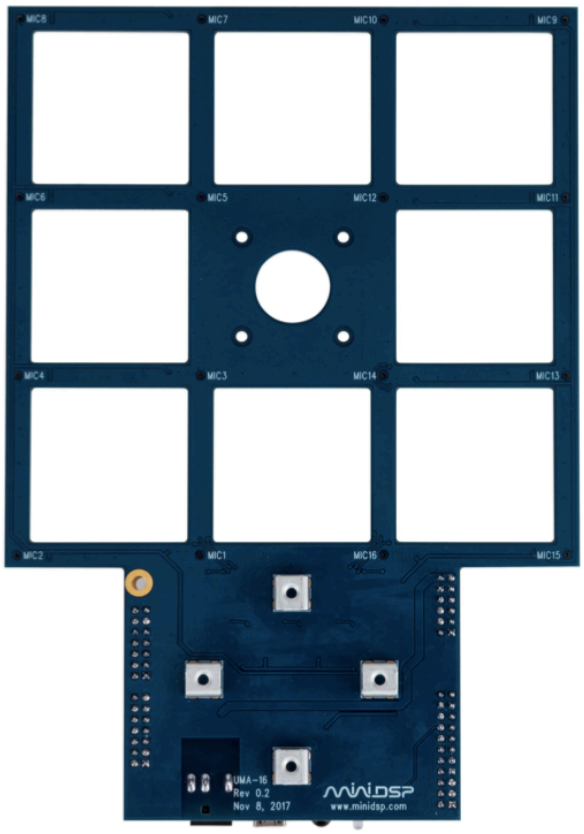

Figure 5 *Mini DSP UMA 16*

Due to its embedded ADC converter, DSP UMA-16 guarantees a high data transfer rate. In addition, the processing power allows for high quality PDM to PCM conversion and present all 16 channels of raw audio to the ASIO USB audio driver.

Moreover, the mini-UMA 16 is equipped with control panel which allows to set and control both the volume of each microphone and master volume i.e. all together.

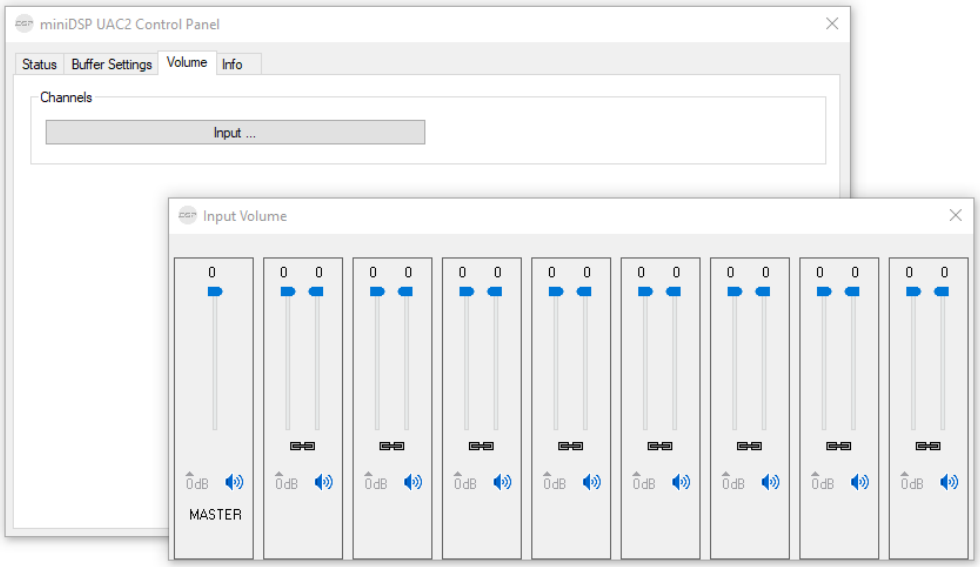

Figure 6 *Mini DSP Control Panel*

Furthermore, its MEMS omnidirectional microphones have a Low Distortion and High SNR with a flat response in the frequency range between 100-10000 KHZ.

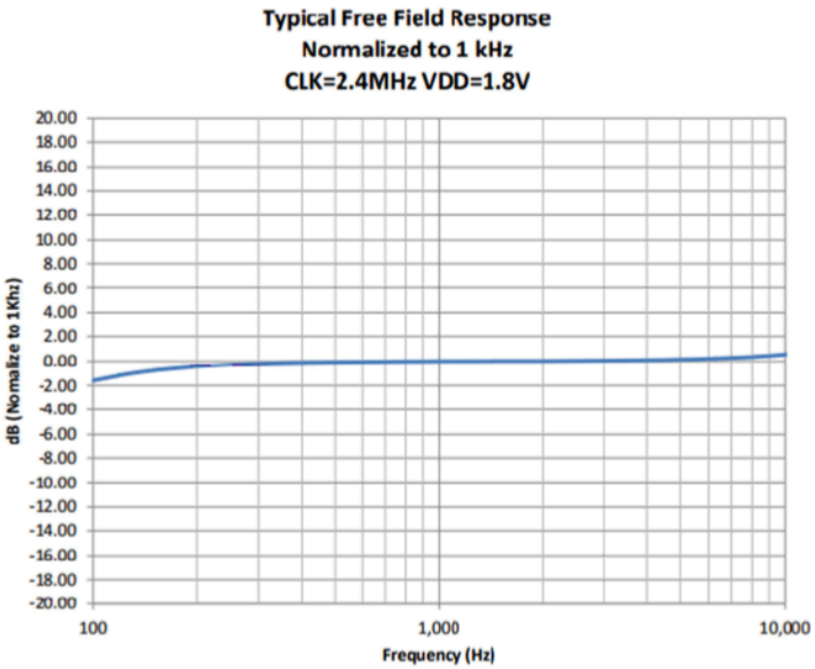

***Figure 7** UMA 16 Frequency Response*

Despite this limited frequency range, the UMA-16 is low-cost microphone array that perfect fit for development of beamforming algorithm and the center hole for USB camera in its center makes the device suitable for acoustic camera project as our case of study.

Item	Description
Digital Signal Processor	32-bit Floating point Analog Devices SHARC ADSP21489 / 400 MHz - Configuration locked
USB audio input	XMOS Xcore200 asynchronous USB audio up to 192 kHz, USB Audio Class 2 compliant • ASIO drivers for Windows • Driverless for Mac OS X
PDM inputs	Up to 16 x MEMS microphone connections (8 x stereo PDM data lines)
MEMS microphone	16 x SPH1668LM4H - Acoustic Overload @ 120dBSPL / High SNR of 65dB / RF shielded
ADC/DAC Sample rate & Resolution	Resolution: 24 bit Sample rate: 14.7k/11.025k/12k/16k/22.05k/44.1k/48k
USB port	USB port type Mini-B for audio streaming and firmware upgrade
Power supply	12 VDC single supply / Header input / 2.5W
Dimensions (H x W x D) mm	132 x 195 x 25 mm
Mounting	4 x M3 holders for front panel mounting / CAD drawings available on demand

***Figure 8** Mini DSP datasheet*

2.10 CMOS OV5640 USB CAMERA

The computation of an acoustic photo depends on the available hardware. The main parameter of a camera is the resolution which influences the quality of the snapshot. More precisely, the resolution determines the number of pixels along X and Y axis. However, the assumed frequency range of the investigating source has to be considered for the choice of the resolution as well. Depending on the frequency of incoming signal and depending on the algorithm, the source may not be found with low camera resolution, or its calculated level may be erroneous.
Another parameter that we have considered is the low dimension of the camera that has to fit the UMA 16.

The CMOS OV5640 is an USB camera module that enables image capturing along with the 720p, 1080p video streaming capability. Its 5MP hardware pixel together with its several resolutions allow to reconstruct the acoustic scene with high quality image.

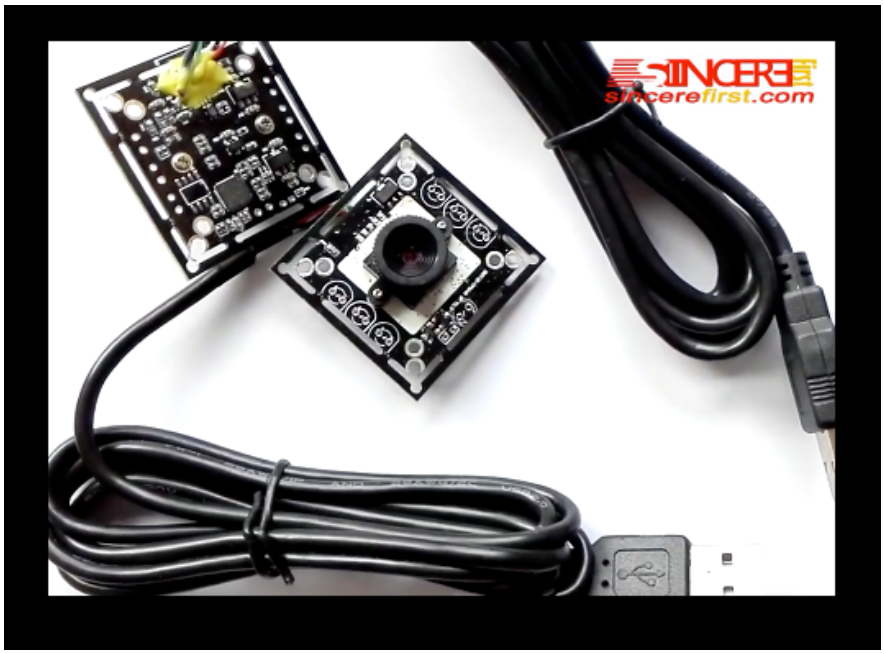

Figure 9 *CMOS OV5640 USB Camera*

Due to the fact CMOS OV5640 perfect fits the UMA 16, and due to its extremely low-cost respect to the quality image that it offers, the choice of the camera fallen it. The devices datasheet is showed in following figure.

Item	Technical Specifications	Remarks
Hardware pixels	500W（5MP）	Free drive
DSP model	SK9032	
Image sensor	OV5640	omnivision
Dynamic image resolution and transmission rate	2048*1536/15fps;1600*1200/15fps;1028*960/30fps;1024*768/30fps ;640*480/30fps;320*240/30fps	MJPG
	2592*1944/3fps;2048*1536/3fps;1600*1200/3fps;1280*960/7.5fps;1024*768/7.5fps;640*480/30fps;320*240/30fps	YUY2
Shooting image resolution	2592*1944;2048*1536;1600*1200;1280*960;1024*768;640*480;320*240	
Colors	24 true colors	
Shooting range	>20mm	
EFL	3.2mm FF	customizable
Lens optical structure	2P2G	
Viewing angle	70°\ 80°	customizable
F.NO	F2.8	
Distortion	<1.0%	
Relative contrast	0.65	
Brightness	automatic	
Contrast	automatic	
Hue	automatic	
Saturation	automatic	
Clarity	automatic	
Gamma	automatic	
White flat horizontal	automatic / manual	
Backlight contrast	automatic	
Exposure contro	automatic / manual	
Color space / compression	YUY2 / MJPG	
Output Pins	5Pin	1.0 mm
Operating voltage	5V	USB
Power consumption	120mW/50μW	
Module dimensions	38mm*38mm or 32*32mm	module height to prevail in kind
Operating temperature	-20°C～70°C	
System compatible	Win 2000、Win XP、Win 7 or Linux2.6.20or later	other systems need to install the driver

Figure 10 *CMOS OV5640 Datasheet*

3 SYSTEM DESIGN

In order to design an acoustic camera, for the sake of convenience, we have implemented the beamforming algorithm in Matlab. It is a high-performance language for technical computing that integrates computation, visualization, and an easy-to-use environment where problems and solutions are expressed in familiar mathematical notation.

3.1 ARRAY MODEL

As shows in chapter 2.4 the operating principle of the beamforming algorithms is based on the delay between microphones respect to reference one. Matlab provides a tool called Phased Array System toolbox very useful in designing and modeling sensor array. The idea is to assign an identification number to each sensor so that the system knows which is the reference one.

Starting from the position of each sensor into the array, we have arranged the microphone array on the X-Y plane and associated a number from 1 to 16 to each microphone.

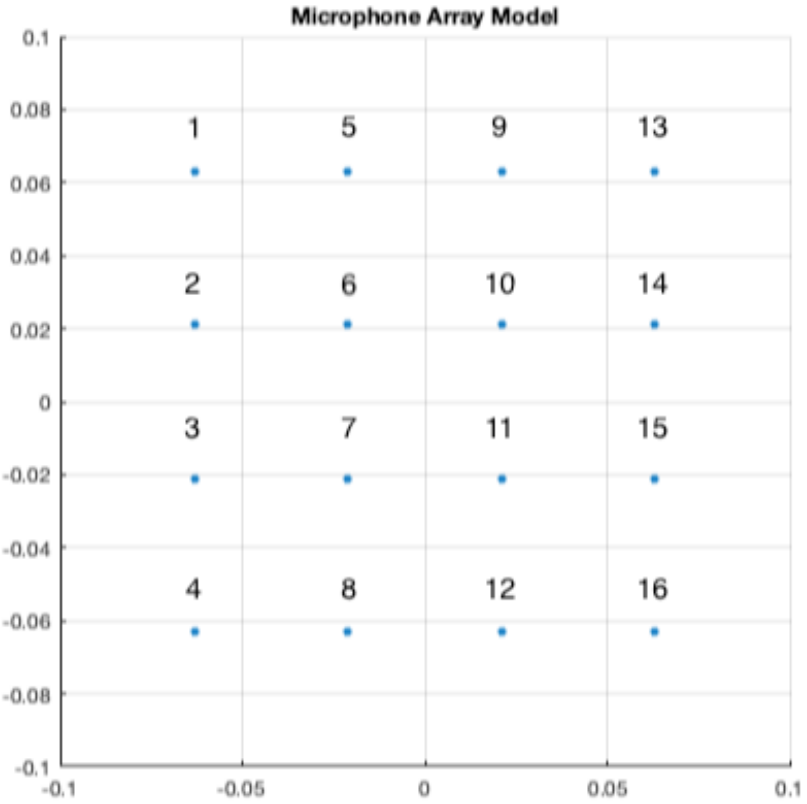

Figure 11 *Microphone positions*

As noticed in chapter 2.3, its relatively small distance between microphone of 0.042 cm, sets the upper limit of the of the incoming signals around 4000 Hz.

3.2 Array Impulse Response

The detection and the localization of a sound source is strictly connected to the environment in which it is located. Generally, a signal in an enclosure space is composed by two components: a direct sound which directly reaches the listener and the reflected sound which is generated by all the room surfaces. As well as our hearing system interprets time, frequency and spatial information from arriving room reflections, the signal received by the microphone array contains not only information about the sound source, but it also contains information about the enclosure space. It means that the source signal contained on the received signal can be affected by

the reflected sound i.e., by the room itself [14]. Since the beamforming techniques are based on the deconvolution of the captured signals, for the beamforming device to be fully functional, a fundamental aspect is the fact the all the sensor of the microphone array must have the same impulse response. But since received signal captured by each microphone is affected from the environment and, even if the microphone positions are known accurately, there may be some phase differences between microphones. This will give the same effect as microphone position errors [19]. The figure 12 shows the impulse response captured by the sixteen channels of the microphone array in a room used for the experiments.

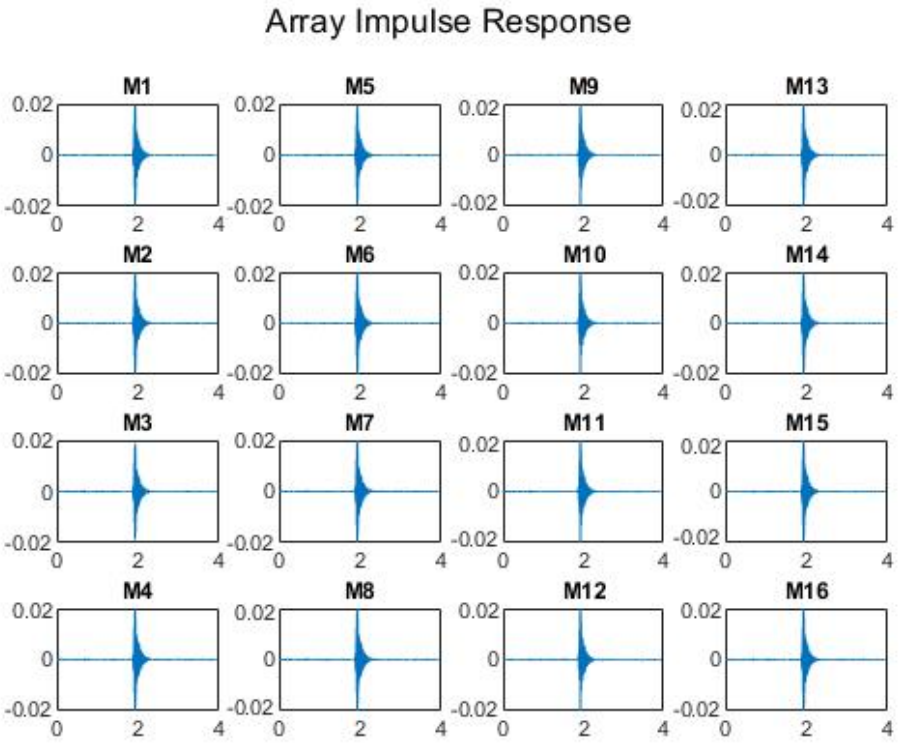

Figure 12 Array Impulse Response

In this measure the speaker is in front of the mini-DSP at 1.5 meters distance. As shown, the microphones present the same amplitude and phase, and this should ensure an identical behavior between them.

3.3 Array Beam Pattern

Another aspect that influences the performance of the beamformer is the corresponding beampattern, which provides a complete characterization of the array system's input-output behavior when the beamformer is steered to a specific direction. It can be used to analyze how the array output is affected by signals different from the focused one [30].

Using the phase array system toolbox in Matlab, we have simulated the beam pattern for our array at different frequencies.

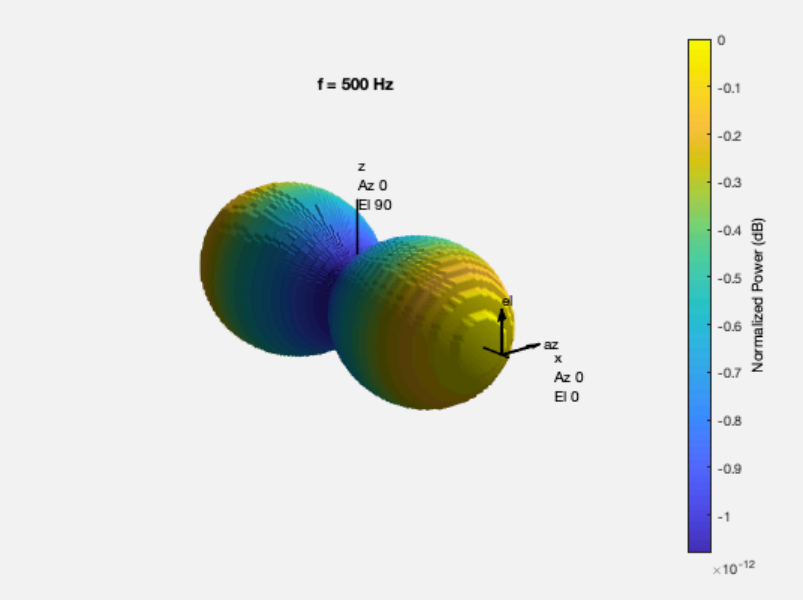

Figure 13 *Array Beam Patter at f=500 Hz*

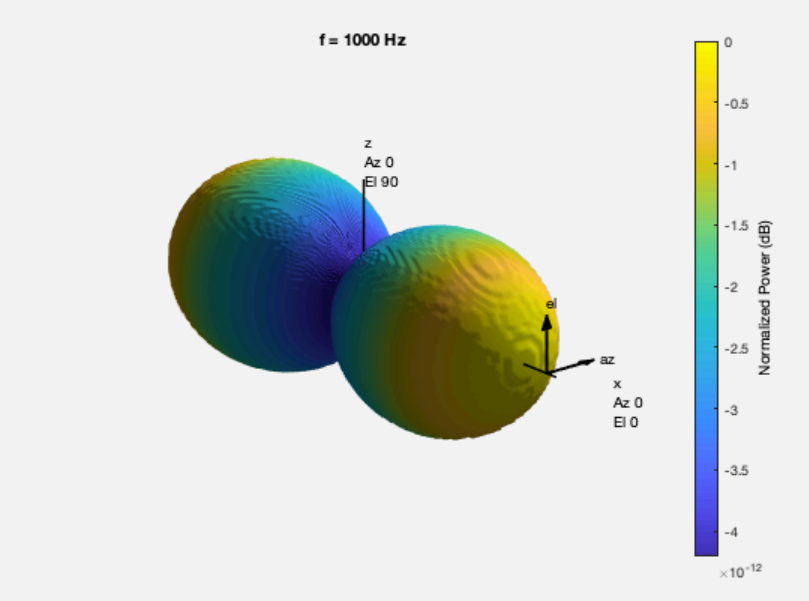

Figure 14 *Array Beam Pattern at f=1000 HZ*

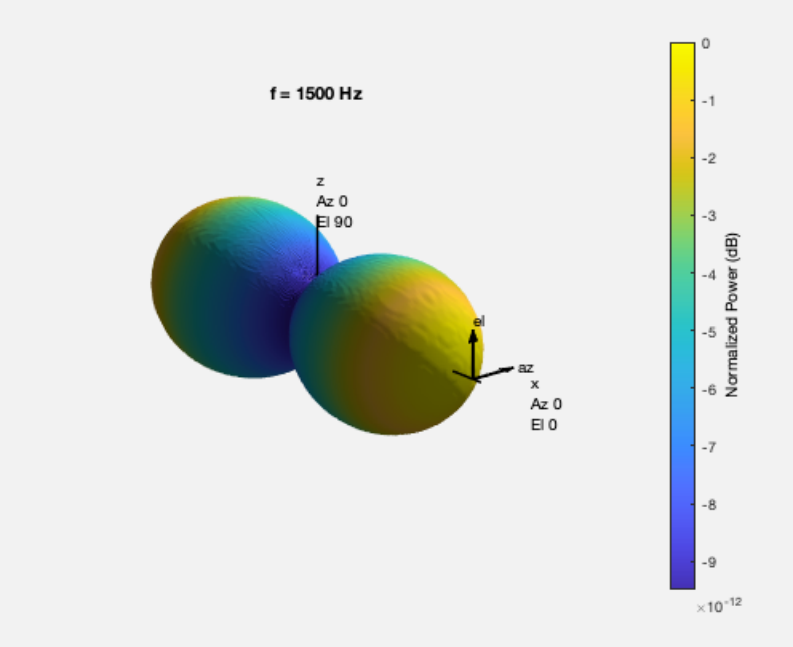

Figure 15 *Array Beam Pattern at f=1500 HZ*

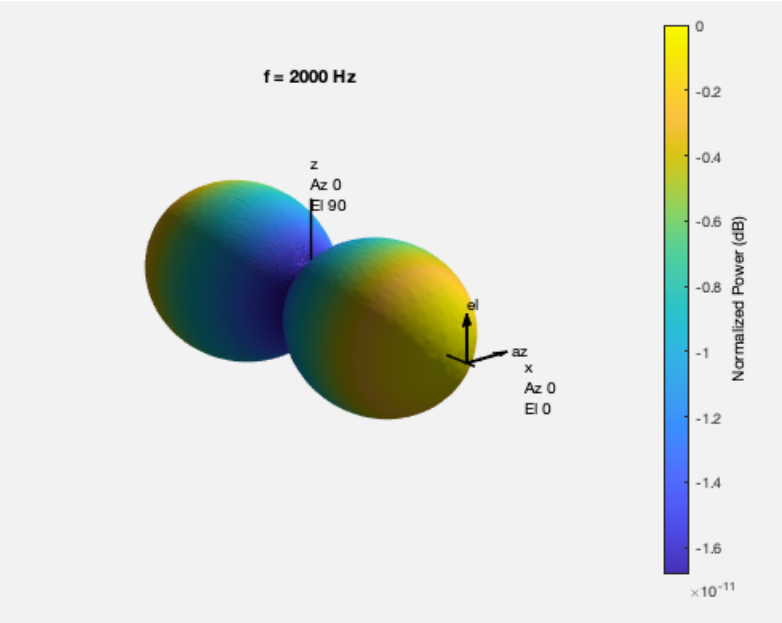

Figure 16 *Array Beam Pattern at f=2000 HZ*

As show in the figures 13-14-15-16, the microphone array has the same behavior at different frequencies because it has identical omnidirectional microphones. But due to its omnidirectionality, putting the array in a X-Y axis, it presents the same beam pattern from signals coming from both the Z and -Z direction. As a result, the system can capture signals coming from the back which theoretically should be a reflected signal. But due to high stability of the beamforming algorithm the system is not so affect by the reflected signal coming from the back when the incoming signal is in front of the device. This will be deeply analyzed in chapter 4 about the test and the experimentation results.

3.4 BEAMFORMING IMPLEMENTATION

In order to create an acoustic image, the acquired signals must be processed through beamforming algorithm. It was firstly developed in a static setup in which the incoming signal was acquired and the collected data were stored into an array. The idea was to verify and test the robustness of the algorithm that was implemented in real-time mode at later time.

3.4.1 Static acquisition

In this section different beamforming algorithms were implemented. In particular, we have implemented both DAS and MUSIC algorithms. The aim was to explore the two algorithms and analyze characteristics as computational time and resolution which are fundamental in the real time application.

The static acquisition process can be resumed into three steps:

- Acquiring data and collect it into matrix
- Processing data and create the pseudospectrum
- Generating the power map

The data was acquired using Matlab Audio Toolbox. It includes a function which allows to capture audio directly from Mini DSP UMA 16 and store all sixteen audio signals coming from the microphone array into an audio file.

Once the data was collected into a matrix, the main process was done by the beamforming algorithm. We have firstly implemented both code algorithms step by step as described in chapter 2.6 and 2.7. Regarding MUSIC method, after some tests we have preferred to employ a function called phased.MUSICEstimator2D of the phase array system toolbox since its computational time is less than own MUSIC algorithms code.

3.4.2 Real-time Process

The core of the project is the implementation of the acoustic camera in real time mode. Different from the static setup, the real time application is based on the continuous streaming of input data without record them. Since the UMA 16 has an integrated DSP the latency of the input signal is almost negligible. But, before locating the sound source and creating the relative acoustic map, the algorithm needs a certain amount of time that can generate a delay. In order to overcome this type of problem, we have decided to implement the MUSIC algorithm which, in the static setup, was resulted the faster implemented algorithm in term of computational time together with a better accuracy.

Obviously, MUSIC does not guarantee a null time delay because it has to analyze sixteen audio channels coming from the microphone array. Since the length of each channel in term of samples plays a fundamental role, the idea was to fill a buffer of a minimum length that guarantees a correct localization.

Another important aspect that we have neglected in the static acquisition is the mapping of the acoustic scene with the usb camera. In doing so, we have created an application with the Matlab toolbox App design which allows to overlay independently the power map created by the algorithm and the image steamed from the camera.

Particular attention was paid to the Angle of View (AoV) of the webcam in the system. This because it specifies steering angle of beamforming. Denoting the AoV by α, it was calculated using a relation of triangle as shown in figure 17.

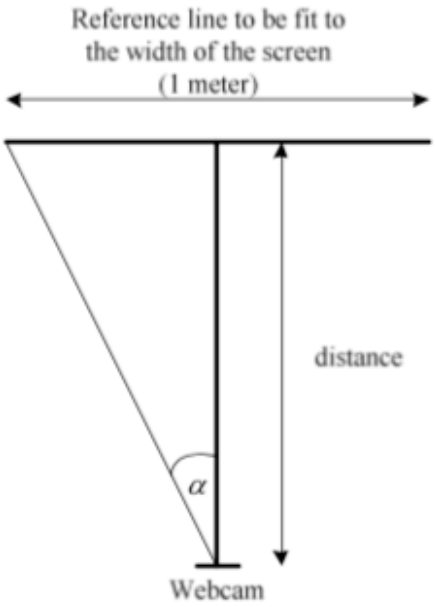

Figure 17 *Measurement of AoV. Image taken from [7]*

In the measurement, a horizontal reference line with 1 meter in length was drown then the position of webcam was adjusted until the line fits to the width of the display. The distance was measured then the AoV can be calculated. AoV must be measured both in vertical and horizontal arrangement. In our case, it was found that the horizontal (Azimuth) and vertical (Elevation) are 80 and 70 degrees respectively.

Once defined all components, we have implemented the application of the acoustic camera. This because the idea was to have a simple graphic interface that looks like a real instrument. In addition, the app simplifies the management of the frequency, which is the main research parameter.

In order to overlay independently the power map created by the algorithm and the image steamed from the camera, we have implemented an application through the App Designer tool. The app allows to create a graphic user interface (GUI) and add all components we need [10]. It is composed by few components which are the axis in which the beamforming algorithm and the webcam plot the images, a control switch which turns on or off the acquisition, a lamp that shows the state of the app (On or Off), and a slider that sets the research frequency of the beamformer, while the

code view of the app design allows to assign a callback function to the components defined in the design view. The result is shown in figure 18.

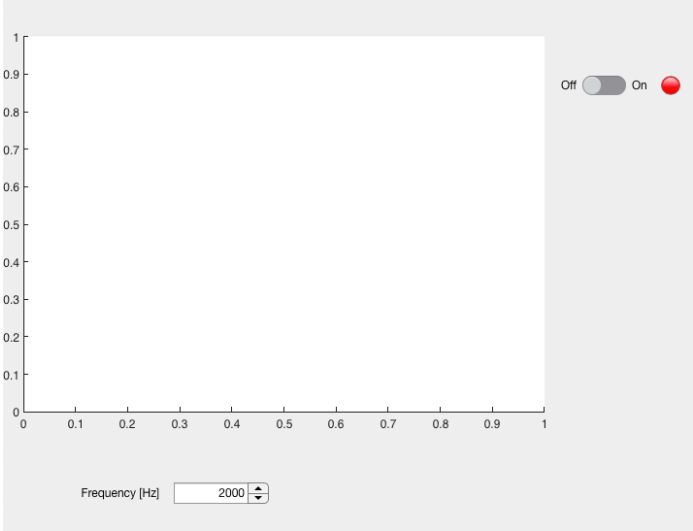

Figure 18 *Acoustic Camera App*

Practically, the audio signals coming from the microphone array fill the audio buffer while the video frame streamed by the cam is sent directly to the app. At the same time Music algorithm takes the audio frame from the buffer, analyzes it and generates a power map which is overlayed to the video frame. The key process is shown in figure 19.

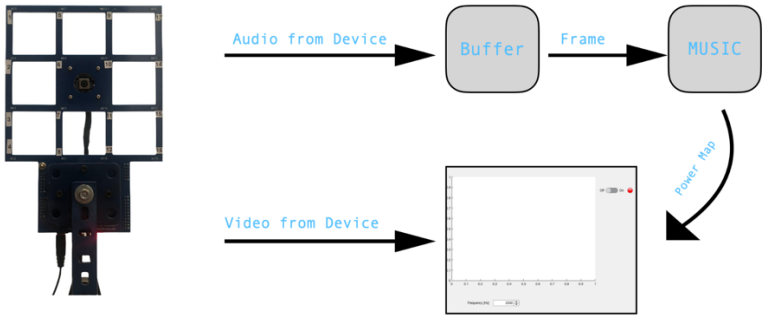

Figure 19 *Designed Flochart*

4 TEST AND RESULTS

In this section we present some experimental results regarding the implemented acoustic camera with the proposed setup. Firstly, we have performed several tests regarding the two implemented algorithm in the static setup. Afterwards, we have moved the device in trhee different room in order to simulate a real time application.

4.1 STATIC SETUP TEST

In this section, different incoming signals have been used with different duration in time. The goal was to find the minimal length of analyzed signals which guarantee the localization of source.
Then we have proceeded by changing the distance between the sound source and the device and changing the position of the sound source.
The performance of the acoustic camera was measured in term of accuracy, resolution and computational time. The accuracy represents how the estimated DoA is close to the real one and it was measured in both Azimuth and Elevation angles. It was also measured with regard the root mean square error (RMSE) between the true source position and the estimated one where the former was measured putting the speaker on a rotate system with reference angle.
All the experiment was carried out in a reverberant environment with T60=1.2 seconds and the test tones were generated from the 5'' Hypnocampus 2C speaker. We will refer to the direction of arrival in both Azimuth and Elevation as [Az El] in degrees [°].
Staring from the static acquisition, a first step was to evaluate the Delay and Sum algorithm using two tone signals of different duration and coming from several directions.
The signals captured by the microphones have a Signal to Noise Ratio between 30 and 54 dB. The results are shown in table 1.

SIGNAL	DOA [AZ EL]	ACCURACY [AZ EL]	RMSE [°]	RESOLUTION	DISTANCE [M]	DURATION [S]
1 KHZ	[-30 0]	[8 20]	3.7	361x361	1.5	4
1 KHZ	[-30 0]	[8 20]	3.7	361x361	1.5	3
1 KHZ	[-30 0]	[8 20]	3.7	361x361	1.5	1
1 KHZ	[-30 0]	[15 25]	4.4	361x361	2.1	1
1 KHZ	[-30 0]	[15 25]	4.4	361x361	2.1	4
1 KHZ	[-30 0]	[8 20]	3.7	901x901	1.5	4
1 KHZ	[30 0]	[0 25]	3.5	361x361	1.5	4
1 KHZ	[30 0]	[0 25]	3.5	361x361	1.5	1
1 KHZ	[30 0]	[12 22]	4.1	901x901	2.1	4
1 KHZ	[30 0]	[12 22]	4.1	901x901	2.1	1
1 KHZ	[0 -30]	[13 45]	5.3	901x901	1.5	1
1 KHZ	[0 -30]	[13 45]	5.3	901x901	1.5	4
1 KHZ	[0 -30]	[20 50]	5.9	901x901	2.1	1
1 KHZ	[0 -30]	[13 45]	5.9	361x361	2.1	1
1 KHZ	[30 30]	[37 45]	6.4	361x361	2.1	1
1 KHZ	[30 30]	[32 40]	6	361x361	1.5	1
1 KHZ	[30 30]	[37 45]	6.4	361x361	2.1	4
1 KHZ	[30 30]	[37 45]	6.4	901x901	2.1	1
2 KHZ	[0 45]	[18 9]	3.6	361x361	1.5	1
2 KHZ	[0 45]	[20 15]	4.1	361x361	2.1	1
2 KHZ	[0 45]	[18 9]	3.6	361x361	1.5	4
2 KHZ	[0 45]	[18 9]	3.6	901x901	1.5	4
2 KHZ	[30 0]	[0.8 7]	1.9	901x901	1.5	4
2 KHZ	[30 0]	[0.8 7]	1.9	901x901	1.5	1
2 KHZ	[30 0]	[0.8 7]	1.9	901x901	1.5	4
2 KHZ	[30 0]	[2 15]	2.9	901x901	2.1	4
2 KHZ	[30 0]	[0.8 7]	1.9	361x361	1.5	4
2 KHZ	[-30 0]	[2 16]	3	361x361	1.5	1
2 KHZ	[-30 0]	[2 16]	3	901x901	1.5	4
2 KHZ	[-30 0]	[2 21]	3.3	901x901	2.1	4
2 KHZ	[0 -30]	[31 55]	6.5	361x361	1.5	1
2 KHZ	[0 -30]	[31 55]	6.5	361x361	1.5	4
2 KHZ	[0 -30]	[35 60]	6.8	361x361	2.1	1
2 KHZ	[0 -30]	[35 60]	6.8	901x901	2.1	1
2 KHZ	[60 30]	[1 17]	3	901x901	1.5	1
2 KHZ	[60 30]	[1 17]	3	901x901	1.5	4
2 KHZ	[60 30]	[5 21]	3.6	901x901	2.1	1
2 KHZ	[60 30]	[5 21]	3.6	901x901	2.1	4
2 KHZ	[60 30]	[1 17]	3	361x361	1.5	1

Table 1 *Table of DaS results*

A first test was performed using 1 KHz tone by changing its duration. The figures 20-21 shows the response of the algorithm to incoming signals of duration 1 second and 4 seconds respectively.

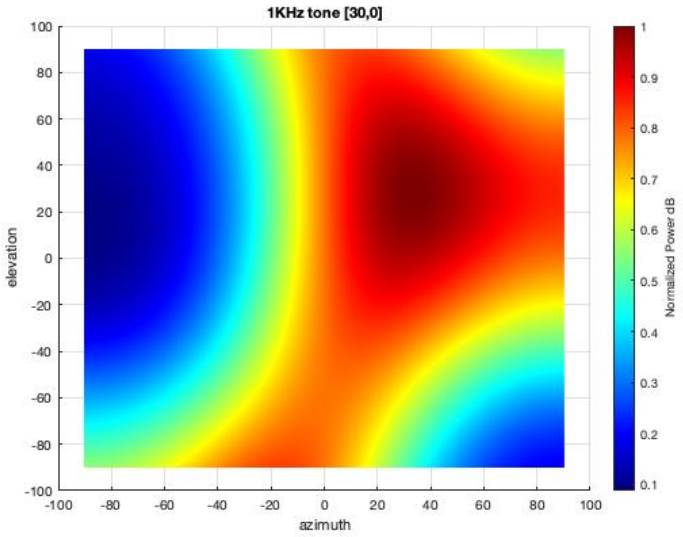

Figure 20 *Power map of 1 KHz tone coming from [30 0] of 1s*

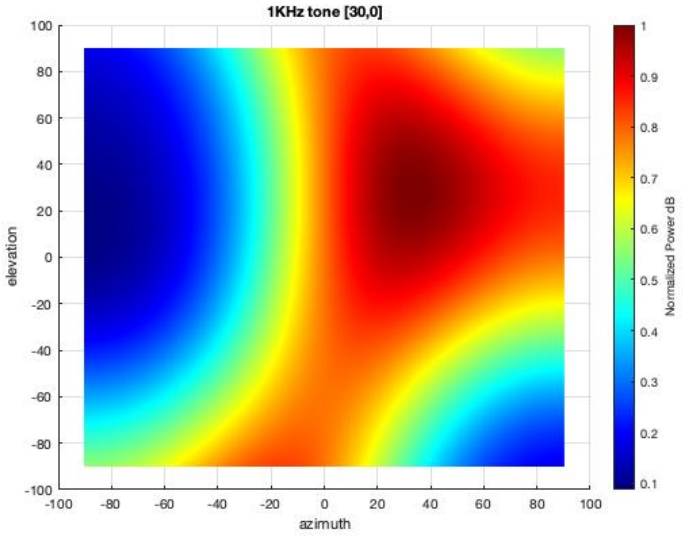

Figure 21 *Power map of 1 KHz tone coming from [30 0] of 4s*

Since the generated power maps present the same accuracy, we have noticed that the DaS is not affect by the length of the signal to be analyzed. The same happens regarding a 2 KHz tone signal.

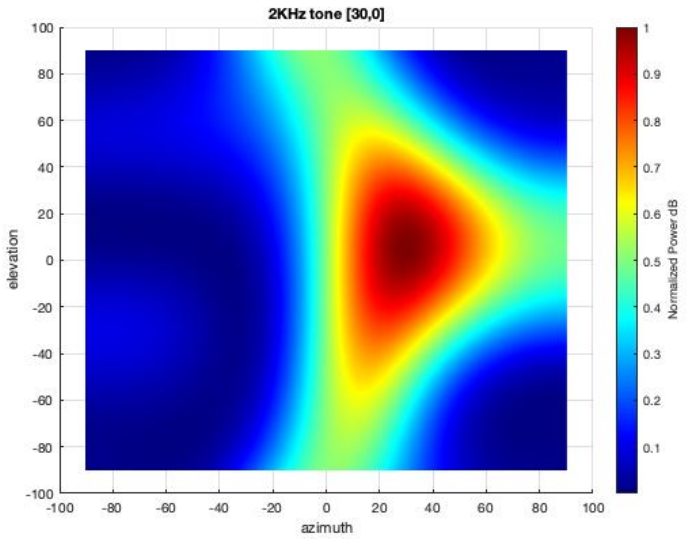

Figure 22 *Power map of 2 KHz tone coming from [30 0] of 1s*

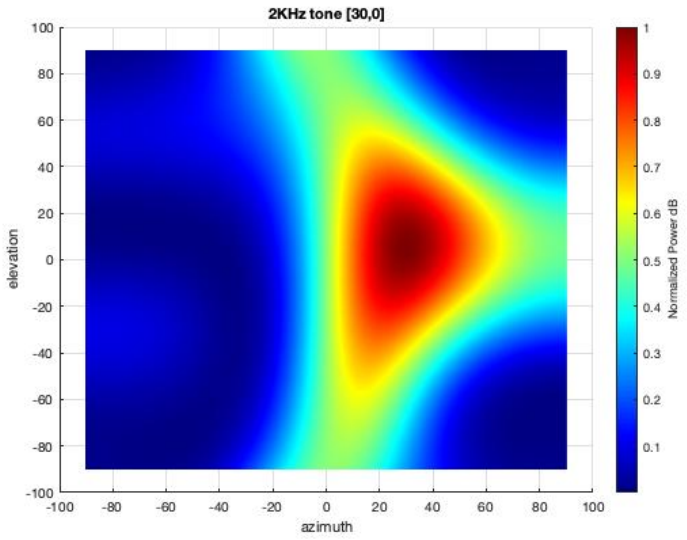

Figure 23 *Power map of 2 KHz tone coming from [30 0] of 4s*

Although the duration of the signal is not relevant, we have noticed that when the number of samples of the incoming signal increases the system takes long to plot the power map. However, the length of the signal depends not only on the duration of acquisition but also it depends on the sampling frequency of the system F_s. The test tones present in figure 20 were recorded at different sampling frequencies starting from 44.1 KHz down to 11.205 KHz. As noticed in chapter 3.1 the distance between the

microphones into the array sets the upper frequency range about 4000 Hz. Since all the selected F_s guarantee the anti-aliasing condition ($F_s>2F_{max}$), we have decided to set the sampling frequency at 11025 Hz in order to process the minimum number of samples.

Afterwards we have proceeded investigating the distance between the microphone array and the sound source. The figures 24-25 show the power maps relative to a 2 KHz tone signal coming from [-30 0] at distance from the array equal to 1.5 and 2.1 meters respectively.

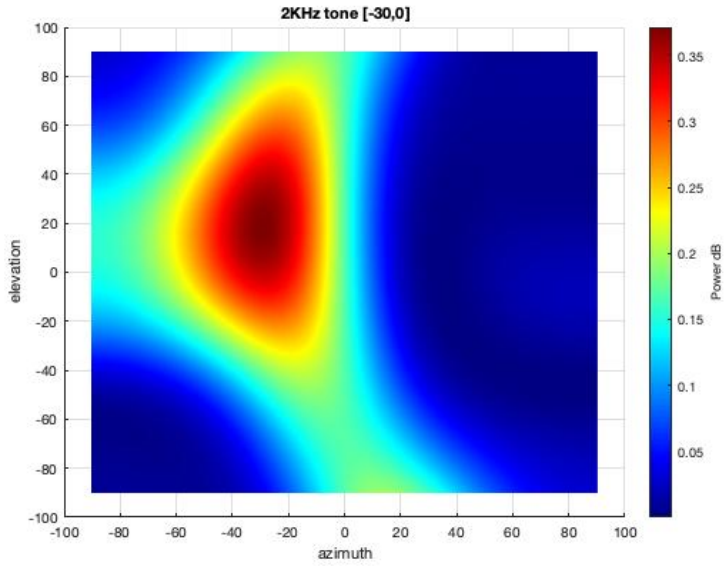

Figure 24 *Power map of 2 KHz tone coming from [-30 0] located at 1.5 m*

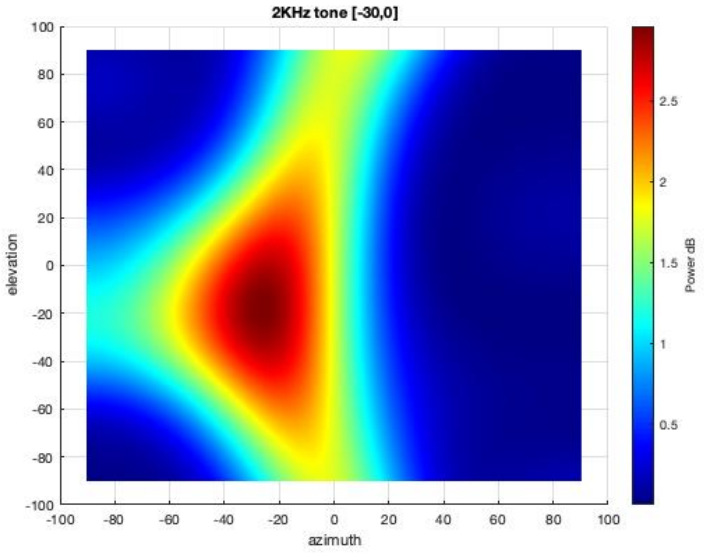

Figure 25 *Power map of 2 KHz tone coming from [-30 0] located at 2.1m*

Comparing the two signals, we have noticed that they are symmetric respect of the X axis, but they also present a similar RMSE which are equal to 3 for the signal positioned at distance 1.5 meter and 3.6 for the signal at 2.1 meters. It means that they present almost the same localization error. A considerable difference is in the Sound Pressure Level of the two signals which are 0.35 dB for the former and 2.8 dB for the latter. In order to relate the distance between the microphone array and sound source we have tested the DaS algorithm considering signals with different SPL. The result was that the algorithm can localize sound sources with maximum distance between 2 and 3 meters with a power between 0.5 and 0.2 dB at least.

The quality of the image power map depends on the resolution. It is related to the number of the scanning point of the steering vector (equation 2.9). The figures 26-27 show the power map related to a 1 Khz tone signal coming from [-30 0] computed with a scan angle between – 90 and 90 for both Azimuth and Elevation with a resolution of 0.2 and 0.5 respectively. As shown, the resolution does not affect the accuracy. The higher the resolution corresponds to the higher quality power map but, on the other hand the price to pay is related to the computational time in terms number of point that the algorithm must draw which are 361x361 for a resolution of 0.5 and 901x901 for a resolution of 0.2.

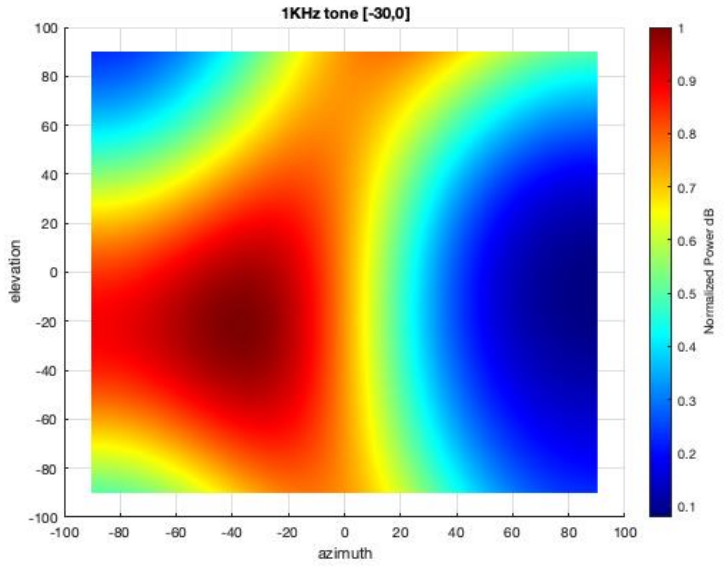

Figure 26 *Power map of 1 KHz tone coming from [-30 0] of 0.2 resolution*

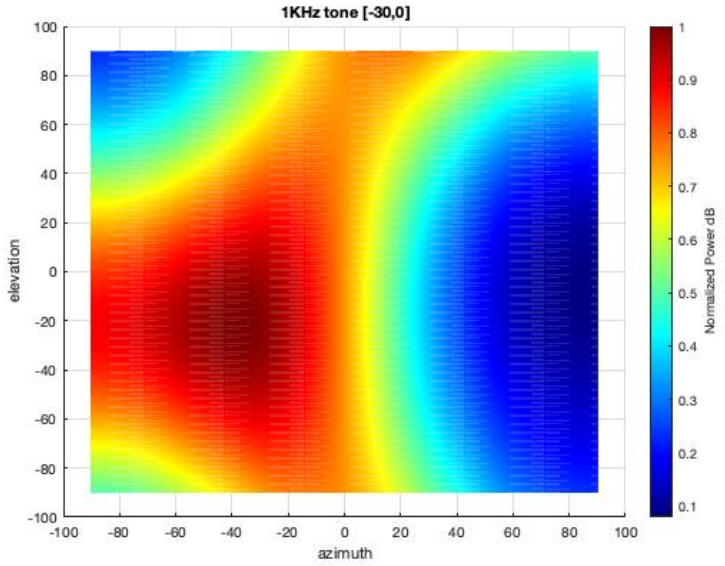

Figure 27 *Power map of 1 KHz tone coming from [-30 0] of 0.5 resolution*

Analyzing the obtained results, we have noticed that the accuracy of the DaS algorithm is a very frequency dependent. Although a discrete localization of the sound source, the performance of the method is influenced by the reflected sound that reaches the device. As a result, the corresponding power map also shows an intensity sound scene that is not related to the source and it might suggest the presence of another sound source.

Under the same condition and using the same test signals we have evaluated the performance of the MUSIC algorithm. The results are shown in table 2.

SIGNAL	DOA [AZ EL]	ACCURACY [AZ EL]	RMSE [°]	RESOLUTION	DISTANCE [M]	DURATION [S]
1 KHZ	[-30 0]	[10 40]	5	361x361	1.5	4
1 KHZ	[-30 0]	[10 40]	5	361x361	1.5	3
1 KHZ	[-30 0]	[10 40]	5	361x361	1.5	1
1 KHZ	[-30 0]	[15 42]	5.3	361x361	2.1	1
1 KHZ	[-30 0]	[15 42]	5.3	361x361	2.1	4
1 KHZ	[-30 0]	[15 42]	5.3	901x901	1.5	4
1 KHZ	[30 0]	[0 15]	2.6	361x361	1.5	4
1 KHZ	[30 0]	[0 15]	2.6	361x361	1.5	1
1 KHZ	[30 0]	[5 17]	3.3	901x901	2.1	4
1 KHZ	[30 0]	[12 22]	4.1	901x901	2.1	1
1 KHZ	[30 0]	[5 17]	3.3	901x901	2.1	1
1 KHZ	[0 -30]	[0 7]	1.8	901x901	1.5	4
1 KHZ	[0 -30]	[2 10]	2.4	901x901	2.1	1
1 KHZ	[0 -30]	[2 10]	2.4	901x901	2.1	1
1 KHZ	[30 30]	[0 15]	2.7	361x361	2.1	1
1 KHZ	[30 30]	[0 12]	2.4	361x361	1.5	1
1 KHZ	[30 30]	[0 15]	2.7	361x361	2.1	4
1 KHZ	[30 30]	[0 15]	2.7	361x361	2.1	1
2 KHZ	[0 45]	[0 15]	3.8	361x361	1.5	1
2 KHZ	[0 45]	[2 18]	3.1	361x361	2.1	1
2 KHZ	[0 45]	[18 9]	3.6	361x361	1.5	4
2 KHZ	[0 45]	[18 9]	3.6	901x901	1.5	4
2 KHZ	[30 0]	[0 3]	1.2	901x901	1.5	4
2 KHZ	[30 0]	[0 3]	1.2	901x901	1.5	1
2 KHZ	[30 0]	[0 3]	1.2	367x361	1.5	4
2 KHZ	[30 0]	[2 5]	1.8	901x901	2.1	4
2 KHZ	[30 0]	[0 3]	1.2	901x901	1.5	4
2 KHZ	[-30 0]	[7 9]	3	361x361	1.5	1
2 KHZ	[-30 0]	[7 9]	3	901x901	1.5	4
2 KHZ	[-30 0]	[10 12]	3.3	901x901	2.1	4
2 KHZ	[0 -30]	[0 5]	1.5	361x361	1.5	1
2 KHZ	[0 -30]	[0 5]	1.5	361x361	1.5	4
2 KHZ	[0 -30]	[0 5]	1.5	361x361	2.1	1
2 KHZ	[0 -30]	[0 5]	1.5	901x901	2.1	1
2 KHZ	[60 30]	[6 18]	3.4	901x901	1.5	1
2 KHZ	[60 30]	[6 18]	3.4	901x901	1.5	4
2 KHZ	[60 30]	[8 21]	3.8	901x901	2.1	1
2 KHZ	[60 30]	[8 21]	3.8	901x901	2.1	4
2 KHZ	[60 30]	[6 18]	3.4	361x361	1.5	1

Table 2 *Table of MUSIC results*

Similar to DaS algorithm, the accuracy of the algorithm is not affected by the length of the incoming signal as shown in figure 29-30 regarding a pure a 2 KHz tone coming from [30 0] of length of 1 second and 4 seconds respectively.

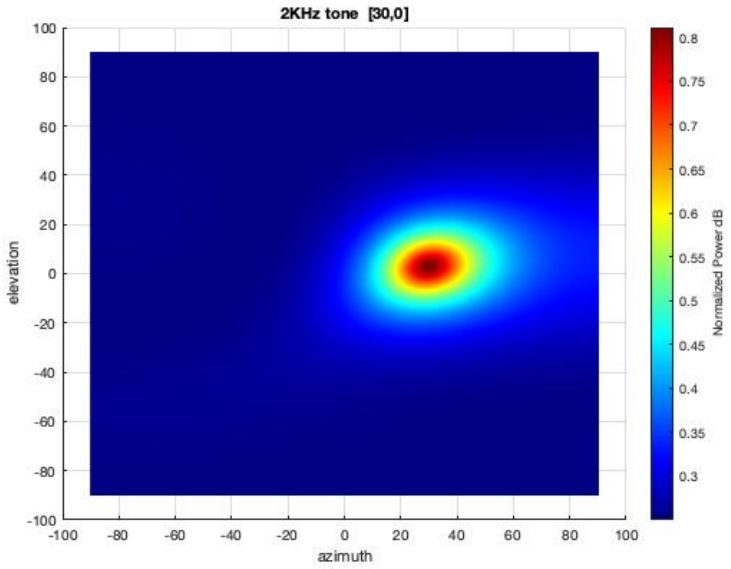

Figure 28 *Power map of 2 KHz tone coming from [30 0] of 1s*

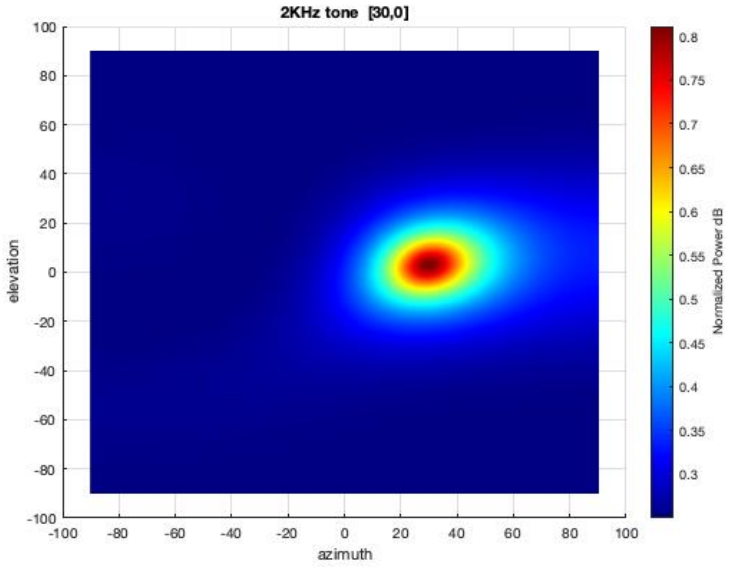

Figure 29 *Power map of 2 KHz tone coming from [30 0] of 4s*

In order to ensure the dependance of the accuracy on the sound source SPL rather than its position respect to the microphone array, we have performed a test using a 2 KHz tone. We have put the sound source at 1.5 meter from the device, we have measured the SPL and we evaluated the MUSIC response. Then we have moved back the sound source at 2.1 meters and we have increase the sound pressure level util it was equal to the SPL measured at 1.5 meters. The result is show in the figure 30-31.

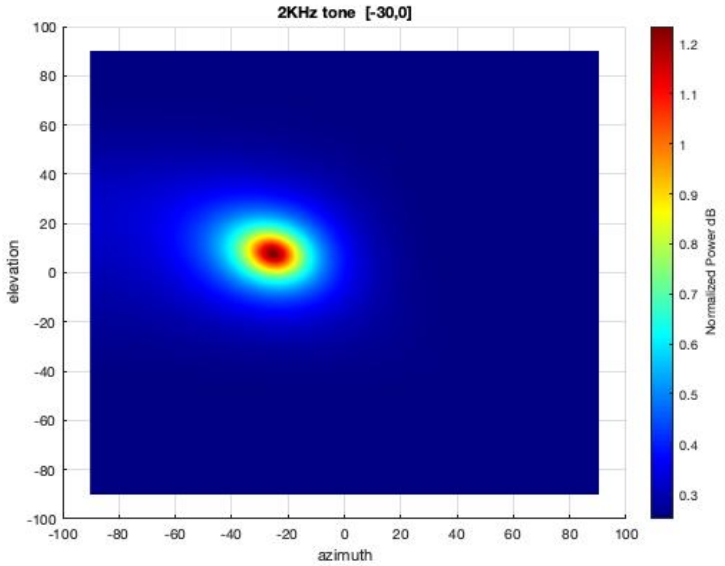

Figure 30 *Power map of 2 KHz tone coming from [-30 0] located at 1.5 m*

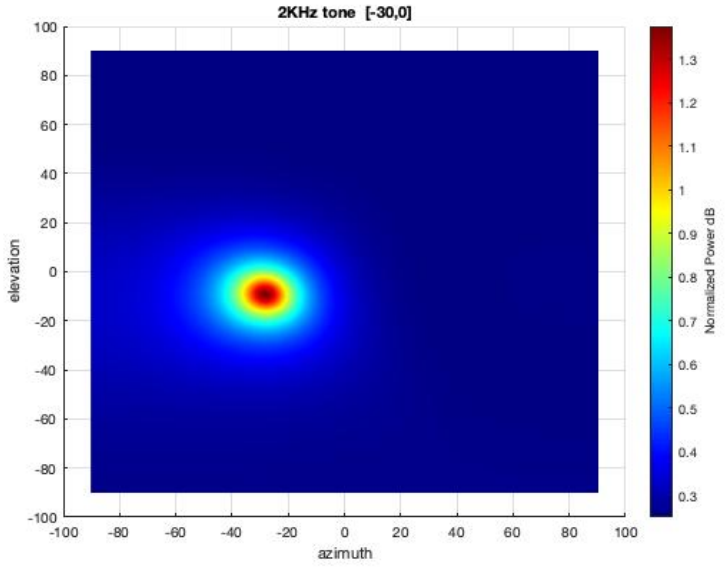

Figure 31 *Power map of 2 KHz tone coming from [-30 0] located at 2.1 m*

The power maps present the same result for signals coming from both directions. In additions we have accomplished a test in presence of two sound sources located at the same distance from the device which generated two signals with equal SPL. In this condition, the system does not locate both sources, but it generates a power map that focuses on the higher energy space. As expected, as the SPL of one source increases the energy map follows the signal generated by that source.

Furthermore, like DaS the resolution does not affect the accuracy. Although the number of drawn point (361x361 for a resolution of 0.5 or 961x961 for 0.2 of resolution both regarding a scan angle between -90 and 90 degrees in Azimuth and Elevation direction) is the same for MUSIC and DaS spectrum, Music performs faster than Das due to the nature of the two algorithms.

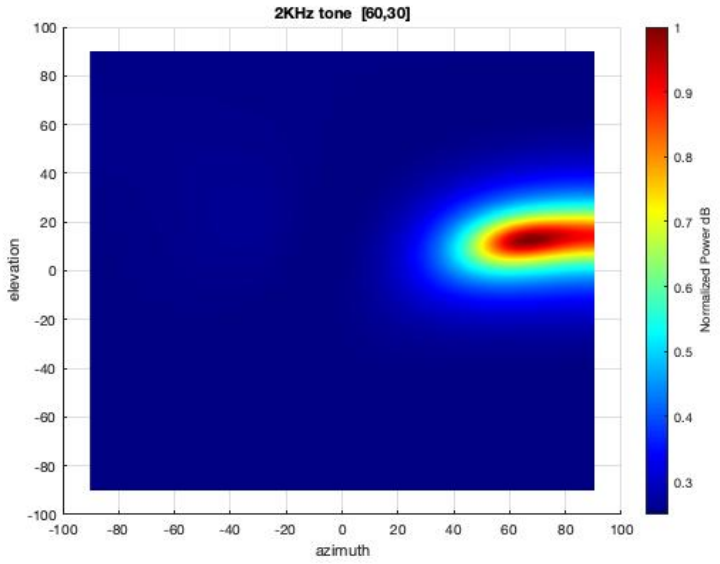

Figure 32 *Power map of 2 KHz tone coming from [60 30] of 0.2 resolution*

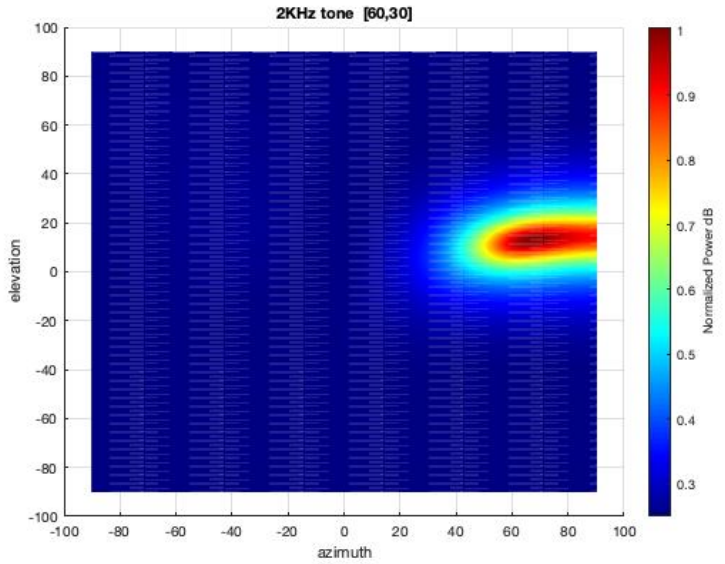

Figure 33 *Power map of 2 KHz tone coming from [60 30] of 0.5 resolution*

In addition, MUSIC algorithm is not affected by the reflected sound and this allows to identify uniquely the sound source.

Furthermore, we have noticed an improvement of the accuracy which drop from 6.8 to 4.1 for the maximum value and from 3 to 1.2 for the minimum value in term of Root Mean Square Error.

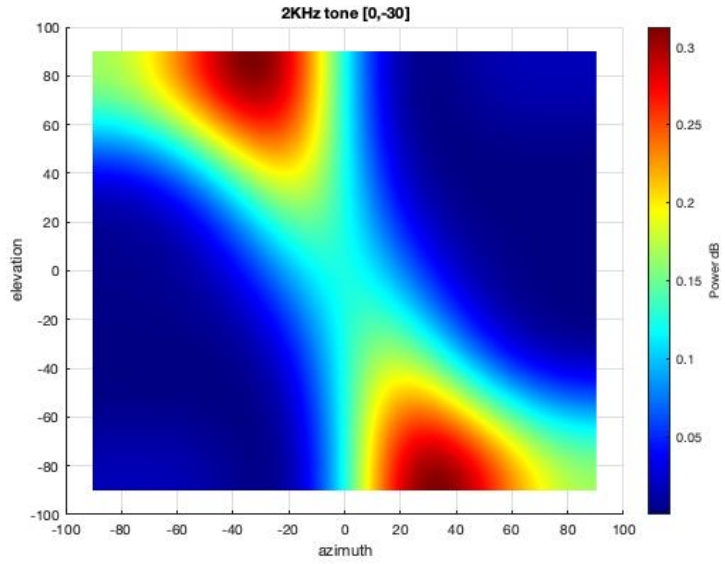

Figure 34 *Power map of 2 KHz tone coming from [0 -30] generated by DaS*

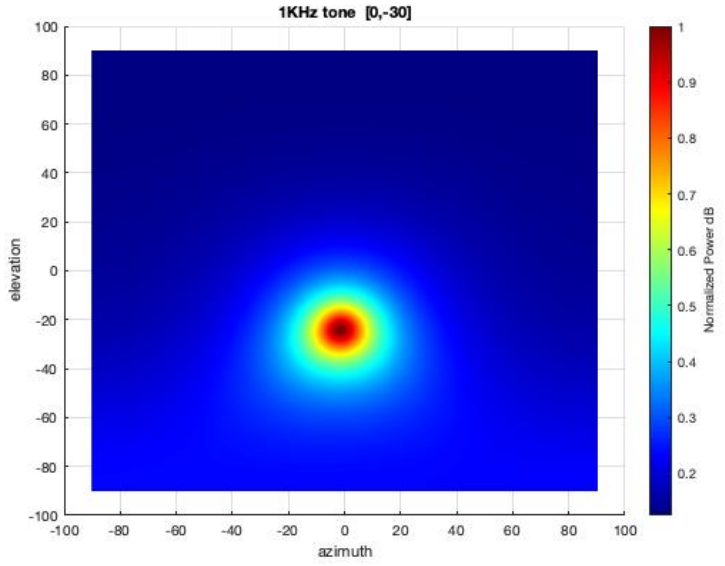

Figure 35 *Power map of 2 KHz tone coming from [0 -30] generated by MUSIC*

4.2 REAL TIME MODE EXPERIMENT

Once evaluated the two algorithms, we have decided to implement MUSIC method in the real time application due to its advantages in term of accuracy and resolution respect to the DaS algorithm.
As anticipated in chapter 3, we have included the streaming video in the app. The goal is to create an acoustic scene through a generated spectrum which maps the whole area seen by the camera [13]. Since the two angles of view of the camera are 80 degrees Azimuth and 70 degrees Elevation, we have set the scan angle of the steering vector equal to the AoV of the CMOS OV5640 camera. The resolution of the camera is 800x600. It was scaled up until 800x700 and multiplying the X and Y axis by a factor of 10, we have perfectly overlayed the image coming from the camera and the power map coming from the MUSIC method [25]. We have also implemented in the app a frequency spinner in order to select quickly a research frequency.
Different from the static tests, we have performed three experiments in three different environments, each one with own reverberation time. We have also used different setup using several speakers each one in a specific room. Furthermore, in order to simulate a real noise machine, we have

performed the experiments using a wideband signals as pink and white noise in addition to a pure tone.

A first test was done in the same environment where we have performed the static acquisition and tests. The reverberation time was unvaried ad equal to 1.2 seconds as the used speaker too.
In this condition the figure 36 shows the result for a pink noise incoming signal.

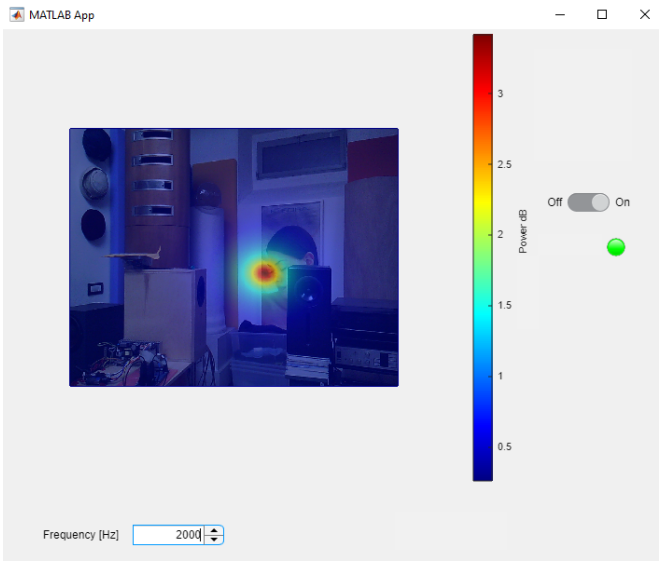

Figure 36 *Acoustic Camera Image related to Hypnocampus 2C at f=2000 Hz*

As we can notice, the system does not locate correctly the sound source with a search frequency equal to 2000 Hz. Tuning the search frequency at 800 Hz, the system locates correctly the sound source.

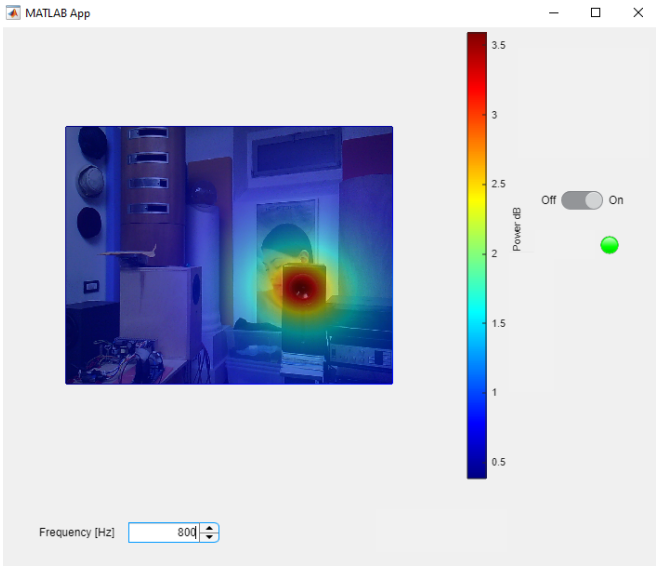

Figure 37 *Acoustic Camera Image related to Hypnocampus 2C at f=800 Hz*

In the second experiment, we have moved in a room a reverberation time equal to 0.5 while the used speaker was a 6.5” Yamaha HS7.

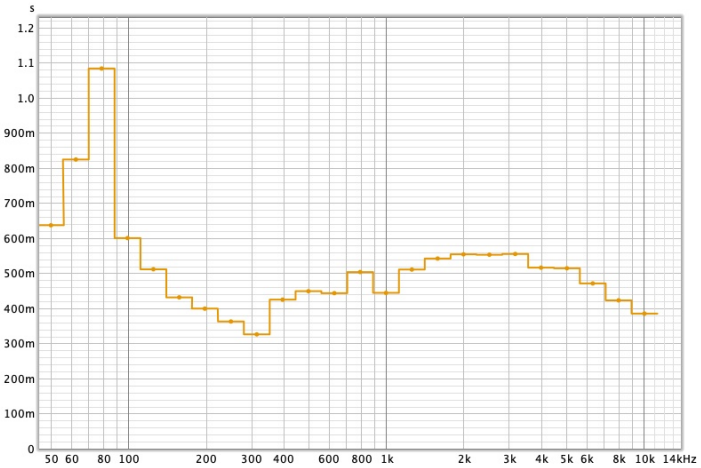

Figure 38 *T60 Room with Yamaha HS7*

As the first test, we have noticed that our system can locate the sound source with a proper search frequency.

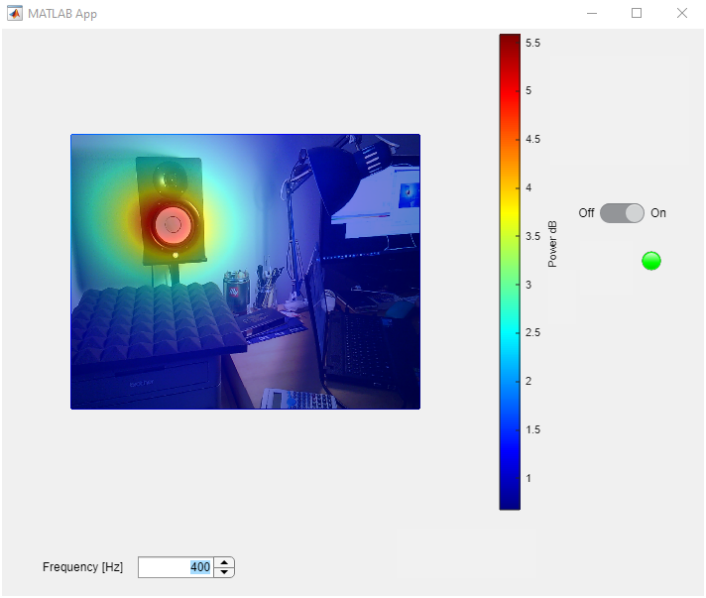

Figure 39 *Acoustic Camera Image related to Yamaha HS7 at f=400 Hz*

In the figure 39, the test signal was a pink noise. In this case the search frequency of 400 Hz is related to the emitting woofer surface. The same behavior was observed using a white noise, as show in figure 40, where the sound also coming from a different position and the search frequency is related to the tweeter area.

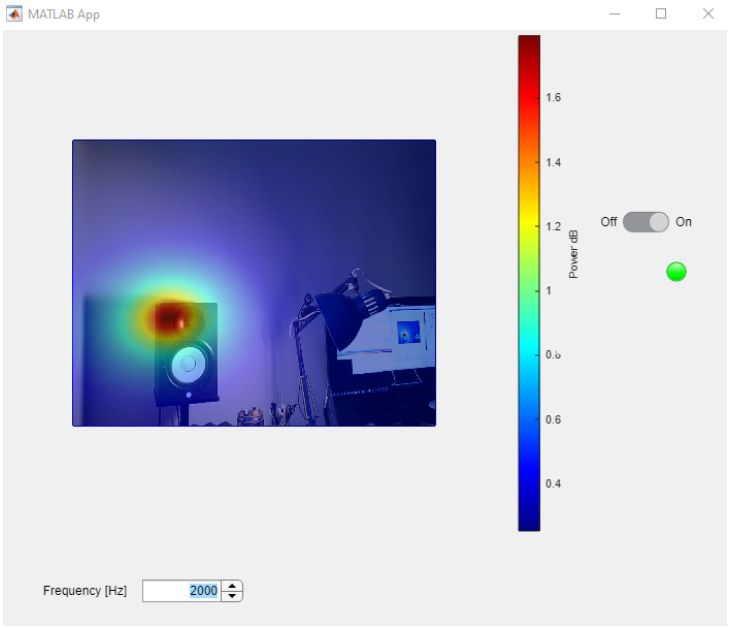

Figure 40 *Acoustic Camera Image related to Yamaha HS7 at f=2000 Hz*

In the last experiment, we have accomplished the test using two different speakers separately which are a woofer 10” Tannoy T 125 and a wideband 2.7” Sony SS-TS20. It was carried out in a third environment with the reverberation time equal to 0.6.

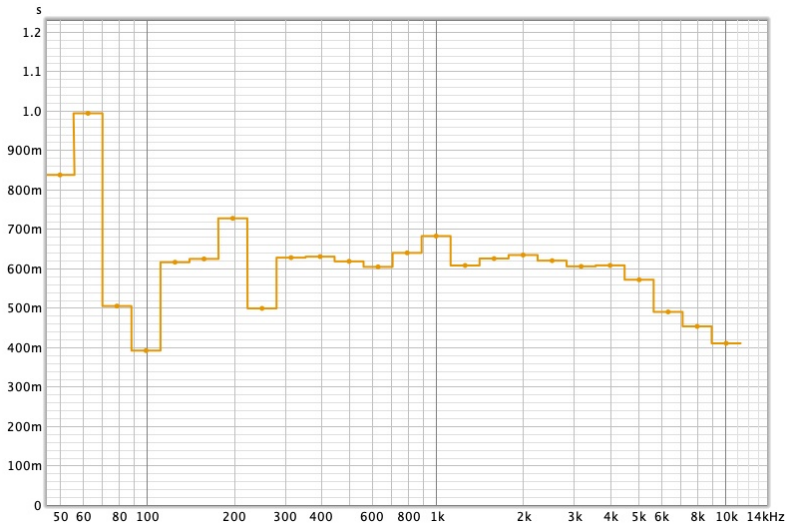

Figure 41 *T60 Room with Tannoy T125*

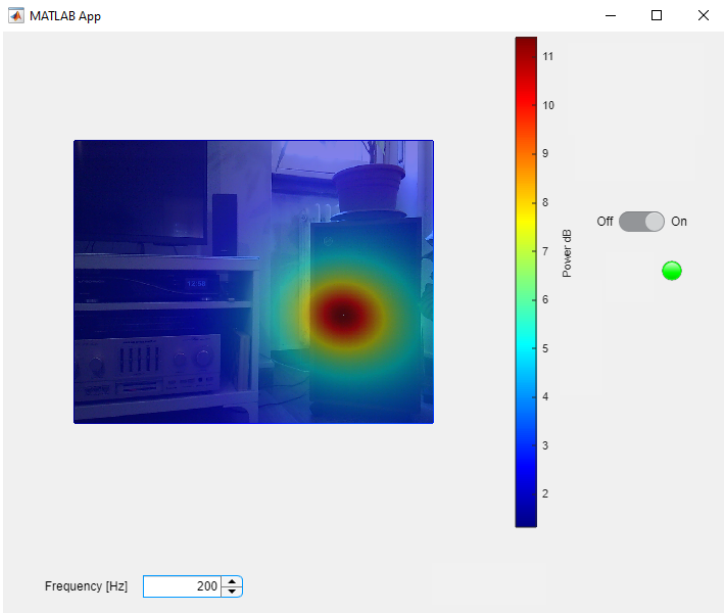

Figure 42 Acoustic Camera Image related to Tannoy T-125 at f=200 Hz

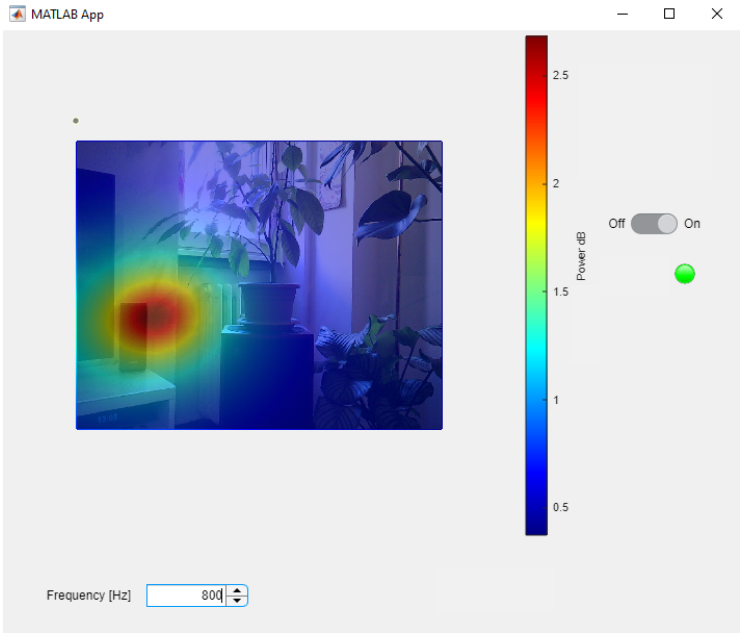

Figure 43 *Acoustic Camera Image related to Sony SS TS-20 at f=800 Hz*

The figures 42-43 show the acoustic scene created for two different sound sources which generate a pink noise signal. As expected, we have noticed that the accuracy of the device depends on the search frequency in case of wideband signals.

In table 3 are collected all the obtained results. The table related the search frequencies that allow to correctly locate the sound source and the corresponding signal and speaker from which they were generated.

SPEAKER	PINK NOISE	WHITE NOISE	T60 [S]
HYPNOCAMPUS 2C	800	1600-2000	1.2
YAMAHA HS7	200-400	900	0.5
TANNOY T 125	200	800	0.6
SONY SS-ST20	800	2000	0.6

Table 3 *Relation between search frequencies, signals and speakers*

In case of wideband incoming signals, we have noticed that the accuracy of the developed acoustic camera depends not only on the search frequency, but it also depends on the emitting surface. In particular, the higher emitting area of sound source the lower tend to be the search frequency. The figure 44 presents the behavior of the acoustic camera in presence of real noise.

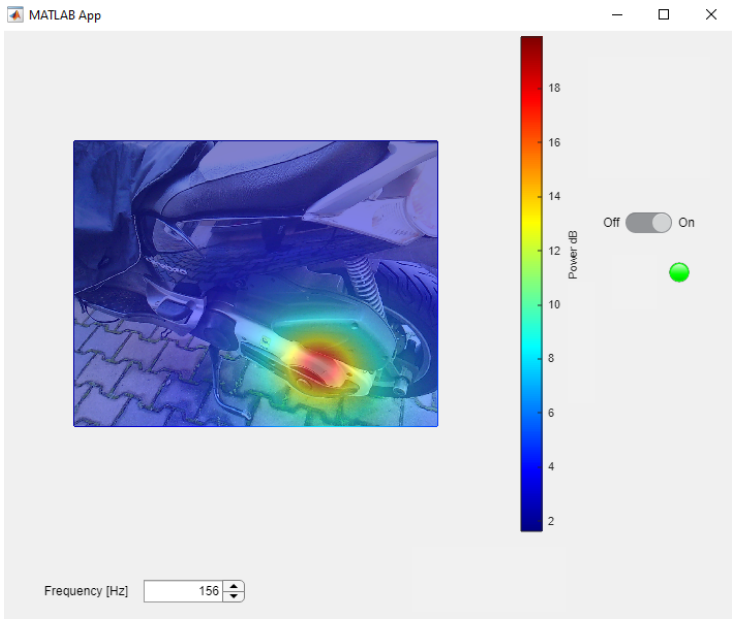

Figure 44 *Acoustic Camera image related to a real noise*

Conclusion and Future works

The goal of this thesis was to design and develop a low-cost acoustic camera. Based on MUSIC algorithm, using a mini-DSP UMA 16 with a CMOS camera, we have implemented a GUI that facilitates its use resulting very intuitive. The operating frequency range is not so extended due to its physical limitation but, depending on the input signal and depending on the emitting area of sound source, the system can detect and locate signals which have energy content from 4000 Hz to 200 Hz. The reproduced acoustic scene depends on both the resolution of the power map and on the resolution of the camera. A right compromise between quality image and time to stream it was found by setting the algorithm resolution at 0.2 and the image resolution at 800x600. In these conditions the system reacts rapidly to an incoming sound source. In addition, the acoustic camera presents a discrete accuracy that, in the best case drops until 1.2 in terms of RMSE. Unfortunately, the device fails when the incoming signals are a low frequency pure tone since it is surrounded by signals that have wavelength greater than its size. Moreover, the acoustic camera is not able to detect and locate more than one sound source but it only detects the source with high SPL.

Although the acoustic camera can detect a wideband signal it is implemented using a narrowband algorithm. A future work could be to implement a wideband algorithm.
Since the microphone array has omnidirectional microphones, which can capture signals coming from the back, our system is almost insensitive to signal coming from the back when there is a presence of sound source in front of it. It could be very interesting evaluate the behavior of the system in presence of signals which also coming from the back and make it insensible to the last one.
Furthermore, to increase the overall performance of the acoustic camera, the Matlab code could be programmed directly on the UMA 16 dividing the audio process and the video process between the dsp and the cpu. This would make the device a stand-alone instrument that could be used without a pc support.

REFERENCE

[1] Acoustic Camera Design with Different Types of MEMS Microphone. Arrays Sanja Grubesa Jasna Stamac, Mia Suhanek Department of Electroacoustics, Faculty of Electrical Engineering and Computing, University, of Zagreb, Zagreb, Croatia.

[2] Computer-steered microphone arrays for sound transduction in large rooms. The Journal of the Acoustical Society of America, 78(5), 1508-1518.Flanagan, J. L., Johnston, J. D., Zahn, R., & Elko, G. W. (1985).

[3] Designing the Acoustic Camera using MATLAB with respect to different types of microphone arrays. J. Stamac, S. Grubesa and A. Petosic Second International Colloquium on Smart Grid Metrology, SMAGRIMET 2019.

[4] Microphone Array and Spatial Method. Augusto Sarti Politecnico Di Milano.

[5] Beamforming: A Versatile Approach to Spatial Filtering. Barry D. Van Veen and Kevin M. Buckley.

[5] The Development and Analysis of Beamforming Algorithms Used for Designing an Acoustic Camera Sanja Grubesa, Jasna Stamac, Ivan Krizanic, Antonio Petosic.

[6] Beamforming algorithms – beamformers. Jørgen Grythe, Squarehead Technology AS, Oslo, Norway.

[7] Design and build of a planar acoustic camera using digital microphones. KhemapatTontiwattanakul. https://www.researchgate.net/publication/335363705.

[8] Fundamentals of Acoustic Beamforming.Leandro de Santana
Department Thermal Fluid Engineering University of Twente. STO-EN-AVT-287.

[9] Display problem on acoustic source identification using beamforming method. Choon-Su Park, Jong-Hoon, Jeon Yang, Hann Kim.

[10] Real-time conversion of sensor array signals into spherical harmonic signals with applications to spatially localised sub-band sound-field analysis. Leo McCormack, Symeon Delikaris-Manias, Angelo Farina, Daniel Pinardi, and Ville Pulkki.

[11] Technical Review beamforming. Brüel & Kjær No.1 2004.

[12] Parametric Acoustic Camera for Real-time Sound Capture, Analysis and Tracking. Leo McCormack, Symeon Delikaris-Manias and Ville Pulkki.

[13] Imaging concert hall acoustics using visual and audio cam- eras. Adam O'Donovan, Ramani Duraiswami, and Dmitry Zotkin, in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*. IEEE, 2008, pp. 5284–5287.

[14] High resolution imaging of acoustic reflections with spherical microphone arrays. Lucio Bianchi, Marco Verdi, Fabio Antonacci, Augusto Sarti, and Stefano Tubaro in *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2015 IEEE Workshop on*. IEEE, 2015.

[15] Applications of Spatially Localized Active-Intensity Vectors for Sound-Field Visualization. Leo Mccormack, *AES Student Member.*

[16] A frequency estimation algorithm based on carrier date detection scheme and MUSIC algorithm for DS/BPSK signals. Biwen Wang, Peng Liu, Jiyan Huang, Bowu, and Guocai Mu.

[17] Acoustic imaging and holography. *Spectrum, IEEE* Korpel, A (1968).

[18] A robust doppler ultrasonic 3D imaging system with MEMS microphone array and configurable processor. Maeda, Y., Sugimoto, M., & Hashizume, H. In *Ultrasonics Symposium (IUS), 2011 IEEE International* (pp. 1968- 1971). IEEE.

[19] A combined microphone and camera calibration technique with application to acoustic imaging. Legg, M., & Bradley, S. *Image Processing, IEEE Transactions on, 22*(10), 4028-4039.

[20] Three-dimensional ultrasound imaging in air using a 2D array on a fixed platform. Moebus, M., & Zoubir, A. M. (2007, April). In *Acoustics, Speech and Signal.*

[21] Signal processing in acoustic imaging. Keating, P. N., Sawatari, T., & Zilinskas, G. (1979). *Proceedings of the IEEE, 67*(4), 496-510.

[22] *Microphone array signal processing* (Vol. 1) Benesty, J., Chen, J., & Huang, Y. (2008). Springer Science & Business Media.

[23] https://www.coventor.com/blog/explanation-new-mems-microphone-technology-design

[24] The fusion of distributed microphone arrays for sound localization. Aarabi, P. (2003). *EURASIP Journal on Applied Signal Processing, 2003,* 338-347.

[25] Combination of microphone array processing and camera image processing for visualizing sound pressure distribution. Goseki, M., Ding, M., Takemura, H., & Mizoguchi, H. (2011, October). In *Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on* (pp. 139-143). IEEE.

[26] Visualizing sound pressure distribution by kinect and microphone array. Goseki, M., Takemura, H., & Mizoguchi, H. (2011, December). In *Robotics and Biomimetics (ROBIO), 2011 IEEE International Conference on* (pp. 1243-1248). IEEE.

[27] Sound positioning using a small-scale linear microphone array. Pei, L., Chen, L., Guinness, R., Liu, J., Kuusniemi, H., Chen, Y., ... & Soderholm, S. (2013, October). In *Indoor Positioning and Indoor Navigation (IPIN), 2013 International Conference on* (pp. 1-7). IEEE.

[28] An acoustic imaging simulation based on microphone array. Jing, Z., Bo, L., LU, D., & Errui, C. (2011, July). In *Cross Strait Quad-Regional Radio Science and Wireless Technology Conference (CSQRWC), 2011* (Vol. 2, pp. 1398-1401). IEEE.

[29] Fundamentals of digital array processing. Dudgeon, D. E. (1977). *Proceedings of the IEEE, 65*(6), 898-904.

[30] Beam patterns from pulsed ultrasonic transducers using linear systems theory. Bardsley, B. G., & Christensen, D. A. (1981). *The Journal of the Acoustical Society of America, 69*(1).

[31] A study on acoustic imaging based on beamformer to range spectra in the phase interference method. Miyake, R., Hayashida, K., Nakayama, M., & Nishiura, T. (2013, June). In *Proceedings of Meetings on Acoustics* (Vol. 19, No. 1, p. 055041). Acoustical Society of America.

[32] Audio location: Accurate Low-Cost Location Sensing. Scott, James, Dragovic, Boris - (Intel research Cambridge 2005)

[33] Microphone Arrays: A Tutorial. Iain McCowan

[34] Design and use of microphone directional arrays for aeroacoustic measurements. W. M. Humphreys Jr., T. F. Brooks, W. W. Hunter Jr., and K. R. Meadows, in 36th Aerospace Sciences Meeting & Exhibit, Reno, NV, Jan 1998, AIAA Paper No. 98-0471.

[35] Array Signal Processing. S. U. Pillai and C. S. Burrus, Springer-Verlag, New York, 1989.

[36] Digital Signal Processing - Principles, Algorithms and Applications 3rd edition. J. G. Proakis and D. G. Manolakis, Prentice Hall International Editions, 1996

[37] Study of DOA Estimation Using Music Algorithm. Bindu Sharma, Ghanshyam Singh, Indranil Sarkar. International Journal of Scientific & Engineering Research, Volume 6, Issue 7, July-2015

[38] Subspace Methods for Direction-of-Arrival Estimation. A. Paulraj. B. Ottersen, R. Roy, A. Swindlehurst, G. Xu and T. Kailath