

# Analysis of Sound Localization and Microphone Arrays

## Análisis de la localización de sonido y arreglos de micrófonos

Recibido: 4 de enero de 2014

Aprobado: 7 de febrero de 2014

Para citar este artículo: Z. E. Hajj, «Analysis of sound localization and microphone arrays», *Ingenium*, vol. 15, n.º 29, pp. 49-70, mayo, 2014.



Zena El Hajj\*

### Abstract

This work describes the principles of sound localization and the capture techniques of microphone arrays for the spatial recording of acoustic signals. The work presents the physical principles of microphone arrays and multidimensional capture techniques, and delivers solutions that conventional systems do not. It is based on the basic principles of surround sound, and holophony, and presents a solution which is based on the MIMO Systems theory, from Radioelectronics.

### Keywords

Sound localization, microphone arrays, multidimensional capture techniques, surround sound, MIMO systems.

### Resumen

El presente trabajo es describir los principios de localización de sonido y arreglos de micrófonos destinados a la grabación espacial de señales acústicas. El trabajo presenta los principios físicos de arreglos de micrófonos y técnicas de captura multidimensional, además de proveer soluciones que los sistemas convencionales no logran. Está basado en los principios básicos de sonido envolvente, y holofonía, y presenta una solución, basada en la teoría de sistemas MIMO, de radioelectrónica.

### Palabras clave

Localización de sonido, arreglos de micrófonos, técnicas de captura multidimensional, sonido envolvente, Sistemas MIMO.

\* M. Sc. Biomedic Engineer, Teesside University, Middlesbrough (UK). Researcher at the University of Teesside, United Kingdom, born in the Lebanon. E-mail: zelhajj@seznam.cz.

## I. Introduction

Array processing involves the use of multiple sensors to receive or transmit a signal carried by propagating waves. Sensor arrays have application in several fields as radar, sonar, radio, seismology, astronomy and tomography.

Application Field	Description
Communications	Directional transmission and reception
Geophysics	Earth crust mapping and oil exploration
Biomedicine	Fetal heart monitoring and hearing aids
Astronomy	High resolution imaging of the universe
Radar	Phased array radar, air traffic control, and synthetic aperture radar
Sonar	Source localization and classification
Imaging	Ultrasonic and tomographic

Table 1. Various Fields of Application for Array Processing.

Phased arrays of antennas are random-access devices whose basic architecture consists of 2 building blocks, namely radiating elements and a beam forming network. They are adapted for reducing severe radiation hazards to the human body, and it is useful for transmitting and receiving signals while taking into account the indoor electromagnetic field strength, using the concept of phase construction and destruction. This concept is inferred to sound waves in order to bear a device based on phased arrays of microphones that capture a sound arriving from a certain direction while ignoring other less significant sources and deriving from other directions.

Many previous researches have shown that this deduction has led to poor results; the prominent reason to this was the wideband characteristic of sound and the poor directivity of the designed phased arrays, especially at the array boundaries.

The project focuses on three principal tasks:

1. To constantly localize a main sound source among many.
2. To capture the sound emitted by that source and amplify it, while ignoring noise and other less significant sources of sound.
3. To assure an automatic transition among the sources.

## II. Phased arrays

An efficient and recent sound localization technique uses the concept of phased arrays of microphones [1]. The array provides the user with a dramatic improvement in speech perception over existing hearing aid designs, particularly in the presence of background noise, reverberation, and feedback. The array enhances signal-to-noise ratio and reduces the effects of reverberation what will describe the basics behind the concept of phased arrays, the constraints to which it is affair and its beam forming effect.

As illustrated in reference [2], let's observe figure 1, representing K- receivers, where a plane wave is incising under an angle  $\theta$  (i.e. the angle formed by the wavefront and the array is  $\theta$ ). Because this wavefront is not parallel to the array, the wave will not attain the elements at the same time. Assuming the receivers are identically spaced by a distance  $d$  and the wave reaches the receiver K, it has to **cross** an additional distance of  $d\sin\theta$  to reach receiver K-1,  $2d\sin\theta$  to attain receiver K-2, and so on.

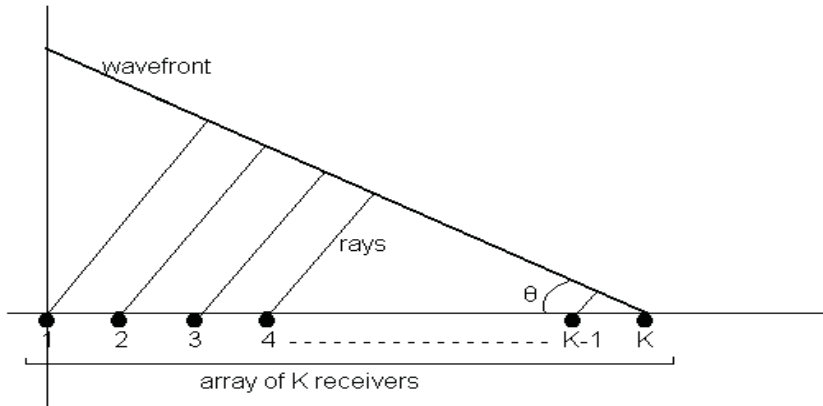


Figure 1. Phased Array Reception.

Given that the waves are periodic in time and space, they will be received at several receivers with diverse phases. For each angle  $\theta$ , a distinct combination of phase differences is obtained. Consequently, the output of each array element can be phase shifted so that the sum of all the outputs of the elements provokes a constructive interference for a given  $\theta$ , while for other angles, the sum will be negligible due to destructive interference.

The receiver spacing has an important role in the performance of the array. Visser [2] affirms that, «the main beam of the linear array antenna gets smaller when the elements fill a larger area», which means that the directivity gets smaller as the **SPACING** increases. However, the spacing cannot be increased indefinitely because surpassing a certain critical spacing will bring additional main beams.

Due to the analogy between electromagnetic waves and sound waves, a phased array of microphones can be nominated.

### A. Far field

The concept annotated above depends on the fact that the incoming sound waves have planar wave fronts. This hypothesis is only valid when the sound wave receiver is located in the far field region of the transmitter. The boundary for this region is at:

$$d = \frac{2L^2}{\lambda} \quad (1)$$

Where:  $L$  is the aperture of the sender and  $\lambda$  is the wavelength of the incoming wave.

## B. Beam forming

The method of directing an array of microphones into receiving all incoming sound sources from a given direction  $\theta$  forms what is called Beam forming.

Beam forming techniques are characterized by [1] :

The data dependence: Beam forming techniques can be both, adaptive or fixed. Adaptive ones modify their parameters based on received signals, while the fixed ones have predefined configurations. The noise assumptions: The result of each beam forming technique is tightly dependent on the type of noise present. Some of them accomplish better with a definite type of noise when others can cover different noise types.

The array configuration: As previously mentioned, the spacing of microphones supports a very important function in the directivity of the array. Each beam forming technique uses a specific microphone spacing which strongly influences the directivity of the array. This phenomenon can be observed in Figure N.º 2.

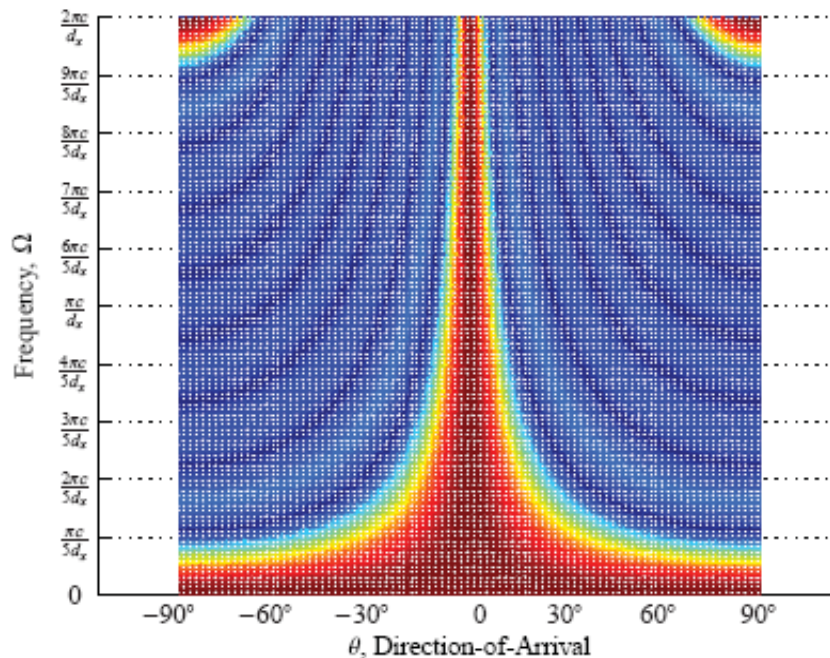


Figure 2. The magnitude of the aperture smoothing as a function of frequency and direction of arrival, showing only the visible region.

## III. Sound localization

The concept of phased arrays can be used for the localization. The mechanisms of sound localization are assessed, as well as the efficacy of previously designed systems.

Sound localization is accomplished by using differences (intensity and level differences) in the sound signals received at different observation points to estimate the direction and eventually the actual location of the sound source. For example, the human

ears, acting as two different sound observation points, enable humans to estimate the direction of arrival of the sound source. Assuming that the sound source is modeled as a point source, two different clues can be utilized in sound localization.

The first clue is the interaural level difference (ILD). Emanated sound waves have a loudness that gradually decays as the observation point moves further away from the source [16]. This decay is proportional to the square of the distance between the observation point and the source location.

Another clue that can be utilized for sound localization is the interaural time difference (ITD), more commonly referred to as the time difference of arrival (TDOA) [3]. Assuming that the distance between each observation point and the sound source is different, the sound waves produced by the source will arrive at the observation points at different times due to the finite speed of sound. Knowledge about the TDOA at the different observation points and the velocity of sound in air can be used to estimate the difference in the distances of the observation points to the sound source location [3, 4, 5, 7, AND 17].

## A. Sound localization mechanism

Directional auditory perception is innate in humans and animals. This sensing ability is greatly dependent on the organs' functioning, especially the two ears, as well as the spatial behavior of sound waves. In order to artificially reproduce this phenomenon, we need to delve into the processes that dictate such a behavior [12]. We are able to recognize sounds coming from the left and right sides, and we are also able to estimate if they are emerging from up or from down. This accordingly reveals that sound localization depends on the azimuth and the elevation angles: this gives a 3D function. If we only take into account the azimuth angle, we get a 2D function [7].

First of all, two important notions are used to localize sound waves: the inter-acoustic differences in phase and differences in intensity level between receptors [12], [22], and [15]. In fact, since the sound waves attain different microphones at different times we can guess their location using the concept of phased arrays.

Data acquisition mainly affects the array's performance as well: in fact, the analog vibrations provoking the sounds recorded by the microphones should be translated into digital information. This conversion commands the use of Nyquist Theorem as well as a proper choice of sampling rate [12], [15].

Once the digital data have been obtained, it should be analyzed: the Fast Fourier Transform is used to detail the frequency content of the recorded speech [12], [22].

Moreover, experiments have also denoted that for a given fixed microphone spacing, rising the number of microphones restricts the acoustic bandwidth of the array [21]. In other terms, as the number of microphones grows, the array becomes more frequency particular, and detects a smaller frequency range in a given direction. What intends that we will have

to select the number of microphones, big enough to serve a narrow directivity, and small enough to serve a bandwidth that suits the human voice range.

Finally, wave diffraction is mostly discounted when the sound source is supposed to be enough faraway to be considered propagating in a planar front.

## IV. External parameters

Failure of most designs is due to the external parameters which room acoustics and noise effects are two of their main factors.

Room acoustics and environment

Any room is usually cut in 4 regions according to frequency-band behavior:

### A. Region X: frequencies between 0 and,

$$f_x = \frac{c}{2L} \quad (2)$$

where L is the largest dimension of the room. Those frequencies cannot be captured by the microphone.

Region A: frequencies between  $f_x$  and,

$$f_A = 11250 \sqrt{\frac{T}{V}} \quad (3)$$

Where V is the volume of the room in feet and T is the reverberation time. This region is characterized by normal modes, which are resonating frequencies that depend on room dimensions.

Region B: frequencies between  $f_A$  and  $f_B = 4 f_A$ . This region is characterized by dominant diffraction and diffusion.

Region C: frequencies larger than  $f_B$  characterized by reflection: sound waves behaving like ray of light (array phasing is the most efficient in this region).

### B. Noise effects

Noise interferes with the information signals, creates distortion, and yields incorrect practical results.

Possible noise sources are:

1. Self-noise from air motor and turbulence within air conditioning system.
2. Mechanical equipment: fans, motors, machine vibration, office machinery and equipment [23].

3. Transmission through walls, ceilings from other neighboring rooms.
4. Noise from external sources: rain, fog, transportations and vehicular noise.
5. Cross talk between rooms.

On the other hand, microphone arrays are typically subject to three main categories of noise.

These categories are defined on the basis of the correlation between noise signals at different spatial locations [10]. This correlation is defined as:

$$\Gamma_{i,j}(f) = \frac{\Phi_{i,j}(f)}{\sqrt{\Phi_{i,i}(f)\Phi_{j,j}(f)}} \quad (4)$$

Where,  $\Phi_{i,j}$  is the cross spectral density between signals  $i$  and  $j$ . It is essentially a normalized cross- spectral measure,

$$(0 \leq |\Gamma_{i,j}(f)|^2 \leq 1)$$

These three types of noise are:

**Coherent noise:** Coherent noise fields are ones in which noise signals travel directly from their source to the microphone array. They are not subject to reflections, dispersion or dissipation due to the acoustic surrounding environment.

In coherent fields, noise inputs at the microphone array are strongly correlated.

In practice, coherent noise fields occur in open-air environments where there are no major obstacles to sound propagation and where wind or thermal turbulence effects are minimal.

**Incoherent noise:** In incoherent noise field, the noise measured at any spatial location is uncorrelated with the noise measured at all of her locations. An example of incoherent noise is the self-noise of the microphone array and processing equipment. This noise, also called spatially white may be assumed as random, thus having a correlation very close to zero.

**Diffuse noise:** Diffuse noise fields receive weakly correlated noise that has approximately the same energy everywhere in the environment. Most incoming noise can be characterized as diffuse (office or car noise can be characterized by a diffuse noise field). Some types of diffuse noise are [21]:

**White noise:** The intensity of the power spectral density is constant and independent of frequency along the whole spectrum.

**Pink noise:** It has a uniform power density for a relative bandwidth generally octave. There is a -3dB/octave frequency response. It is also called  $1/f$  noise.

## **C. Microphones**

The last step of our review focused on microphone design and characteristics. In this section, we analyze microphone technology and microphone directivity.

The most frequently used types of microphones are the condenser microphone and the dynamic microphone. The former operates based on a capacitor whose plate vibrates when exposed to sound waves, creating changes in voltage, whereas the latter is made of a coil inside a magnetic field, which will vibrate, generating electric current, whenever exposed to sound waves. The condenser microphone is much more sensitive to small pressure variation, much more fragile, and much more expensive. Because in our project we will be dealing with average human voice, we do not need to take advantage of the high sensitivity offered by the condenser microphone, so we will basically use dynamic microphones.

A microphone array is a set of closely positioned microphones. Microphone arrays achieve better directionality than a single microphone by taking advantage of the fact that an incoming acoustic wave arrives at each of the microphones at a slightly different time. The chief concepts that are used in the design of microphone arrays are beam forming, array directivity, and beam width.

An array of microphones can do a better job for isolating a sound source and rejecting ambient noise and reverberation.

Microphone array solutions should follow these basic design guidelines:

- Implement the characteristics of the tested microphone array geometries that are supported in Windows Vista as summarized in Table 1 in this paper.
- Meet the design requirements for low noise, directionality, and low manufacturing tolerances for microphones and preamplifiers, as summarized in Table 2 in this paper.
- Follow the recommendations for analog-to-digital converters for sampling rate, sampling synchronization, and anti-aliasing filters as summarized in Table 3 in this paper.

### **1. Beam Forming**

By combining the signals from all microphones, the microphone array can act like a highly directional microphone, forming what is called a «beam.» This microphone array beam can be electronically managed to point to the originator of the sound, which is referred to in this paper as the «speaker».

In real time, the microphone array engine searches for the speaker position and acts as if it points a beam at the current speaker. The higher directivity of the microphone array reduces the amount of captured ambient noises and reverberated waves. More details about the algorithm for beam form design can be found in reference [13, 14]. Typical examples of array responses can be observed in Figure N.º 3.



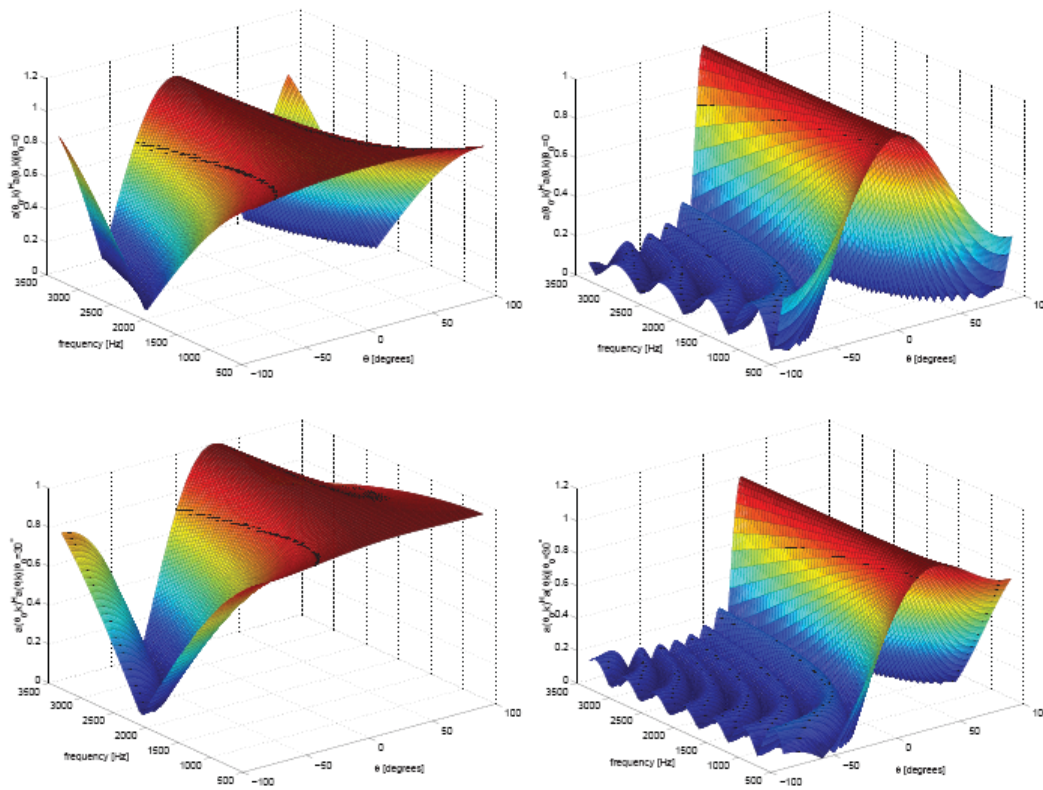


Figure 3. Normalized array response evaluated at  $f = 350$  Hz to  $f=3500$  Hz.

### 1. a) Array Directivity

Because the microphone array output contains less noise and has less reverberation than a single microphone, the stationary noise suppressor does a better job than it would with a signal from a single microphone. A typical directivity pattern of a microphone array beam for 1,000Hz is shown on Figure5. The pattern is more directive than even that of an expensive, high-quality hypercardioid microphone

Directivity is another factor that is necessary to mention. The following figures show different types of microphone directivity.

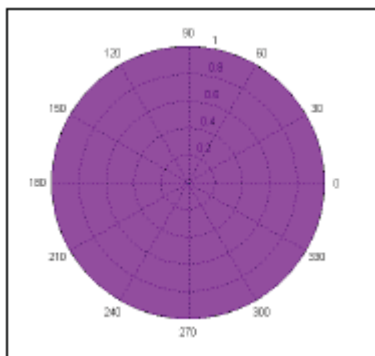


Figure 4.a: Omnidirectional Directivity

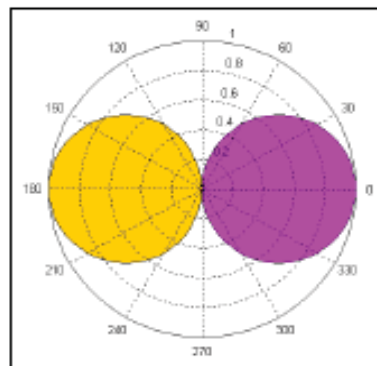


Figure 4.b: Bidirectional Directivity

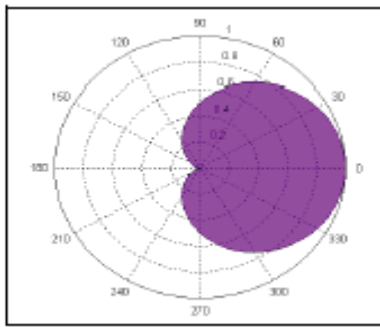


Figure 4.c: Cardioid Directivity

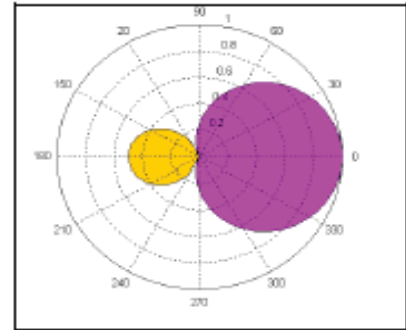


Figure 4.d: Hypercardioid Directivity

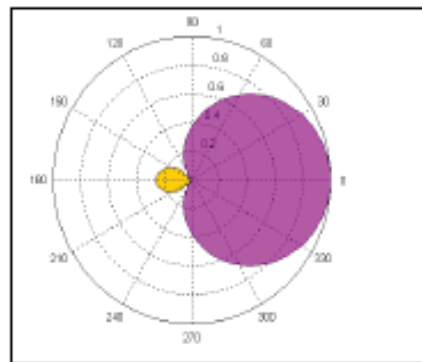


Figure 4.e: Supercardioid Directivity

As it can be observed in Figure n.º 4, directivity increases as frequency increases.

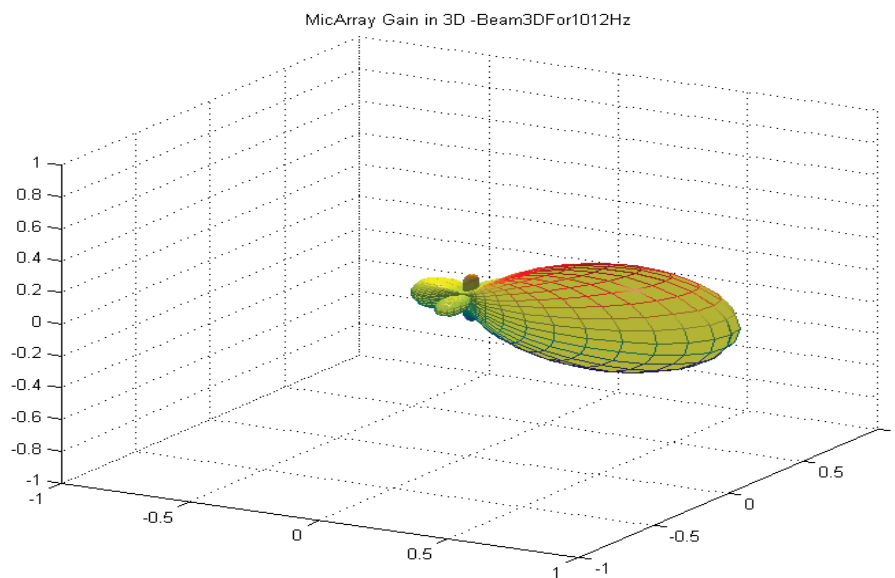


Figure 5. Microphone array directivity pattern in three dimensions

During sound capturing, the microphone array software searches for the speaker's position and aims the capturing beam in that direction. If the person speaking moves,

the beam will follow the sound source. The mechanical equivalent is to have two highly directional microphones: one constantly scans the work space and measures the sound level, and the other—the capturing microphone—points to the direction with highest sound level; that is, to the speaker.

### 1. b) Constant Beam Width

The expression that characterizes the beam width is,

$$b \sim \lambda d / (s \sin \theta)$$

- $\lambda$  = wave length in observer coordinate system
- $d$  = distance of moving source
- $s$  = size of array
- $\theta$  = emission angle rel. velocity vector
- Beam width inversely proportional to frequency, however resolution for high frequencies limited by coherence loss between microphone signals.
- Beam width must be constant for good results.

The normal work band for speech capturing is from 200 Hz to 7,000 Hz, so wavelengths can vary by a factor of 35. This makes it difficult to provide a constant width of the microphone array beam across the entire work band. The problem is somewhat simpler in a typical office environment, where most of the noise energy is in the lower part of the frequency band—that is, below 750 Hz. Reverberation is also much stronger at these low frequencies and is practically absent above 4,000 Hz.

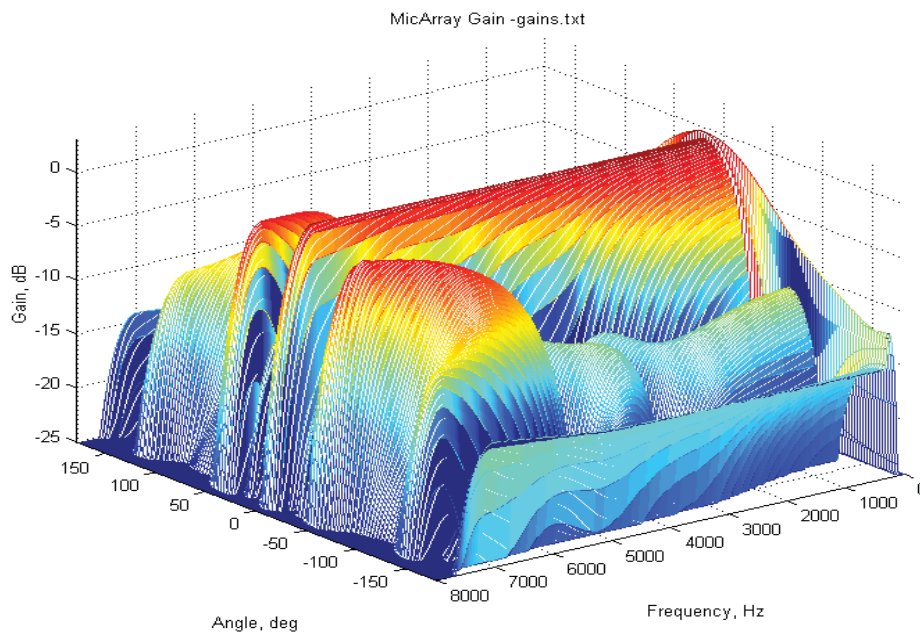


Figure 6. Microphone array directivity as a function of frequency, horizontal plane

## 2 Microphone Array Characteristics

The parameters of a microphone are summarized in the Figure N.º 6. The acoustical parameters of a microphone array are measured like those of any directional microphone. This section defines a set of parameters that are used later to compare different microphone array designs. Because of their directivity, microphone arrays offer better signal-to-noise ratio (SNR) and signal-to-reverberation ratio (SRR) than a single microphone can.

### 2. a) Ambient Noise Gain

The isotropic ambient noise gain for a given frequency is the volume of the microphone array beam:

$$G_{AN}(f) = \frac{1}{V} \iiint_V B(f, c) dc \tag{5}$$

$$c = \{\varphi, \theta, \rho\}$$

Where:  $V$  is the microphone array work volume, that is, the set of all coordinates (direction, elevation, distance).

$B(f, c)$  is the microphone array beam directivity pattern-that is, the gain as a function of the frequency and incident angle. An example for one frequency is shown on Figure 1. An example in one plane is shown on Figure 2.

The total ambient noise gain  $NG$  in decibels is given by:

$$NG = 20 \log_{10} \left[ \int_0^{F_s/2} N_A(f) \cdot H(f) \cdot G_{AN}(f) \cdot df \right] \tag{6}$$

Where:

$N_A(f)$  is the noise spectrum.

$H(f)$  is the preamplifier frequency response (ideally flat between 200 and 7,000 Hz with falling slopes from both sides going to zero at 80 and 7,500Hz respectively).

$F_s$  is the sampling rate (typically 16 kHz for voice applications).

Ambient noise gain gives the proportion of the noise floor RMS in relation to the output of the microphone array and to the output of an omnidirectional microphone. A lower value is better, and 0 dB means that the microphone array does not suppress ambient noise at all.

### 2. b) A-Weighted Ambient Noise Gain

Because humans hear different frequencies differently, many acoustic parameters are weighted by using a standardized A-weighting function. The A-weighted total ambient noise gain  $NGA$  in decibels is given by:

$$NGA = 20 \log_{10} \left[ \int_0^{F_s/2} A(f) \cdot N_A(f) \cdot H(f) \cdot G_{AN}(f) \cdot df \right] \quad (7)$$

where:

$A(f)$  is the standard A-weighting function; other parameters are the same as above.

A-weighted ambient noise gain gives the proportion of the noise floor in relation to the output of the microphone array and to the output of an omnidirectional microphone as they would be compared by a human. In this case,  $-6$ dB NGA means that a human would say that the noise on the output of a microphone array is half that of an omnidirectional microphone.

## 2. c) Directivity Index

Another parameter to characterize the beam-former is the directivity index  $DI$ , given as:

$$P(f, \varphi, \theta) = |B(f, c)|^2, \quad \rho = \rho_0 = \text{const} \quad (8)$$

$$D = \int_{f=0}^{F_s/2} \frac{P(f, \varphi_T, \theta_T)}{\frac{1}{4\pi} \int_0^\pi \int_0^{2\pi} P(f, \varphi, \theta) d\theta d\varphi} df \quad (9)$$

$$DI = 10 \cdot \log_{10} D \quad (10)$$

Where:

$P(f, \varphi, \theta)$  is called power pattern.

$\rho_0$  is the average distance (depth) of the work volume.

$(\varphi_T, \theta_T)$  is the steering direction or main response axis (MRA).

The directivity index characterizes how well the microphone array detects sound in the direction of the MRA while suppressing sounds, such as additional sound sources and reverberation that comes from other directions. The  $DI$  is measured in decibels, where  $0$ dB means no directivity at all. A higher number means better directivity. An ideal cardioid microphone should have  $DI$  of  $4.8$  dB, but in practice cardioid microphones have a  $DI$  below  $4.5$ dB.

## V. Supported Microphone Array Geometries

The proper microphone array geometry (number, type, and position of the microphones) is critical for the final results.

Microphone array	Elements	Type	NG [dB]	NGA [dB]	DI [dB]
Linear, small	2	uni-directional	-12.7	-6.0	7.4
Linear, big	2	uni-directional	-12.9	-6.7	7.1
Linear, 4el	4	uni-directional	-13.1	-7.6	10.1
L-shaped	4	uni-directional	-12.9	-7.0	10.2
Linear, 4 el second geometry	4	integrated	-12.9	-7.3	9.9

Table 2. Characteristics of supported microphone arrays.

Table 2 shows only the noise reduction due to microphone array processing. The stationary noise suppressor in the audio stack will add 8 to 13dB of noise reduction. Themicrophone array not only reduces the amount of ambient noise, but it also helps this noise suppressor to do a better job. Suppose that the signal-to-noise-ratio (SNR) in the room is 3dB when captured with an omnidirectional microphone. With this input SNR, a stationary noise suppressor cannot do much noise reduction without introducing heavy nonlinear distortions and adding audible artifacts called musical noises. The noise reduction can add around 3dB as well, so in this case the output SNR is 6dB.

Under the same conditions, the microphone array reduces 13dB of the ambient noise, and now the noise suppressor has a 16dB SNR on its input. It can easily reduce an additional 13dB of stationary noise without significant distortion in the signal and audible musical noises. The output SNR in this case will be 29dB, which is 23dB better than a system with an omnidirectional microphone. The total noise reduction of the audio stack reaches an impressive 26dB, creating high sound quality with a very low level of distortion and artifacts.

### A. Two-element arrays

Two-element microphone arrays can cover a quiet office or cubicle with good sound capturing when the speaker is less than 0.6 meters (2 feet) from the microphones. These arrays are suitable for integration into laptops and tablets or as standalone devices. Both microphones point directly forward to the speaker. These microphone arrays steer the beam in a horizontal direction only, in the range of  $\pm 50^\circ$ . The directivity in a vertical direction is provided by the uni-directional microphones. The geometries are shown on Figure 7.

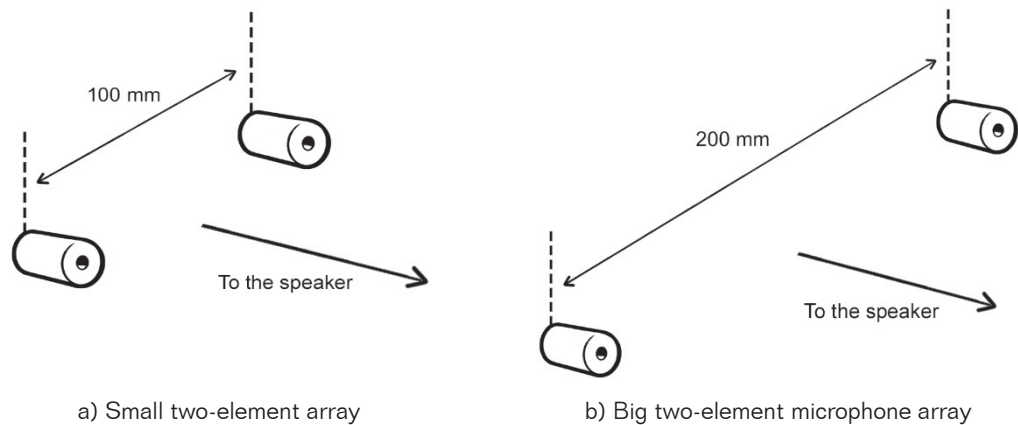


Figure 7. Two-element microphone arrays.

## 1. Small Two-element array

A small two-element array uses two uni-directional microphones 100 millimeters (4 inches) apart. Work area:  $\pm 500$  horizontally,  $\pm 600$  vertically. The channels order is: 1 –left, 2 –right, looking from front of the microphone array.

### 1.2 Big Two-element array

A big two-element array uses two uni-directional microphones 200 millimeters (8 inches) apart. Work area:  $\pm 500$  horizontally,  $\pm 600$  vertically. The channels order is: 1 – left, 2 – right, looking from front of the microphone array.

## 2. Four-element arrays

A four-element microphone array can cover an office or cubicle with good sound capturing when the speaker is up to 2.0 meters (6 feet) away.

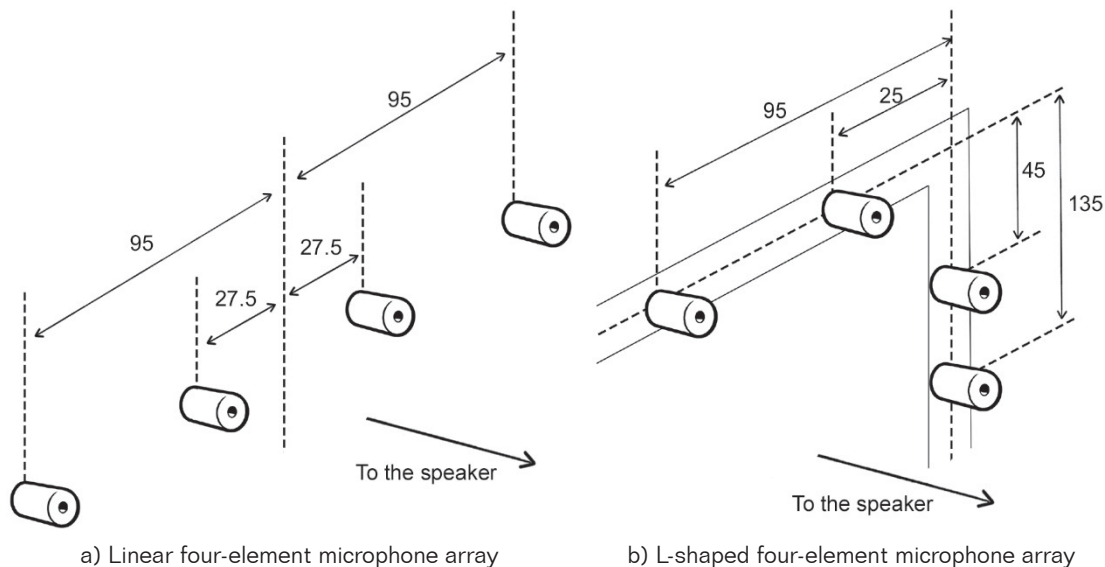


Figure 8. Four-element microphone arrays.

These arrays are suitable for both external microphone array implementation and for integration into laptops and tablets working under normal noise conditions in the office or on the go. All microphones point directly forward to the speaker, as shown in Fig. 8.

### 2.1 Linear Four-element array

A linear four-element array uses four uni-directional microphones 190 millimeters and 55 millimeters apart. Work area:  $\pm 50$  degrees horizontally,  $\pm 60$  degrees vertically. This microphone array implemented as external USB device is shown on Figure 9. It steers the beam in a horizontal direction only, in the range of  $\pm 50$  degrees. Directivity in a vertical direction is provided by the uni-directional microphones. The channels order is from left to right looking from front of the microphone array, that is, 1 is the most left, 4 is the most right microphone.



Figure 9. Linear four-element USB microphone array.

## 2.2 L-shaped Four-element array

An L-shaped four-element array uses four uni-directional microphones and is designed especially for laptop/tablet convertibles. Two of the microphones are positioned on the upper bezel of the screen at 95 and 25 millimeters from the right edge; the second pair is positioned on the right bezel of the screen at 45 and 135 millimeters from the upper horizontal edge.

With this geometry, the microphones are not covered by the hand for either left-handed or right-handed users, and they are away from the keyboard and the speakers. These positions are for laptop mode. Work area:  $\pm 500$  horizontally,  $\pm 500$  vertically. This microphone array integrated into a tablet is shown on Figure 10. Note that in tablet mode, after the screen rotation, the microphones are positioned along the left bezel of the screen. L-shaped four-element arrays steer the beam in horizontal direction only, in the range of  $\pm 500$ , but having microphones with different vertical coordinates improves the directivity in vertical direction. The channels order: 1 and 2 are the left and right microphones in the horizontal side, 3 and 4 are respectively the upper and lower microphones in the vertical side.



Figure 10. L-shaped microphone array integrated into a tablet PC.

## 2.3 Linear Four-element Array Second Geometry

A linear four-element array uses four omnidirectional microphones 160 millimeters and 70 millimeters apart. Otherwise it looks the same as the one on Figure 11a). The channels



order is from left to right looking from front of the microphone array, that is, 1 is the most left, 4 is the most right microphone.

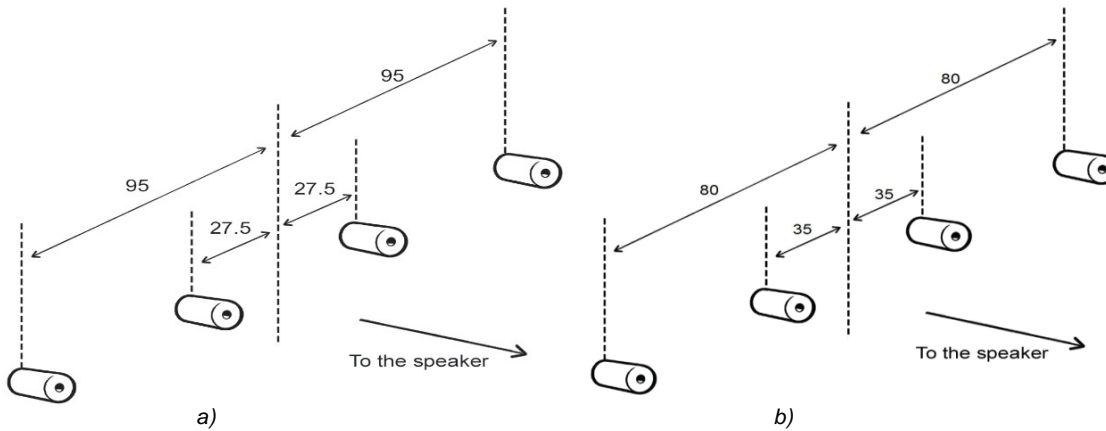


Figure 11. Linear four-element microphone array second geometry.

### 3. Requirements for microphones and preamplifiers

The microphone elements used in a design solution should be low noise, directional, and with low manufacturing tolerances. Under any circumstances, ribbon, carbon, crystal, ceramic, and dynamic microphones must not be used for building microphone arrays.

For microphone arrays integrated into tablets and laptops, it is acceptable to use omnidirectional microphones. The laptop, tablet body, or screen provides certain directivity of the microphone; however, in general such microphone array will have lower noise suppression. For external microphone arrays, using omnidirectional microphones is unacceptable.

To prevent adding electrical noises, shielded cables should be used to connect the microphones with preamplifiers. The preamplifiers should be low noise and should provide high-pass filtering.

The tolerances of elements used should provide close matching of the preamplifiers' parameters.

Component	Requirement
Work band	200-7,000 Hz
Microphone type	Uni-directional
Microphone SNR	Better than 60 dB
Sensitivity tolerances, microphone and preamplifier	In the range of $\pm 4$ dB or better for all frequencies in the work band
Phase response tolerances, microphone and preamplifier	In the range of $\pm 10^\circ$ for all frequencies in the work band
High-pass filter cut-off	150 Hz at $-3$ dB
High-pass filter slope	Better than 18dB/oct

Table 3. Requirements for microphones and preamplifiers.

The device itself is nothing but the microphone array algorithm moves signal processing inside the pc and enables the manufacture of very inexpensive microphone array devices.

The device captures the signals from the microphones converts them to digital form and send them to the computer.

#### 4. Requirements for ADCs

We assume that ADCs used in microphone arrays have integrated anti-aliasing filters and do not place any requirements for low-pass filtering.

For capturing voice for telecommunication and speech recognition, a bandwidth of 200 to 7,000Hz is considered sufficient. This means that a minimum sampling rate of 16kHz is required. Increasing the sampling rate beyond this minimum only increases the processing time without bringing any additional benefits.

The sampling rate of the ADCs used should be synchronized for all microphone channels. Not only should all ADCs use the same sampling frequency (a common clock generator can ensure this), but they should also all be in the same phase with a precision of better than 1/64th of the sampling period. This last point requires taking some additional measures that depend on the ADCs used. For a four-channel microphone array, a typical solution uses two stereo ADCs with a synchronized sampling rate.

The requirements for ADCs are specified in Table 4.

Component	Requirement
Sampling rate	16,000 Hz, synchronized for all ADCs
Sampling synchronization	Better than 1/64th of the sampling period
Anti-aliasing filter	Integrated

Table 4. Requirements for the ADCs.

#### 5. Number of Microphones

The number of microphones depends on the scenario. More microphones mean better noise suppression and greater ability to provide good sound quality in noisy environments. On the other hand, it also means a more expensive array and more CPU power needed for the real-time computations.

- For capturing a speaker voice from up to 2 meters away in office or cubicle noise conditions and  $\pm 500$  work area, *four microphones* should be sufficient and provide a good trade-off of price versus performance
- A *two-element* microphone array can be used for capturing the PC user's voice in a quiet office when the user is very close to the microphones-for example, when the user is sitting in front of the PC display monitor.

Figure N.º 12 shows comparison of directivities between systems of different number of microphone elements.

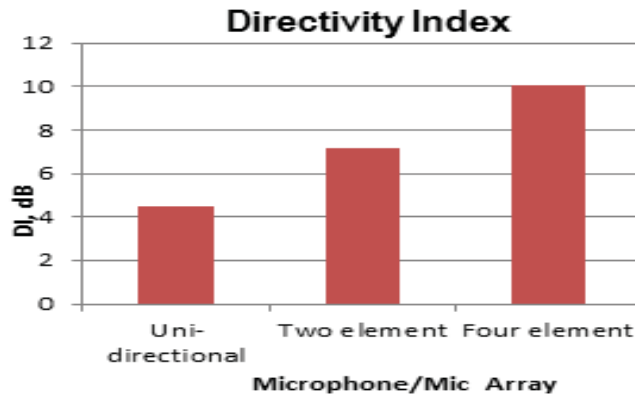


Figure 12. Directivity Index comparison.

## 6. Microphone Array Geometry

The Microsoft generic microphone array system can work with an arbitrary number of microphones and can provide close-to-optimal sound capturing with a given number, type, and position of microphones; together, these factors are referred to as the microphone-array geometry. The microphone-array geometry is critical for achieving good sound capturing. In general, poorly designed and analyzed microphone-array geometry leads to bad results.

## 7. Placement of the Microphone Array

In general, follow these guidelines for placing the microphone array:

- As far as possible from noise sources such as loudspeakers, keyboards, hard drives, and the user's hands.
- As close as possible to the speaker's mouth.

Figure N.º 13 shows these main ideas for placement of the microphones.

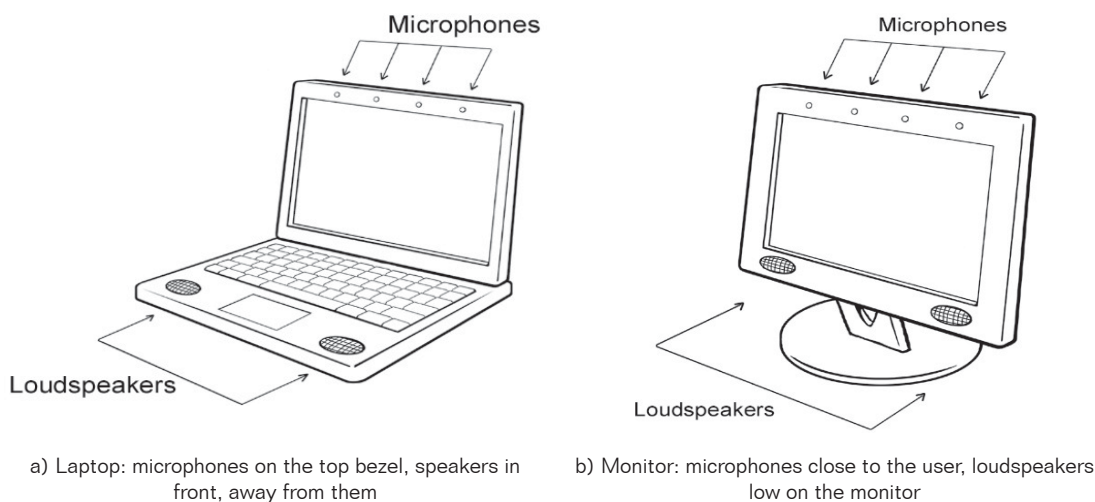


Figure 13. Several microphone array placement ideas for laptops, monitors, and tablets.

## 8. Acoustical Design and Construction

All microphones should be acoustically insulated from rattles and vibrations of the laptop, monitor, or enclosure by placing them in rubber holders. In case of strong mechanical noises or vibrations, additional measures should be taken, such as putting the microphones in foam beds.

Directional microphones need the sound to reach them from both front and back. Ideally, a microphone's enclosure should be acoustically transparent; that is, acoustically the microphones should appear to be hanging in the air in their respective positions. This condition can be simulated with proper vents in front and back of the microphone array enclosure.

Dust and humidity are other factors that can lead to changes in the microphone parameters. The acoustic vents should be covered with acoustically transparent fabric to protect the array.

For integrated solutions in laptops and monitors, the loudspeakers should be acoustically insulated from the chassis so they transmit as little sound and vibration as possible inside the body of the laptop or monitor. There should not be vibrating elements in the loudspeakers' holding or attachment construction. Glue, foam, and rubber should be used to minimize rattles and vibrations and to improve playback sound quality.

If the system microphone array and loudspeakers are going to be used for real-time communications, the loudspeakers should have less than 5% total harmonic distortion (THD) in the work band (typically 200to7,000Hz). Higher quality loudspeakers are needed, and the gain of the microphone preamplifiers has to be adjusted to prevent clipping on the capturing side. These things are required because the acoustic echo canceller (AEC) expects a linear transfer function between the loudspeakers and the microphones. Thus, the AEC will not suppress all harmonics due to nonlinear distortions.

For the real-time communications scenario, it is important that the microphones be placed as far as possible from the loudspeakers. Suppose that AEC suppresses 30dB of the sound from the loudspeaker. In a real-time communication session, the loudspeaker level is usually set to provide sound as loud as a human speaker. If the user is 60 centimeters away from the microphone and the loudspeaker is only 5 centimeters away from the microphone, then the signal that the microphone captures from the loudspeaker will be 12 times (21.5dB) louder than the signal it captures from the human speaker.

## Conclusions

Microphone arrays are valuable tools for high-quality sound pick up in reverberant and noisy environments the identification of the dominating sound sources on moving objects: Sound sources can be determined in each desired frequency band (large arrays for long wave lengths small arrays for small wavelengths) .

The beam forming capabilities of microphone arrays system allow highly directional sound capture, providing superior signal-to-noise ratio (SNR) when compared to single microphone performance.

Most reported results for microphone array technology have been theoretical, proposed algorithms with predicted performance, or based on ideal or near-ideal simulations of microphone-array systems and room environments.

Here are two aspects in microphone array system performance: The ability of the system to locate and track sound sources, and its ability to selectively capture sound from those sources. Both aspects of system performance are strongly affected by the spatial placement of microphone sensors. A method is needed to optimize sensor placement based on geometry of the environment and assumed sound source behavior. The objective of the optimization is to obtain the greatest average system SNR using a specified number of sensors. A method is derived to evaluate array performance for a given array configuration. A framework for optimum placement of sensors under the practical considerations of possible sensor placement and potential location of sound sources is also characterized (Number of required microphones increases with measuring distance).

## References

- [1] Mcgowan, I. Microphone Arrays: A tutorial, April 2001. [www.idiap.ch/~mccowan/arrays/tutorial.pdf](http://www.idiap.ch/~mccowan/arrays/tutorial.pdf)
- [2] Visser, H. *Array and Phased Array Antenna Basics*. New Jersey: Wiley, 2005.
- [3] Gregory, E., Cole, J. Audio Localization. Connexions Project, 2004. [cnx.rice.edu/content/col10250/latest](http://cnx.rice.edu/content/col10250/latest)
- [4] M. S. Brandstein, A framework for speech source localization using sensor arrays, Ph.D. thesis, Brown University, Providence, RI, USA, 1995.
- [5] P. Aarabi and S. Zaky, «Iterative spatial probability based sound localization», in Proc. 4th World Multi-Conference on Circuits, Systems, Computers, and Communications, Athens, Greece, July 2000.
- [6] M. S. Brandstein and H. Silverman. «A robust method for speech signal time-delay estimation in reverberant rooms», in Proc. *IEEE Int. Conf. Acoustics, Speech, Signal Processing*.
- [7] J. Flanagan, J. Johnston, R. Zahn, and G. Elko. «Computer steered microphone arrays for sound transduction in large rooms». *Journal of the Acoustical Society of America*, vol. 78, pp. 1508-1518, November 1985.
- [8] Tauchi M. An analysis of sound localization performance provided by the present sounds of audible traffic signals for visually impaired pedestrians.
- [9] P. Aarabi. «The integration of distributed microphone arrays», in Proc. *4th International Conference on Information Fusion*, Montreal, Canada, July 2001.
- [10] J. Dibiase, H. Silverman, and M. S. Brandstein. «Robust localization in reverberant rooms», in *Microphone Arrays: Signal Processing Techniques and Applications*, Eds., pp. 131-154, Springer Verlag, New York, USA, September 2001.
- [11] P. Aarabi. «Robust multi-source sound localization using temporal power fusion», in Proc. *Sensor Fusion: Architectures, Algorithms, and Applications V*, vol. 4385 of SPIE Proceedings, Orlando, Fla, USA, April 2001.
- [12] Kushleyev A. Sound Localization. University of Maryland, Institute for Systems Research, 2004. [www.isr.umd.edu/ISR/education/REU/REU-2005/REU2004/2004reports/Kushleyev\\_Vohra%20paper.pdf](http://www.isr.umd.edu/ISR/education/REU/REU-2005/REU2004/2004reports/Kushleyev_Vohra%20paper.pdf)
- [13] I. Tashev, H. Malvar. *A New Beam former Design Algorithm for Microphone Arrays*. *Proceedings of ICASSP*, Philadelphia, PA, USA, March 2005.
- [14] I. Tashev. *Gain Self-Calibration Procedure for Microphone Arrays*. *Proceedings of ICME*, Taipei, Taiwan, July 2004.
- [15] Basu, S., Schwartz, S., Pentland, A. Wearable Phased Arrays for Sound Localization and Enhancement. MIT Media Laboratory, IEEE 2000. 111 [doi.ieeecomputersociety.org/10.1109/ISWC.2000.888471](http://doi.ieeecomputersociety.org/10.1109/ISWC.2000.888471)
- [16] K. Guentchev and J. Weng, «Learning-based three dimensional sound localization using a compact non-coplanar array of microphones», in Proc. *AAAI Spring Symposium on Intelligent Environments*, Stanford, Calif, USA, March 1998.

- [17] Intel High Definition Audio specification at [www.intel.com/standards/hdaudio/](http://www.intel.com/standards/hdaudio/).
- [18] D. Rabinkin, R. J. Ranomeron, A. Dahl, J. French, J. L. Flanagan, and M. H. Bianchi. «A DSP implementation of source location using microphone arrays», in *Proc. 131st Meeting of the Acoustical Society of America*, Indianapolis, USA, May 1996.
- [19] Widrow, B. A. *Microphone Array for Hearing Aids*. Stanford University.
- [20] M. S. Brandstein, J. Adcock, And H. Silverman. «A practical time-delay estimator for localizing speech sources with a microphone array», *Computer Speech & Language*, vol. 9, n.º 2, pp. 153-169, 1995.
- [21] Loeppert, P. & Wickstrom, T. *Advanced Microphone Technology*. Emkay Innovative Products.
- [22] Kazuhiro, N., Hiroshi, O., Hiroaki, K. Auditory Fovea Based Speech Enhancement and its Application to Human-Robot Dialog System. Kyoto University, Graduate School of Informatics. [www.symbio.jst.go.jp/~nakadai/papers/icslp02-1.pdf](http://www.symbio.jst.go.jp/~nakadai/papers/icslp02-1.pdf).
- [23] Beranek, L. *Noise Reduction Prepared for a Special Summer Program at the Massachusetts Institute of Technology*. New York: McGraw-Hill, 1960.

# INGENIUM

REVISTA DE LA FACULTAD DE INGENIERÍA



UNIVERSIDAD DE  
SAN BUENAVENTURA  
BOGOTÁ



2014