Research and Analysis of Methods for Localization of Audio Sources

Viktor Hristov[1] and Snejana Pleshkova[2]

Abstract – **In this paper was done research and deep analysis of existing methods for localization of audio sources. After the precise mathematical description of microphone arrays, are prepared simulations of some chosen types of microphone arrays, for example with three microphones, using the Matlab resource and it was applied the Time Delay Of Arrival method for localization of the audio source. Therefore, the goal of this article is to present the microphone array description as mathematical model and to use this model for simulation in Matlab environments.**

Keywords – **microphone array, beamforming, simulation, time delay of arrival.**

I. INTRODUCTION

Discussing and researching of microphone array processing is one of the most interesting topics nowadays. The main problem in simulating and constructing following devices is localization of audio source and define angel at which sound wave has arrived. It has many important applications including systems for extracting voice input from ambient noise (speech recognition systems, hearing aids and telephones), mobile robots [1, 2], video-conferencing systems [3], surround sound and locating objects for police and military purposes [4].

Methods for localization of sound sources can be defined in three main groups: dependent from time of arrival, fixed beamforming and adaptive (hybrid) beamforming using correlation functions.

The most popular and investigated method for sound source localization (SSL) is based on measurement of the time of delay of arrival (TDOA) [5].

Beamforming is spatial filtering based on microphone characteristics and microphone array configuration. It can be classified in two groups: fixed (classical) and adaptive (hybrid) depending of the input data [6].

There are two scenarios dependent of the distance from source to the microphone: far-field and near-field. In first one situation, the destination from source to the each microphone in microphone array is many times bigger than the distance between microphones and it can assume that waves are planar. In near-field, wave fronts are spherical. It is included attenuation of the signals [7].

[1]Viktor Hristov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: hristov.viktor@gmail.com.

[2]Snejana Pleshkova is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: snegpl@tu-sofia.bg.

There are two scenarios dependent of the place, were experiment was done – open space, without any reflections and in closed space, also called multipath propagation, because each reflection from the wall or object can be treated as a separate source [6].

The goal of this article is to present the microphone array description as mathematical model and to use this model for simulation in Matlab environments. As an example to prepare experimental tests of the proposed simulation model is chosen the most popular method for sound source localization (SSL) based on measurement of the time of delay of arrival (TDOA).

II. METHODS OF INVESTIGATION

A. Mathematical model of sound localization method using time delay of arrival

On Fig. 1 is shown the general microphone array processing block scheme suitable for sound localization. Sound reaches the microphone array in form of plane wave. It was made an assumption that the sound source is at reasonable distance from microphone sensors, therefore a flat sound wave falls in the place of linear arranged microphone array. Each microphone receives sound waves in different moment. Output signals from all microphones are processed with appropriate filters $w_1(f)$… $w_N(f)$ necessary to adjust time delay of microphones from M_2 to M_N according to first microphone M_1. After that all from filters output sound signals are summed to obtain final output signal of the microphone array y(f).

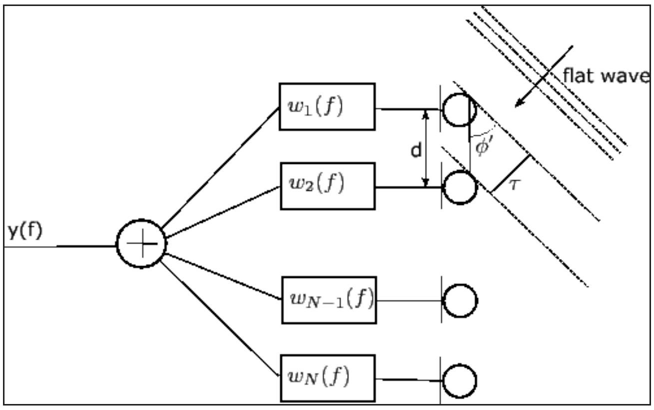

Fig. 1. Microphone array processing

Shown on Fig. 1 block scheme can be used to describe the mathematical model of a linear microphone array. This model can be used to analize different types of structures in situation of linear microphone array, choosing and changing *M* as number of microphones in microphone array, which can be situated on a distance d from each other. The choice of this distance *d* also can be subject of changes to investigate the sound localization capabilities of microphone array from values of distance *d*.

Localization of the sound source can be done using the hypothesis that it should be found the point in the space where energy is the highest one.

From the microphone array structure presented on Fig.1 is can be determined the mathematical equations described the work of a linear microphone array.

If *M* is the number of microphones in an array, then the signal received at the microphone m, where m=1...M at the moment n is Eq. (1):

$$x_m(n) = h_m(n)s(n) + n_m(n) \quad (1)$$

where, $n_m(n)$ is additive noise and $h_m(n)$ is a room impulse response.

It is possible to assign *E(l)* in Eq. (3) as sound energy in point *l* and *p(l)* in Eq. (2) as an arbitrary space location. These two microphone array characteristics are presented with the following equations:

$$\mathrm{p}^*(\mathrm{l}) = \arg\max\{\mathrm{E}(\mathrm{l})\} \quad (2)$$

$$\mathrm{E}(\mathrm{l}) = \left|\sum_{m=1}^{M} x_m(n - t_m)\right|^2 \quad (3)$$

where, t_m is the time that takes sound to travel from the signal source to the microphone *m*.

When use microphone array with only two microphones, it is applicable one-step TDOA. The main advantage of that method is the fact that calculation can be done in one step and there are no need of approximations and loss of information. The disadvantage of this method is in the complicated calculations.

When the array has more than two microphones, the conventional TDOA sound source localization is in two steps. In first step calculation is done for each pair of the microphones. The weighting function is added to improve the quality in process of estimation for localization. In the second step, results from the first one are taken and summed in order to obtain the final localization of the sound source. The main advantage of that method is that this well-known domain has a lot of weight functions, which can be used. This is the reason for easier calculations on the second step.

The basic disadvantages are the approximations for the average time of delay in arrival, which leads to loss of information.

B. Mathematical model of sound localization method using classical beamforming

Classical beamforming is data-independent technique. Some of the fixed beamforming methods are Filter-and-Sum beamformer design, superdirective beamforming, matched filtering as white noise gain maximization, Delay-and-Sum beamforming and Delay-and- Subtract algorithm.

The simplest configuration of all microphone arrays is Filter-and-Sum beamformer design. It is so named because the sensors' inputs are first delayed by time τ_n in time domain on Eq. (4), and then summed to give a single array output.

$$\tau_n = \frac{(n-1)d\cos\phi}{c} \quad (4)$$

Usually, each channel is given equal amplitude weighting in the summation. That means the directivity pattern demonstrates unity gain in the desired direction. This leads to the complex channel weights on Eq. (5):

$$\omega_n(f) = \frac{1}{N} e^{j\frac{-2\pi f}{c}(n-1)d\cos\phi} \quad (5)$$

The delay-sum beamformer belongs to a more general class known as filter sum beamformers, in which both the amplitude and phase weights are frequency dependent. In practice, most beamformers are a class of filter-sum beamformer.

One simple method of covering broadband signals is to implement the array as a series of sub-arrays, which are by themselves linear arrays with uniform spacing. These sub-arrays are designed to give desired response characteristics for a given frequency range [8].

Super directive beamforming method has the highest directivity (DI) but very poor white noise gain (WNG). This is the reason for the problems with robustness (sensor noise) [9].

C. Mathematical model of sound localization method using adaptive (hybrid) beamforming

In previous point [II.B] were described most popular fixed beamforming techniques and now we continue with adaptive beamforming, which is data dependent.

We will focus in LSMV beamforming, Frost beamforming and Generalized sidelobe canceller.

Linearly Constrained Minimum Variance (LCMV) beamforming. The main purpose is to minimize noise output power, while maintaining a chosen response in a given look direction which is closer to the operations of superdirective array or delay-and-sum, but now noise field is unknown.

Frost beamforming is the adaptive version of LCMV. It is a long standing (1970's) beamforming technique that reduces noise whilst maintaining the look direction signal. It is an adaptive system that aims to minimal noise energy by adjusting the beamformer filter weights. Although it was investigated and published a while ago, a great deal of current work builds upon and references this work, so it's well worth looking at.

Generalized sidelobe cancerler (GSC) is alternative adaptive filter formulation of the LCMV-problem: constrained optimisation is reformulated as a constraint pre-processing, followed by an unconstrained optimisation, leading to a simpler adaptation scheme [10]. GSC consists of three parts: fixed beamformer, which satisfying constraints but not yet minimum variance, creating `speech reference'; blocking

matrix, placing spatial nulls in the direction of the speech source (at sampling frequencies), creating `noise references' and multi-channel adaptive filter (linear combiner) LSM.
GSC is an algorithm shown on Fig.2 is with complex mathematics calculations in Eq. (6).

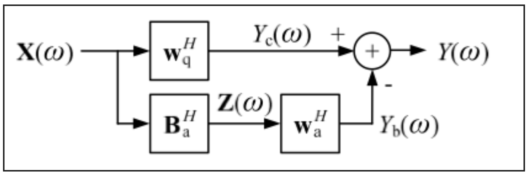

Fig. 2. Microphone array processing

The main problem is the fact of impossibility to reduce noise from look-direction.

$$\mathrm{Y} = \left(w_q - B_a w_a\right)^H X \qquad (6)$$

In this situation reverberation effects cause signal leakage in noise reference adaptive filter should only be updated when no speech is present.

III. MATLAB SIMULATION OF TIME OF DELAY OF ARRIVAL (TDOA) METHOD FOR SOUND LOCALIZATION

The simulation model based on the mathematical description (equations 1-5) is prepared and is shown on Fig. 3. It is a special case of Fig.1 with constant distance d between each element. We used a real microphone to record an audio file and apply this sound record as test signal simulating the sound signal from first microphone in microphone array (assuming relative zero time delay for this microphone).

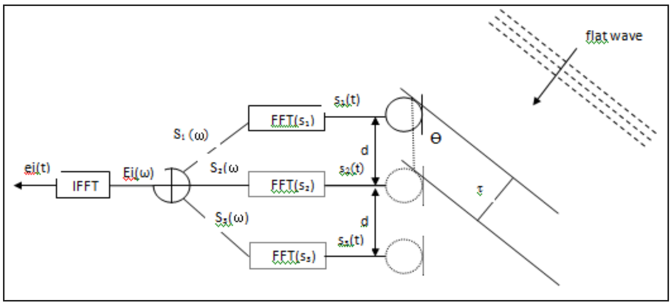

Fig. 3. The simulation model based on the mathematical description (equations 1-5)

Then we simulated microphone array with following initial conditions, which are input parameters in simulation of microphone array example:

- M = 3 microphones (1 real and 2 simulated)
- c = 330 m/s – propagation speed of voice
- d = 0.03 m – distance between microphones
- θ = 30 deg - angel from which sound reaches microphone.
- Fsig = 4000 Hz –signal frequency

The simulation model is arranged as an appropriate algorithm shown in Fig. 4 and applying the mathematical description (equations 1-5).

In block 1 we received sound signal in first microphone of array. In block 2 we set microphone array parameters to simulate it. After that we define time delay of arrival of the signals for the different microphones in block 3. We apply Fourier transform for signals in different microphones in block 4 to make possible sum them in block 5.

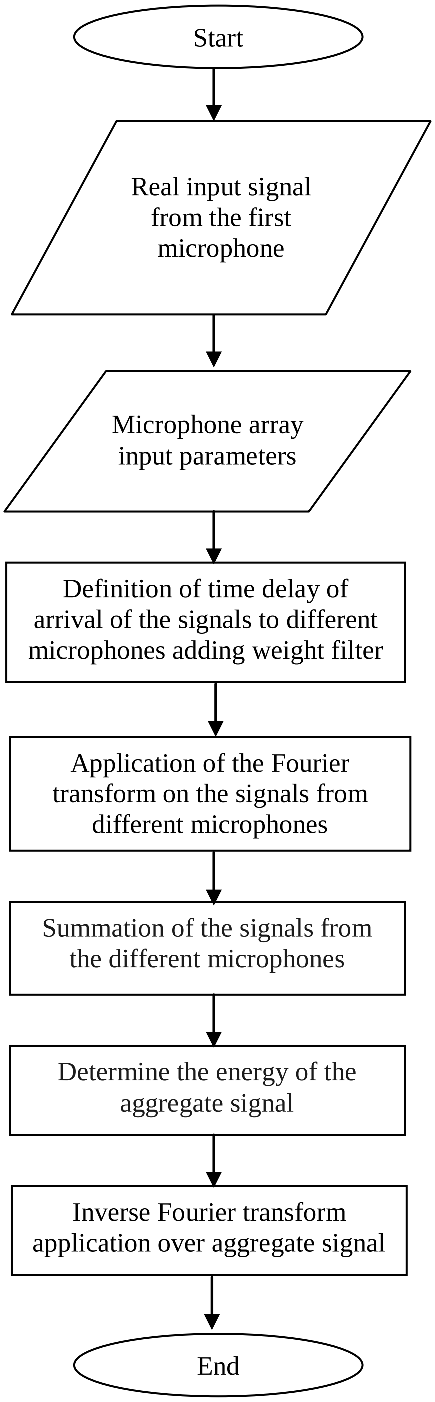

Fig. 4. The simulation model is arranged as an appropriate algorithm

In the next block we determine the energy of the aggregate signal to apply the hypothesis for localization in point of the space where it is the highest one. Finally in last block we apply Inverse FFT over the aggregate signal to find time of delay and following the Eq. (7) to calculate the angel of arrival.

The results from execution of the described algorithm above are used to calculate time delay in each sound waves arrival situation and after that the desired direction of arrival indicating sound source localization.

The results from simulation as values of the parameters in the following equations to calculate time of delay of the signal to each microphone:

$$\tau = abs\left|\frac{d\cos(\theta)t_{sig}}{c}\right|. \tag{7}$$

Sampling period can be calculated from sampling rate:

$$T_{sig} = \frac{1}{f_{sig}}. \tag{8}$$

Number of bits delay on the graphic is:

$$N = \frac{\tau}{T_{sig}}. \tag{9}$$

The experiment was carried out in far-field zone therefore we can make the assumption that angle for all microphones in the array is the same. The algorithm of Fig.3 for Time Delay Of Arrival and simulation of microphone array was implemented by Matlab.

First we get a real signal from microphone and record it in audio file. Then we load this file and input the array parameters. We calculate the time of delay and using it we simulated the second and third signal from microphone array. Using a Fast Fourier Transformation (FFT) we receive spectrum of all signals. Following the physical diagram next step is to summarize them in one signal and find the point where the amplitude is maximal. After that using the Inverse FFT we transfer the received signal spectrum in time domain.

IV. EXPERIMENTAL RESULTS

The simulation model arranged as an appropriate algorithm shown in Fig. 3 is used in experiment carried out in this article. The experiments are performed and experimental results are presented below.

On Fig.5 are shown three sound signals, received in microphone array. The first one is real sound signal from microphone and the other two below are simulated with proper delay, described in previous section.

On the next Fig. 6 are presented three signals on one graphic to be clear shown their delay. Graphic in blue shows the original signal, received from microphone, green – shows the simulated signal with delay τ and with red – simulated signal with delay 2τ.

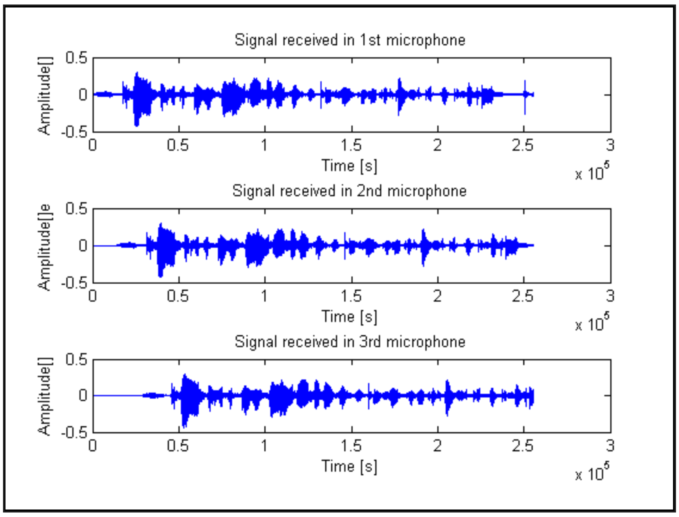

Fig. 5. Sound signals received from microphone array

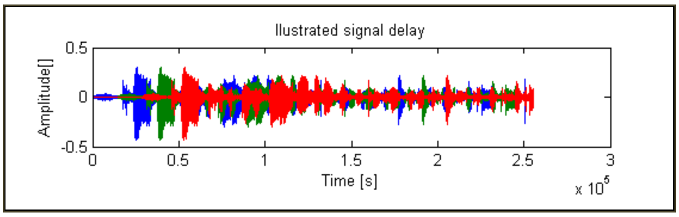

Fig. 6. Sound signals received from microphone array on one graphic

According to the algorithm in Fig. 7 and in Fig. 8 are shown spectra of the individual and aggregate microphone signals after Fast Fourier Transformation and their Inverse Fast Fourier Transformation (Fig. 9), respectively.

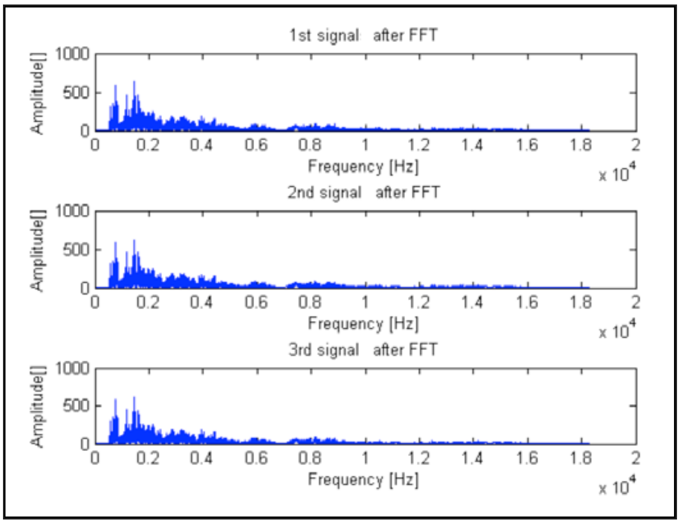

Fig.7. Sound signals spectrum applying Fast Fourier Transformation

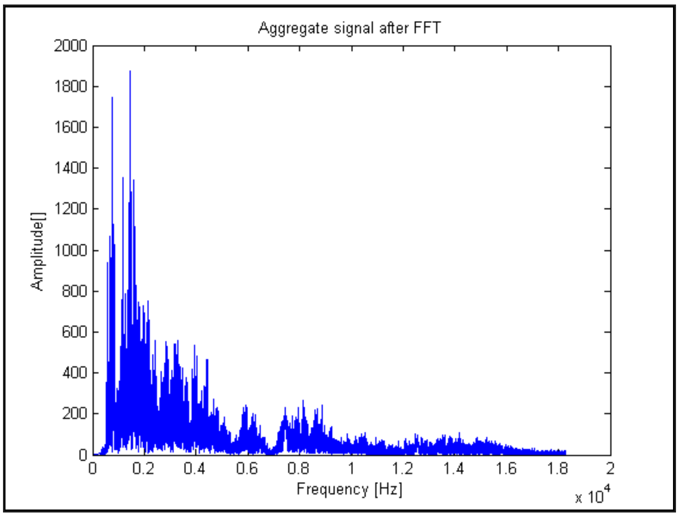

Fig. 8. Spectrum of aggregate signal after Fast Fourier Transformation

We can see that the maximum amplitude of the signal is observed between 2-4 kHz. Also it is possible to conclude that in the spectral domain there isn't exist an important information, which can be used to sound direction of arrival calculation applying the used in this and investigated in this article time of delay of arrival (TDOA) method.

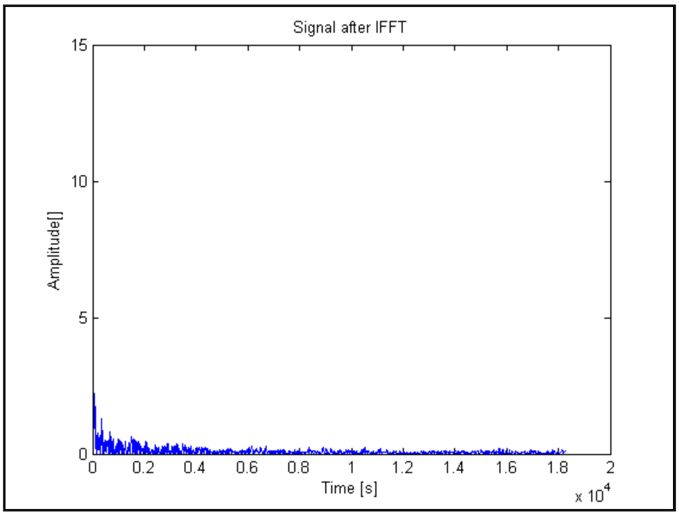

Fig. 9. Inverse Fast Fourier Transformation from microphone signals spectrum

The results from experiments with algorithm are used to calculate time delay in each sound waves arrival situation and after that the desired direction of arrival indicating sound source localization.

These calculations are explained below for a chosen example of simulations.

Following the MatLab workspace we use discrete values

$$N = 1.389.10^{4} \ ,$$

which will be different for the real micrphone array and setted sample frequency 44100 kHz.

Using the Eq. (7), we can calculate the Ts:

$$T_{sig} = \frac{1}{f_{sig}} = 2{,}268.10^{-5}\, s \ .$$

After that using the Eq. (8) we can calculate time of delay:

$$\tau = NT_{sig} = 0{,}315 s \ .$$

Finally appling the Eq. (6), we received:

$$\theta = arc\left(\frac{\tau.c}{f_{sig}\, d}\right) = 29{,}974 \deg \ .$$

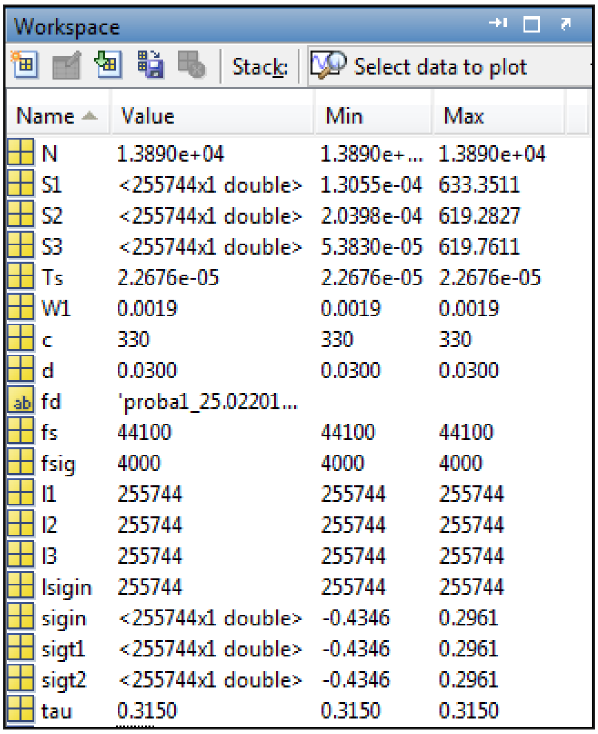

Fig. 10. Values in Matlab Workspace after execution of one simulation example to use in angle of sound arrival θ calculations

The presented values above are delivered from Matlab Workspace after execution of one simulation example (Fig. 10) and are used in angle of sound arrival θ calculations.

The values of calculate angle θ represent direction of sound waves arrival from source of the sound. This means that the sound localization is done properly.

V. Conclusion

In this paper was described microphone array simulation and its realization in Matlab framework. They are based on presented here mathematical equation as models of more important sound localization methods. After an analysis carried out it is chosen to simulate in Matlab environments Time Delay Of Arrival method. An block algorithm is prepared first to describe the necessary operations in Matlab simulation program, which is executed for some defined as concrete values microphone array parameters. The results from execution of Matlab simulation program are used to calculate time delay of sound waves arrival and from this calculated time delay determined direction of arrival indicating sound source localization. The results of these calculations show the ability of the proposed in this article algorithm and Matlab simulation program to properly determine direction of arrival of sound waves from sound source solving the problem of sound source localization, when applying the investigated in this article Time Delay Of Arrival method and algorithm.

ACKNOWLEDGEMENT

This paper was supported by Technical University – Sofia inner program to support PhD research projects under Contract 142 PD 0018-07: "Development of methods and tools to locate audio sources in information and communication networks".

REFERENCES

[1] J.M. Valin, J. Rouat, F. Michaud, "Enhanced Robot Audition Based on Microphone Array Source Separation with Post-Filter", Université de Sherbrooke, Sherbrooke (Quebec) CANADA, 2004.

[2] A. Bekiarski, S. Pleshkova, "Microphone array beamforming for mobile robot", CSECS'09 Proceedings of the 8th WSEAS International Conference on Circuits, systems, electronics, control & signal processing, 2009, p. 146-149.

[3] Powerful microphone for the Polycom Group and HDX Series , http://www.polycom.com .

[4] M. Okandan, E.P. Parker, N.A. Hall, K. Peterson, P. Resick, D. Serkland, "Ultrasensitive Directional Microphone Arrays for Military Operations in Urban Terrain", 2007.

[5] Y. Rui, D. Florencio, "New direct approaches to robust sound source localization", January, 2004.

[6] I. McCowan, Microphone Array, April 2001.

[7] I. McCowan, "Multi-channel sub-band speech recongnition", Speech Research Laboratory, RCSAVT, School of EESE, Queensland University of Technology, GPO Box 2434, Brisbane QLD 4001, Australia, Feb 2001.

[8] E. Fisher, B. Rafaely, Near-field spherical microphone array for speech, France, May 2010.

[9] S. Doclo, M. Moonen, "Superdirective beamforming robust against microphone mismatch", Katholieke Universiteit Leuven, Dept. of Electrical Engineering (ESAT-SCD), Kasteelpark Arenberg 10, 3001 Leuven, Belgium.

[10] J. McDonough, "Generalized Sidelobe Canceller", Spoken Language Systems Saarland University, Jan 2009.