SOUND SOURCE LOCATION METHOD

Michal Mandlik[1], Vladimír Brázda[2]

Summary: This paper deals with received acoustic signals on microphone array. In this paper the localization system based on a speaker speech processing and extraction of sibilant sequence is suggested. It is shown, that this method leads to good position estimation accuracy in indoor systems.

Key words: localization, sound, microphone array.

INTRODUCTION

Nowadays, acoustic systems are used not only for a pure speech processing but also frequently for a localization of sound sources. More principles are offered for Radio Frequency emitter positioning but those used for acoustic systems in indoor applications suffer typically from multiple signal reflections from obstacles. The sound source localization methods in general can be divided into Difference Time Arrival (5), Time Difference of Arrival and Interaural Time Difference (4) based methods. The measurement of time delay of two signals based on the evaluation of reciprocal correlation functions are used commonly for the localization in TDOA systems without the time synchronization between the signal source and receivers. Modern acoustic localization systems are used generally for military applications, for example in automatic systems like moving digital video camera or turning recording microphone around towards a speaker in a conference room.

1. POSITIONING SYSTEM

1.1 General Description

Sound source (e.g. speakers, accidental shot) position localization system consists of a few (three or more) microphones, distributed in a line or a square microphone array. Each microphone with a known position receives signals in the full frequency acoustic band, because human speech consists of syllables containing frequencies over 5 kHz. Received signals are amplified in an analog linear amplifier and then collected in the central station.

Time differences of arrivals of signals emitted by the individual sound sources received by the receiving stations are evaluated at the central station. The position of source is then estimated. An increase in the number of microphones increases the accuracy of source location estimation. All receiving microphones and amplifiers must have the same frequency characteristics. Their role is to receive signals and to amplify dynamic signals without any changes. The microphones of our system receive not only a direct signal from each the acoustic source present in the station range, but also reflected signals from walls, ceiling and

[1] Ing. Michal Mandlik, University of Pardubice, Faculty of Electrical Engineering and Informatics, Department of Electrical Engineering, Tel.: +420466037109 E-mail: michal.mandlik@student.upce.cz
[2] Ing. Vladimír Brázda University of Pardubice, Faculty of Electrical Engineering and Informatics, Department of Electrical Engineering, Tel.: +420466037109 E-mail: vladimir.brazda@student.upce.cz

floor in the room etc. (1), (2). In the case of special outdoor application (military), the signals are reflected from buildings and terrain. These reflected signals differ from the direct signal in the amplitude and the delay.

All direct and reflected signals give the total signal $s[n]$ at the n^{th} microphone, which can be expressed as:

$$y_n(k) = g_n * s(k) + v_n \qquad \text{for n = 1,2…N} \qquad (1)$$

Where

y_n is n^{th} received signal,
g_n is an impulse response from source to n^{th} channel,
$s(k)$ is an impulse transmitter signal,
v_n is an additive noise into n^{th} channel.

1.2 Microphone array

The microphone array used in our localization system consists of four capacitor microphones with a set of linear amplifiers. The microphone array is user-configurable into two basic forms – line and square. The line array with microphone spacing of about one meter is applicable to a sound source position in the range of approximately 20 meters. The square array with a side of one meter or shorter is suitable only to estimation of azimuth. The main advantage of microphone array in square configuration is that an estimation error remains the same in every direction. A direct estimation error can be computed using Dilution of Precision. The measurement chain contains of four microphones, low noise amplifiers, analog to digital converters and a signal processing unit (e. g. computer or laptop).

1.3 Acoustic signal source

A speech signal is used as a signal source for our localization system. The main goal of signal processing in a sound system is to detect significant markers in a speech of speaker and locate them in the signal. A speech contains of very short time slots (frames) with syllables.

Therefore a received signal can be divided into frames. A frame length should be short enough to consider speech as a stationary signal, but it should be long enough to estimate parameters. A speech organ has inertia typically 20 to 25 ms. An overlap around 10 ms can be used for a better detection. Frames can be selected by windows (e. g. a rectangular window or a hamming window). A received signal is influenced by a chosen window.

A speech can be divided to vibrant and voiceless syllables. A vibrant syllable has typically high energy, low frequency spectrum (less than 2 kHz) and a small number of zero crossings. A voiceless syllable is different, it has as low energy as a noise and spectrum of signal is high (around 5 kHz). These properties could be used for detection in each channel (8).

2. TDOA

There are varieties of ways in which a position can be derived from measurement of signals, and can be applied to any sound source localization. The most important measurement is propagation in time. In this paper a time difference of arrival (TDOA) method will be discussed (7).

The microphones "listen to" speech and measure a difference between each pair of arrivals. For example, if there are three microphones, two independent TDOA measurements can be made. Each TDOA defines a hyperbolic locus on which speaker has to lie. The intersection of the two hyperbolic loci will define the position of the speaker (1).

This method has typically two steps. The first one is a time delay estimation and the second one is a position estimation.

2.1 Time Delay Estimation

Having signal which is propagated through general space with noise, the received signal can be expressed by:

$$x_i(n) = g_i * s(n - \tau_i) + v_n \,, \tag{2}$$

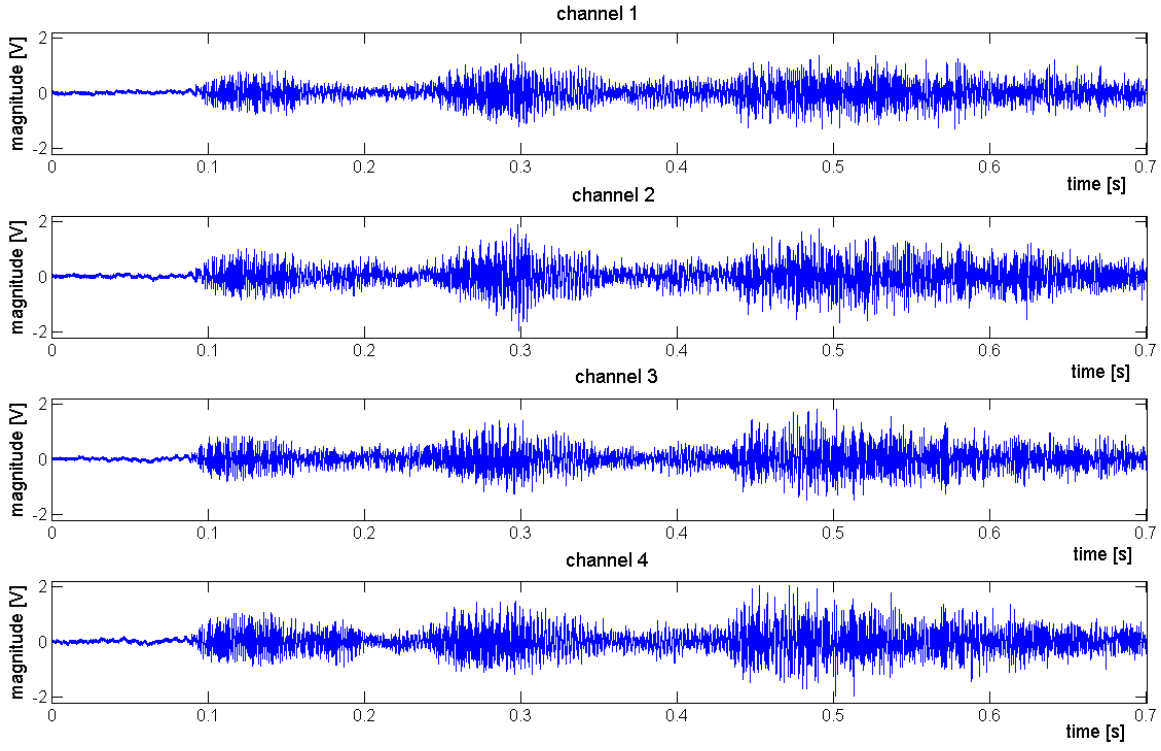
Where

τ_i is propagation time,

g_i is an impulse response between transmitter and an i^{th} microphone,

v_n is an additive noise.

Signals which were received in our case are shown in Fig. (1). Each receiver corresponds to its channel.



Source: author

Fig. 1 – received signals

For the time delay estimation between one pair of microphones the Generalized Cross Correlation function (GCC) can be used (1), (2), (5),

$$R_{ij}(\tau) = F^{-1}\{\Psi(f)X_i(f)X_j^*(f)\} \tag{3}$$

Where

$R_{ij}(\tau)$ is the Generalized Cross Correlation Function,

$X_i(f), X_j^*(f)$ are Fouriere transforms of received signals,

$\Psi(f)$ is the Fourier transform of filter.

PHAT (Phase Transform) filter could be used for better time estimation. PHAT filter can be written as follows.

$$\Psi(f) = \frac{1}{\left|X_i(f)X_j^*(f)\right|} \tag{4}$$

While $i = j$ R_{ij} is the autocorrelation function and while $i \neq j$ R_{ij} is the cross correlation function between x_i a x_j (1, 2, 3, 6). The peaks of the correlation are determined from their sample indexes which are noted as,

$$\alpha_{\alpha,c} = \arg max\left(R_{ij}(\tau)\right) \tag{5}$$

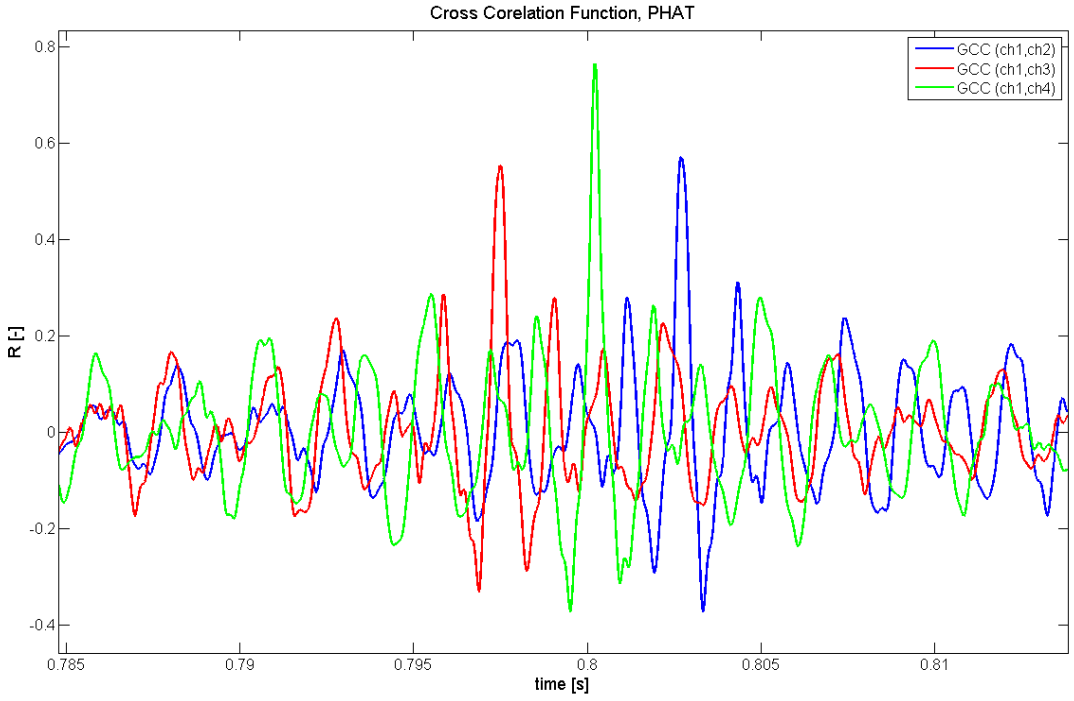
where the subscripts α, c stand for and auto and cross correlation respectively. The sample delay is

$$\tau_s = \alpha_\alpha - \alpha_c \tag{6}$$

and the time delay between microphones i and j, is obtained from

$$\tau_{ij} = \frac{\tau_s}{f_s} \tag{7}$$

where f_s is the sampling frequency for received signals. Figure (2) shows result GCC between channel 1 and channel 2.



Source: author

Fig. 2 – GCC with PHAT

2.2 Position Estimation

A general model for the three dimensional estimations of source using N microphones is expressed:

$$ct_i = \sqrt{(x_i - x_0) + (y_i - y_0) + (z_i - z_0)} \quad \text{for i = 1 to N} \tag{8}$$

Where

$[x_n, y_n, z_n]$ are coordinates of i^{th} receiving microphone,

$[x_o, y_o, z_o]$ are coordinates of the acoustic source,

c is velocity of sound,

t_i is a time between the transmitter and i^{th} receiver.

We can rewrite the equation

$$c(t_j - t_i) = |R_i T| - |R_j T| \quad (9)$$

Where

R_i are coordinates of i^{th} receiving microphone,

T are coordinates of i^{th} receiving microphone.

There are many methods for solving these nonlinear equations e.g. LS method, Newton method, Taylor-Series Method etc... The method which is used in this paper is based on brute force approach. The beginning of coordinate system is placed into the centre of explored space without losing generality. The space is divided into finite numbers of points. For each of this points, the distance from particular receiver is calculated (10), where n is number of receiver

$$\boldsymbol{X}_n = |\boldsymbol{A}_{ij} R_n| \qquad \text{For n = 1..N} \quad (10)$$

Where

N is the number of receivers

$\boldsymbol{A}$ is the matrix of grid,

R_n is n^{th} receiver,

$\boldsymbol{X}_n$ is the matrix of distance between n^{th} receiver and every index in matrix **A.**

Difference (11) is calculated for all combinations of matrices $\boldsymbol{X}_n$ and saved to $\boldsymbol{Y}_{ij}$, where i, j are each particular receivers i and j. These operations are very time consuming and for the given array configuration can be pre-computed and stored in a lookup table.

$$\boldsymbol{Y}_{ij} = \boldsymbol{X}_i - \boldsymbol{X}_j \quad (11)$$

Where:

$\boldsymbol{X}_i, \boldsymbol{X}_j$ are matrixes for i^{th} and j^{th} receiver,

$\boldsymbol{Y}_{ij}$ is the differential matrix for pair i^{th} and j^{th} receiver.

Then the pseudo-range delay between microphones i and j is subtracted from $\boldsymbol{Y}_{ij}$ (12).

$$\boldsymbol{Z}_{ij} = \boldsymbol{Y}_{ij} - \tau_{ij} * c \quad (12)$$

Where

τ_{ij} is the time delay between microphones i and j,

c is the speed of sound,

$\boldsymbol{Y}_{ij}$ is the differential matrix for pair i^{th} and j^{th} receiver,

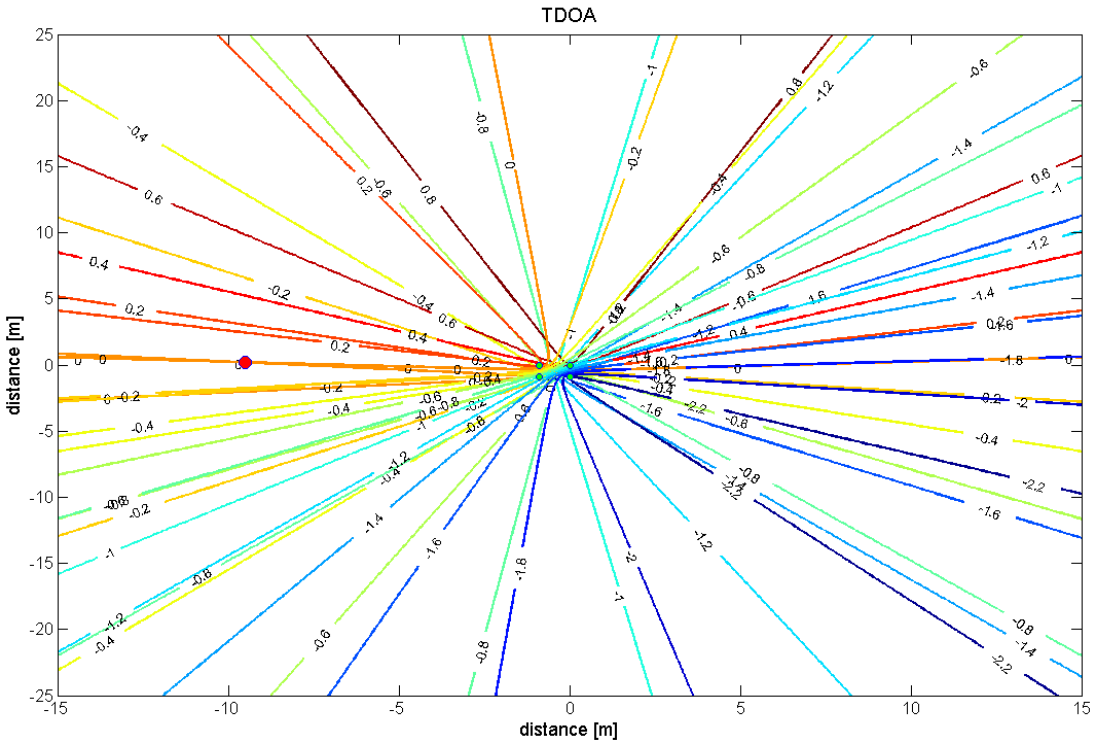
$\boldsymbol{Z}_{ij}$ is the final ij^{th} matrix.

The last step is a coalescence of all matrices. The transmitter lies on intersection of the zero value elements of particular matrices. The result is the point with the transmitter position estimation.

$$\boldsymbol{T}_e = min\left(\sum_{m=1}^{M} abs(Z_{ij})\right) \quad (13)$$

Figure of one case is shown in Fig. 3. This figure shows simulation of the brute force

approach with the square array receivers (green points) in a centre of room and the transmitter (red point). The intersection of the many hyperbolic loci defines the position of the transmitter.



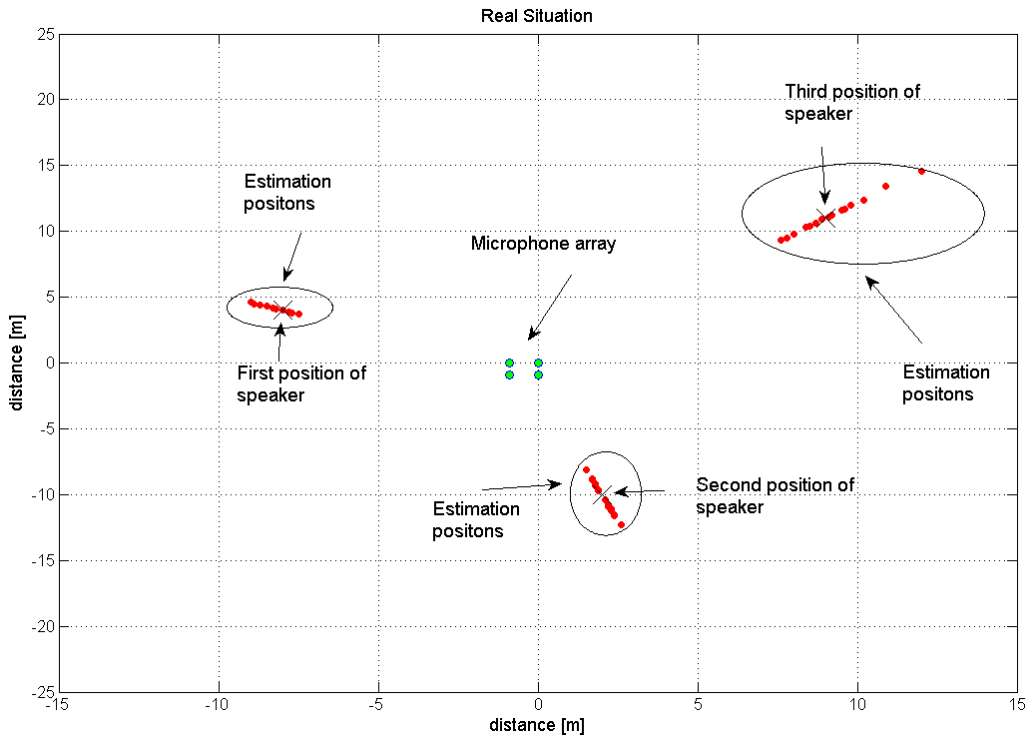
Source: author

Fig. 3 – TDOA

3. REAL SITUATION

Measurements were performed in the hall of size 50x30m. The microphone array was situated in the center of the hall. The array had a square shape and the distance between sides of microphones was 1m. The measurement chain for each channel contained: a capacitor microphone, a pre-amplifier and an oscilloscope with sample frequency 200ks/s. Signal processing was done offline.

In total, 20 measurements were done for every speaker position. Figure 4 shows three random positions and twenty measurements for each position. The green points mark the position of receivers. The blue cross shows the real position of the speaker and the red points are estimations of a position for each measurement.



Source: author

Fig. 4 – Real Situation

CONCLUSION

We have shown a case where the sound source localization is described by method which uses Lookup Table for estimation of time difference of direct signals arrival from transmitter to receivers. We used GCC with PHAT for time estimation and Lookup Table for estimation position. Those methods were tested in simulation also in real situation and results were approximately same for both cases.

ACKNOWLEDGMENT

The results presented in this article were supported by the specific research project of the IGA (SGFEI05/2011), University of Pardubice and the GACR grant P102/11/1376.

REFERENCES

(1) BRANDNSTEIN M. S.; SILVERMAN H., “A practical methodology for speech source localization with microphone arrays,” Computer Speech and Language, 11(2):91-126, April 1997

(2) BRANDNSTEIN, M., WARD D.. 2001. MicrophoneArrays: Signal Processing Techniques and Applications. New York: Springer, 2001. ISBN-3-540-41953-5

(3) BENESTY, Jacob, CHEN, Jingdong, HUANG, Yiteng. 2008. Microphone Array Signal Processing. Berlin: Springer, 2008. ISBN-978-3-540-78611-5.

(4) POURMOHAMMAD, A.; AHADI, S. M. TDE-ILD-based 2D half plane real time high accuracy sound source localization using only two microphones and source

counting. Electronics and Information Engineering (ICEIE), 2010 International Conference On [online]. 1-3 Aug. 2010, 1, [cit. 2011-11-29].

(5) JIAN, M; KOT, A.C.; ER, M. H. DOA Estimatino of Speech Source with Microphone arrays. Circuits and Systems, 1998. ISCAS '98. Proceedings of the 1998 IEEE International Symposium on [online]. 31 May-3 Jun 1998, 293 - 296 vol.5 , [cit. 2011-11-03].

(6) VALIN, J.-M, et al. Robust sound source localization using a microphone array on a mobile robot. Intelligent Robots and Systems, 2003. (IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on [online]. 27-31 Oct. 2003 , 1228 - 1233 vol.2 , [cit. 2011-11-15].

(7) NĚMEC, Z.; BEZOUŠEK, P. A TDOA system using received signal decomposition on delayed replicas. Radar Symposium, 2008 International . May 2008, p. 1 - 4.

(8) ZANNINI, C.M., et al. Improved TDOA disambiguation techniques for sound source localization in reverberant environments. Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium. June 2010, p. 2666 - 2669.

(9) PSUTKA, J. 1995. Komunikace s počítačem mluvenou řečí. Praha : Academia Praha, 1995. ISBN 80-200-0203-0.