Technical Sciences
Academy of Romania
www.jesi.astr.ro

Journal of Engineering Sciences and Innovation
Volume 3, Issue 1 / 2018, pp. 51-58

E. Electrical and Electronics Engineering

Received ***6 June 2017*** ***Accepted 10 January 2018***
Received in revised from ***14 September 2017***

Novel method for determining the position of speakers in a room using beamforming

DOBRE ROBERT-ALEXANDRU*, DUMITRU STANOMIR

Telecommunications Department, Politehnica University of Bucharest,
Iuliu Maniu Blvd. no. 1-3, sector 6, Bucharest, Romania

Abstract. People location detection has many applications in fields starting from healthcare, conference systems, antiterrorist protection and even home comfort through smart homes. Speaker position localization is also useful in situations like theatre plays. The current paper presents a speaker localization system that can easily be installed in any room thanks to its implementation using omnidirectional microphones uniform linear array. Since the microphones have small dimensions, the arrays can be easily hidden behind sound transparent walls. Three speaker position identification situations are analysed, and the signal processing solutions are indicated for each.

Keywords: beamforming, direction of arrival estimation, sensor array, speaker position detection.

1. Introduction

Although a popular application of people position detection is in meetings [1] and audioconferencing, it has also become increasingly used for law enforcement and intelligence activities [2] and in areas such as healthcare, comfort or smart rooms [3]. Beamforming [4], also known as spatial filtering, represents an efficient way, compared to radio or video-based methods, to implement a solution for speaker position localization since its hardware part is reduced to a sensor array and a digital signal processing system.
The paper is organized as follows: in Section 2 the proposed principle and system are thoroughly described. Section 3 contains the detailed signal processing methods, the Simulink model and the discussion of results while Section 4 concludes the paper.

* Correspondence address: rdobre@elcom.pub.ro

2. Description of the proposed solution

A person's position in a room can be completely characterized by two coordinates. Because the position of a speaking person is to be determined, estimating the direction of arrival (DOA) [5] of the speech relative to two axes can be used to solve this problem. Two uniform linear microphone arrays with 15 elements spaced at 25 cm measuring a total 3.75 m were considered as placed on two orthogonal walls of one room. The arrays are considered built with omnidirectional microphones and are arranged so they are orthogonal on the height of the room to properly estimate directions that are parallel with the floor and ceiling. A typical situation is presented in Fig. 1.

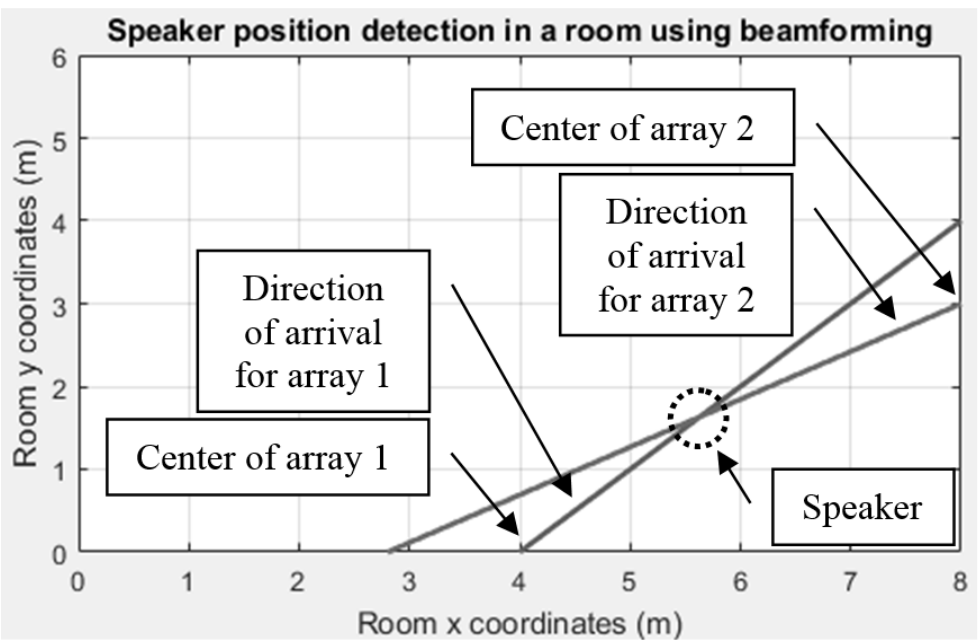


Fig. 1. The principle of detecting the position of a speaking person by determining the direction of arrival of the speech signal using two microphone arrays.

The height of the speaker is not considered important in this application so only the DOA's azimuth will be determined. Because the arrays are placed on walls, the possible values for azimuth are contained within the (-90°, 90°) interval. The directivity characteristic of the considered microphone arrays is presented in Fig. 2 for incident signal frequency equal to 563 Hz (the average value of the first formant for speech). It can be observed that the main lobe's width is equal to 8° if

the azimuth values where the directivity drops 3 dB are considered and around 14° if a 10 dB drop is considered (notable attenuation). Therefore, sources with a deviation larger than 10° from the main lobe's direction can be ignored.
The signals collected by the 15 microphones feed 19 time delay beamformers, each one steered to a particular angle (-90°, -80° … 80°, 90°) in order to cover the whole possible azimuth space available in this application.

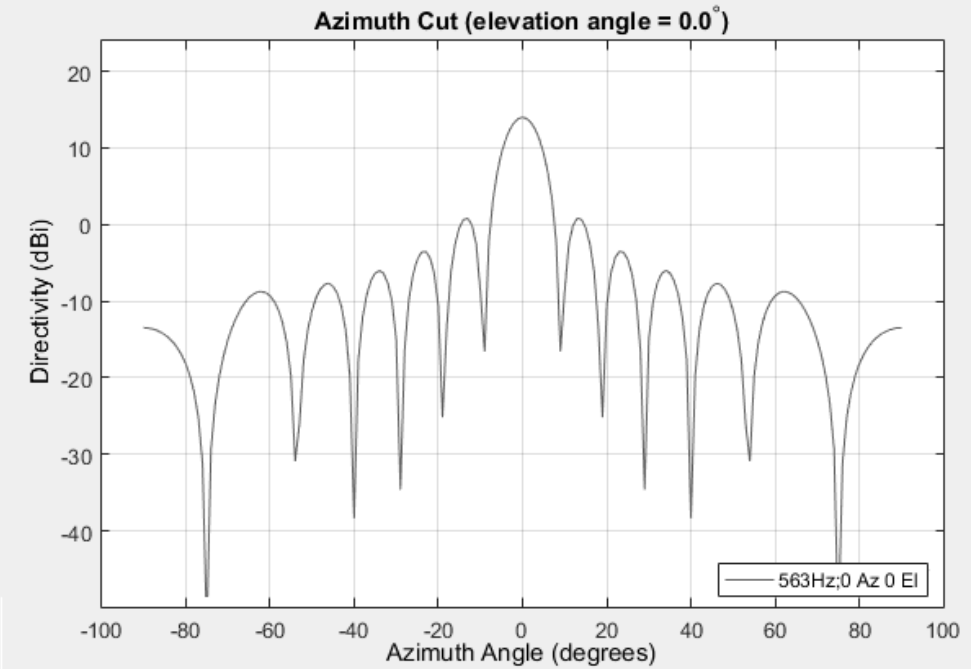


Fig. 2. The directivity characteristic for the considered uniform linear microphone array.

We must note that, in this study, the room was considered to have perfectly sound absorbing walls so reflections [6] were not taken into consideration.

3. Results

The system discussed above was implemented using Simulink. The project for one array is illustrated in Fig. 3. There are three situations that can be taken into consideration: there is only one speaking person in the room, there are multiple speakers in the room but they are not placed along the same DOA for any of the two arrays (no line can be drawn so that it crosses the center of one array and the position of two or more speakers) and the situation in which two or more persons are placed along the same DOA for one array.

The detection of the speaker's position in the first situation is done using only the energy of the received signals. The energy is calculated for the signals at the output of the beamformers of each array and the signals with maximum energy are identified. Each signal corresponds to a DOA, so two directions, one for each array, are determined. At their intersection the position of the speaker will be found. If the speaker is placed exactly on the border between the 10° spatial section discussed above, the signals at the output of two beamformers will have the same energy, so the localization will present a 5° position uncertainty will be present, being also the maximum uncertainty for this proposed solution. This limitation comes from the microphone arrays' structure – the arrays' number of elements is limited by the spatial dimensions since they must fit in rooms.

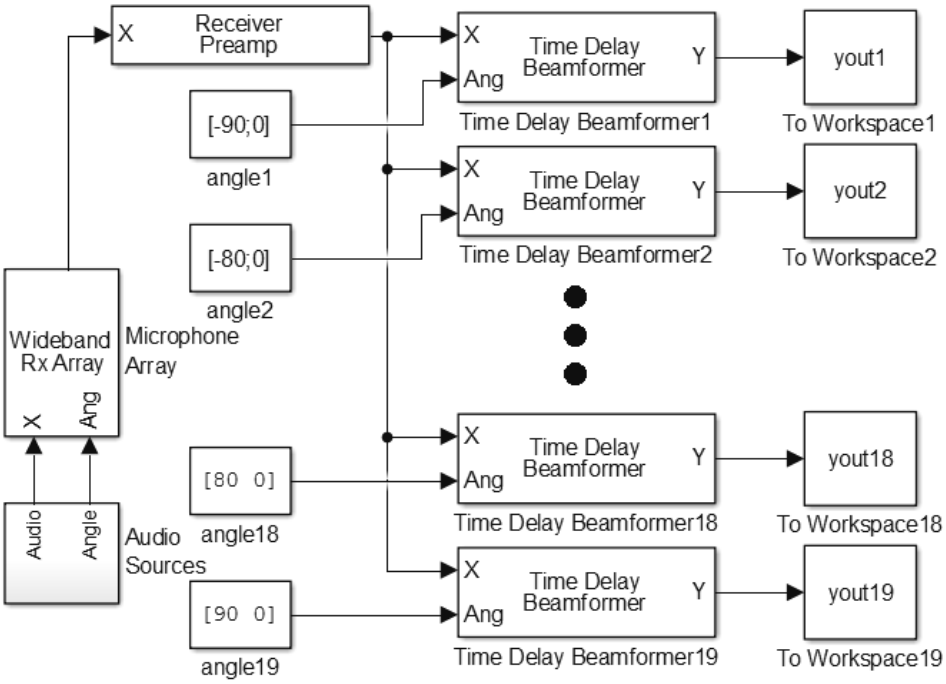


Fig. 3. The Simulink model for the signal processing system at one microphone array.

The second situation [7] is: two persons are speaking at the same time considering that both speakers determine similar loudness at the locations of the arrays. If a speaker is placed along the -10° DOA relative to one array and the other is placed along the 50° DOA, in the considered conditions, the signals at the output of the beamformers will be the ones presented in Fig. 4. The plotted amplitude range is from -0.5 to 0.5. The degrees indicate the steering angles of the beamformers.

At first the signal energy is computed for all the 19 signals. The normalized results for the example are presented in Fig. 5. Only signals with energies larger that a threshold are taken into consideration further. If the threshold is set to 0.6, 5 signals are qualified for further processing. A correlation coefficient is calculated between each of these signals to help determine how many different signals are present. This coefficient represents the maximum value of the cross-correlation function between each of the qualified signals. Before computing the cross-correlation function, the signals are normalized to one another, so their maximum values will be equal. The coefficients are then normalized to the value of the greatest one. These correlation coefficients between the signal coming from the −10° DOA and the other signals are presented in Fig. 6. It can be observed that the signal coming from −10° direction has similar energies with the neighboring signals, but also is greatly correlated with them meaning they are versions of the same signal, so they come from the same speaker.

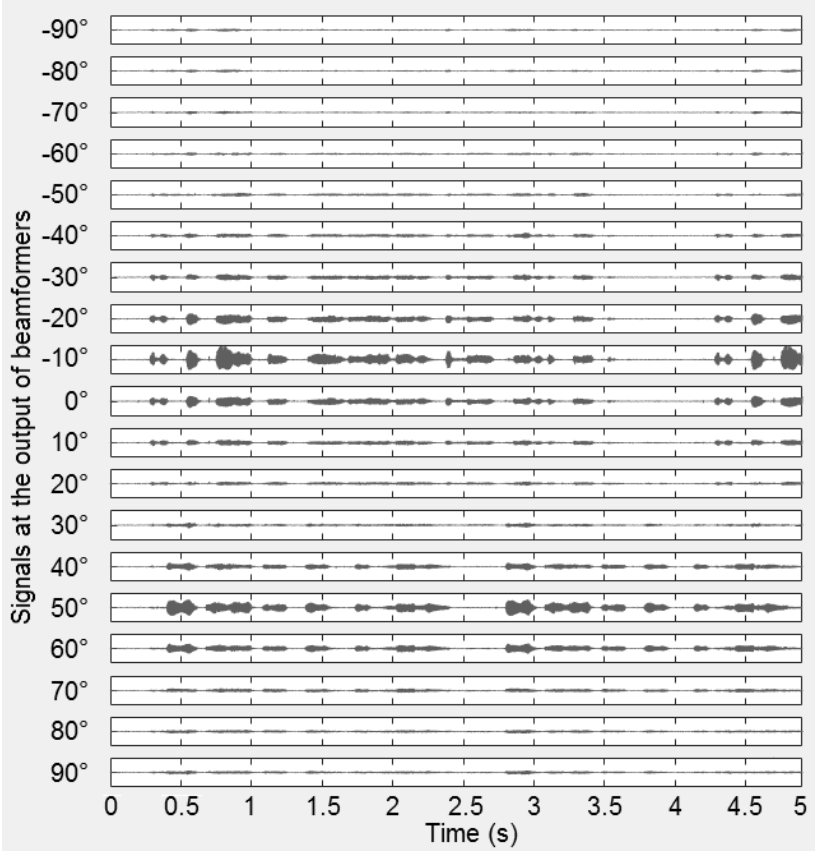


Fig. 4. The signals at the output of the 19 beamformers in a situation in which two persons peak simultaneously. It can be observed that the two signals of interest have -10° and 50° directions of arrival.

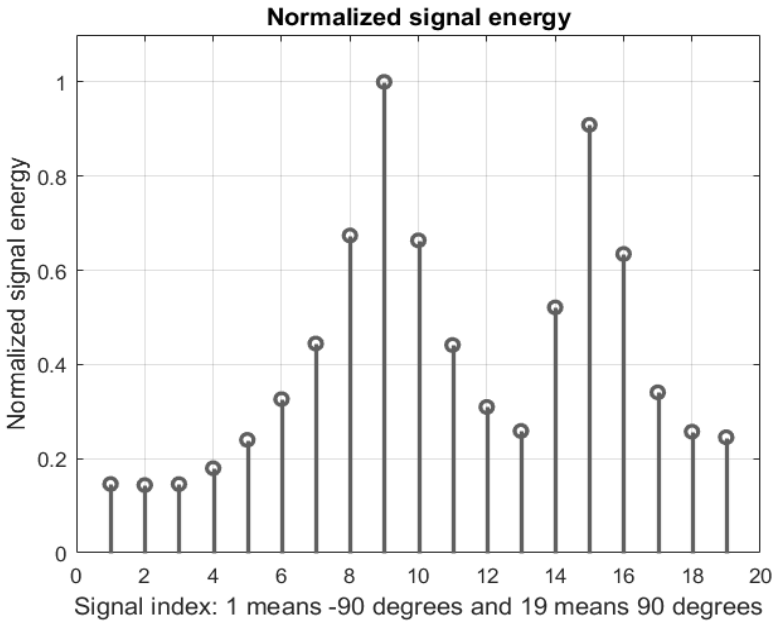
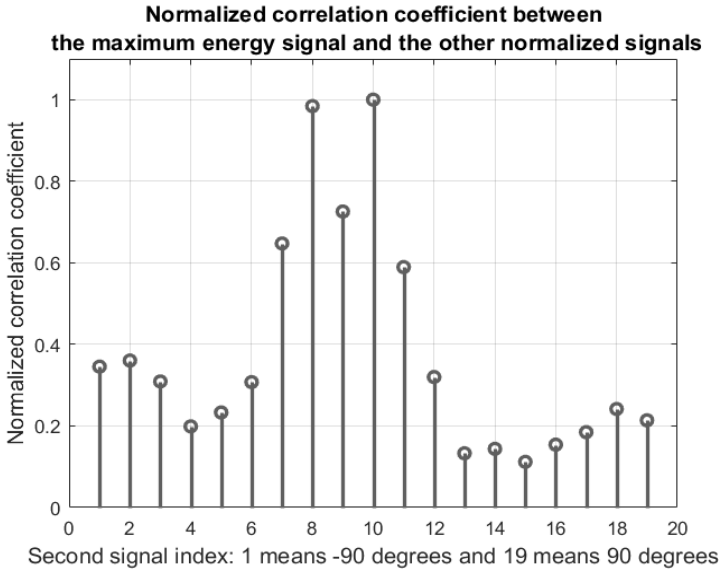


Fig. 5. The normalized energy for the signals presented in Fig. 4.



\Fig. 6. The correlation coefficient between the largest energy signal and the other signals.

Furthermore, it can be observed that the correlation coefficient between the signal coming from −10° direction and the one coming from 50° direction is very low, while the energy of the latter signal is comparable to the first. This helps concluding that 2 speakers say something at the same time, and they are placed along the −10° and 50° directions relative to the first array. The same analysis will be made by the second array and another two directions will be determined. Another similar correlation analysis between the signals for whom the DOA was already determined can be used to tell how to combine in pairs the four directions of arrival (two for each array). For example, if the second array similarly determined the directions of arrival being −50° and 80° the values −10°, 50°, −50° and 80° must be combined in pairs. If the correlation coefficient between the signal coming from the −10° direction of the first array and the signal coming from the 80° direction of the second array is large and the other coefficient between the same first signal and the one coming from −50° direction of the second array is smaller than the first coefficient, it will be concluded that the pairs are (−10°, 80°) for the one speaker and (50°, 80°) for the other. Using this information, the position of each speaker is determined as the intersections of the paired directions.

In the third situation, because the speakers are considered aligned with the same DOA for an array [8], only one signal will be dominant in terms of energy. So the signal processing explained above will help retrieve only one DOA. For the other array a similar situation to the one presented above will happen and two directions of arrival will be determined. The latter directions will be both paired with the direction given by the first array. For example, if both speakers are aligned with the 0° direction for the first array, and it is determined by the second array that relative to it they are placed along 40° and −40° directions, the final results will be (0°, 40°) for a speaker and (0°, −40°) for the other speaker.

4. Conclusions

The principle of localizing speakers in a room using beamforming was presented. A Simulink model for a microphone array part of a speaker localization system was proposed along with the signal processing methods that need to be applied to the signals at the output of the beamformers in order to correctly determine the position of the speakers in all possible situations. The maximum uncertainty at direction determination is 5° for each of the two directions. The system was simulated, and results were presented.

A limitation of the proposed system is given by the energy analysis. If simultaneous speakers are to be detected, their loudness must be comparable to each other. The method does not sense a difference between speech signals and other audio signals so if in a room is placed a turned-on radio, television or other similar sound sources, their position will also be detected.

References

[1] Gatica-Perez, D., Lathoud, G., Odobez, J. M., McCowan, I., *Audiovisual Probabilistic Tracking of Multiple Speakers in Meetings*, IEEE Transactions on Audio, Speech, and Language Processing, **15**, 2007, p. 601-616.
[2] Rose, P., Forensic Speaker Identification, CRC Press, 2003.
[3] Busso C. et al., Smart room: participant and speaker localization and identification, Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005, p. ii/1117-ii/1120
[4] Van Veen, B. D., Buckley, K. M.: Beamforming, A versatile approach to spatial filtering, IEEE assp magazine, **5**, 1988, p. 4-24.
[5] Huang, G., Chen, J., Benesty, J., *Direction-of-arrival estimation of passive acoustic sources in reverberant environments based on the Householder transformation*, J. Acoust. Soc. Am., **138**, 2015, p. 3053-3060.
[6] Paleologu, C., Benesty, J., Ciochina, S., *A Variable Step-Size Affine Projection Algorithm Designed for Acoustic Echo Cancellation*, IEEE Transactions on Audio, Speech, and Language Processing, **16**, 2008, p. 1466-1478.
[7] Souden, M., Affes, S., Benesty, J., *A two-stage approach to estimate the angles of arrival and the angular spreads of locally scattered sources*, IEEE Trans. Signal Processing, **56**, 2008, p. 1968-1983.
[8] Ajder, T., Kozintsev, I., Lienhart, R., Vetterli, M., *Acoustic source localization in distributed sensor networks, Proceedings of Asilomar Conference on Signals*, Systems, and Computers, 2004, p. 1328-1332.