Three Microphones Embedded System for Single Unknown Sound Source Localization

Flaviu Ilie Bob
Technical University of Cluj-Napoca
The Faculty of Electronics, Telecommunications and Information Technology
Baritiu Street 26-28, 400027, Room B7, Cluj-Napoca, Romania
ilie.bob@bel.utcluj.ro

***Abstract* — This paper presents an embedded system for localizing a sound source in a two-dimensional plane. It uses an ensemble of three microphones placed in a straight line, and a digital signal processor for capturing and processing the signals. The controller is equipped on a development board, along with a display which shows the calculated results. The signal processing consists of a slightly modified Cross Correlation method used to obtain the delays of arrival of the three microphone signals. To obtain the distance and angle, inversed mathematical equations were used, determined in Matlab from the geometric direct equations. The obtained results show a minimal error for short distances and the error increases when the distance between the sound source and the microphones gets bigger. A large distance also shows some flaws in obtaining the right time delays, thus some erroneous results can be observed. This is caused by the limited power of the speaker used for tests, but also by the environmental noise and the reverberations. The system could be improved by using a faster approach for extracting the time delay information from the signals.**

***Keywords* — embedded, microphone, sound localization, sound ranging, sound source localization, time delay of arrival, general cross correlation.**

I. Introduction

A. History, present and future

Sound localization has been first reported during the World War I, as a system that captured the sounds emitted by the enemy artillery as it fired, and sent them to a head quarter where the calculations were done. Microphones were spread over the battle field, thus the system was not easily portable. It was also easy to be affected in case of communication lines interruption. The army has not lost its interest into this topic, continuing even nowadays to develop different sound localization systems.

Teleconferencing systems are the most known sound localization devices to the general public. The person who speaks is detected and the camera is automatically pointed towards him.

Maybe, the greatest efforts in the sound localization field have been done in robotics, trying to teach robots having a maximal manlike behavior or to do special tasks concerning sound localization [1].

Next step to be done by sound localization is to be implemented in the surveillance and security domain, as these methods are less invasive than video surveillance, and they are effective even in low or no visibility conditions.

Human prosthesis is also very likely to use sound localization consequent with further development of the technology.

The current state of research mostly provides systems that are able to localize only the direction of the sound source. The few systems that are able to determine also the distance of the source require complex computing power or complex structures of statistic data for prediction. Portability and simplicity is not their strong point.

B. Problem statement and attributes

A sound source has some characteristics that influence the localization. Directivity and constancy are two of them, but own source noise is not to be neglected either. Sound emitted from the source also has parameters like frequency and power that can bring prejudice to the localization.

Presuming the source and sound are viable, the environment is the next factor. It also has humidity, temperature and texture as influencing attributes. Noise present in the environment is also susceptible of perturbing the sound through ambient noise, other sound sources and reverberations.

Sound transducers are also perturbing the signal, as there are no two identical transducers. But there are also many types to choose from, that have different directional and bandwidth characteristics. Components that amplify or repeat the signal in order to get acquisitioned by an analog to digital converter also insert noise in the system. In digital appear the conversion errors or mathematical round errors.

Computational power is required by most of the sound localization systems, as they are of high complexity. Databases containing pre-recorded samples or environment geometric attributes, models or propagation characteristics are the ones that load the memory, and the processors are kept busy by the advanced calculations that have to be done.

Along with complexity comes the size, as the main used processing unit is the PC. Also wires, advanced sensors, or other related equipment increase the volume of the systems.

C. System outline

In the following, the system desired objectives are to be defined. In matter of size, the system will be compact and portable, which also means a low power consumption, to be able to power it with batteries. Most suitable for this is an embedded system.

Movement of the sensors in the process of localization will not be possible, and the localization will be based only on the received sound. Databases do not have a place in a low power consumption system. Calculations thus must be done, so the processor must be able to have signal processor characteristics.

For best performance in terms of mobility and simplicity, the number of sensors must be reduced to minimum. For direction determination, two microphones are enough, but for determining the distance, at least three microphones are required. In the interest of minimalism, only three sensors will be used.

After the historical overview, problem statement and system outline presented in the first introduction section, the chosen method and its software implementation will be presented in the second section. A hardware design presentation is exposed in the third section, followed by the experimental setup described in the fourth section. Experimental results follow closely in the immediately following section, and in the end the conclusions are drawn.

II. Method and Firmware

A. Principle

Speed of sound plays an important role in this principle. Defined by it, a sound emitted from a source travels a certain distance to the sound receptors. Different positioned microphones mean different distances for the sound to travel. Differences in terms of distance can be measured using the time of arrival of the sound. Based on these delays, implicitly on the differences of distance, the position can be calculated using two differences.

Considering a linear setup of the three microphones, with equal distances between the two pairs, delays differ depending on the position of the source regarding the microphone setup.

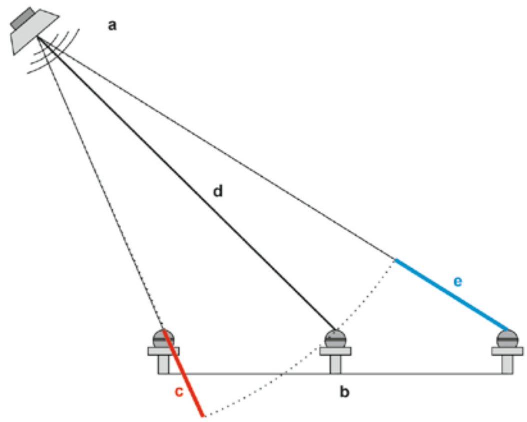

Figure 1. Microphone setup: a) Sound source b) Measuring device c) Difference between Left and Center microphones d) Distance form source to device e) Difference between Center and Right microphones

If the distance between the source of sound and the microphone setup remains constant, and only the angle changes, the ratio of the delays changes. Similar, if the distance changes and the angle remains constant, the sum of the delays changes.

Using these observations, based on the ratio and sum of the delays, the exact position of the source can be calculated. The method is thus based on computing time delays of arrival. For determining the time differences, the General Cross Correlation (GCC) is the most reliable method.

B. Algorithm and firmware

Implementing the cross correlation algorithm on the signal processor requires some changes, as the memory of the processor is limited. As a result of this limitation, from the result of the correlation only the maximum is retained, its position in the resulting vector and a counter for the number of appearances of it. The result vector of the cross correlation is not saved in RAM to preserve it for the acquired signal.

The processor performs the signal acquisition using 3 channel 10 bit simultaneous conversions at almost 1.1 MHz sample rate. The conversions are made in the auto mode, with DMA data transfer through a 64 Kbytes buffer. The voltage reference is VDD, the same voltage level of 3.3 V that is divided equally to create the continuous component of the microphone signal. The 10 bit recorded signal has values between 0 and 1023.

The detection of the sound is done by monitoring the inputs, checking the 64 Kbytes buffer delivered to the processor core through DMA. This consists of comparing each sample to a predefined threshold that has a value of 511±164. When this level is exceeded, the recording of the signal is triggered. The threshold value is a theoretical one, calculated based on the microphone amplifier circuit schematic.

The recording buffer is only 2 kilo samples large on each channel. At the sampling rate of almost 1.1 MHz, the length of the recorded signal is about 7.8 milliseconds. Even for this short recording, the processor working at 40 Mega Instructions per Second (MIPS) is taking 29 seconds to process, due to the complexity of the firmware, especially the cross correlation function.

After recording, the signal suffers a simple filtering for eliminating the continuous component created for the possibility of bipolarity. So the mean value on each channel is calculated and subtracted from each sample. The 10 bit recorded signal now has values between [-512, +512] instead of [0, 1023] when recorded.

After this, the cross correlation is calculated, between the center and the left channel, and between the center and the right channel. The correlation function used is not the one included in the Microchip Libraries, but a modified one, due to RAM memory limitations.

The results of the two cross correlation functions, annotated *Samples_R* and *Samples_L*, are measured in ADC samples. To obtain the measured value in microseconds, multiplication by 3.9 is performed. The factor 3.9 us represents the actual time in between two consequent samples of the same ADC channel. Then the speed of sound is taken into consideration to obtain a meter measurable value. It is all divided by to obtain a result measured in meters.

$$DR = \frac{Samples\_R \cdot 3.9 \cdot 346.18}{10^6}. \quad (1)$$

$$DL = \frac{Samples\_L \cdot 3.9 \cdot 346.18}{10^6}. \quad (2)$$

After having the two differences measurable in meters, two intermediary variables are used: the sum and the ratio of the two differences. The sum is calculated in (3).

$$S = DL + DR. \quad (3)$$

To calculate the ratio, the next step is to establish the *orientation* of the sound source, meaning determining in which half-plane is located the sound source. This is realized by comparing the values of *DL* and *DR*. If they are equal, the source is located right ahead. If *DR* is greater, then the source is located in the left half-plane, otherwise it is located in the right half-plane.

The ratio can be calculated depending on the *orientation*, as follows. If the sound source is located in the left half-plane (4) is used and (5) if not.

$$R = \frac{DL}{DR}. \quad (4)$$

$$S = \frac{DR}{DL}. \quad (5)$$

In (6) parameter *a* appears. It is equal to the distance between the microphones; how far are apart the left and the center microphones, respectively the center and the right ones.

$$Dt = \frac{-2 \cdot a^2 - 4 \cdot R \cdot a^2 - 2 \cdot R^2 \cdot a^2 + S^2 + S^2 \cdot R^2}{2 \cdot S \cdot (1+R)^2}. \quad (6)$$

Having determined the distance, an angle determination is possible using (7).

$$An = \frac{360}{2\pi} \cdot \arcsin\left(\frac{(2 \cdot Dt - S)\sqrt{-S^2 + 4 \cdot Dt \cdot S + 4 \cdot a^2}}{2 \cdot a \cdot Dt} \right). \quad (7)$$

Equations (6) and (7) are inversed mathematical equations determined in Matlab from the geometric direct equations.

Now the obtained information are shown on the display, and left there for interpretation until a new acoustic emission event occurs, and new information are available. The display also shows information regarding the status of the processor, as "*Waiting*", "*Processing*", or "*New data!*". A led was used to signal each passing second, and also a counter on the display that measures the elapsed processing time.

III. HARDWARE

Three ROM-2238P-NF-R electret omnidirectional microphones were used. The operating voltage is 2 V of direct current and the impedance is 2.2 kΩ. The sensitivity is -38±3 dB and the performance characteristic is almost linear from 30 Hz to 19 kHz [2].

The schematic presented in Figure 2 is used to power the microphones and to amplify their signal. The microphone is powered from 3.3V through *R7* to meet its specifications. *C3* assures that only the alternative component enters the first Operational Amplifier (OA). *R5* and *R6* are used to set the continuous component for the microphone signal reaching the first AO, and *C1* and *C2* are used to filter the continuous component, so only the alternative component of the signal is amplified. The amplification is established by *R1* and *R2* respectively *R3* and *R4*. The amplification is large and is the reason why it is done in two stages, the secondary one being optional and selectable by means of *J1*. *C4* and *C5* are used to filter the power voltage. The *LM358N* OA is powered with 5 V so it will not enter into limitation due to insufficient power supply.

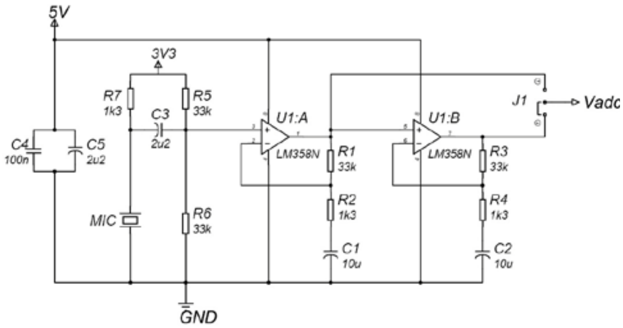

Figure 2. Microphone schematic [3]

The signal is processed by a Microchip dsPIC33F256GP710A processor, equipped on an Explorer16 Development Board. The processor has built-in analog to digital converters (ADCs) capable of up to 1.1 Mega Samples per Second (MSPS) at a resolution of 10 bits. It benefits of Direct Memory Access (DMA) that speeds up the system. The development board has an equipped display that shows the information.

IV. EXPERIMENTAL SETUP

The microphones were fitted at a distance of 40 cm between them, on a linear mount, at a height of 1.5 meters. Experiment room was a concrete wall classroom, with desks on sides. The length of the room was of 8 meters, the width of 6 meters and the height of 3 meters. The sound source was mounted at a height of 1.7 meters.

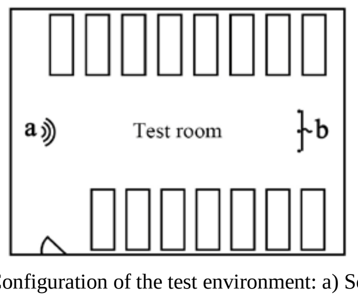

Figure 3. Configuration of the test environment: a) Sound source b) Acquisition system

Two mono test tones were emitted from the regular desktop speaker used as a sound source. The first test tone is high frequency and has a duration of 1 ms. The second test tone has a duration of 6 ms and it is composed of both high and low frequencies, as shown in Figure 4.

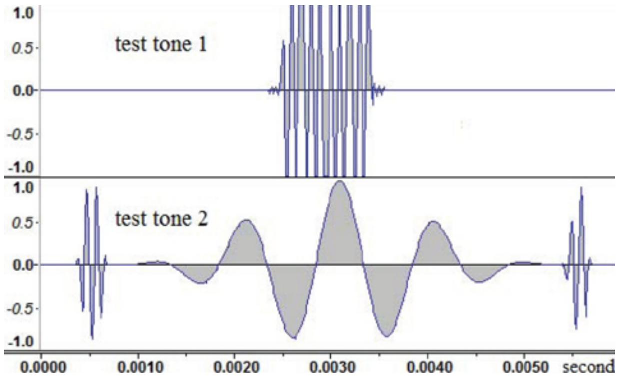

Figure 4. Emitted test tones

Recordings with a mono sound recorder were done from the location of the microphone system, to highlight the losses through the sound source and the environment. In Figure 5 is visible how at big distances the sound is affected by reverberations and is strongly attenuated. For close ranges, the sounds are denaturized only by the sound source, which is unable to reproduce exactly the test tones.

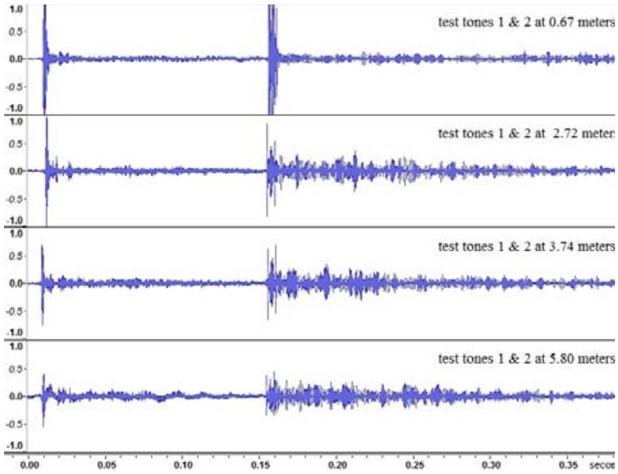

Figure 5. Recorded test tones

The tests were done at different angles: 35°, 0° and −23°, where −23° means that the sound source orientation was towards the left side, at an angle of about 23°. For each angle value, tests were done also at different distances: 5.80, 4.77, 3.74, 2.72, 1.71, 1.15, 1.00 and 0.71 meters. For each distance a set of measurements were performed for each one of the two test tones. For each test instance (angle, distance and test tone) around 25 measurements were done for conclusive tests. In cases of obvious results, the number of measurement was reduced.

V. RESULTS

Obtained results are being highlighted in this section. Exposed in the following figures are the results of the tests, all the performed measurements on a specific setup are shown. Invalid measurements are shown as zeros and are not to be confused with measurements affected by errors that are represented exactly as they were taken.

In Figure 6 is exposed a set of 26 consecutive measurements using test tone 2. The sound source was positioned at a distance of 3.74 meters, and on a direction deviated to the right by 35 degrees from the straight ahead direction. There can be observed an invalid measurement on the index, and an error affected measurement at the index.

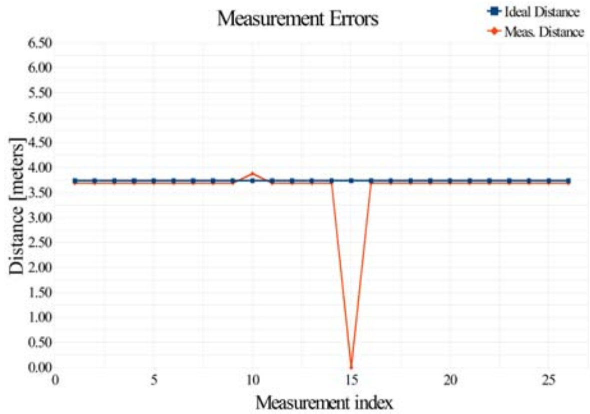

Figure 6. Measurements of distance taken on test tone 2, at +35 degrees and at a distance of 3.74 meters

In the case of the angle, in Figure 7 can be seen that index is an error affected measurement, instead of an invalid one. Figures 5 and 6 are pointing out the system's ability to localize a sound.

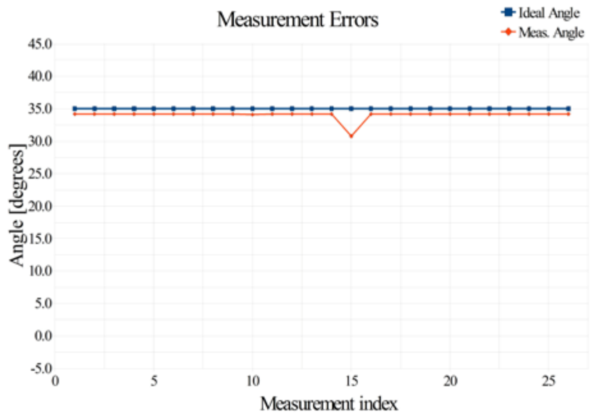

Figure 7. Measurements of angle taken on test tone 2, at +35 degrees and at a distance of 3.74 meters

Worst case scenario is when measurements are made on test tone 1, at 0 degrees and at a distance of 5.80 meters. At 0 degrees, when the source is located straight ahead, the time delays are at minimum values. On top of this, test tone 1 has reported worst results. The worst case scenario is shown in terms of measured distance in Figure 8 and in terms of measured angle in Figure 9.

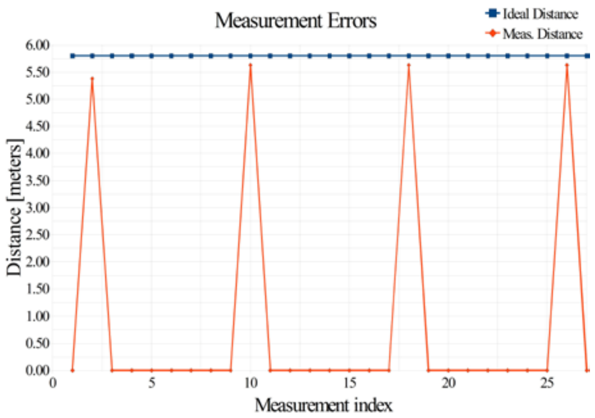

Figure 8. Measurements of distance taken on test tone 1, at 0 degrees and at a distance of 5.80 meters

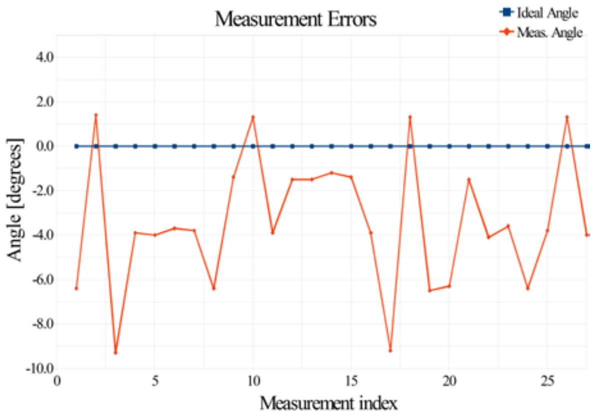

Figure 9. Measurements of angle taken on test tone 1, at 0 degrees and at a distance of 5.80 meters

Even in worst experimental instance the system was able to take some correct measurements, proving that it is capable of localization, but still needs some improvement.

Precise measurements were obtained also, as an example being shown Figures 10 and 11 that represent 26 consecutive measurements of distance and angle taken on test tone 2, at +35 degrees and at a distance of 1.71 meters. The system obtained a maximum error of 4 cm in distance calculation, and a maximum of 0.2 degrees in angle calculation.

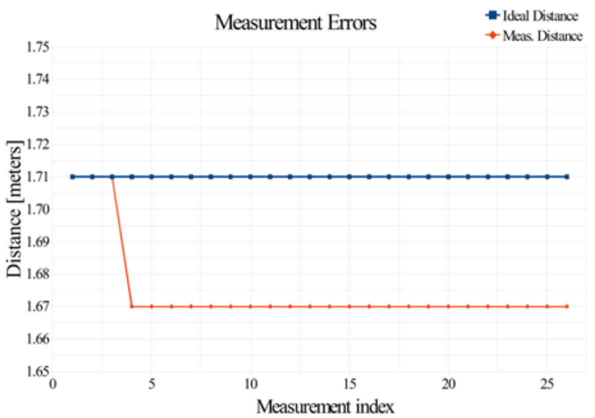

Figure 10. Measurements of distance taken on test tone 2, at +35 degrees and at a distance of 1.71 meters

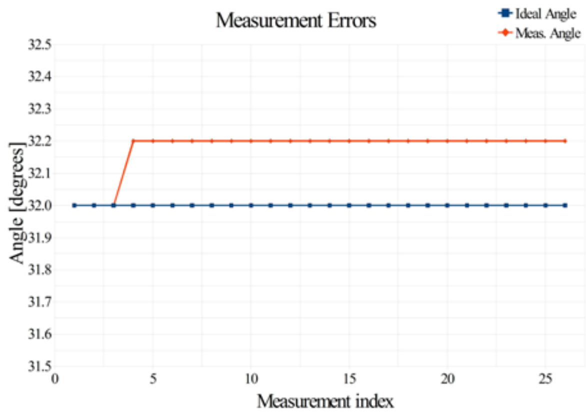

Figure 11. Measurements of angle taken on test tone 2, at +35 degrees and at a distance of 1.71 meters

Figures 12 and 13 present measurements of distance and angle at -23 degrees and at a distance of 3.74 meters.

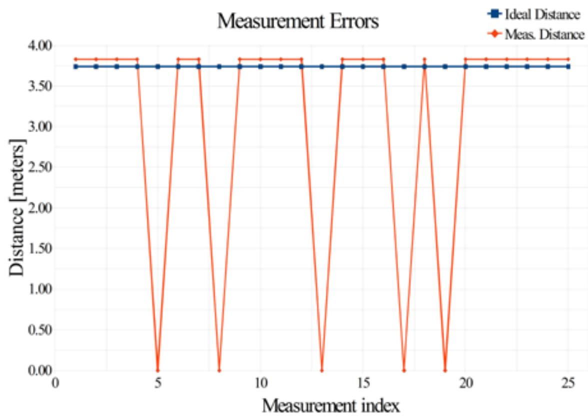

Figure 12. Measurements of distance taken on test tone 1, at -23 degrees and at a distance of 3.74 meters

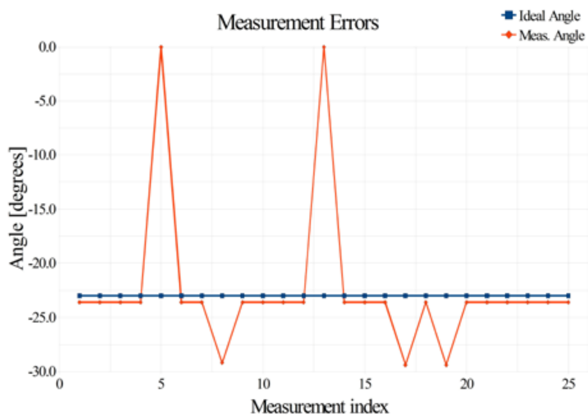

Figure 13. Measurements of angle taken on test tone 1, at -23 degrees and at a distance of 3.74 meters

Here are some invalid measurements in distance, which are not always translated into invalid angle measurements. To be noted that the measurements were taken on test tone 1, the test tone that has reported worst results.

VI. Conclusions

Sound source locating systems have some limitations. Perfect solutions are not possible, since the accuracy depends on several factors:

- Geometry of microphone and source;
- Accuracy of the microphone setup;
- Uncertainties in the location of the microphones;
- Not identical microphones and amplifiers;
- Inexact propagation delays;
- Bandwidth of the emitted pulses;
- Presence of noise sources;
- Numerical round-off errors.

The performed experiments show that the system is able to localize an unknown sound source in an unknown environment, with a certain degree of accuracy. Results are affected by all the limitations stated before. Even in these conditions, the designed embedded system has managed to obtain correct and useful data, so the experiments confirm the expectancies of the designed system.

Future work is necessary as the results are not perfect and still need improvement. An USB interface could be used by a PC to acquire the signals and do more complex signal processing, but this solution is limited by the amount of data that has to be transmitted, around 12 Mbps in case that the same settings are used for the converter and processor. As the last stated one has no USB interface, transferring this volume of data to another controller that is equipped with USB is problematic.

Another approach would be to develop a faster method for determining the time delays of arrival, thus maintaining the portability and the robustness of the system. Also further signal processing can be used for eliminating noise, if it is simple enough not to delay the computation time of the results.

Acknowledgment

This paper was supported by the project "Improvement of the doctoral studies quality in engineering science for development of the knowledge based society-QDOC" contract no. POSDRU/107/1.5/S/78534, project co-funded by the European Social Fund through the Sectorial Operational Program Human Resources 2007-2013.

References

[1] C. Lenz, *Localization of sound sources*, Master's thesis, ETH Zurich, 2009.

[2] ROM-2238P-NF-R datasheet available at ://www.farnell.com/datasheets/40095. .

[3] Schematic adapted from ://www.minidisc.org/mic_preamp/Simple%20Stereo%20Electret%20Mic%20Preamp. .

[4] C. Knapp and G. Carter, “The generalized correlation method for estimation of time delay”, *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 24, Issue: 4, pp. 320-327, 1976.

[5] N. Yousefian, M. Rahmani, and A. Akbari, “Sound source localization in 3-d space by a triple-microphone algorithm”, in *Asia-Pacific Conference on Communications (APCC)*, October 2008, pp. 1-5.

[6] A. K. Tellakula, *Acoustic source localization using time delay estimation*, Master of Science (Engineering) in Faculty of Engineering, Supercomputer Education and Research Centre, Indian Institute of Science, Bangalore, India, August 2007.

[7] B. Choi, *Acoustic source localization in 3D complex urban environments*, submitted dissertation, Blacksburg, Virginia, 2012

[8] A. M. Ali et al., “An empirical study of collaborative acoustic source localization”, Proceedings of the 6th international conference on *Information processing in sensor networks*, April 25-27, 2007, Cambridge, Massachusetts, USA

Biography

Bob Flaviu Ilie is a PhD student at the Technical University of Cluj-Napoca. He has 2 years experience in a multinational company as a software engineer. He also has a MsC diploma in Business Management obtained at the Babes-Bolyai University of Cluj-Napoca. His interests are microcontrollers, embedded software, converters and signal processing.