Tracking Sound Source Localization for a Home Robot Application

Gil Lopes, Andreia Albernaz, Hélder Ribeiro,
Fernando Ribeiro
Algoritmi Research Center
Dpt. Industrial Electronics, University of Minho
Guimaraes, Portugal
e-mail: gil@dei.uminho.pt
andreialbernaz13@gmail.com
a58795@alunos.uminho.pt
fernando@dei.uminho.pt

M. S. Martins
CMEMS, University of Minho
Campus of Azurém, Guimarães, Portugal
LARSyS, University of Algarve
Campus de Gambelas, PT-8005-139 Faro, Portugal
e-mail: mmartins@dei.uminho.pt

***Abstract*—The future of robotics is now trending for home servicing. Nursing homes and assistance to elder people are areas where robots can provide valuable help in order to improve the quality of life of those who need it most. Calling a robot, for a person of age, can be a daunting task if the voice is failing and any resort to battery operated devices fails to comply. Using a simple mechanical apparatus, such as a Click trainer for dogs, a person can call a robot by pressing the button of a powerless device. The high pitch sound produced by this device can be captured and tracked down in order to estimate the person's location within a room. This paper describes a method that provides good accuracy and uses simple and low cost technology, in order to provide an efficient positional value for an assistance robot to attend its caller. The robot does not need to search for the person in a room as it can directly travel towards the Click's sound source.**

Keywords-localization; sound source; interaural sound difference; time difference of arrival

I. INTRODUCTION

The use of home robots is in demand specially in tasks such as dust cleaning and food cooking. The future is promising and an increase of research is being held in areas of robotic assistance in industry, hospitals and also at home. In the latter case, the Robocup@Home competition [1] is contributing with valuable research and development of robotic solutions for home assistance with demanding tasks that increase in difficulty and complexity every year.

One of the main targets for home assistance is the help for elderly people, where normal daily activities could be improved if a personal assistant was always present. This is the case of nursing homes, where usually this task is taken care of by the regular staff. They are in charge of responding to calls of elder people when any type of assistance is necessary (to get hold of some object such as a book, TV remote, food, beverages, etc.). An assistance robot can be the helping hand 24/7.

In that sense, calling a robot can be performed in different ways. The first approach is vocal and therefore it is still a viable solution for calling someone or a machine if the person's voice is healthy. That is not the case generally for elder people. A second approach is via electronic means, such as battery operated remote controller or a button on a wall. Electronic devices need energy to operate and both present weaknesses. Batteries on a remote controller can run or dim out and the assistance cannot be called. This builds up stress on the caller that keeps pressing the button without any response from the assistant. A button on the wall does not rely on batteries to operate but on the ability of the caller to walk to it. For an elder person, this is often a major issue they have to deal with everyday.

A third approach is then necessary that can ease the calling process, providing the localization of the caller inside the room. In this case, the person can even be lying down on the floor and thus, difficult to be tracked down by the robot when it gets into the room. By providing an accurate localization, the robot can travel directly to the place where the call was originated from. It can then proceed with any reconnaissance procedures in order to find the person in a shorter range.

This paper describes a method for calling a robot that can be easily used by elder people in any situation. It does not require batteries and provides sufficient accuracy of its localization in a room. It is based on a device (Click trainer) (Figure 1) that sends a high pitch mechanical tone when pressed and another when released. This system only uses sound waves as the high pitch tone propagates within the room walls. A method is described that uses the generated acoustic signal in order to track the caller's position.

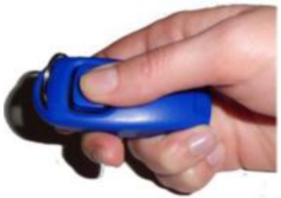

Figure 1. Click device that produces a high pitch mechanical tone

Section II describes existing methods found in literature and Section III describes the objectives of this work, followed by some theoretical background on Section IV. Section V describes how the system was implemented and Section VI shows the methodology and obtained results in the experimentation, finishing with the conclusions on Section VII.

II. SOUND SOURCE LOCALIZATION

Tracking the localization of a sound source is an area of research that is well exploited. Authors have taken different approaches but the Time Difference of Arrival (TDOA) method is recurrently used. For that, an array of microphones is necessary and different authors use different methods and applications.

Mandlik [2] used an array of four microphones displaced in a square 1 m apart from each other in the center of a room (50 x 30 m). The sound of a speech is recorded by the microphones and then it is processed offline in order to calculate its source localization. The authors used signal processing, by using the Generalized Cross Correlation function (GCC), Fourier transform, Fourier transform filtering and Phase Transform filtering (PHAT), to estimate the time delay of the sound received between the four microphones. Based on the position of the microphones, a model was developed to estimate the 3D position of the sound source. According to the presented graphical results, the direction of the sound source was very accurate, although the position of the source (distance from the speaker to the microphone array) showed estimation points of up to 5 m apart from each other on the experimented results (~2.5 m error from the real speaker position).

Using TDOA and Direction Of Arrival (DOA), a group or researchers [3] developed an acoustic source localization system in order to trace sound at the band of 100 Hz to 4 kHz. Using two sets of microphone pairs (1 m apart) arranged on two perpendicular horizontal walls, they combined the two processes (TDOA and DOA) to estimate the time difference (on each microphone of a pair) with the angle (between pairs), thus providing a 2D position of the sound source. Signal processing is used such as Power Spectral analysis, Fast Fourier Transform and phase difference computation with a Finite Impulse Response filter. The presented experimental results show angle estimation errors (DOA) from 3º to 30º on the worst angle scenario (45º from the center of the microphone array), and errors below 1º for the best scenario (90º) with time delay differences of up to 0.2 ms for the various tested angles.

Combining signal processing (GCC and PHAT) on a TDOA system with the use of Artificial Neural Networks (ANN), a group of researchers [4] used an array of microphones to estimate the position and orientation of a sound source. Experimental results show estimation positional errors with an average of 0.341 m. With the application of a phase transform method they obtained positional errors with an average of 0.298 m in 3D space.

In general, it can be concluded that signal processing applied to an array of microphones and using the TDOA method, is the process many researchers implemented for sound source localization systems.

III. OBJECTIVES

By using the Click device of Figure 1, the objective is to locate its position when operated inside a room, as shown in Figure 2.

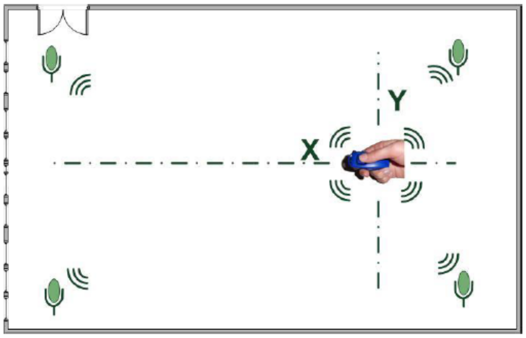

Figure 2. Click device localization when operated

A rectangular room was considered since it generalizes different room configurations (square, circular and rectangular). Four microphones were displaced in known positions of the room. They were placed near the corners and the ceiling since this was the best chance to avoid obstacles. Other configurations are planned to be experimented in the future, such as half way on each wall making a cross positioning. This paper only describes the results obtained with the microphones placed in corners.

Other two important objectives were defined: cost and accuracy. The system would have to be of low implementation cost for wide spreading in all rooms of nursing homes. The accuracy was defined to be less than 1 m radius around the caller, since it was considered sufficient for a good close visual detection of the caller from the robot.

IV. THEORETICAL BACKGROUND

Since this work uses sound waves, the first premise was the sound speed when propagating through air. At room temperature of 20º C the speed of sound is defined to be 343.21 m/s with 315.77 m/s at -25º C till 351.88 m/s at 35º C. When the sound is created in a certain spatial position, it is expected to travel in all directions at the same speed thus reaching each sensor (microphone) at a different time period. Sound waves at a temperature of 20º C take 2.91 ms to travel 1 m. If a sound starts at a distance of 1 m from one microphone and at 2 m from a second microphone there will be a difference of 2.91 ms of the sound arrival between microphones. This is TDOA as it is also graphically shown in Figure 3 (t_i and t_j is the time taken from the source s to microphones i and j respectively).

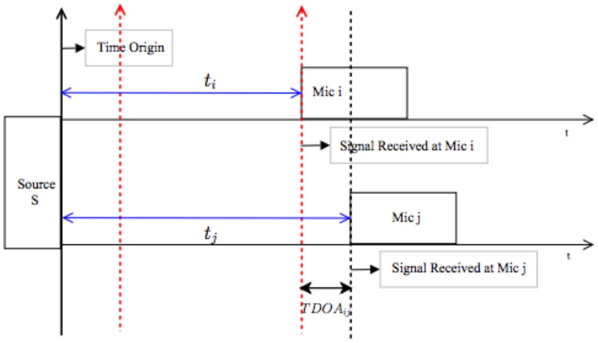

Figure 3. Signal receiving time in TDOA [5]

Although TDOA provides the time difference between two signals, it is still not a straight forward process to calculate the distance based on two microphones. There is no other way to communicate that a sound started at a given time. Therefore, one can only rely on the first samples of the sound signal, when they arrive, to start the clock ticking. An approach based on the TDOA is the DOA, or also Interaural Time Difference (ITD), which resembles the human ears. It provides the ability to track an angle where a sound is coming from. This angle (θ) is based on the distance that separates the two ears (*x*), the relative time difference of the sound arrival at the two ears (Δ*t*) and the speed of sound (*c*), as shown in (1) [6].

$$\theta = \operatorname{asin}\left(\frac{\Delta t\, c}{x}\right) \tag{1}$$

With the ITD angle calculated, it is then possible to compute the intersection between different angles in order to estimate a possible position of a sound source, as shown in the next section.

V. Implementation

Considering the ITD process, a pair of ears will be considered as a set of two microphones, separated by a *x* distance. Since four microphones are used near the corners of a rectangular room, each two microphones side by side will become 'ears' of that wall. In other words, a rectangular room will have then four sets of ears. Hence, four angles will be generated when a sound is created within the room, as shown in Figure 4.

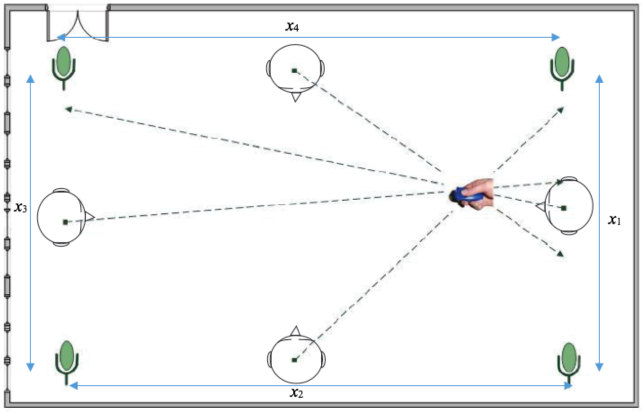

Figure 4. ITD process at work where each pair of microphones emulates the head ears with the different obtained angles from the sound source

After obtaining the θ angle for each pair of microphones, two points are then calculated. The first point (P_1) is on the "head" position (center point between two microphones). The second point (P_2) is obtained when the line crosses the opposite and parallel axis, as shown in Figure 5. This second point is obtained by multiplying the tangent of θ by the distance between the two parallel axes (*x*).

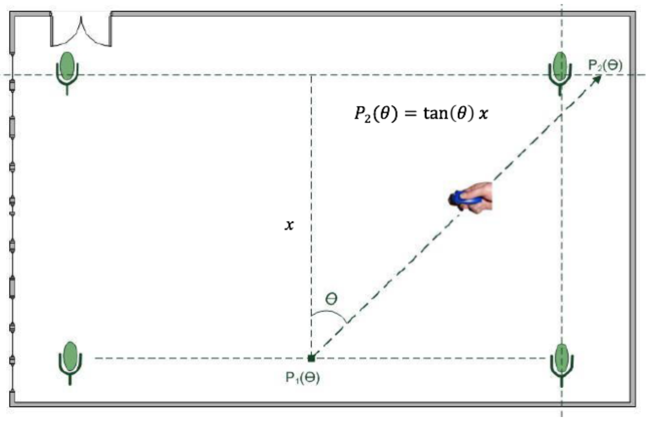

Figure 5. Obtaining P_2 from P_1 and θ angle

For each θ angle, two points are calculated and therefore, a total of height points (four lines) are obtained in the room. The intersection point between these four lines is the sound source (*x*, *y*) position. This point can be calculated using the determinant of each pair of lines, as shown in (2).

Considering a generic pair of obtained lines (lets call them line *a* and line *b*), points (x_1, y_1) and (x_2, y_2) are the points P_1(θ) and P_2(θ) of line *a*. Points (x_3, y_3) and (x_4, y_4) are the points P_1(θ) and P_2(θ) from line *b*. The point (P_x, P_y) is the intersecting point of line *a* and line *b*.

$$(P_x, P_y) = \left(\frac{(x_1y_2 - y_1x_2)(x_3 - x_4) - (x_1 - x_2)(x_3y_4 - y_3x_4)}{(x_1 - x_2)(y_3 - y_4) - (y_1 - y_2)(x_3 - x_4)}, \right. \\ \left. \frac{(x_1y_2 - y_1x_2)(y_3 - y_4) - (y_1 - y_2)(x_3y_4 - y_3x_4)}{(x_1 - x_2)(y_3 - y_4) - (y_1 - y_2)(x_3 - x_4)} \right) \tag{2}$$

Although this intersection can be calculated from any pair of obtained lines, from experimentation, only two lines shown consistently lower deviations on the calculated position. They are the opposite lines from the closest microphone (opposite quadrant of the room) to the Click device (Figure 6). In practice the closest is the microphone that firstly receives the sound signal.

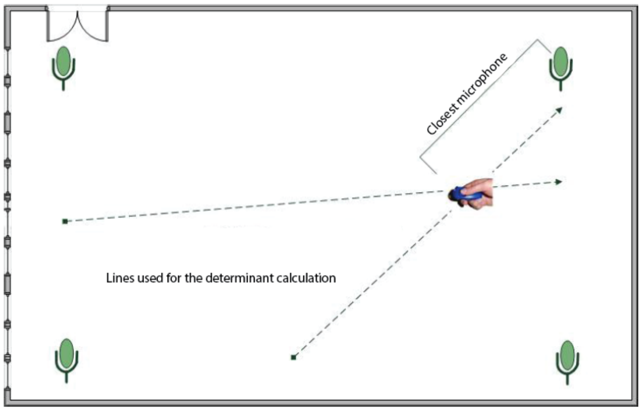

Figure 6. Closest microphone to the Click device and oposite lines used for the determinant calculation

Another important aspect on selecting the opposite quadrant, is the fact that on the same or adjacent quadrants, the line's intersection may produce a singularity: the lines are almost parallel to each other and therefore, an intersection point can fall outside the room. A small angle calculation deviation can move the intersection point outside the room, as shown in the example of Figure 7. This occurrence was found during trials.

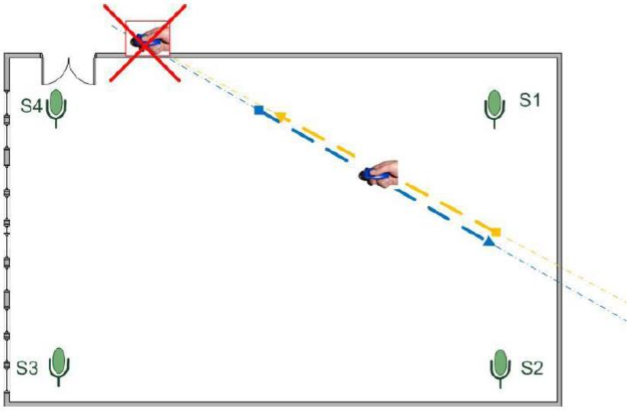

Figure 7. Example of an occurring singularity

The developed system was implemented in two separate blocks: Acoustic detection block (Adb) and Control block (Cb)(Figure 8).

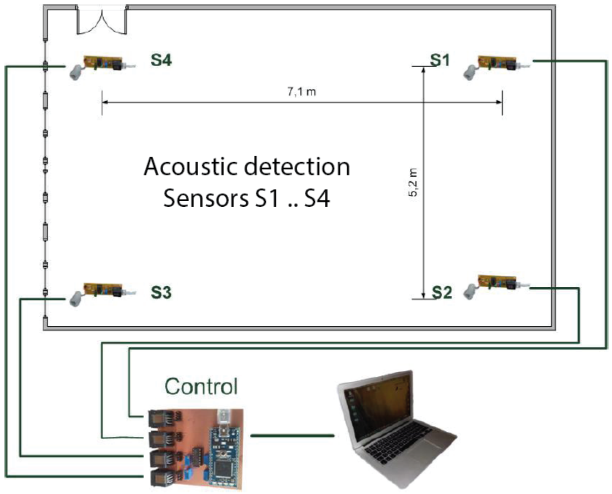

Figure 8. Developed system and the two blocks of operation

The Click device utilized on the experimental tests generates an acoustic tone at around 5 kHz (+/- 500 Hz). Each sensor (S1-S4) is based on an electret microphone with a pre-amplifier, a 2nd order high-pass filter tuned to 5 kHz, an amplifier and a threshold comparator. The latter produces the 5 V level pulses that are supplied to the Cb. These operations are performed by a single low cost chip with four operational amplifiers on the Adb side. More details of the developed system can be found in [7].

Each Adb is connected to the Cb via a twisted pair cable (Ethernet cable). The cable uses one pair for the signal (Adb to Cb) and one pair for powering the Adb (Cb to Adb). The Cb contains an mbed NXP LPC1768 microcontroller board (ARM® Cortex™-M3 Core) with 96 MHz clock speed. It also contains a threshold comparator to regenerate the incoming signals from each sensor. These signals are then injected into four digital input ports of the microcontroller. At each incoming signal (pulsed signal as shown in Figure 9), a hardware interrupt is generated in the microcontroller that uses its internal timestamp to tag them. The timestamp is in microseconds. The system only reacts to the first pulse received per port and it ignores subsequent interrupts from the same port, until a valid point is calculated or a timeout is generated.

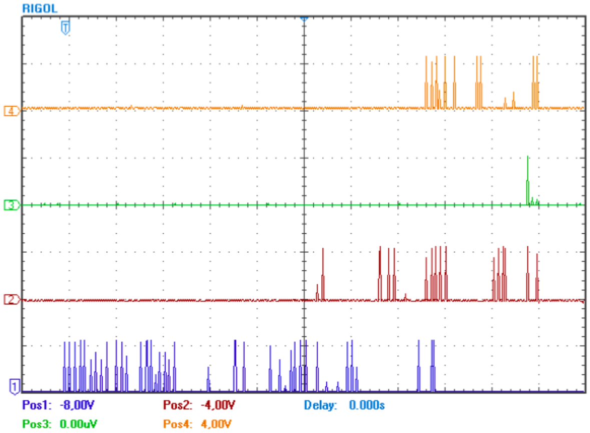

Figure 9. Example of obtained pulses on the four sensors and their time differences

After four valid signals are received, the microcontroller calculates the ITD angles. As explained before, the signals used are of the opposite microphones from the first received signal. From the angles, each line points (P_1 and $P_{2)}$ are calculated followed by the determinant of the pair of lines. The end result is the intersection point (P_x, P_y) that is sent to the robot via serial port of the microcontroller.

Room setup and sensor location information is configured in the microcontroller algorithms so the tracked position is relative to the real room length and width.

VI. EXPERIMENTATION

In order to test the accuracy of the developed system, trials were conducted where the Click device was positioned at different pre-determined positions in the room. A constant height of 1 m from the floor was used. On each position, three clicks were made at intervals of 2 s each. Figure 8 shows the room setup where the microphones are placed apart 7.1 m on the *x* axis and 5.2 m on the *y* axis. Two sets of tests were done on each round of trials: a) 12 positional diagonal points; b) 5 positional orthogonal points.

Figure 10 shows the results obtained on the diagonal trial positions. The blue diamond shape marks the intended real position where the clicks were performed. The obtained calculated positions are the other different encircled shapes where each circle is a set of three clicks. Figure 11 uses the same approach but for orthogonal trial positions from the sensors.

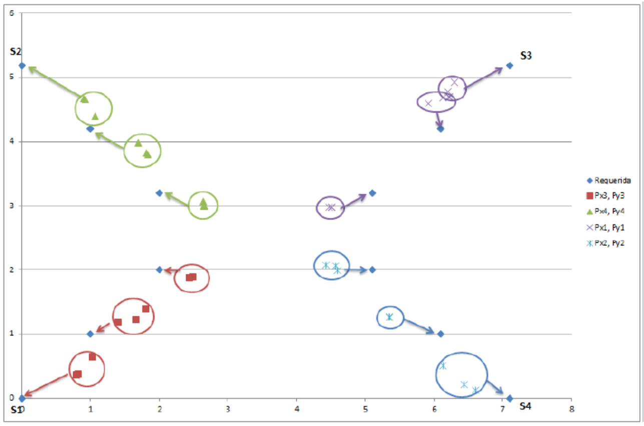

Figure 10. Trial results obtained at the diagonal of the sensors

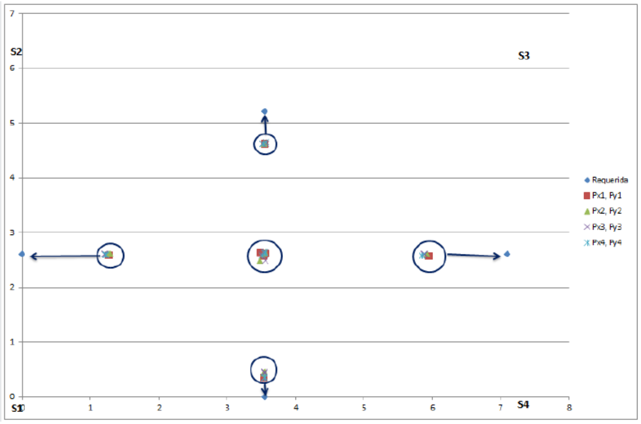

Figure 11. Trial results obtained at the orthogonal of the sensors

The results for the diagonal of the sensors show that, between each three clicks at the same position, a deviation of 0.229 m was found on the *x* axis and a deviation of 0.193 m was found on the *y* axis. For the positioning deviations on the diagonal tests, they were divided by quadrants and Table I summarizes the results.

TABLE I. STANDARD DEVIATION OF THE OBTAINED RESULTS

Quadrant	X(m)	Y(m)
1st	0.684523312	0.330336495
2nd	0.81212981	0.429232324
3rd	0.635961332	0.461925981
4th	0.913362087	0.706452613

The average deviation is then 0.761 m for the *x* axis and 0.482 m for the *y* axis. The absolute deviation (measured by the shortest distance between the real and the obtained points) is 0.901 m. For the orthogonal tests, the results demonstrated a lower deviation between each click on the same position (0.01 m), although they show a higher deviation on the *x* axis (1.191 m) and on the y axis (0.458 m). At the room center, the values were typically below 0.1 m. It is clear though, that as the Click device moves closer to a sensor, the deviation from the real value increases. On the other hand, as the device moves towards the quadrant borders, the values tend to be more consistent. They show very low differences at the same position, but a higher difference to the real value as it moves away from the center.

Another set of trials was conducted, in order to estimate the influence on the results of the Click device at different heights. Starting from the floor and with increments of 0.5 m up to a maximum of 2 m, tests were performed in the room center. This was where the lowest deviations were achieved at a fixed height of 1 m. At each height three clicks were performed. Table II shows the obtained deviation results in meters from the room center position (3.55 m, 2.6 m). As it is shown in the table, the influence of height in the deviation accounts for less than 5% in absolute terms and only in one axis.

TABLE II. TRIALS AT DIFFERENT HEIGHTS AND OBTAINED DEVIATION RESULTS

Height	$\bar{x}$	$\bar{y}$	*x* deviation	*y* deviation
0	3.57	2.66	0.02	0.06
0.5	3.56	2.59	0.01	-0.01
1	3.54	2.55	-0.01	-0.05
1.5	3.55	2.56	0.00	-0.04
2	3.55	2.53	0.00	-0.07

A descent trend in the obtained values is visible on the graph of Figure 12, from the floor level up to 1 m. Then, a levelling trend for heights above 1 m is achieved, showing that around this floor distance (1 m +/- 0.5 m) the best results are produced with the developed solution.

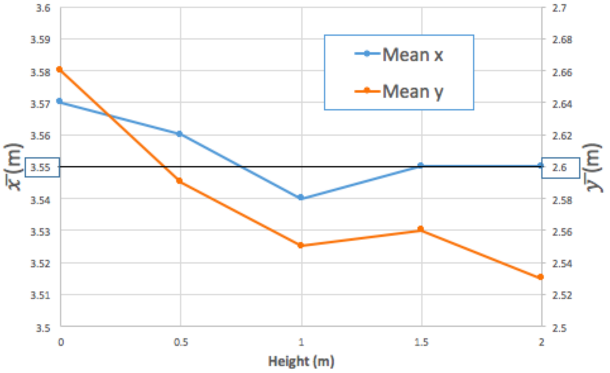

Figure 12. Trials graphical results at different heights

Further investigation is necessary in order to identify the influence of occlusions and reflections to the sound signal and the deviations caused by them.

VII. CONCLUSIONS

This paper presents a system for tracking the sound source localization of a Click trainer device. It describes a solution using simple and low cost devices that produces good results in terms of accuracy and simplicity. It has direct application on a robotic system's implementation, to localize a caller by an acoustic signal. The results show an accuracy below 1 m, fulfilling the original objective of localizing the person that called the service robot. The influence of the device position in height, showed a small deviation between the obtained position with the real one. Several improvements have to be addressed in the future, nonetheless the achieved accuracy demonstrated other possible applications of the developed system in different areas.

REFERENCES

[1] RoboCup@Home. (2015, December 15). Retrieved from http://www.robocupathome.org/

[2] M. Mandlik and V. Brázda, “Sound source location method,” Mandlik Magazines, vol. 6, nº 5, Dec. 2011, pp. 197-204.

[3] S. Paulose, E. Sebastian and B. Paul, “Acoustic Source Localization,” International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, vol. 2, nº 2, Feb. 2013, pp. 933-939.

[4] A. Nakano, S. Nakagawa and K. Yamamoto, “Estimating the position and orientation of an acoustic source with a microphone array network,” em INTERSPEECH 2009, 10th Annual Conference of the International Speech Communication Association, Brighton, United Kingdom, September 6-10, 2009, pp. 1127-1130.

[5] K. A. Tellakula, “Acoustic Source Localization Using Time Delay Estimation,” Indian Institute of Science, Bangalore, India, 2007.

[6] R. M. Warren, “Auditory Perception,” Cambridge University Press, Cambridge, UK, 2008.

[7] A. F. G. Albernaz, “Sistema de localização através de ondas de som no interior de edifícios,” University of Minho, Guimaraes, Portugal, 2013.