Analysis of Sound Source Localization and Tracking for the @Home Service Robot in Multiple Distances

Thiago Spilborghs Bueno Meyer and Plinio Thomaz Aquino-Junior

Abstract—Robots have been used in several areas, and in many of them it is common to see an interaction between a human and the robot. Human-Robot Interaction (HRI) became a crucial part on the process of understanding and creating machines, which will communicate using voice as the main factor. Speech is used as an input and an output of the communication between machines and humans, in the way that commands and answers are performed using this method. There are several Sound Source Localization and Tracking (SSLT) methods that may be used in an ambiance and many systems can be used to complete this task. To succeed in the usage of this knowledge source, establishing an appropriate system, which allows doing the calculus and implementation in a service robot, is necessary and, using the past work of the author, it was concluded. The main objective of this paper is to introduce some of the available methods for SSLT in the literature, and present results of the implementation and implantation of methods for a Domestic Service Robot, with the objective of having a more natural interaction with humans, allowing to interact the human facing front, instead of facing any direction.

I. INTRODUCTION

Voice interaction is applied in software, in self-services machines, vehicles, domestic devices, and smartphone apps. Robots can also interact with voice commands and this type of interaction is an excellent format because it is equal to the people scenario affordance. However, the interaction cycle with humans in the robot is very complex, since the robot can be in surroundings and use context which may require the robot to move around. For this reason, when a human-robot interaction cycle occurs, the user command can be given while the robot is facing another way, rather than the user’s face.

The robot used in this paper is the Robot HERA (Home Environment Robot Assistant), seen in figure 1. This robot was built by the RoboFEI Team [1] and competes in the RoboCup @Home league [2] and in the Brazilian Robot Competition [3]. It is an autonomous mobile robot designed to perform actions in domestic environments interacting with people.

The way that the robot, which was used for this work, receives a command occurs by a speech-based communication. The sound captured by the microphones can be used for speech recognition, speech emotion recognition and even sound signals recognition from the robot surroundings. The user may execute a command while it is located behind the robot, and this information is an important challenge in the human-robot interactive project.

Sound localization is a listener’s ability to identify the origin or position of a detected sound in distance and direction. The sound localization mechanisms of the human auditory system have been extensively studied. The auditory system consumption’s several signals for sound source localization, including time- and level-differences (or intensity-difference) between both ears, timing analysis, correlation analysis, spectral information, and pattern matching.

This paper’s objective is to introduce the methods to evaluate the user’s location in the ambiance and create an analysis of the method chosen to be used in the @Home robot while performing in multiple distances. This paper continues the past work, seen in [4], evaluating the usage...
of methods alongside the chosen system for Sound Source Localization and Tracking and analyzing the precision of the method while using it in multiple distances, creating a process with a higher human interaction potential.

This article is organized by starting with the sound source localization, followed by the section in which the tests are described. It is concluded with the discussion of the results and future works that can be realized based on this paper.

II. SOUND SOURCE LOCALIZATION AND TRACKING

One of the challenges for a robot is to locate a sound source in an environment, in the case of direct interaction with users, the robot must know where the active operator, the one who is talking to the robot, is and answer the request. To conclude this task the robot must, through the sound, locate the active user. The same operation protocol may be executed when the robot is interacting with objects, when searching a broken glass which fell on the ground or even looking for something like a sound system.

To complete the user localization through sound challenge, the robot must have a Sound Source Localization and Tracking system, which are based on the definition of position having 3 dimensions, with 2 angles and 1 distance where one of the angles is the horizontal angle, or azimuth angle, 0 to ± 180 degrees, and the vertical angle, 0 to ± 90 degrees, and distance, depth, is measured from 0 to ±∞. Three main physical parameters are used by the auditory system to accomplish sound source localization: time, intensity and the spectral shape [5].

On the azimuthal plane, the angle if defined by binaural factors, involving both ears or sound receptors, such as, e. g. Interaural Time Difference (ITD). In the vertical plane, the height is defined by monaural factors, verified by only one sound receptor, like the sound spectral shape variance, and distance is also proved with monaural factors [5].

In [6] is presented a method for locating a sound source with variation by wind currents. These researchers use a junction between the improved mapping of acoustic correlated sources (IMACS) and the Amiet method, called Amiet-IMACS. This method is proposed so that the Amiet is used for the correction of the 3D sound propagation and the IMACS mapping makes the corrections by changing the parameters for a greater convergence, changing it from a simple mapping to a mapping with a self-improvement. They obtained a reduction of approximately 8% in the localization error verified when compared to previous methods, reducing the error percentage to approximately 3.8% [6].

In [7], an architecture for real-time localization and tracking of a sound source applied to robots is presented. The implemented system is based on the auditory system present in the mammalian brain, it is a neuromorphic real-time sound tracking system, aiming to model the functionality of the Lateral Superior Olive to locate and track high frequency sounds.

When compared to previous works, the proposed system obtained a better result having as error in the worst possible condition less than 10 degrees of the real location of the source. When in a median condition getting errors smaller than 5 degrees from the actual position. So the proposed system can be used in tasks involving robotics since the error is acceptable [7].

In TACHIKAWA, T. et al. work [8] it’s proposed a system that can locate the angulation of the sound location, as well as the distance to the microphones used by the system. The method Direction-of-arrival (DOA) is used to locate the direction of sound, but since there is a problem with the continuity of system values, a grid of available space was used to prevent DOA system initialization errors. The variables used were discretized to reduce such errors and set up the grid used by the system. It is demonstrated that the variation of points in space when one more dimension is added is affected at exponential levels.

The increase in the number of points according to the angle at the adjacent points decreases, is exponential when comparing the 3D environment to the 2D, so the use of a 2D environment reduces the complexity of location analysis of the sound emitter in question. The use of the system for 3D localization of sound emitters is possible and uses a dictionary of 3D points and a modified convex clustering for localization [8].

MURRAY, J. C. et al. [9] defines a sound relation system based on the intersection of information and data. It is defined problems with systems based on height of sound waves, since in different acoustic environments false information can occur. The architecture developed is inspired by the central hearing system present in the mammalian brain. Two models based on biology are presented. One using the Interaural Time Difference (ITD) method, which uses the difference of time of arrival in the ears to identify the angle of arrival of the sound in the sound system. The other model presented uses interaural cues for the Licklider triplex model, this model creates a sketch of the auditory signals by checking the signal delay when the receiver signals are compared, thus creating a line of possible location of the sound source.

It is seen that the solution in [9] was effective, and the hybrid model described had an effective result in locating and tracking sound sources, with accuracy of 1.5 degrees around center point 0, but around the range of -90 degrees to +90 degree the system presented accuracy of 7.5 degrees.

In [10], a sound localization system based on the use of easily accessible and low-cost equipment, such as the Microsoft Kinect, is defined in conjunction with the execution of Steered Response Power using the PHAse Transform (SRP-PHAT) algorithm, making use of collected data. A prototype was developed using two Kinect devices to evaluate the proposed solution using the channels of each of the microphones present in the environment. The algorithm identifies a possible position of the sound source, based on the SRP-PHAT algorithm that is based on the differential in the arrival delay of the sound signal.

The solution proposed in [10] does not present precision in the identification of the location of the emitting source, but it can present the sound direction with efficiency. For future
work, it was desire to use other Microsoft Kinects devices for better identification of the location of the sound source, being, than, one of the main studies for this paper, since the same method was used and a bigger microphone array was used instead of only one Microsoft Kinect.

In the same sense [11] define a location system based on the use of a multi-channel spherical microphone and an embedded video card. The proposed sound location is 3D based. The SRP-PHAT algorithm was used for localization in conjunction with the GPU and it was concluded that the use of a GPU for processing, transforms the processing in real time, even with a GPU found in mobile devices. The system was considered as efficient for locating a sound source in real time.

Location recognition is a very important area of research that can be decisive in real environments with situations of danger, such as in fire situations for the localization of people using rescue robots and for this, techniques of multiple sound sources localization and tracking are used [12].

As seen in [12], a very popular method for estimating location with only one dimension is the arrival time difference, in conjunction with the cross-correlation vector calculation. However, this method is very sensitive to reverberations and other sounds that may occur while picking up the audio. For these problems to not create disturbances in the calculation, the cross-correlation method based on the dominant frequency can be used.

Based on the data generated in the estimate, a trace of sound sources can be performed. The estimation is done using a single time span, where the location of a source is calculated in a single time window, and when these calculations are generated, it is possible through filtering methods to know the tracking of the sound source being located by the system.

Based on the author previous work, verified in [4], the Sound Source Localization and Tracking was implemented using the SRP-PHAT method based on modules used in the robot system. Tests were made based in distances provided by the RoboCup@Home rulebook with some changes comparing the precision of the localization in multiple distances and angles.

III. TESTS

For the tests, a continuous sound with an average intensity of 63 dB was used, which according to [13], represents a normal conversation. Two angles were chosen to perform the precision check according to the distance, the angles of 85° and 115° that represent, in the used system, a point between microphones and another in the direction of a microphone, respectively, and for the distance, distances of 30 centimeters to 190 centimeters, varying 20 centimeters between measurements. For averaging and estimated error in calculating the location of the sound source, 30 tests were performed for each previously defined position.

With the data being collected, two graphs were generated to check the approximation of the collected data with the expected value. Graph 1 represents the desired value for the recognition and the real value during the tests to 85° for the system. Graph 2 represents the desired value for the recognition and the real value during the tests to 115° for the system.

IV. CONCLUSION

From the data collected in the tests, an estimated error of 5.73 % in its entirety was verified for 85 degrees, verified in the figure 4, and when the result of the data between the distances of 70 and 130 centimeters was verified, which correspond to the distance exemplified in the requirements of the project, we obtain an estimated error of 1.29 %. For the 115 degree angle, an estimated error of 5.53 % was obtained, also verified in the figure 4, and by performing the same check made for the test at 85 degrees, an error of 1.89 % was found.

It is possible to conclude, therefore, that as the distance between the system and the sound source increases, the recognition of its location has a tendency to lose its accuracy, however, most of the results were satisfactory, obtaining a value very close to the real value expected.

Regardless of whether the sound source is located in the direction of a system microphone or between them, the location can be performed and the robot will have the possibility to interact directly with the user. It is desired for future work, to locate the user’s height, in addition to making use of other sensors available in the system chosen as the basis of the project.
REFERENCES


Fig. 4. Comparison Table for 115 and 85 Degree test.