# A sound source localization sensor using probabilistic occupancy grid maps

Kenn, Holger
School of Engineering and Science
International University Bremen
Bremen, Germany
Email: h.kenn@iu-bremen.de

Pfeil, Andreas
School of Engineering and Science
International University Bremen
Bremen, Germany
Email: a.pfeil@iu-bremen.de

***Abstract*— Sound source localization can be used in the Robocup Rescue Robots League as a sensor that is capable to autonomously detect victims that emit sound. Using differential time of flight measurements through energy cross-spectrum evaluation of the sound signals, the angular direction to multiple sound sources can be determined with a pair of microphones for SNRs better than -8dB. Assuming that the robot pose is known, this information is sufficient to create probabilistic occupancy grid map of the sound sources in the environment and thus localize the victims in a global map.**

## I. Introduction

In the field of teleoperated and autonomous robotics, a new competition-based benchmark has emerged in the recent years in the form of the Robocup Rescue Real Robots league. The goal of this competition is to encourage the design of mobile robotic systems that are of use in desaster scenarios, helping with the assessment of the situation by gathering sensoric data from areas inaccessible by humans.([2],[3],[4])

The IUB Robocup Rescue team has been competing in the Robocup Rescue Robot League competitions in Fukuoka [5] and Padova [6]. In this competition, the goal of the competing teams is to locate victim dummies in different disaster scenarios in limited time using mobile robots. The scenarios range from office areas that are easily accessible to mobile robots to pancake-collapsed buildings that simulate a major earthquake disaster and pose a significant challenge to the sensing and manoeuvering capabilities of the mobile robots used. All scenarios contain a number of victim dummies that have human appearance and are equiped with other detectable features such as movement, sound, body heat and $CO_2$ emission. The performance of each team is evaluated through a common scoring function that analyzes the quality of the information gained and scores it against the number of human operators used by the team. This function thus both rewards autonomy and high-quality localization and multi-sensoric assessment of victims and their state.

In both competitions, the IUB Robocup Rescue team used a single operator for the competition runs and relied mostly on cameras for teleoperation and victim sensing. However, another sensor has been used in both competitions that has proven to be very useful for assessing the state of victims found. This sensor wasn't planned in the original design of the robots and came ”for free” in the form of the buildin microphones of the USB cameras used.

A simple network audio transmission prototype was built on site. A standard audio conversion utility (sox) was used to read the data from the audio input device driver of the Linux operating system, to encode it and transmit it via a simple TCP connection to the operator station where the network stream was decoded and played back using the same conversion utility.

During the 2002 competition, it became clear that this sensor was a useful addon to the existing robot and was the only sensor that was capable of locating invisibly trapped or entombed victims. This finding has been one of the reasons for the extension of the robots with additional sensors. In 2003, an additional thermographic camera has been added that proved be very sucessful for victim identification, but the acoustic sensor remained a valuable tool.

With the introduction of the new control middleware FASTRobots [7] in the 2003 robot system, it became possible to add more sensors to the robot platform, first for non-victim related tasks such as localization and environment mapping. The generated LADAR-based map provided to be useful for robot localizaion [8]. However, as no automatic victim localizing sensor was available, the localization and identification of the victim dummies was still performed manually by the operator. In order to do this, the operator would carefully analyze all available sensors including the sound from the microphones, then note down the perceived signs of the presence and state of the victim in a paper victim sheet and then mark the victims location by using a mouse to click on the approximate position of the victim next in the LADAR map that is displayed on his operator control station screen together with the robots current position and orientation. This process is time-consuming and error-prone.

For automatic victim localization with bitmap sensors such as the visible light and thermographic cameras, computer vision based approaches may be used. However, these approaches are computationally expensive and their performance in the highly unstructured environment of a robocup rescue scenario is hard to predict.

Fortunately, there is another approach using sound localization. It's advantage is that it can automatically create a map of sound sources on the operator display. These still

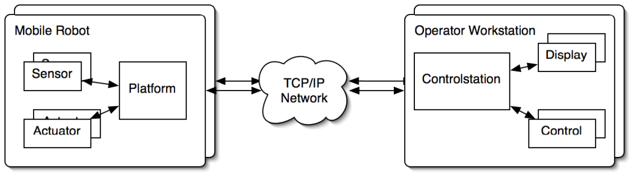

Fig. 1. An overview of the FAST-Robots middleware

have to be manually identified using the onboard cameras and other sensors but as their location is known now, their localization on the environment map will be much more precise. Another advantage of this approach is that its performance can be simulated beforehand as it will be shown in this paper.

The remaining part of this paper is structured as follows: The second section gives an overview over the system setup and an introduction into the theory of sound source localization. The third section describes an experiment to estimate the performance of the obtained sound source localization and shows that the measurements obtained are close to the theoretical boundaries. These results are then used to simulate the performance of a probabilistic map based on a occupancy grid. The last section discusses the results obtained so far.

## II. System overview and theoretical analysis

A typical robot system used for Robocup rescue consists of one or more mobile robots and one or more operator control stations. For the IUB Robocup Rescue system, the communication between the mobile robots and the control station is implemented through the FAST-Robots middleware[7]. Each mobile robot runs an instance of the platform component of the framework that instantiates multiple sensor and actuator driver objects that communicate with the corresponding sensor and actuator hardware. The platform communicates with its counterpart, the controlstation component running on the control stations through a TCP/IP network. The controlstation visualizes sensor data coming from the platform and transmits control commands to the platform.

The sensor described in this paper can easily be accomodated by this framework. Figure 2 shows an overview of the components that are part of the sound localization sensor system.

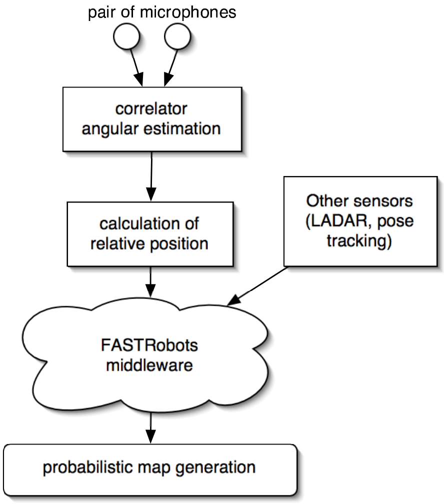
pair of microphones
correlator angular estimation
calculation of relative position
Other sensors (LADAR, pose tracking)
FASTRobots middleware
probabilistic map generation

Fig. 2. An overview of the sound source localisation system.

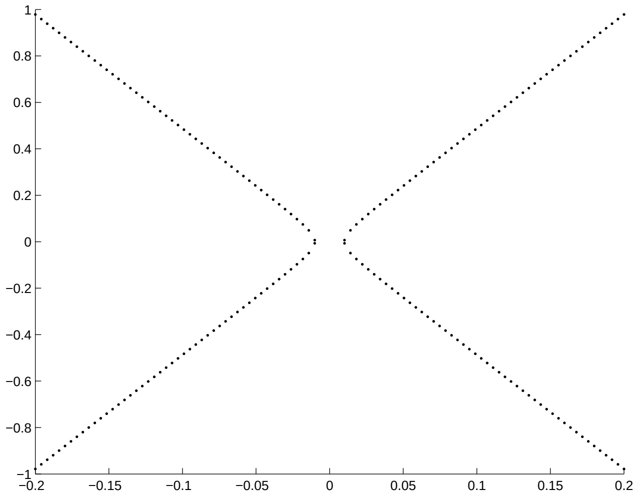

Fig. 3. Two hyperbolas indicating the possible location of the sound source for a given time-of-flight difference $\delta t$ and $-\delta t$.

### A. Microphone Phase array

A simple way to model the localisation of a sound source (sometimes also called “passive sonar”) by using multiple microphones is the so-called linar microphone phase array. In this model, a number of microphones are located equidinstant along the x axis of our coordinate system. It is then possible to determine the position of a sound source in the coordinate system using differential time-of-flight measurements, i.e. the time difference for the signal of the same sound source to arrive at different microphones. This system however cannot detect the correct sign of the y coordinate, i.e. it cannot distinguish between sound sources in positive or negative y direction.

Information obtained by only one microphone pair allows to restrict the position of a sound source to a hyperbola. These hyperbolas are solutions to the equations given by evaluating the time-of-flight (and thus path-length) difference between the paths from the source to the microphones. These solutions are instable, therefore it is desirable to approximate the sound source position with the asymptotes to the hyperbola.

In this approximation the localization information consists of the angle between a) the line connecting the two microphones and b) the vector pointing from the center of the microphone pair towards the sound source.

By using two pairs of microphones with different center-points, it is possible to determine the location of the sound source in a plane through triangulation.

Hence the problem is split up into two parts, the orientation measurement using a pair of microphones, and using the orientation information to determine the position of a sound source.

### B. Signal Detection

The problem of sound source localization can be solved in different ways. One approach is sound source localization based on beamforming techniques, such an approach has for example been presented in [1]. However, our approach is using the cross-energy spectrum of signals recorded at microphone pairs to evaluate the sound source directions. This is computaionally less expensive for a small number of microphones and allows for easier detection of multiple sound sources.

To determine the time delay between two incoming sound signals at the microphones, the cross-energy spectrum of the two signals is evaluated. Identical, but shifted signal portions produce peaks in this spectrum; the position of the peak is an indicator for the delay between the first and the second occurance of the same signal portion in the different signals. To avoid ambiguous results, the signal portion to be detected should not be periodic. If it is possible to choose the signal form, white noise would be the best, as it creates a single peak in a cross-energy spectrum. Should there be several distinct sound sources with different relative delays, one peak for every sound source can be detected. For the remainder of this section, it will be assumes that only a single sound source is being localized. It will be shown later that with multiple sound sources can be dealt with using probabilistic occupancy grids.

### C. Orientation measurement

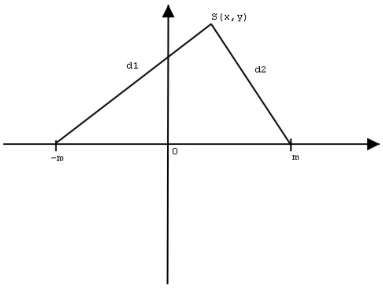

Fig. 4. Two Microphone Setup

Two microphones are located at $(-m, 0)$ and $(m, 0)$, a sound source at point $S(x, y)$.

If the sound source is further away than $2 \cdot m$, we can approximate the incoming beams to be parallel. Hence the signal path length difference $\delta$ can be found in an orthogonal triangle, and the angle $\alpha$ between the incoming beams and the y-axis is described by:

$$\alpha = arcsin\left(\frac{\delta}{m}\right) \tag{1}$$

In order to process the signal from the microphones, it is sampled with the highest possible time resolution that the hardware offers. For the standard audio interfaces of PCs, this is typically 48khz, i.e. 20.8 microseconds between two samples. This consequentially is the shortest delay $\Delta t_{min}$ that the system can distinguish. Together with the speed of sound $c$, this results in a quantization of the differential distance measurements into $c\Delta t_{min}$.

As the distance difference $\delta$ is quantized, the angular resolution of the microphone pair detector significantly differs for different angular areas. Angles near the direction of the normal vector can be measured with a fine resolution and angles in the direction of the line connecting the microphone pair, i.e. outside of the ’focal region’ can only be measured with a high uncertainty.

### D. Angular Resolution

The resolution of the localization is strongly depending on the resolution of $\delta$. Given a sampling frequency $f_s$ and the speed of sound $c$, the maximal time difference of $k$ samples is reached, when a signal is coming from the x-axis outside of the microphone pair. It is:

$$k = \frac{m \cdot f_s}{c} \tag{2}$$

$\delta$ varies from $-k$ to $k$ samples.

Evaluating the angles for all possible $k$, the half-circle in front of the microphone pair can be divided into different zones, a sample resolution is shown in figure 5.

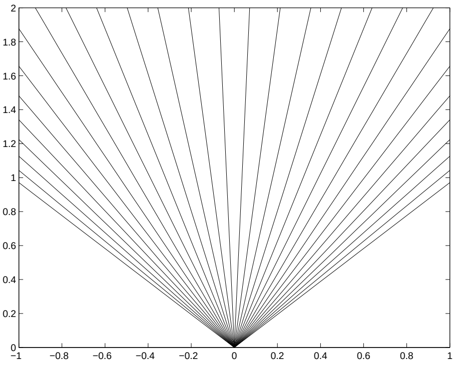

Fig. 5. The angular sectors that can be distinguished by one microphone pair 10cm apart, using 48kHz sampling frequency

### E. Triangulation

Using two microphone pairs, it is possible to use triangulation to determine the position of the measured sound source. The accuracy of this position measurement is strongly depending on the angular accuracy and the distance of the two pairs (base length). As shown, the angular resolution cannot be increased without changing the hardware. Therefore, the base length should be maximized. Using this system on a mobile robot puts a limit to the possible baselength.

Given two lines passing through the points $(-n/2, 0)$ and $(n/2, 0)$ respectively and knowing their angle against the x-axis, their intersection point can be determined. Assuming quantisation in the angles again, the plane in front of the two microphones can be divided into quadrangular sectors. The size of these sectors is a measure for the

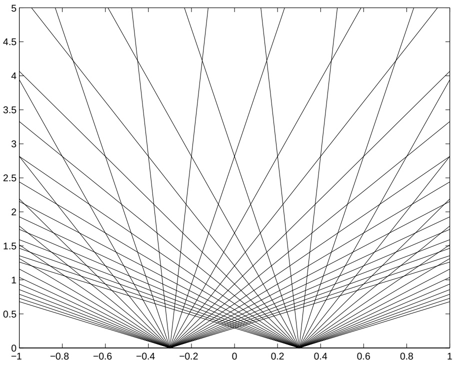

Fig. 6. The sectors that can be distinguished through triangulation, using 2 microphone pairs of the type in Fig. 5, centerpoints 60cm apart

precision of the position localization. Such a sector map can be seen in figure 6.

The base length can be considerably increased using two robots, that can be apart by several meters and whose relative position and orientation is known. By using this information, the location of the sound source can be determined in a common world coordinate system. No particular time synchronization or correlation of the signals of the two robots is needed, as they simply have to measure at approximately the same time and provide the angle measured.

Assuming that the sound source is immobile, these angle measurements can as well be done sequentially by one robot only. The robot needs to measure at one point, move for a precise distance and measure the angle again. Using both angles and the base length, triangulation can be performed.

Assuming that the current pose of the robot is known, the system has sufficient information to create a probabilistic occupancy grid map [12] of the environment in a world coordinate system. Unlike the occupancy grid map used for robot navigation[8], this map does not contain information about the probability of cells being occupied by obstacles but with the probability of cells being the location of a sound source.

This type of map has been chosen over other approaches to probabilitstic mapping ([9],[10],[11] or see [13] for an excellent overview of the topic) as we assume that it is hard to extract features from the sensory input that could be redetected in the future. Moreover, the location of sound sources will not provide much structure as the location of walls in an office environment would give us. As this sensor is not intended for robot self-localization but only for sound source localization, it is assumed that an accurate estimation of the current pose of the robot is provided by other means. Note that this information could be provided through other means of probabilistic mapping and localization such as SLAM[11], but this mapping would then use other sensors such as LADAR.

The probabilistic map building algorithm is implemented in a straight-forward way: For every grid cell a value is calculated that represents its change in probability of being a sound source based on the current sensor data and this value is added to the current value stored for the grid cell.

The calculation of this change in probability does not only depend on the current sensor value but also on the properties of the sensor, i.e. on a sensor model. Here we assume that the sensor only gives good information for sound sources that are neither too faint (i.e. far away) nor are outside of the focus area where angular information is unreliable. Information concerning these areas is ignored.

If a sound source is located within the focus area of the sensor, its signal energy level is compared against a threshold $T$. If there is no signal higher than the treshold, the angular area reaching from the robot to a constant maximum reliability distance $D$ is considered free of sound sources and every cell that has its centerpoint in this angular area receives a negative probabiltity change $-\Delta$. If a sound source is detected in the angular area, every cell receives a positive probability change $\Delta$. The probability values in the cells are then updated and limited to reasonable positive and negative maxima $P_{MAX}$ and $P_{MIN}$.

Initially, all cells are initialized with a value of 0 that corresponds to a maximum of uncertainty for this cell, we neither know that it is a sound source nor we know that it is one.

It can easily be seen that the occupancy grid can solve the triangulation from two different robot poses provided that the sound source is within the detection range from both poses. If the robot is in the first pose A, it will increase the probability value of all cells between its location and its detection range in the direction of the sound source. All other cells within the detection range will receive a decrease in probability value. After a number of sensor readings are analyzed, the probability value for all cells between the robots current position and the sound source will converge to a value of $P_{MAX}$ and all other cells of the grid will either remain 0 or will converge to $P_{MIN}$. If the robot is now moved to a pose B and if the sound source is still in the detection range of the robot, it will further increase the probability value in all cells in between of the current position of the robot and the sound source and it will decrease the probability value for all cells that are not in the direction of the sound source, thus the probability value of all cells in the proximity of the sound source will remain at $P_{MAX}$ and all other cells will either converge to $P_{MIN}$ or remain 0.

Unfortunately, a sector that has received a positive probability from pose A and is not in the detection range from pose B will remain with $P_{MAX}$ probability value. However, this value is misleading as it only depends on a single measurement and therefore is not a true triangulated value. These sectors would lead to false positives, i.e. the detection of a sound source when there is none. In order to eliminate these false positives, additional measures have to be taken. A true triangulation consists of two measurements that use different angular directions to establish the triangulation. To distinguish true triangulations from false positives, the robot taking the measurement and incrementing the probability value in a cell additionally computes an angular sector ID in world coordinates. This angular sector ID is an integer that numbers the angular sectors of the semicircle from 0 to $AS_{MAX}$ so that every direction gets a distinct ID. If a robot finds a different sector ID in the grid cell it is about to increment, it sets a flag in the cell indicating that it contains the result of a true triangulation.

This algorithm uses a number of parameters. The parameters that specify the size of the distinguished angular areas are determined by the geometric properties and the sampling frequency of the sensor. The treshold energy $T$ and the reliability distance $D$ parameters are dependent on the properties of the transmission system formed by the sound sources to be detected, the transmission medium and the microphones. The Parameters $\Delta$, $P_{MAX}$ and $P_{MIN}$ determine the number of iterations that are needed for convergence. Additionally, the model described here assignes the same probability value increase to all grid cells in a sector. This does not reflect the real probabilities as the sector becomes wider when the cells are further away from the sensor. Consequently an individual cell that is further away should receive a linearily lower probability increase than a cell that is close to the sensor, but the simulations have shown that for rather small angular sensors, a fixed value is a reasonable approximation.

## III. Experimental Results

The perfomance of the whole system was evaluated using a combination of simulations and measurements. After the performance of the detector with the presence of (white-gaussian) background noise was simulated and showed the receiver rather immune to this kind of disturbance, the predicted angular resolution of the sensor was verified in a real experiment. Finally, the probabilistic mapping algorithm was again implemented and tested in a simulation.

### A. Sensor performance under noise

The required Signal-to-Noise Ratio (SNR) has been determined by overlaying the incoming two signals with white noise individually.

An input signal of 2000 samples (corresponding to ca. 0.04s at 48kHz sampling frequency) of white noise was created. The signal was delayed to create a signal for the second microphone. Finally, to both microphone signals, additional white gaussian noise was added with different Signal-to-Noise ratios. For each SNR, 100 measurements were performed, and a detection rate of more than 95 per cent required. It could be shown, that the system performes well under these conditions, as long as the added noise keeps the SNR above $-8dB$ at the receiver. The simulation results are shown in figure 7.

Hence we conclude that unstructured background noise has little impact on the sensor system.

### B. Angular resolution

Using a sampling frequency of 48kHz (standard audio input on a soundcard), a microphone apart by a distance of $m = 10cm$, and sound velocity of $c = 341m/s$, a theoretical resolution of 31 distinct steps can be achieved, as shown in figure 5. This setting has been proven to work in a real-life experiment.

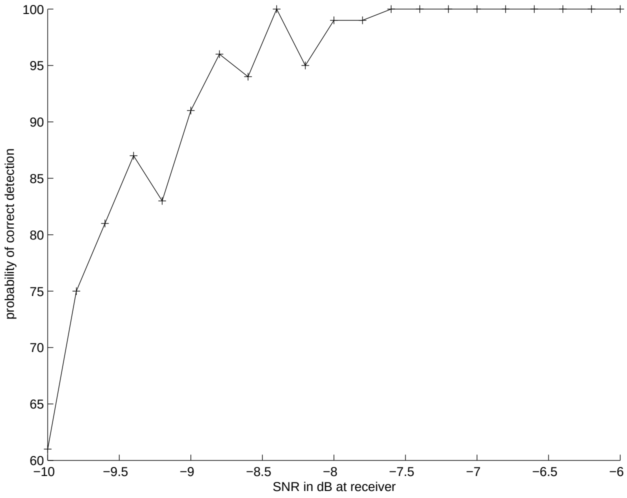

Fig. 7. The simulated detection results for different SNR levels.

The whole aparatus was placed in an office environment, which meant computer background noise, people talking, walking around. A plane in front of the microphones was sampled and the respective delay calculated for each point. The theoretically predicted results could be verified: all different sectors could be distinguished and consistently found in the measurements.

### C. Probabilistic mapping

In order to to estimate the performance of the sensor in a probabilistic grid map, a sensor model has been derived from the data gained so far. The sensor model has 31 different zones that can be distinguished, each zone consisting of two angular sectors in positive and negative y direction as shown in Figure 8. To produce this figure, a sensor in the origin with a normal orientation of 45 degrees and a sound source at position x=2/y=1 was simulated. As the sensor cannot distinguish the exact position of the sensor in the zone, the probability of the presence in the zone is uniformly increased (red areas) and the probability of it not being in any other zone is uniformily decreased (green areas). The sensor is assumed to have a fixed range and for sources that are further away, it is assumed that the source is lost in the background noise, so it will not be detected. From the simulation result, it can be seen that a single sensor measurement is quite ambiguous.

In Figure 9, the simulation results for a sound source from three different sensor positions are shown. In this case, the sound source at position x=1/y=1 is clearly indicated with a positive probability. However, there are other parts of the map that receive positive probability. This occurs due to the fact that these areas are only covered by a single sensor reading, so the probability is increased by the sensor model of that one sensor reading, but is never decreased by the model of another sensor reading.

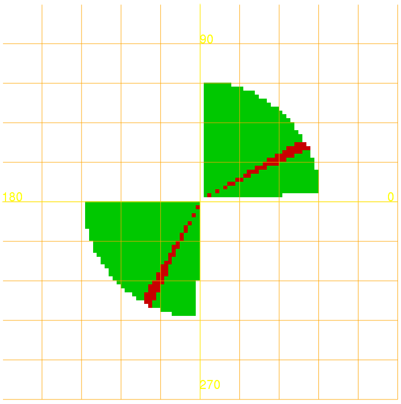

Fig. 8. One of the zones that can be distinguished by the sensor.

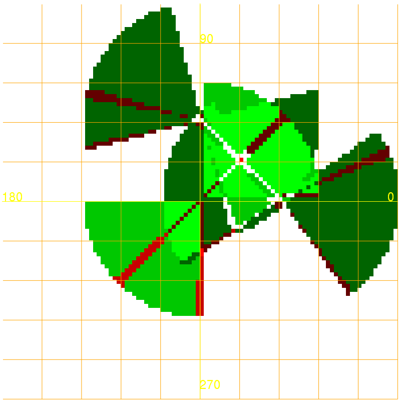

Fig. 9. A simulation of three sensor readings of a single sound source with a simple sensor model.

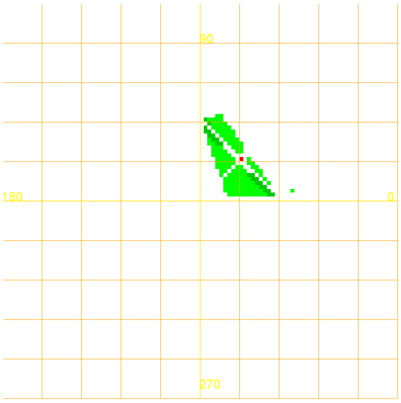

Fig. 10. A simulation of three sensor readings of a single sound source with a better sensor model using a touchcount filter.

This sensor model is formally correct, as there could be indeed three independent sources that are each only detectable by a single sensor. However, it is much more likely that only a single source creates the sensor readings. Therefore, we add an additional counter to each cell of the probabilistic map that counts how many sensor readings have contributed to the final value of the cell. By comparing this value against a threshold and filtering the result by this, we obtain the simulation result shown in Figure 10. Here the sound source can clearly be distinguished as the single point with positive probability that remains.

## IV. Conclusion

The problem of automatic victim localization in RoboCupRescue has been presented. A solution using microphones mounted on mobile robots and differential time-of-flight measurements of sound has been simulated and its accuracy shown to be sufficient in a simple experiment. A mapping algorithm using occupancy grids has been presented based on the experimental finding and it has been shown in simulation that is able to localize a sound source in a global map.

The next step will be the implementation of the sensor on a robot of the IUB Robocup Rescue team and the comparision of the simulation results with the real performance of the sensor. This will be an improvement over current victim localization techniques that are entirely based on human operators.

## References

[1] Mattos, L., Grant, E.:Passive Sonar Applications: Target Tracking and Navigation of an Autonomous Robot. Proceedings of the IEEE International Conference on Robotics and Automation. IEEE Press, 2004.

[2] Kitano, H., Tadokoro, S.: Robocup rescue. a grand challenge for multiagent and intelligent systems. AI Magazine 22 (2001) 39-52

[3] Takahashi, T., Tadokoro: Working with robots in disasters. IEEE Robotics and Automation Magazine 9 (2002) 34-39

[4] Osuka, K., Murphy, R., Schultz, A.: Usar competitions for physically situated robots. IEEE Robotics and Automation Magazine 9 (2002) 26-33

[5] A. Birk, H. Kenn et al., The IUB 2002 Smallsize League Team, Gal Kaminka, Pedro U. Lima and Raul Rojas (Eds), RoboCup 2002: Robot Soccer World Cup VI,LNAI, Springer, 2002

[6] A. Birk, S. Carpin and H. Kenn, The IUB 2003 Rescue Robot Team RoboCup 2003: Robot Soccer World Cup VII, LNAI, Springer, 2003

[7] H. Kenn, S. Carpin et al., FAST-Robots: a rapid-prototyping framework for intelligent mobile robotics, Artificial Intelligence and Applications (AIA 2003), ACTA Press, 2003

[8] S. Carpin, H. Kenn and A. Birk, Autonomous Mapping in the Real Robots Rescue League, RoboCup 2003: Robot Soccer World Cup VII, LNAI, Springer, 2003

[9] McLachlan, G., Krishnan, T.: The EM Algorithm and Extensions. Wiley-Interscience, 1996

[10] Kalman, R.: A new approach to linear filtering and prediction problems. Transactions of ASME. Journal of Basic Engineering 83, 1960

[11] Dissanayake, G., Newman, P., Clark, S., Durrant-Whyte, H., , Csorba., M.: A solution to the simultaneous localisation and map building (slam) problem. IEEE Transactions of Robotics and Automation 17, 229-241, 2001

[12] Moravec, H.P.: Sensor fusion in certainty grids for mobile robots. AI Magazine, 1988

[13] Thrun, S.: Robot mapping: a survey. Technical Report CMU-CS-02-111, Carnegie Mellon University, 2002