Auditory Occupancy Grids with a Mobile Robot

Submitted 7th July 2011; accepted 30th March 2012

Brian P. DeJong

Abstract:

This paper presents the use of auditory occupancy grids (AOGs) for mapping of a mobile robot's acoustic environment. An AOG is a probabilistic map of sound source locations built from multiple measurements using techniques from both probabilistic robotics and sound localization. The mapping is simulated, tested for robustness, and then successfully implemented on a three-microphone mobile robot with four sound sources. Using the robot's inherent advantage of mobility, the AOG correctly locates the sound sources from only nine measurements. The resulting map is then used to intelligently position the robot within the environment and to maintain auditory contact with a moving target.

Keywords: *auditory occupancy grids, sound localization, occupancy grids, mobile robot*

1. Introduction

Robots are becoming ever more common in our homes, offices, factories, military, and emergency-response units. In every application, the more information a robot has about its environment, the more versatile it is. Robots are equipped with cameras, ultrasonic sensors, laser range finders, accelerometers, and microphones. The information obtained is then used for mapping, locating, tracking, and informed decision-making.

While robotic vision has seen great advances, robotic audition is still in its infancy. Yet hearing – specifically, sound location – is not unimportant for robots. Assistive robots need to be able to respond to verbal commands such as "Come here" by finding where "here" is. Mechanic robots need to listen to factory machinery or car engines to detect unwanted or troublesome sounds. Sentry and security robots need to recognize suspicious noises, locate their origin, and investigate. Search and rescue robots need to locate the sound of survivors in smoky buildings and piles of debris.

Compounding the issue is the myriad of noises surrounding the robot. Human bystanders, cars, televisions, radios, plumbing, air vents, and machinery create unwanted noises. The robot itself generates noises from its internal fans, motors, and drive systems. Furthermore, these noises echo off of nearby surfaces, creating phantom sources and locations.

This paper explores how to use one of the robot's unique advantages: its mobility. As a robot moves, it can keep track of the auditory information to generate a sound-based map of the environment. The effect of location-specific echoes and self-generated noise is diminished when information from multiple locations is compiled, while the stationary sound sources materialize out of the noise. What results is a probabilistic contour map of the auditory scene, such as that in Figure 1. Given these *auditory occupancy grids* (AOGs) [1], a robot is better equipped to listen, such as by moving to a quiet location, maintaining a line-of-hearing to a target, or comparing two time-different AOGs of the same area.

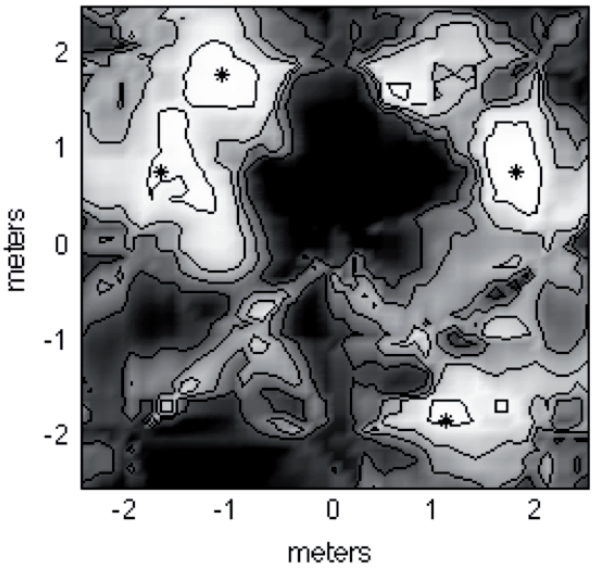

Figure 1. Auditory occupancy grid for the robot's workspace, generated from nine measurements. Lighter color corresponds to higher probability. The four sound sources are designated by asterisks, and are correctly located

The following sections cover the background, simulation, robustness, and implementation of these AOGs. First, we begin by briefly reviewing techniques used in this paper, and discuss related research in robotics. Then, we present the creation of these grids from multiple measurements, both in simulation and implementation with a mobile robot. Given these maps, we explore several of their uses.

2. Background

AOGs combine techniques from two well-established research fields. The mapping component comes from probabilistic robotics while the auditory processing is based on sound localization. We briefly review fundamentals of each, and then discuss the limited research that intersects them.

The term *auditory occupancy grid* is necessary to distinguish these maps from traditional *noise maps*. For

many years, acoustical and safety engineers have used noise maps for modeling potentially hazardous environments or determining noise impact levels for proposed infrastructure [2]. Unlike AOGs, noise maps are straightforward contour maps of sound pressure levels, and do not directly reflect source location.

2.1. Probabilistic robotics

Probabilistic robotics is the application of probability theory to robotic sensing and mobility [3]. For example, suppose a robot is using a laser range finder to locate an object in its workspace. Traditional, deterministic robotics assumes the object to be wherever the sensor measures the strongest return. Probabilistic robotics, on the other hand, represents the sensory information as a probability distribution that inherently includes the uncertainty of the measurement. Standard probability mathematics can then be used to combine multiple measurements from multiple sensors to define a more confident *belief* about the environment.

The first step in this process is the converting of raw sensor data into a probability distribution. This algorithm is called the *inverse sensor model* [4] because it is uses the data measured by the sensor to describe a model for that sensor. A simple example of a sensor model is one that scales the sensor data to a range of probabilities between arbitrary bounds:

$$\left[data_{\min} : data_{\max}\right] \rightarrow \left[p_{\min} : p_{\max}\right]. \quad (1)$$

The bounds represent the highest confidence assigned to any single sensor measurement, since the strongest sensor return is mapped to p_{max} and the weakest to p_{min}. More complicated models exist, such as filtering algorithms. For example, Grabowski *et al.* [5] developed a dynamic inverse sensor model to improve sonar mapping amid specular reflection. Similarly, Thrun [6] used maximum-likelihood filtering of sensor data to improve the mapping and reduce possible conflicts between multiple measurements. Note that the probability distribution is not a probability density function – the probabilities do not necessarily sum to unity.

Once the sensor data is in terms of probabilities, it can be combined with other measurements to improve the robot's belief. Combining multiple probabilities is generally done by a Bayes filter [7]. Given a *prior* belief and a *conditional* sensor measurement, the updated *posterior* belief is

$$p_{posterior} = \frac{p_{prior} \cdot p_{cond}}{p_{prior} \cdot p_{cond} + \left(1 - p_{prior}\right) \cdot \left(1 - p_{cond}\right)}.$$

In numerical algorithms, this equation becomes unstable with values near zero and one, so the log odds (*LO*) method [8] is often used:

$$LO_i = \ln \frac{p_i}{1 - p_i}$$

$$LO_{posterior} = LO_{prior} + LO_{sensor} - LO_{initial}$$

where i is any variable and $LO_{initial}$ is the log odds of the default probability (usually 0.5, or 50%). Then, the posterior probability can be recovered by

$$p_{posterior} = 1 - \frac{1}{1 + e^{LO_{posterior}}}.$$

This is also called the binary Bayes filter.

Probabilistic beliefs are then used for various purposes, such as mapping the robot's environment, location of the robot within that environment, tracking of objects, or planning of paths or control.

In this paper, however, we focus on the mapping and assume accurate knowledge of the robot's pose. Mapping of the auditory scene is done by a grid of cells where each cell has a probability of being occupied by an object (first introduced by Elfes [9]). This grid is called an *occupancy grid* [3], *certainty grid* [10], or *evidence grid* [11], and is usually used for mapping of the physical environment.

Occupancy grids are most useful for stationary, *i.e.*, static, environments, although there has been some recent research in applying them to dynamic situations. For example, Wolf *et al.* [12] used separate static and dynamic occupancy grids to map a robot's environment, while using a landmark-occupancy grid to locate the robot within it.

2.2. Sound localization

While occupancy grids come from probabilistic robotics, the auditory processing techniques used in AOGs come from the well-established field of sound localization.

Sound can be localized based on two or more recordings at known locations. Suppose a source is emitting a sound, $s(t)$, which is recorded by two microphones of known locations. The microphones record different sounds, $m_a(t)$ and $m_b(t)$, each containing an attenuated

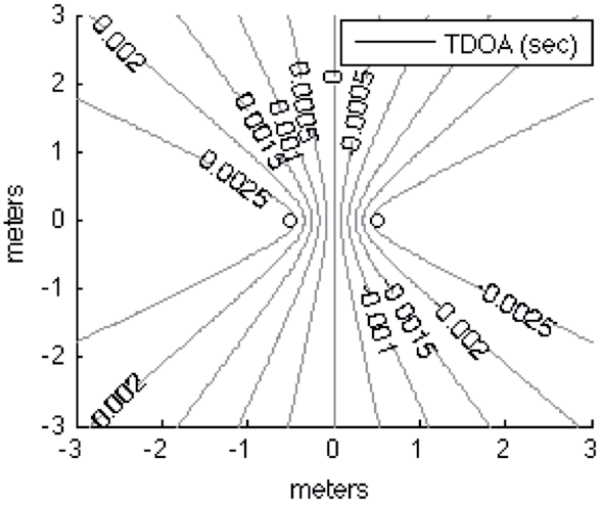

Figure 2. Sample Time Difference on Arrival (TDOA) for two microphones spaced 1 meter apart, assuming the sound travels 343 m/s

[1] The frequency domain equation follows directly from the time domain equation given properties of Fourier transforms and cross correlations.

(due to distance, medium, directionality, etc.) and time-delayed version of the source with unknown noise. Given the time difference between the two recordings, called the Time Difference on Arrival (TDOA), the source must be located somewhere on a surface in space that corresponds to that TDOA. In two dimensions, this surface is a parabola – see Figure 2; in *n*-dimensional space, one microphone pair restricts the source to a one-less-dimensional ((*n*-1)-dimension) space.

The TDOA between two microphones is found by performing a cross correlation of the signals, either in the time domain,

$$corr_t(\tau) = \int_\tau m_a(t) m_b(t-\tau) d\tau ,$$

or in the frequency domain,

$$corr_t(\tau) = \int_\omega M_a(\omega)\overline{M_b(\omega)} \cdot e^{-j\omega\tau} d\omega ,$$

where τ is a time shift.[1] Here, $M_a(\omega)$ and $M_b(\omega)$ are the Fourier transforms of microphone recordings $m_a(t)$ and $m_b(t)$, $\overline{M_b(\omega)}$ is the complex conjugate of $M_b(\omega)$, and $j = \sqrt{-1}$. The best estimate for the TDOA is thus the τ that maximizes the cross correlation.

The frequency domain technique is standard because of its accuracy. The accuracy in the time domain is restricted by the sampling frequency: $m_i(t\text{-}\tau)$ must be known, so τ must be in timestep increments. Figure 3 shows the effect of this increment for a sample microphone placement and a sampling time of 5000 Hz. On the other hand, any τ can be tested in the frequency domain, so the frequency domain analysis allows for more accurate calculation of the TDOA. However, the frequency analysis requires a Fourier transform of the signals, which can be computationally intensive in near-real-time applications.

In most applications, three or more microphones are used. Given *n* microphones, there exist (*n*-1) independent pairings and respective TDOAs. Thus, with three microphones in two dimensions, the source can be triangulated to the intersection of the two parabolas. In three

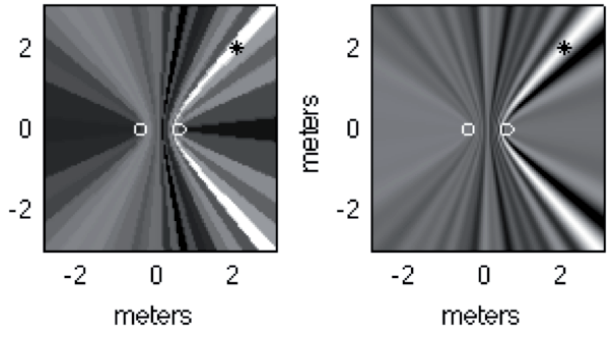

Figure 3. Comparison between (a) time domain and (b) frequency domain cross correlations for 5000 Hz signals. The color bands correspond to points in space that have the same corr(τ) values. The time domain is restricted by the sampling frequency while the frequency domain can be more detailed given finer increments of τ

dimensions, four microphones are needed. Additional microphones (*e.g.,* eight [13,14], thirty-two [15], or even 128 [16]) are often used to provide more comparisons and improve robustness.

Another way of approaching the localization is to assign a likelihood value to each position in space based on the time delay corresponding to that position. The Steered Response Power (SRP) [17] for microphone pair (*a*,*b*) at each robot location $\boldsymbol{x}$ is defined as

$$\boldsymbol{F}_{ab}(\boldsymbol{x}) = \int_\omega M_a(\omega)\overline{M_b(\omega)} \cdot e^{-j\omega\tau_x} d\omega ,$$

where τ_x is the expected TDOA for that location. These correlations add to give the SRP given all microphone pairs:

$$\boldsymbol{F}(\boldsymbol{x}) = \sum_{a=1}^{n}\sum_{b=1}^{n}\int_\omega M_a(\omega)\overline{M_b(\omega)} \cdot e^{-j\omega\tau_x} d\omega . \quad (3)$$

This *likelihood function* thus maps the probabilities around the microphone array.

The SRP can be tuned to certain frequencies by including a prefiltering weighting in the cross correlation, $W(\omega)$:

$$\boldsymbol{F}(\boldsymbol{x}) = \sum_{a=1}^{n}\sum_{b=1}^{n}\int_\omega W(\omega) M_a(\omega)\overline{M_b(\omega)} \cdot e^{-j\omega\tau_x} d\omega . \quad (4)$$

This is called the Generalized Cross Correlation (GCC) [18]. The most common weighting is the Phase Transform (PHAT) [14,19],

$$W_{PHAT}(\omega) = \frac{1}{|M_a(\omega)|\,|\overline{M_b(\omega)}|} ,$$

because it reduces the effect of echoes – the echoed frequencies appear stronger in the microphone recordings but are normalized by the weighting term. The PHAT weighting also reduces the effect of unequal preamplifiers on the microphone signals. With a PHAT weighting, the GCC reduces to an integral of phase-based terms.

In the literature, there has been much work done on sound localization, although usually with set of stationary microphones distributed around a room. For example, Stillman and Essa [20] used a four-microphone array for localization in a smart room. Mungamuru and Aarabi [21] used models of source and microphone directivity to improve upon the GCC PHAT algorithm for stationary microphone arrays distributed about a simulated room. Also, Aarabi [19] combined ten weighted likelihood maps from two-microphone arrays to locate three sound sources.

2.3. Auditory sensing in robots

There has been some research on sound localization using human-like robots. For example, Nakadai *et al.* [22] built a biologically inspired two-microphone humanoid robot that localizes sound sources using inter-aural phase and intensity differences. Unfortunately, they found that the front-back ambiguity couldn't be solved without active audition [23], such as by rotating the microphone array [24]. Huang *et al.* [25] used a three-microphone array and biomimetic processing to determine direction

to two sound sources. Implementing it on a mobile robot, they then combined visual and auditory information to localize a human speaking [26].

For tracking sound sources from a mobile robot, Valin *et al.* [13] used an eight-microphone array, beamforming, and particle filtering. (In a previous work [27], they implemented a simplified method for determining the angle to a source based on a far-field assumption.) Likewise, Sasaki *et al.* [15] used a thirty-two-microphone array on a mobile robot to map two moving sound sources.

Using auditory evidence grids as an intersection between probabilistic robotics and sound localization was first proposed by Martinson and his colleagues. They implemented auditory evidence grids on a mobile robot via a four-microphone array. They were successful in locating two sources, but needed post-processing to locate three [28]. They were also able to determine source volume and directionality by having the robot approach each source to investigate it [29]. In addition, they have successfully applied their mapping to human tracking and speech recognition, for improved human-robot interaction [30].

3. Creating AOGs on a mobile robot

Giving a mobile robot the ability to localize sound is not necessarily easy. In many applications, the robot's environment does not have a previously distributed static microphone array, and the mobile robot must carry the array with it. All sound measurements must be taken at the robot's location with limited distance between microphones, greatly reducing the accuracy of localization. Figure 4 shows a comparison of sound localization for fixed three-microphone arrays of two different radii. The figure also illustrates another common issue with closely spaced microphone arrays: the array can locate the angle to the source accurately, but not the distance.

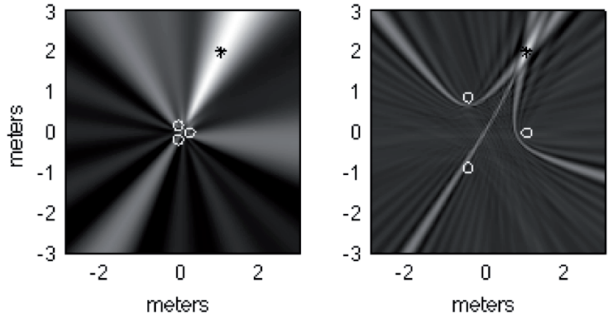

Figure 4. Sound localization for two arrays of different radii. The closer array has poorer localization, especially along the radial direction

A mobile robot, however, has the advantage of mobility over static microphone arrays. As the robot traverses its environment, it can take multiple sound measurements that can be fused together to make a clearer, more accurate occupancy grid of the auditory scene. This AOG is similar to (but not directly analogous to) an elevation map of an area.

The following sections detail a validation of the AOG concept. First, we *simulate* the robot, microphones, and environment to verify the AOG algorithm. Second, we map randomly-located sound sources to get a sense of its *robustness*. Third, we *implement* it on a mobile robot for one environment, as a proof of concept. While these results are not meant as an in-depth experiment, they do show that AOGs are successful in many environments. For both simulation and implementation, we use a three-microphone array for simplicity; however, the algorithm holds for any number of microphones.

3.1. Simulation

Let us illustrate the procedure via the simulation we created in MATLAB. Suppose we have a mobile robot in a two-dimensional 5x5 meter workspace with four omnidirectional, dimensionless sound sources (see Figure 5). We grid the workspace into 0.1x0.1 meter cells; for each cell we want to determine the probability that it holds a sound source. Thus, each cell initially has a probability of 0.5. This grid is the AOG.

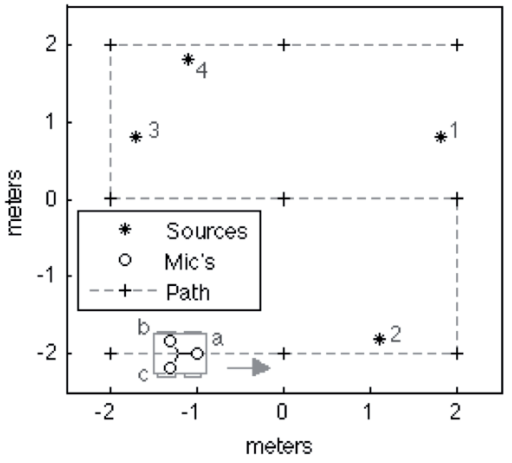

Figure 5. The three-microphone (labeled a-c) mobile robot's workspace, showing sound source locations (numbered 1-4) and path (plus signs represent measurement locations).

For the simulation, the four sound sources were modeled as omnidirectional signals consisting of a sum of 250 sinewaves (p) with uniformly distributed frequency (ω_p) and phase (ϕ_p), and with pressure attenuation based solely on distance between source k and microphone i ($d_{k,i}$). We numbered the sources as *1-4* (*e.g.* Figure 5). For each microphone recording, the sources were individually time-delayed (τ) based on distance.

The robot carried a three omnidirectional microphone array, equally distributed on a 0.2-meter flat circle centered at the robot's center (Figure 5). We labeled the microphones as *a-c*. Because the sound signals were sums of sinewaves rather than recordings, they could be time-shifted analytically to create omnidirectional microphone signals ($m_a(t)$, $m_b(t)$, and $m_c(t)$). For each microphone, the four source signals were then combined along with uniformly distributed random noise ($n(t)$). The general magnitudes of the sources (c_k) were 1.0, 0.5, 0.5, and 1.0, with a noise magnitude of 1.0 for each. Mathematically, source k was

$$s_k(t) = c_k \sum_{p=1}^{250} c_p \sin\left(\omega_p t + \varphi_p\right),$$

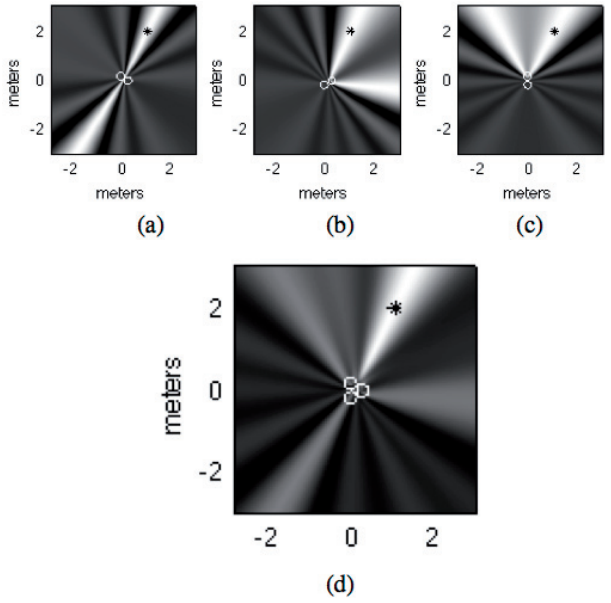

Figure 6. Likelihood maps for one source, for (a) microphones a and c, (b) microphones a and b, (c) microphones b and c, and (d) combined

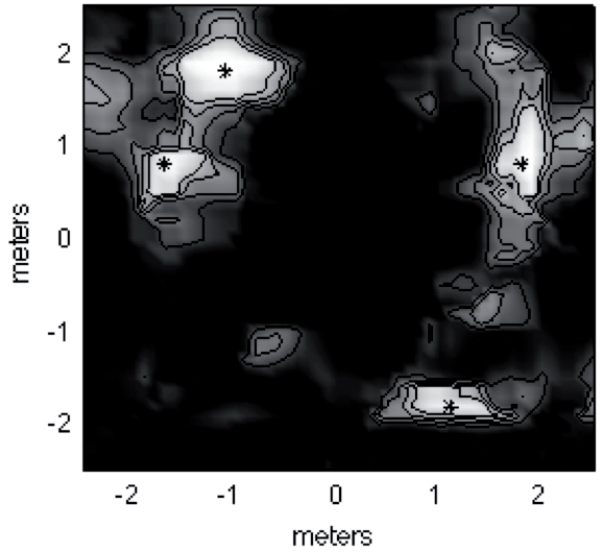

Figure 7. Auditory occupancy grid for simulated robot and sources (represented by asterisks)

where c_k is the general magnitude of source k (1.0, 0.5, 0.5, or 1.0), and c_p is the magnitude of sinewave p (randomly distributed). Thus, microphone recording for microphone i was

$$m_i(t) = \text{noise}(t) + \sum_{sources} \text{attenuated\_and\_delayed\_}s(t)$$
$$= n_i(t) + \sum_{k=1}^{4} \frac{c_k}{d_{k,i}} \sum_{p=1}^{250} c_p \sin\left(\omega_p \left(t - \tau_{i,k}\right) + \varphi_p\right),$$

These source and microphone models are simplistic, but serve the purpose of our simulation. The simulation also ignores any robot-generated noise (such as from onboard motors or fans).

At each robot pose throughout the workspace, we obtain a 3-meter-radius likelihood map for sound sources, $F(\boldsymbol{x})$ (Equation 4), for the surrounding area based on 1024-sample simulated microphone recordings. This is approximately equal to a sampling rate of 5000 samples per second (the maximum sampling rate for our physical implementation discussed later). Thus, frequencies above a frequency of 2500 Hz (which includes much of the audible range) can cause aliasing. The likelihood map is centered on the robot and covers a subset of the overall workspace. We chose a radius of 3 meters based empirically on the useful range of our microphone array. To calculate $F(\boldsymbol{x})$, we used the GCC with PHAT weighting (Equation 4) using Fast Fourier Transforms. Because the position-dependent time delays, τ_x, are independent of the microphone recordings, we calculated them beforehand.

Figure 6 shows a sample likelihood map for a situation with only one sound source. Each microphone pair defines a parabola of high likelihoods. When combined, the result is one area of high likelihood. The three weaker arcs in Figure 6d are artifacts of the individual pairings. Note that the simulation involves more complicated likelihood maps, since the multiple sources create multiple parabolas in each microphone pairing.

Next, the likelihood values are converted to probabilities to form a single-measurement probabilistic map, which we call the *sensor grid*. Our inverse sensor model scales the likelihood values to a range between 0.1 and 0.9 probability (as in Equation 1), based on the minimum and maximum values seen at that the corresponding robot pose. This in an arbitrary model and has implications on the quality of the AOG, as we will discuss later. Mathematically, our scaling for any likelihood, $F(\boldsymbol{x})$, is

$$p_{sensor}(\boldsymbol{x}) = 0.1 + \frac{0.9 - 0.1}{F_{\max} - F_{\min}} \cdot \left(F(\boldsymbol{x}) - F_{\min}\right),$$

where F_{max} and F_{min} are the maximum and minimum likelihoods.

The simulated robot traversed the space as shown previously in Figure 5, generating sensor maps and fusing them together using the binary Bayes filter (Equation 2).

The resulting AOG is shown in Figure 7. In the figure, color corresponds to probability, with white representing highest probability and the highest contour lines representing 90% confidence (contours are at 23, 45, 68, and 90%). Visually, the algorithm successfully located the four sources, even though the microphones are tightly grouped on the mobile robot. Furthermore, while each likelihood map is relatively coarse (Figure 6d), the resulting AOG is quite accurate.

To localize the sound sources from the AOG, we used a threshold of 70% probability to isolate peaks. This threshold can significantly affect the number and locations of the sources – in general, increasing the threshold decreases the number of sources found, and can increase or decrease the location accuracy depending on the shape of the peaks. The 70% used here was chosen empirically from the robustness test discussed later as a robust and consistent threshold across various workspaces. Here, using averages for the peaks weighted based on each

nearby point's probability,

$$\boldsymbol{x}_{peak} = \frac{\sum p_{\boldsymbol{x}_i} \boldsymbol{x}_i}{\sum p_{\boldsymbol{x}_i}}.$$

we calculate each peak's weighted center. The four sources were localized within 0.19, 4.3, 4.1, and 4.3 cm, respectively.

3.2. Robustness of the algorithm

As a test of the robustness of the AOG algorithm to source location and distribution, we simulated 100 four-sound-source environments and located the sources in them. For each environment, the four sound sources were randomly placed among the nine possible locations shown in Figure 8. The AOG simulation mapped each acoustic environment, and then located sources based on the resulting peaks. From initial tests, we empirically chose a threshold of 70% and a minimum peak weight of 1.5 (units of probability).

The results show that AOGs are relatively robust to source positioning in the environment. Of the 100 tested trials, the algorithm correctly found four sources 65% of the time, and incorrectly found three or five sources 19% and 16% of the time, respectively. However, in the three-source results there were no false positives (all sources found were valid), and in the five-source results there were no false negatives (all valid sources were found). Thus, the AOG located 381 out of the 400 sources (95%).

For the correctly located sources, the AOGs averaged a maximum localization error per environment of 8.7 cm, with the worst localization error of 22 cm. Figure 9 shows the percent of sources found with each error, and within each error (probability and cumulative density functions). The majority of sources were found with very small errors: over 50% were found within 3 cm, while 90% of the sources were found within 9 cm. This is surprisingly accurate for only nine sound measurements.

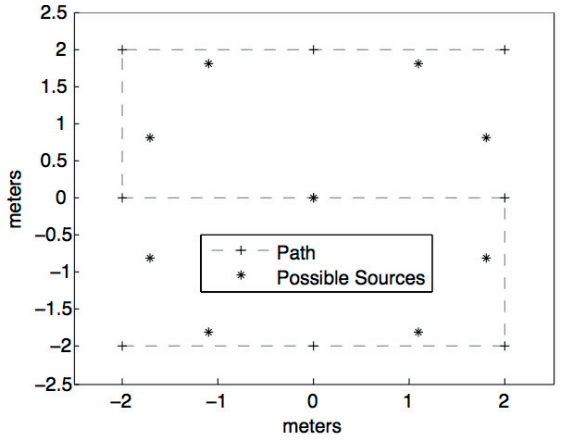

Figure 8. Robot's path (same as before) and possible source locations for the robustness test

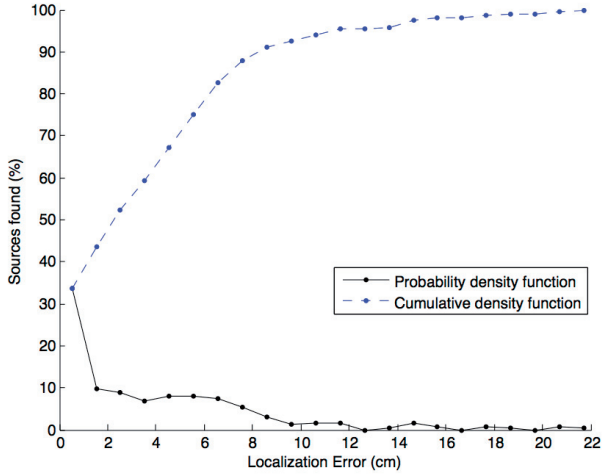

Figure 9. Number of sources found in the robustness test with each error value (probability density), and within each error value (cumulative density)

3.3. Implementation

Next, we implemented AOGs on an in-house wireless mobile robot (see Figure 10) as a proof of concept. Our goal was to validate the simulation and algorithm, not to conduct a full experiment.

The robotic system consists of two computers running MATLAB's xPC Target software: a *host* desktop where the algorithms are programmed, and the robot's onboard *target* PC-104 stack. The robot's computer includes an Advantech PCM-3375 CPU (533 MHz) carrying a 512 MB CompactFlash card as a hard drive, a Sensoray 526 DAQ board, and miscellaneous input/output/protection circuitry. Algorithms are programmed on the host computer in Simulink, compiled into C code via Real-Time Workshop, ported to the robot via xPC, and run there as a kernel. Data can be ported back to the host computer for additional analysis and plotting.

The robot uses three Audio-Technica MT830c omnidirectional condenser lavalier microphones and drive motor encoders. The microphones are attached to 20-cm arms extending radially, as modeled in the simulation, with onboard preamplifiers.

The four sound sources were selected to give a range

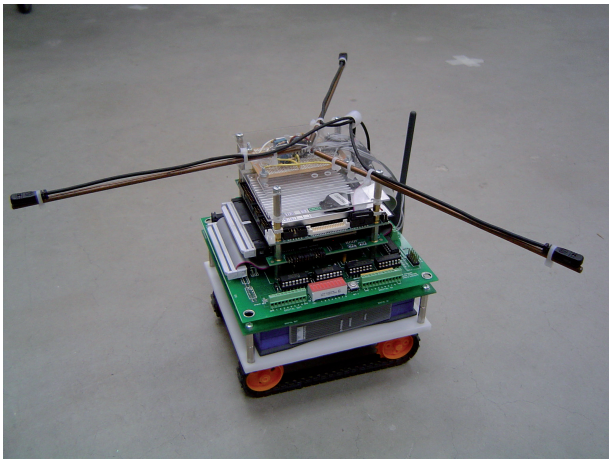

Figure 10. Photograph of the mobile robot, with PC-104 stack, three-microphone array, and wireless Ethernet bridge

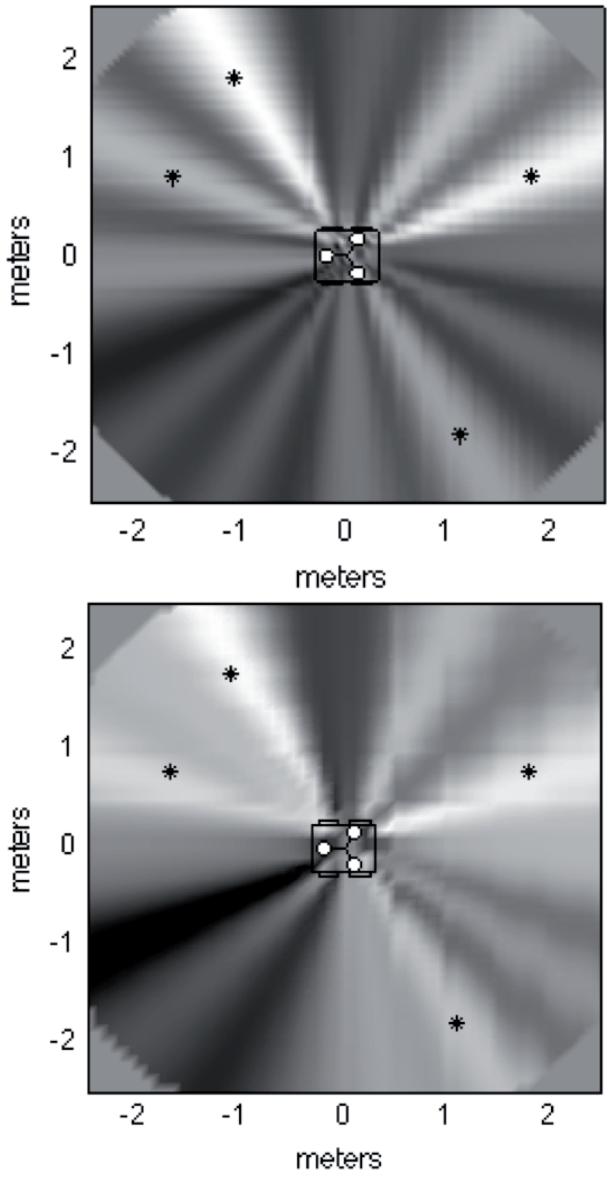

Figure 11. Sensor maps for the central location in the workspace, for (a) the simulation and (b) the implementation. Color corresponds to probability, from white for 90% to black for 10%. The white beams correctly point to the sources (asterisks)

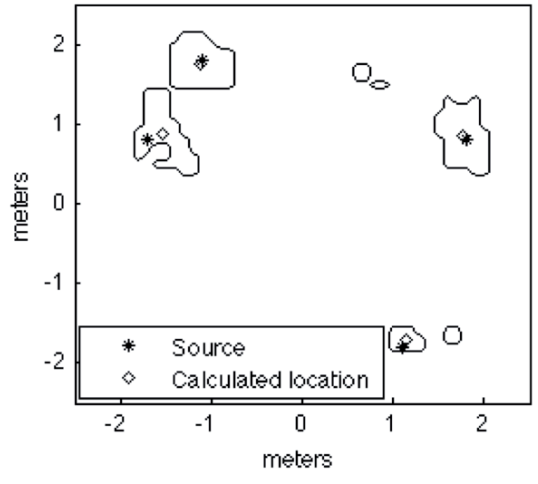

Figure 12. Sources' locations as calculated from the AOG. All sources were correctly located within 20 cm

of sounds typical of a mobile robot's environment. They were (recall the numbering as shown in Figure 5):

1. A knocking at approximately 5 Hz, to create distinct high-frequency beats in the microphone recordings.
2. A recording of classical music, with a mix of long tones and regular beats.
3. A recording of a human speaking, with irregular patterns and frequencies.
4. An electric shaver, with relatively constant frequencies.

The sound sources were not all at the same volume – source 1 (tapping) and source 4 (shaver) were both louder than the other two (as in the simulation). The sources were positioned facing up (perpendicular to the robot's workspace) to minimize directionality of the sound, since the current algorithm assumes omnidirectionality.

The robot traveled around the 5x5 meter workspace as shown previously in Figure 5. At each location, it briefly stopped moving and recorded 0.2 seconds of sound at 5000 Hz, generated a sensor grid for that location from a 1024-sample correlation window, and updated its AOG.

Figure 11 shows a comparison between sensor grids from the simulated and real robots, when the robot is in the center of its workspace. In the plots, the color corresponds to probability of a source, with white representing 90% probability, to black representing 10% probability. The sensor maps correctly locate the sounds sources since the white beams point to them. The sensor maps in Figure are similar, giving credence to our simulation. Recall that the simulation used simple models for the sound sources and microphones.

The overall AOG is shown previously in Figure 1, where lighter colors correspond to higher probability. The map correctly locates the sound sources, with the highest contour lines representing 97% confidence (contours are at 24, 49, 73 and 97%). Even though the four sources had various volumes, they were all strongly located (see Figure 12): distance errors were 10.3, 7.3, 8.4, and 0.4 cm, respectively. This AOG is similar to the one from the simulation.

In addition, two small peaks can be seen in the bottom right and upper right of the grid (in both the simulation and the implementation) – these are false peaks can be eliminated with further measurements, or simply based on their size. From analyzing the map as it was generated, it appears that these peaks are artifacts of the sensor grids radial-direction inaccuracy.

The AOGs are still effective at locating the sound sources in this application when using time-domain correlations. Figure 13 shows the robot's AOG using time-domain correlations (same contours). While not as accurate as the previous GCC version, it still locates the four sources from nine measurements. The sensor maps generated during the time-domain analysis are coarser (recall Figure 3), and unweighted like the GCC, yet the AOG algorithm is still successful. This result is promising for applications where the Fourier transforms needed by the GCC are too computationally intensive.

The AOGs show that the robot successfully located the four various sources with only nine measurements. The robot now has an understanding of the auditory land-

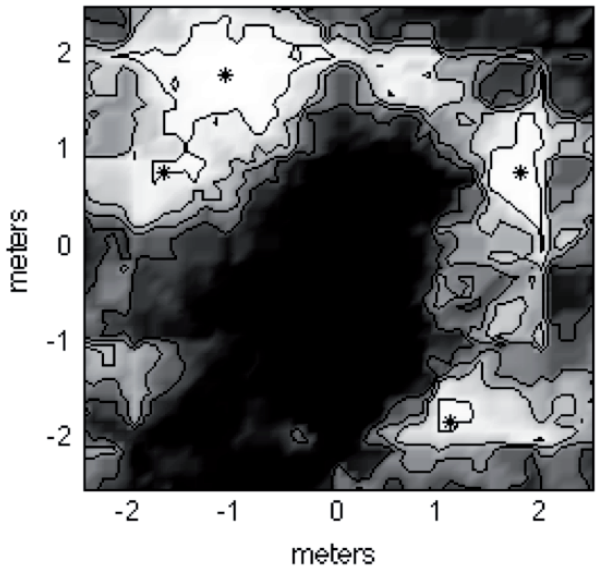

Figure 13. Auditory occupancy grid for the robot's workspace using time-domain correlations. The sound sources are still located

scape and can use that knowledge for various tasks. In the following section, we present and implement several uses for the AOG.

4. Using AOGs

We now discuss several uses for the AOG, using the real-world AOG found via the implementation.

4.1. Moving to a sourceless location

In some situations, such as when the robot is listening for commands or suspicious sounds, it may be beneficial for the robot to move to a position of low source-probability. This is not the same as moving to the quietest location in the workspace, but it is moving to the area that is least likely to contain an interfering source. We can find this location from the AOG, using a similar method to the source locating. Using a threshold of 5% probability and weighted average of the largest space, we get the map

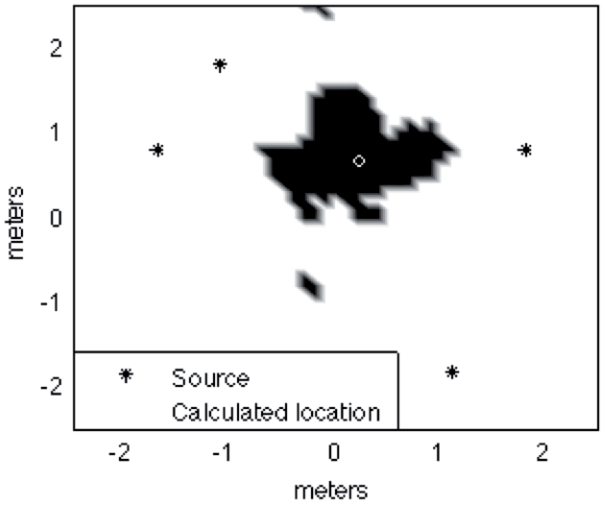

Figure 14. Sourceless location as calculated from the AOG

and location shown in Figure 14. When commanded, the robot moved to this location.

4.2. Line of hearing

Another situation that benefits from the acoustical knowledge is when attempting to track a moving source that the robot is listening to. Suppose the robot is trying to listen to a source of interest, such as a human speaking. Clearly, the best method for hearing the human is to move next to it. But suppose the human is moving in the given workspace across the bottom border from left to right, while the robot is constrained to the top border – the robot is listening to the human from across the room.

In this application, the robot can use standard beamforming techniques [13] to focus the listening towards the human, but it is also advantageous for the robot to position itself along its border such that there are no sources between it and the human. That is, sources between the robot and human will hurt the robot's listening capability.

The robot can use its mobility and the AOG to maintain a low-source-probability "line-of-listening" with the source of interest.

We implemented this with a discrete path-step algorithm. A human moved across the bottom border from left to right at a constant speed. The robot started on the left of the top border and moved intelligently along that border, at twice the speed of the human. At each time step, the robot had several possible locations to move to – to the left or to the right – based on its speed (and bounded by the edges of the map). For each possible location, there exists a line-of-listening, $l(\boldsymbol{x})$ with n points, from robot to human made up of map locations and their corresponding probabilities. The robot assigned a weight, $w(\boldsymbol{x})$, to each possible location based on the probability

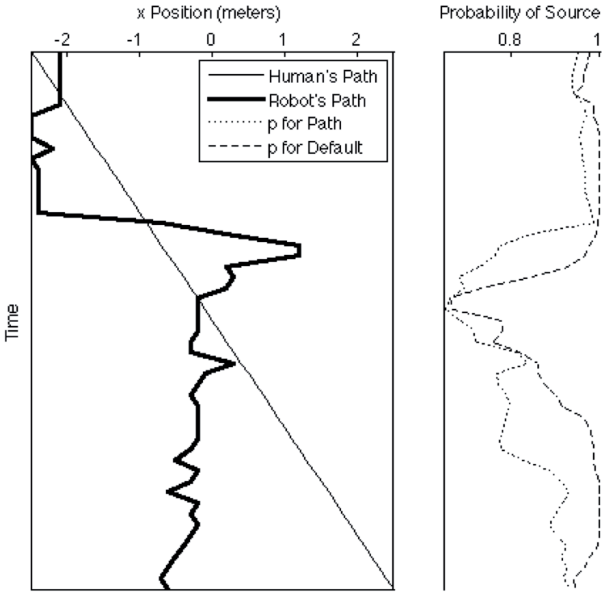

Figure 15. Line-of-listening tracking of human speaker. The human moved along one border, while the robot moved along the opposite so as to minimize the probability of a sound source between them. The second plot compares the intelligent tracking to default tracking

that a source exists somewhere on that line, by combining each point's probability via standard statistics:

$$p(1 \cup 2) = p(1) + p(2) - p(1)p(2)$$
$$w(\boldsymbol{x}) = p\big(l(\boldsymbol{x})\big) = p(1 \cup 2 \cup 3 ... \cup n) \ .$$

It then moved to the position with the lowest weight. This weighting could be of another form, such as maximum probability seen, average probability, etc. Figure 15 shows the path taken by the human and robot. It also shows the probability of a source along the line-of-listening for both a default exact-position tracking and the intelligent weighted tracking. The weighted tracking reduced the probability of sound interference, and thus improved the ability of the robot to listen to the human.

5. Conclusions and Future Work

We have successfully mapped four sound sources in two dimensions using a three-microphone mobile robot. The simulation, robustness test, and implementation show that high quality, accurate AOGs are achievable from only nine measurements, for various source locations. We have also successfully demonstrated several uses for these auditory maps.

We are continuing to research AOGs. For example, the AOGs presented here were created when our mobile robot paused its motion to maintain position and eliminate self-generated noise while the sound segments were recorded. This procedure may not be desirable in all applications. Can the maps be generated on the fly without any pauses?

In addition, the robot's workspace did not include any sound-echoing surfaces or obstacles that typically degrade a robot's listening capability. The AOGs should be robust enough to strong reverberations – are they?

AOGs could prove useful in some three-dimensional applications, such as factories or car engines. Once again, the AOGs should be easily extendable – are they?

One limitation of all occupancy grids is that they are designed to record static information. As mentioned earlier, work by Wolf [12] and others (*e.g.,* [31]) suggest that occupancy grids can be used for dynamic environment, and there has been limited robotic sound localization of moving sources (*e.g.,* [13,15]). How can AOGs be applied to dynamic environments?

The inverse sensor model used here was a simple scaling from likelihood values (from the cross correlation) to a probability range. This model has limitations when the likelihood map is near unity. For example, if the robot is far from all sources, the likelihood map will be near zero – each cell receives a low likelihood from the cross correlation. With our sensor model, however, the slightly-more-likely cells are scaled to high probabilities, even though the likelihood map implies that they probably don't contain sources. One way to mitigate this effect is to use the maximum and minimum likelihoods seen at any robot pose as the boundaries for the scaling. However, this is *post hoc* information, and we found it gives too much weight to only one or two likelihood maps. How can the inverse sensor model be improved?

Finally, in this paper we have demonstrated only a few sample applications for AOGs, although many more exist. What lessons can be learned from the application of AOGs to new situations?

AUTHOR

Brian P. DeJong – Central Michigan University, Mount Pleasant, MI 48859, USA;
e-mail: b.dejong@cmich.edu

References

[1] DeJong B.P., "Auditory occupancy grids: Sound localization on a mobile robot". In: *IASTED International Conference on Robotics and Applications,* 2010.

[2] Martinson E., Arkin R., "Noise maps for acoustically sensitive navigation". In: *Society of Photo-Optical Instrumentation Engineers*, 2004.

[3] Thrun S., "Probabilistic algorithms in robotics", *AI Magazine,* vol. 21, no. 4, 2000.

[4] Thrun S., Burgard W., Fox D., *Probabilistic Robotics,* MIT Press, Cambridge, MA, 2005.

[5] Grabowski R., Khosla P., Choset H., "Autonomous exploration via regions of interest". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2003.

[6] Thrun S., "Learning occupancy grid maps with forward sensor models", *Autonomous Robots,* vol. 15, no. 2, 2003.

[7] Leon-Garcia A., *Probability and Random Processes for Electrical Engineering,* 2nd ed. Addison-Wesley, 2004.

[8] West M., Harrison J., *Bayesian Forecasting and Dynamic Models,* 2nd ed. Springer-Verlag, New York, 1997.

[9] A. Elfes, "Sonar-based real-world mapping and navigation", *IEEE Transactions of Robotics and Automation,* 1987.

[10] Moravec H.P., "Sensor fusion in certainty grids for mobile robots", *AI Magazine,* vol. 9, no. 2, 1988.

[11] Schultz A.C., Adams W., "Continuous localization using evidence grids". In: *IEEE International Conference on Robotics and Automation*, 1998.

[12] Wolf D.F., Sukhatme G.S., "Mobile robot simultaneous localization and mapping in dynamic environments", *Autonomous Robots*, 2005.

[13] Valin J.-M., Michaud F., Rouat J., "Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering", *Robotics and Autonomous Systems,* vol. 55, no. 3, 2007.

[14] Rabinkin D.V., Renomeron R.J., Dahl A., *et al.*, "A DSP implementation of source location using microphone arrays", *The Journal of the Acoustical Society of America,* vol. 99, no. 4, 1996.

[15] Sasaki Y., Kagami S., Mizoguchi H., "Multiple sound source mapping for a mobile robot by self-motion triangulation". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2006.

[16] Kagami S., Tamai Y., Mizoguchi H., and Kanade T., "Microphone array for 2D sound localization

and capture". In: *IEEE International Conference on Robotics and Automation*, 2004.
[17] DiBiase J.H., Silverman H.F., Brandstein M.S., "Robust localization in reverberant rooms*", Microphone Arrays: Signal Processing Techniques and Applications,* Springer, 2001.
[18] Knapp C.H., Carter G., "The generalized correlation method for estimation of time delay", *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1976.
[19] Aarabi P., "The fusion of distributed microphone arrays for sound localization". *Journal of Applied Signal Processing,* vol. 4, 2003.
[20] Stillman S. , Essa I., "Towards reliable multimodal sensing in aware environments". In: *ACM International Conference,* 2001.
[21] Mungamuru B. , Aarabi P., "Enhanced sound localization", *IEEE Transactions on Systems, Man, and Cybernetics,* vol. 34, no. 3, 2004.
[22] Nakadai K., Matsuura D., Okuno H., Kitano H., "Applying scattering theory to robot audition system: Robust sound source localization and extraction". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2003.
[23] Nakadai K., Lourens T., Okuno H., Kitano H., "Active audition for humanoid". In: *National Conference on Artificial Intelligence*, 2000.
[24] Nakadai K., Hidai K., Okuno H., Kitano H., "Epipolar geometry based sound localization and extraction for humanoid audition". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2001.
[25] Huang J. , Ohnishi N., Sugie N., "A biomimetic system for localization and separation of multiple sound sources", *IEEE Transactions on Instrumentation and Measurement,* vol. 44, no. 3, 1995.
[26] Huang J., Supaongprapa T., Terakura I., Wang F., Ohnishi N., Sugie N., "A model based sound localization system and its application to robot navigation", *Robotics and Autonomous Systems*, no. 27, 1999.
[27] Valin J.-M., Michaud F., Rouat J., Letourneau D., "Robust sound source localization using a microphone array on a mobile robot". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2003.
[28] Martinson E., Schultz A., "Auditory evidence grids". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2006.
[29] Martinson E., Schultz A., "Robotic discovery of the auditory scene". In: *IEEE International Conference on Robotics and Automation*, 2007.
[30] Martinson E., "Improving human-robot interaction through adaption to the auditory scene". In: *ACM/IEEE International Conference on Human-Robot Interaction*, 2007.
[31] Graves K., Adams W., Schultz A., "Continuous localization in changing environments". In: *IEEE International Symposium on Computational Intelligence in Robotics and Automation*, 1997.