

# Robotic Discovery of the Auditory Scene

E. Martinson and A. Schultz

**Abstract** - In this work, we describe an autonomous mobile robotic system for finding and investigating ambient noise sources in the environment. Motivated by the large negative effect of ambient noise sources on robot audition, the long-term goal is to provide awareness of the auditory scene to a robot, so that it may more effectively act to filter out the interference or re-position itself to increase the signal-to-noise ratio. Here, we concentrate on the discovery of new sources of sound through the use of mobility and directed investigation. This is performed in a two-step process. In the first step, a mobile robot first explores the surrounding acoustical environment, creating evidence grid representations to localize the most influential sound sources in the auditory scene. Then in the second step, the robot investigates each potential sound source location in the environment so as to improve the localization result, and identify volume and directionality characteristics of the sound source. Once every source has been investigated, a noise map of the entire auditory scene is created for use by the robot in avoiding areas of loud ambient noise when performing an auditory task.

**Index Terms** – *Sound Source Localization, Evidence Grid, Mobile Robots, Sound Mapping.*

## I. INTRODUCTION

In the future, audition is likely to play a large role in robotics. A companion robot will need speech recognition. A mechanic robot might need to listen to the machines it is fixing. A security robot will listen for unexpected sounds. What all of these scenarios assume, however, is that the robot can automatically separate out the signal of interest from the myriad of noise sources that fill our daily lives and mask the target signal. Cars, plumbing, air vents, computers, etc., are all things that these robots dependent on audition must learn to ignore, and possibly work around in order to do their job. But how can the robot filter out this excess noise given the complex and dynamic nature of the signals to which it is listening? It is our supposition that overcoming this masking noise can be accomplished by making the robot aware of its acoustic surroundings, i.e. the auditory scene. If the robot builds models of those ambient noise sources that fill an environment, then the robot will become aware of the masking sounds present at any location. Then the robot can more effectively filter those sounds out, or try to re-position itself where the signal to noise ratio is higher.

Manuscript received January 30, 2007. This research was supported by the Office of Naval Research under work order number N0001406WX30002.

E. Martinson is with the Georgia Institute of Technology, Atlanta, GA 30332. (ebeowulf@cc.gatech.edu).

A. Schultz is with the U.S. Naval Research Laboratory, Washington, D.C. 20375. (schultz@aic.nrl.navy.mil)

In this work, we explore attaining awareness of the auditory scene by a mobile robot through exploration and discovery. Tasked with listening for “interesting” auditory events, a mobile robot uses a two-step process to build a representation of those sound sources that might interfere with its acoustic task. The first step is to move through all areas where the robot might need to listen for events, recording ambient noise along the way using a microphone array. This recorded data allow localization of pertinent sound sources by combining multiple sound source location estimates with robot pose in an auditory evidence grid [1]. The second step is to then investigate detected sources using an area-coverage heuristic in the vicinity of each source. For a medium to long duration source, this second set of data now allows the robot to construct near-field models of sound propagation through the environment, possibly identifying secondary weaker sources, and constructing models of volume and directionality to predict the effects on the auditory scene beyond the sampled area.

The remainder of this paper is organized as follows. The first section discusses related work in robot audition and auditory scene analysis. The second section describes algorithms used in this work for sound source localization, mapping, and sound source modeling. This is followed by a description of the robotic implementation, and, finally, results of the sound source discovery process.

## II. RELATED WORK

The goal of building models of the auditory scene is to combine movement with sensory information to better overcome the effects of noise on auditory processing. Models of how this can be accomplished are loosely inspired by biological systems. In animals, the mechanism for overcoming noise appears to be a neuronal spatial map of the auditory scene, constructed in the inferior colliculus. Individual neurons become attached to specific locations in the surrounding environment, only firing when a noise is detected to originate from that location. These neuronal spatial maps, however, are not being constructed from auditory information alone. Visual [2] and/or body pose [2, 3] data are equally critical in providing additional spatial information to the creation of the map. Without the extra senses, the localization error introduced through just the auditory system is simply too large, and the neural maps become misaligned over time.

In robotics, researchers have only recently begun to explore these advantages. Work by Nakadai et al. [4] demonstrated the combination of movement with auditory information. By simply rotating the microphone array, they could overcome internal noise interference to accurately indicate the direction of a source. Other work, by Huang et.

# Report Documentation Page

Form Approved  
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE <b>2007</b>		2. REPORT TYPE		3. DATES COVERED <b>00-00-2007 to 00-00-2007</b>	
4. TITLE AND SUBTITLE <b>Robotic Discovery of the Auditory Scene</b>				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) <b>Naval Research Laboratory, 4555 Overlook Avenue SW, Washington, DC, 20375</b>				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT <b>Approved for public release; distribution unlimited</b>					
13. SUPPLEMENTARY NOTES <b>The original document contains color images.</b>					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES <b>6</b>	19a. NAME OF RESPONSIBLE PERSON
a. REPORT <b>unclassified</b>	b. ABSTRACT <b>unclassified</b>	c. THIS PAGE <b>unclassified</b>			

al. [5], demonstrated a multi-modal approach using sound and vision to localize a human in an environment.

Extending the biological model even further, however, the robot can also create maps of the auditory scene. Using recent developments in tracking relative robot position, a mobile robot can now capture auditory information from a number of positions in the environment and combine these data together. If the collected information are source localization estimates, then combining them together accurately triangulates upon multiple simultaneously operating source locations, despite the presence of robot ego-noise (motors, wheels, etc) and echoic environments [1,6]. Other work in robotic mapping samples the auditory scene over a large area, and interpolates across all data to construct a noise contour map of the auditory scene [7]. These maps can be used to guide the robot in re-positioning itself to maximize signal-to-noise ratio.

The work presented in this paper continues this last line of research with a number of specific advances: (1) the sound source localization process has been automated, (2) an algorithmic approach to extracting the source positions has been developed, (3) the resulting source position estimates are used to guide a robotic investigation of the source, and (4) models of sound source directivity and local noise maps are constructed from the investigation results.

### III. MODELING SOUND SOURCES

In this section, we summarize the three algorithmic tools that the robot uses to discover sound sources: (1) Auditory Evidence Grids, (2) Volume and Directivity estimation, and (3) Noise Contour Maps. Each of the three serves a purpose in identifying how the surrounding area has changed by correspondingly localizing the sound source, characterizing its acoustic properties, and finally measuring the effects of the environment on the flow of sound. Each of these tools is designed to be used in conjunction with guided robotic movement to gather the necessary data (Section 4).

For the remainder of the paper, we will be largely focusing on discovering sound sources that are medium to long in duration. Examples of such sources include engine and/or machine noise, fan noise, HVAC systems, etc. Such sources are very common in indoor environments, and can be measured repeatedly by a robotic system that takes time to move from place to place. While identifying such transient noises as speech and alerts is equally important to an auditory system, these noises, by necessity, have to be treated differently from sound sources that remain stationary and relatively constant over time.

#### A. Auditory Evidence Grids

The basic algorithm we use for estimating sound source positions from microphone array data are spatial likelihoods [8], an algorithm based on the principle of time difference on arrival. As the speed of sound can be assumed constant, and the microphones are physically separated in space, the signal received by each microphone from a single source will be offset by some measurable time. If the value of these offsets can be determined, then the location of the sound source will be constrained to all positions in the room whose geometry relative to the array corresponds to the measured time

differences. Spatial likelihoods are then a maximum likelihood approach utilizing these time differences to estimate the likelihood associated with every possible location in the room.

In theory, given enough microphones in an array, it should be possible to exactly localize the source using spatial likelihoods. In practice, however, given the small distances between microphones in an on-robot array, as well as the levels of ambient noise and echoes from the environment, spatial likelihoods tend to be better at estimating angle to the sound source rather than distance. So to overcome these errors in distance estimation, multiple spatial likelihood measurements are collected at different points in the environment so as to triangulate the source position. The algorithm used for combining the spatial likelihood measurements together is that of auditory evidence grids[1].

An auditory evidence grid is an evidence grid representation that combines spatial likelihood measurements and robot pose estimates using Bayesian updating to estimate the probability of a sound source being located in a set of predetermined locations (i.e. a grid cell center). Initially, we assume that every grid cell has a 50% probability of containing a sound source. Then, as each new spatial likelihood measurement is added to the evidence grid, the likelihood for each grid cell is adjusted. For simplicity in adding measurements together, we use log odds notation when updating the evidence grid. Equation 1 demonstrates this additive process:

$$\log\left(\frac{p(SS_{x,y} | z^t, s^t)}{1 - p(SS_{x,y} | z^t, s^t)}\right) = \log\left(\frac{p(SS_{x,y} | z_t, s_t)}{1 - p(SS_{x,y} | z_t, s_t)}\right) + \log\left(\frac{p(SS_{x,y} | z^{t-1}, s^{t-1})}{1 - p(SS_{x,y} | z^{t-1}, s^{t-1})}\right) \quad (1)$$

In this equation,  $p(SS_{x,y} | z^t, s^t)$  is the probability of occupancy given all evidence (sensor measurements  $z$ , and robot pose  $s$ ) available at time ( $t$ ), and  $p(SS_{x,y} | z_t, s_t)$  is the probability that a single grid cell contains the sound source based on a single measurement. To then actually localize sources and extract coordinates, we use an iterative nearest-neighbor clustering algorithm to identify the centroid of those clusters most likely to contain a sound source.

To prepare the map for clustering, it is first scaled so that the most likely grid-cell is no more than 99% likely, and the least likely cell is no less than 1% likely. A threshold of 75% is then applied to the map (or 0.5 in a log-likelihood grid) to eliminate cells unlikely to contain a sound source. Nearest neighbor clustering then collects all adjacent cells together in a single cluster, calculating a weighted centroid of the cluster using the likelihood at each grid cell as the weight. Only those clusters with an area larger than a few grid cells are identified as potential sound sources, with their centroids used as likely source positions.

#### B. Determining Volume and Directionality

Provided with enough data in the vicinity of the localized sound source, the next logical step is to construct a model of volume and directivity. The challenge with this step, however, is the difference between the ideal method for

constructing such a model and the actual nature of the data from which to construct it. In the ideal method for determining source directivity, the sound source would be located in an anechoic chamber where the magnitude of any reflections is negligible, and the sound could be measured at a constant distance from the source. With the robot, however, we are in a real environment where there is a substantial reverberant component to measured sound. Furthermore, the collection of data gathered comes from an arbitrary set of distances and angles to the source. How do we overcome these differences?

The first step in overcoming these differences is to separate the measured signal into each of its component parts (direct and reverberant sound), and identify the loudness of each component:

$$p_s^2 = p_{direct,s}^2 + p_{reverb,s}^2 \quad (2)$$

where  $p_s$  is the rms pressure of the sample ( $s$ ),  $p_{direct,s}$  is the rms pressure due to un-reflected sound, and  $p_{reverb,s}$  is the rms pressure due to reflected sound waves. The loudness of the direct sound is the quantity we are the most interested in, but before we can estimate  $p_{direct,s}$  we need to first identify  $p_{reverb,s}$ .

Estimating the loudness of the reverberant component requires making some simplifying assumptions. The first assumption is that the loudness due to reverberant sound will remain constant over the entire room. Since reverberant sound describes the contribution of reflected sound waves, and sound waves will reflect many times all over the room before either decaying to nothing or reaching a receiver, this is a good approximation often used by acousticians for quickly estimating the reverberant field [9].

Even with this simplifying assumption, however, the direct sound contribution needs to be further defined before the reverberant component can be determined. As the direct field describes the volume of un-reflected sound emanating from the source, the energy of the direct field will decay with the square of the distance. So the farther away the robot is from the source, the greater the energy coming from reverberant sound and the less from direct sound. Equation (3) demonstrates this energy relationship, given a sampled distance  $d_s$  and an arbitrary distance  $d_0$ .

$$d_0 p_{direct,d_0}^2 = d_s p_{direct,s}^2 \quad (3)$$

Using this same equation as a guide, our second simplifying assumption is that after some distance the contribution due to the direct field is minimal, and that we can estimate  $SPL_{reverb}$  as the mean volume of the sampled data taken beyond  $d_r$  meters from the source. In this work, we used 3m as a good approximation, where the volume of the direct field will have dropped 9.5dB from the volume at 1m from the source.

Now that we have estimated the contribution of the direct field, the final step is to combine all of the data collected from arbitrary distances and angles into a single model with a specified distance and angle. For this purpose, we first use equation (3) to calculate  $p_{direct,d_0}$  at the specified distance  $d_0$ , and then we apply a Gaussian smoothing function centered

on the desired angle ( $\omega$ ). After combining the earlier equations, the final equation for the model is:

$$p_{direct}(d_0, \omega) = \frac{\sum_s e^{-(\theta_s - \omega)^2 / 2\sigma^2} d_s p_{direct,s}}{d_0 \sum_s e^{-(\theta_s - \omega)^2 / 2\sigma^2}} \quad (4)$$

$(d_s, \theta_s)$  is the position of the sample relative to the center of the source, and  $\sigma$  is the standard deviation of the applied Gaussian. Figure 1 demonstrates a measured directivity plot across all angles, for a distance of 1m from the source.

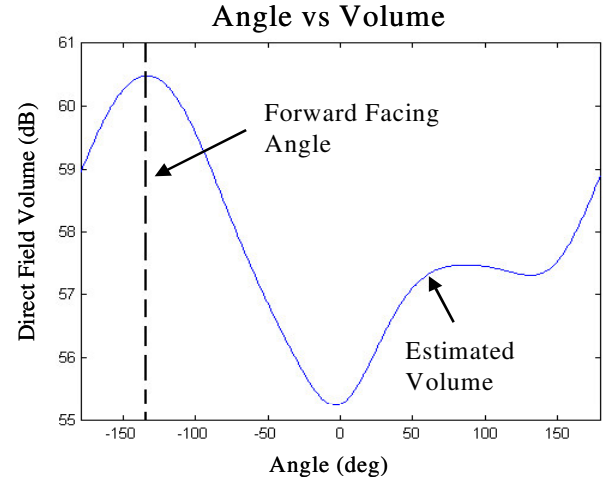


Fig 1. Directivity plot showing angle vs. volume for the direct sound coming from a pc-speaker. This plot is centered on the speaker coordinates, and the volume is estimated at a distance of 1m.

### C. Noise Contour Map

Once a set of sources has been identified, the method for estimating their combined effect on the auditory scene is the Noise Contour Map. In general, noise maps are a graphical tool commonly employed by acousticians to plot the average levels of noise found throughout an environment. In mobile robotics, a noise map provides a robot with a guide to the auditory scene. In previous work[7], noise maps were used to build gradient fields that a mobile robot could follow to decrease the level of ambient noise to which it is exposed. In this work, we focus on the creation of the noise map using the provided source location and directivity models discovered by the mobile robot.

Given a set of known sources with coordinates  $\{x_s, y_s\}_i$  and known directivity models, the theory of super positioning says that the total rms pressure squared at a given location  $(p_{x,y}^2)$  can be estimated as the sum of the reverberant pressure ( $p_{reverb}^2$ ) plus the sum of the direct sound due to each sound source ( $i$ ) on that location:

$$p_{x,y}^2 = p_{reverb}^2 + \sum_{i=1}^n p_{direct,i}^2(d_{x,y,i}, \omega_{x,y,i}) \quad (5)$$

where  $n$  is the number of known sources, and  $\{d_{x,y,i}, \omega_{x,y,i}\}$  are the distance and angle from location  $[x,y]$  to source ( $i$ ). Using this equation, a map estimating the loudness due to ambient noise can be constructed for the entire area traveled by the robot, creating a gradient for the robot to follow when it needs to escape a noisy area. Figure 2 demonstrates an example noise map for two sources.

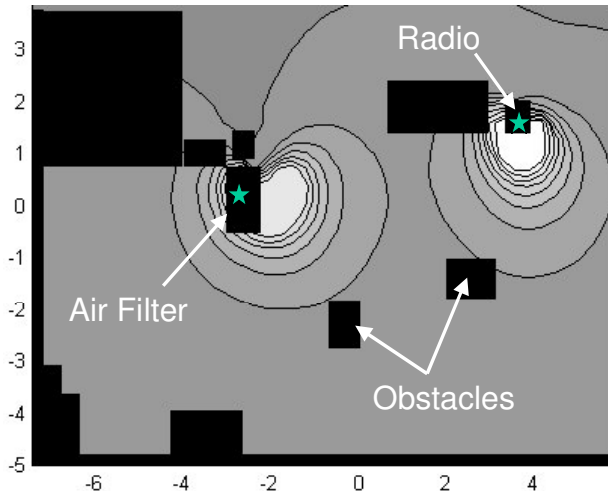


Fig 2. A predicted noise map created from the models of two sound sources: an air filter and a radio. In this map, white depicts areas of higher ambient noise volume

#### IV. ROBOTIC DISCOVERY

Now that the robot is equipped with a set of tools for modeling sound sources and mapping environmental sound flow, the robotic discovery process still needs a movement strategy for collecting the necessary data. For this purpose, we propose a two-step process. In the first step, the robot patrols the environment, collecting data while following a waypoint path through areas where knowledge of the auditory scene is necessary. These data are then used in an auditory evidence grid to extract sound source positions. The second movement step is then a guided investigation of the detected sound sources using an area-coverage heuristic. The data collected from the second step can then be used to refine localization estimates and build models of directivity.

##### A. Experimental Setup

This work was tested on two different indoor robots: (1) a Pioneer-2dx robot over a 10m x 15m area, and (2) an iRobot B21R over a 12m x 12m area. Both robots can be seen in Figure 3. The sensor suite used for this task was the same on both robots: odometric position sensors, a SICK laser measurement system and a 4-element (Audio-Technica ATR35S lavalier mics) microphone array attached to battery-powered preamps and an 8-channel data acquisition board. Robot pose was then calculated by comparing laser range finder and odometric readings to a robot created obstacle map.

On the Pioneer robot, robot control was implemented using the Player/Stage [10] robot control software. Built in drivers provided obstacle avoidance and path planning. An adaptive monte-carlo localization algorithm (amcl), also native to Player/Stage, then provided robot pose estimates by comparing laser range finder results to a robot created obstacle map. Due to processor limitations, the path-planning and amcl algorithms were run on a desktop Linux machine over a wireless network. Also because of computational limitations, all auditory processing was performed on a separate laptop mounted beneath the microphone array.

On the B21R, an NRL proprietary implementation of continuous localization [11] provided robot pose estimates, while a Trulla path-planner [12] guided the robot from location to location. As with the Pioneer, auditory processing was performed on a separate laptop mounted on top of the robot base, below the monitor. However, localization and path planning was performed on internal machines so the wireless network was not required for testing the discovery process.

##### B. Patrol Task

The first phase of autonomous movement is the patrol task, described by a set of ordered waypoints in the environment for the robot to visit. The purpose of this phase is to expose the robot to as much of its environment as possible so that it will be able to detect any significant ambient noise sources. In our implementation, spatial likelihoods are only calculated for each sample over a 3m radius so we selected a set of waypoints that should bring the robot within 3m of every location in the environment. This 3m requirement was selected empirically based on the requisite loudness of a source being detected. Beyond 3m, sources less than 60dB are not reliably detected using spatial likelihoods.

Provided with this waypoint path, the robot then uses a path-planner to guide it from its current position to each waypoint in turn while dynamically avoiding obstacles. Upon arriving within some threshold distance (0.4m) of the desired waypoint, the robot selects the next waypoint in the specified order as a target, and the cycle repeats. To account for inconsistencies between the real world and the map, a timeout mechanism monitors the robot progress and forces it to move on to the next waypoint after 3 minutes. The task is finished when the robot has successfully visited or tried to visit all specified waypoints. An example waypoint path can be seen as a solid line in Figure 4.

After completing one loop through the environment, the robot then processes its auditory data using the auditory

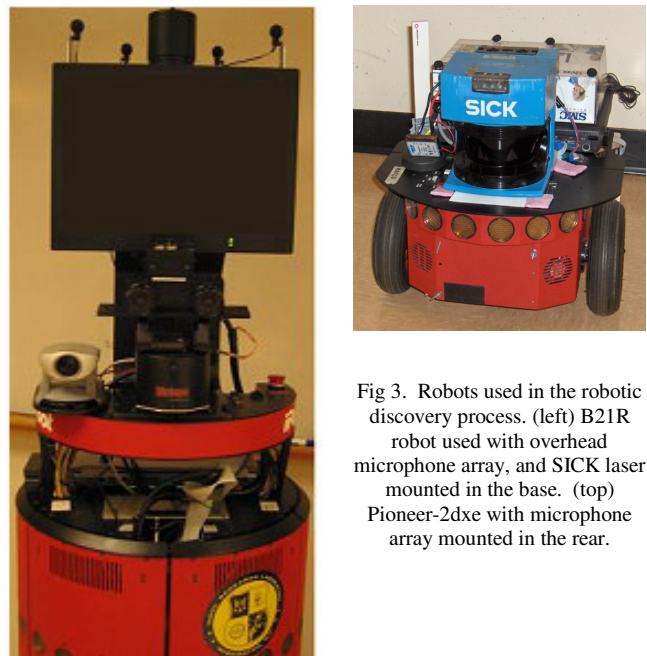


Fig 3. Robots used in the robotic discovery process. (left) B21R robot used with overhead microphone array, and SICK laser mounted in the base. (top) Pioneer-2dx with microphone array mounted in the rear.

evidence grid and clustering process to search for likely source position candidates. Given the scarcity of data over any given area, the resulting source localization estimate can have relatively high error. So the next step is to investigate this source and refine that localization result.

### C. Robotic Investigation of a Source

The second phase of autonomous movement begins after the robot has successfully completed a patrol loop, and has identified one or more areas potentially containing a sound source. The purpose of this phase is to actively investigate each of those identified areas in order to determine whether a sound source is actually located there, where exactly it is located, and what is the directivity of that source.

Provided with a target set of coordinates to investigate, a set of unobstructed locations is identified within a 3.5m radius of the target using the obstacle map of the environment. These unobstructed locations become waypoints for the robot to visit, effectively performing an area coverage task in the vicinity of the suspected sound source. Unlike the waypoint task, however, visiting these waypoints does not need to be done in any particular order, and so the robot will always travel to the nearest waypoint. The circles in Figure 4 show a set of waypoints to be used for investigating a single source.

When investigating a source, a different sample collection strategy is utilized. During the patrol task, the robot collected samples while moving along its route. Movement, however, is both acoustically and algorithmically noisy, resulting in poorer accuracy when localizing a source. When identifying possible locations that need further investigation, such as during the patrol phase, lower accuracy is fine. But during the investigation phase, the robot needs to use a pause and sample methodology to achieve higher accuracy. By stopping the robot to sample the auditory scene whenever it reaches a waypoint, the volume of robot ego-noise is reduced. Additionally, position error is reduced as more data are available from which to accurately estimate position.

After completing the investigation of a single source, the robot now has enough data to re-fine the position of the source and determine its directivity. To refine the position, an auditory evidence grid is constructed using just the data collected from the area-coverage task, and the clustering algorithm is reapplied. Discarding smaller clusters, the coordinates of the largest cluster's centroid are used as the global location of the source when estimating directivity of the source. This value is necessary for determining the position of the sample relative to source center ( $d_s, \theta_s$ ).

Following the completion of investigatory phase for a single source, the robot then repeats this phase for all other sources detected during the patrol phase. Only after all suspected sources have been detected is a noise map constructed to predict the sound flow in the area beyond the investigated areas.

## V. RESULTS

Testing the discovery process was divided into three stages. In the first stage, we tested the accuracy of the directivity model by investigating one source with known

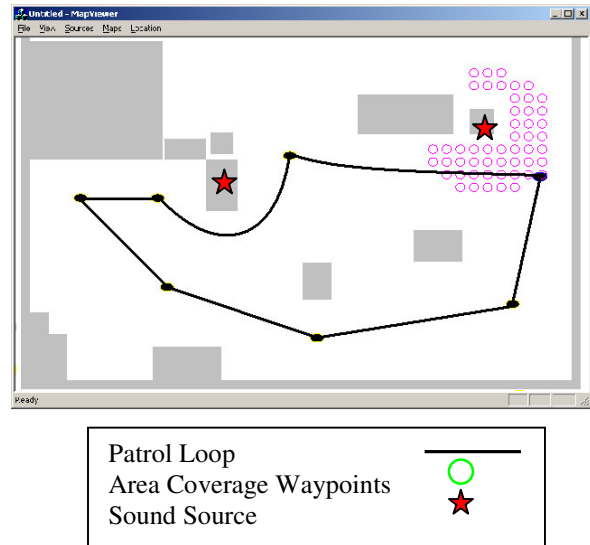


Fig 4. Screen capture of an obstacle map used by the pioneer for navigation in an environment with two sources. The black line shows the waypoint path followed during a patrol, while the circles illustrate a set of targets reached by the robot to complete an investigative task for a single source.

$\{x,y\}$  but unknown  $\theta$ . In the second stage, we tested the entire discovery process for a single source of unknown  $\{x,y,\theta\}$ . Finally, in the third stage, we tested the ability of the discovery process to localize multiple simultaneously operating sources of unknown  $\{x,y,\theta\}$ .

During each of these three stages, we applied a 10<sup>th</sup> order highpass FIR filter (300Hz cutoff frequency) to every sample before analyzing the data. Since the ambient noise sources being measured had significant high frequency components, this had little effect on the auditory evidence grid creation. What the filter did do, however, is reduce the impact of robot motor noise on determining directivity. Since the robot's motor was in close proximity to the microphone array, it could overpower the weak volumes measured farther away from the source.

### A. Stage 1 - Known $\{x,y\}$ , Unknown $\theta$

In this testing stage, a single source of known centroid position was rotated through 7 different angles in 45° increments. One angle was not tested due to the source pointing at a solid pillar where the robot could not investigate. The sound source was a pc-speaker playing nature sounds (rain) measured as being 65 dB 1-m from the source (including both direct and reverberant sound). Provided with the ground truth source location, the B21r was used to investigate the source once for each angle using just the area-coverage algorithm with a 3.5m range. Over 7 trials, the mean error of peak directivity was 0.2rad of ground truth with a maximum error of 0.5rad. Given that the source itself is a pc-speaker with a wide frontal lobe, this approximation should be adequate to guide the robot away from the loudest areas surrounding the source.

### B. Stage 2 - Unknown $\{x,y,\theta\}$

In the second stage testing, the B21r was used to localize each of three pc-speakers with unknown  $\{x,y,\theta\}$  5 times. During any one test, only one speaker was playing and all

speakers played the same nature sounds track (rain) at a 65dB volume. For each test, the robot first moved along the same patrol route, localizing the active source. Then it would dynamically choose where to center its investigation. After investigating the area, the sound source position was re-estimated along with the map of the surrounding environment and the source orientation. Table 1 shows the mean error for each source.

	Localization Error (m)	Orientation Error (rad)
Source 1	0.32	0.22
Source 2	0.2	0.18
Source 3	0.22	0.32
Combined	0.24	0.24

Table 1. Mean localization and orientation error as produced by the discovery process.

These results demonstrate the reliability of the discovery process in accurately finding and modeling sources. Sources 1 and 2 were located in areas where the robot could completely encircle the source, and therefore gather data from all directions. Source 3, however, was against a wall, so the robot was limited to gathering data in the 180° foreground. Due to this limited area, as well as the proximity to the wall and its echoic effects, the orientation error is highest for this third source.

### C. Stage 3 – Multiple Sources

The final stage of robotic testing demonstrated the ability of the robot to detect multiple simultaneously operating sources. Two sources, an air filter (0.5m x 0.3m x 0.3m) and a two-speaker radio generating static noise, were placed 5.8m from each other. Figure 4 shows their relative placement. The pioneer-2dx robot was then used to localize and model each source. Following the initial patrol phase, the robot identified two potential clusters, corresponding to each of the two sources. Both initial clusters were within 1m of the actual source location. Upon further investigation, the robot improved the localization accuracy for the air vent to within 0.2m, and to 0.4m for the radio. The orientation accuracy for each source was 0.64rad and 0.4 radians respectively. Using these source locations and their directivity models, a noise map was created to estimate their combined effect on the environment (Figure 2). Despite some inaccuracy in the orientation estimates, likely due to the simplifying assumption regarding constant reverberant sound levels over such a large area, this resulting noise map can still easily guide a robot away from interfering ambient noise sources.

## VI. CONCLUSION

What this work has demonstrated is the ability of a robot to construct models of the interfering ambient noise in the

auditory scene. By first listening for sources, and then actively investigating possible locations, a robot gradually discovers where the sources are located and how the sources are oriented. The same process has been demonstrated to work on very different robotic platforms, and work in the presence of simultaneously operating sources. Once the knowledge has been collected, a robot can then extrapolate from the models to the shape of the auditory scene as a whole.

In future work, equipped with these general processes for mapping and modeling ambient noise sources in the auditory scene, the next step of the robotic discovery process is to take advantage of these tools to improve the signal-to-noise ratio in a surveillance operation. Given a robot that is listening either for a particular sound, or trying to determine what sound is new to the environment, knowledge of where the ambient noise is coming from is critical to task performance.

## VII. REFERENCES

- [1] E. Martinson and A. Schultz, "Auditory Evidence Grid," presented at Int. Conf. on Intelligent Robots and Systems (IROS), Beijing, China, 2006.
- [2] P. Hyde, and Knudsen, E.I., "A topographic projection from the optic tectum to the auditory space map in the inferior colliculus of the barn owl," *Journal of Computational Neurology*, vol. 421, pp. 146-160, 2000.
- [3] M. Rucci, G. Tononi, and G. M. Edelman, "Registration of neural maps through value-dependent learning: modeling the alignment of auditory and visual maps in the barn owl's optic tectum," *Journal of Neuroscience*, vol. 17, pp. 334-352, 1997.
- [4] K. Nakadai, K. Hidai, H.G. Okuno, and H. Kitano., "Epipolar Geometry Based Sound Localization and Extraction for Humanoid Audition," presented at IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Maui, Hawaii, 2001.
- [5] J. Huang, T. Supaongprapa, I. Terakura, F. Wang, N. Ohnishi, and N. Sugie, "A Model Based Sound Localization System and Its Application to Robot Navigation," *Robotics and Autonomous Systems*, vol. 27, pp. 199-209, 1999.
- [6] Y. Sasaki, S. Kagami, H. Mizoguchi "Multiple Sound Source Mapping for a Mobile Robot by Self-motion Triangulation" presented at Int. Conf. on Intelligent Robots and Systems (IROS), Beijing, China, 2006.
- [7] E. Martinson and R. C. Arkin, "Noise Maps for Acoustically Sensitive Navigation," *Proceedings of SPIE*, vol. 5609, 2004.
- [8] B. Mungamuru and P. Aarabi, "Enhanced Sound Localization," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 34, 2004.
- [9] D. R. Raichel, *The Science and Applications of Acoustics*. New York, NY: Springer-Verlag, 2000.
- [10] B. Gerkey, R. Vaughan, and A. Howard, "Player User Manual v2.1," in *Player/Stage Project*, <http://playerstage.sourceforge.net>, February 2006.
- [11] A. Schultz and W. Adams, "Continuous localization using evidence grids," presented at IEEE International Conf. on Robotics and Automation, Leuven, Belgium, 1998.
- [12] K. Hughes, A. Tokuta, and N. Ranganathan, "Trulla: An algorithm for path planning among weighted regions by localized propagations," presented at Int. Conf. on Intelligent Robots and Systems (IROS), Raleigh, NC, 1992.