Using a Mobile Robot to Detect Changes to the Auditory Scene

Eric Martinson
Харьковский национальный университет радиоэлектроники
61166, Харьков, пр. Ленина, каф. технологии и автоматизации производства РЭС и ЭВС
E-mail: ebeowulf@cc.gatech.edu

This work discusses the autonomous detection of change in the auditory scene by a mobile surveillance robot through the comparison of sampled-data classification results with ambient noise volume levels predicted by a simplified wave equation.

Introduction

Mobile robots are a great source of supplementary information about the environment. For instance, a security system would generally use cameras to monitor large sections of the environment, rather than rely entirely on robots. If, however, a human monitor detects something on the security camera, then a mobile robot can be used to investigate the phenomenon in more detail. In such a manner, most benign security threats can be easily handled with a reduced human presence. The acoustic monitoring of an environment holds similar advantages for using a mobile robot. Given, for example, a factory with a number of stationary microphones mounted in and around the equipment to listen for unexpected activity, such as broken machinery or human intrusions, a robot can be used to investigate in further detail auditory phenomenon detected by stationary microphones, but which is significantly masked by current ambient noise conditions. Before a robot can detect changes to the auditory environment, however, it first needs to make predictions about how the environment should sound so that there exists something with which to compare acoustic measurements.

In this work, we explore a knowledge-based approach for making predictions using the wave equation to model sound propagation. Previously, we demonstrated that an autonomous mobile robot could be used to first localize active sound sources in the environment, estimate volume and directivity for each, and then estimate the relative volume of each source in the environment in the form of a noise map [1]. Now we expand upon that work by also building classification vectors using mel-frequency cepstral coefficients for each of the detected sound sources. As will be demonstrated, the relative sound source volumes predicted from robot-collected data can also be used to predict the sampled-data classification results using nearest-neighbor classification to the autonomously collected sound source classes.

Methodology

This paper is focused on first predicting the dominant sound source in the environment, and then measuring that sound source. As such, the theoretical parts of the problem can be broken into two sections. In the first section, we discuss the estimation of relative sound source volumes for arbitrary locations of the room given some robot determined data about active sound sources. In the second section, we discuss the creation and use of a classification vector for identifying the loudest sound source in a recorded sample.

The robot that we used for testing the following work is a Pioneer2-dxe mobile robot equipped with a 4-element microphone array (Figure 1). Robot control was implemented using the Player/Stage [2] 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 (Figure 2). Due to processor limitations, the path-planning and amcl algorithms were run on a desktop Linux machine over a wireless network; all auditory processing was performed on a separate laptop mounted beneath the microphone array.

Predicting Sound Source Volume – Determining the expected sound source volume requires a priori information, much of which is obtainable autonomously by a mobile robot. In particular, the robot needs to know its position relative to the sound source, and the sound source volume at some known distance from the source. This knowledge allows the robot to

Figure 1. Pioneer2-dxe mobile robot.

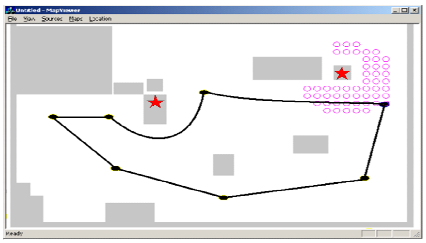

Figure 2. Obstacle map overlaid with sound source positions (stars) patrol route (solid line) and sampling area for building source class (circles)

predict the volume of the direct field, i.e. unreflected sound, assuming no obstacles block the path from the sound source to the robot. Unfortunately, this simple direct field model is not very accurate either in the near-field (approx. < 1m) or in the far-field (>1-m). In the near field, the directionality of both the sound source and the receiver will have a strong impact on the detected volume of the sound source by the robot. In the far field, the reverberant field (reflected waves) will have a greater impact on perceived volume than the direct field.

To improve the quality of the sound estimates, especially in the near-field, we use more robot determined information about the sound source. As was demonstrated in previous work [1], the robot can also build two-dimensional models of sound source directivity while accurately localizing the sound source and estimating a constant volume reverberant field effect. Architectural acoustics commonly assumes a constant reverberant field when accurate information is otherwise unavailable [3]. This information, along with a pre-determined microphone directivity model, can be used for predicting the effect of the sound source on the environment is:

$$D_s = \frac{Q_r(\theta_r) \cdot Q_s(\theta_s)}{l} \cdot V_s + R_s \qquad \text{Equation 1}$$

In this equation, Q_r and Q_s are unitless two-dimensional receiver and sound source directivity models, θ_r and θ_s are angles of incidence upon the receiver and sound source respectively, l is the distance from the sound source to the robot [m], V_s is the max rms pressure [Pa] generated by the sound source at 1-m from the sound source, and R is the estimated constant reverberant field volume [Pa].

The advantage of using this simplified model of sound propagation through the environment, assuming constant reverberant field and no obstacle interference, is that a robot can gather all of the sound source information autonomously without assistance from a human. This is important because the auditory environment is constantly changing as sound sources are turned on or off, or simply adjusted in volume. Therefore, when the robot determines that the environment has changed using these characteristics, the robot can also re-investigate the altered environment without human assistance to predict further changes in the future.

Detected Sound Source Classification – The assumption behind this work in detecting change to the auditory scene is that the loudest sound source in the environment at the location where samples are being collected should be the sound source that is most often detected using a classification algorithm. Given that the classification algorithm does correctly identify the loudest sound source, then our predictive models of source volume should correctly identify the sound source which will be detected most often for each region of the room. The classification feature set that we chose for this purpose was mel-frequency cepstral coefficients (MFCC's).

MFCC's are based on the mel-scale filter bank, a filter-bank designed to group frequencies in a manner similar to human perception. Low frequencies are grouped in equal spaced bands, while higher frequencies are grouped in bands that increase logarithmically in size with the frequency. MFCCs are then calculated by taking the discrete cosine transform of the energy of these mel-scale frequency bands [4]. For this work, the first 8 coefficients are calculated over two successive 250-msec samples (10-msec frame, 31-msec window) and

averaged together. Then the first coefficient, which primarily reflects sample volume, is discarded to build a 7-coeffecicient vector describing the sample[5].

To build the source class, the robot needs to collect samples from the vicinity of the sound source. As with the volume estimation model, it was very important that a robot be able to build the sound source class autonomously so that new sound sources, and other future changes to the model could be handled without assistance. Therefore, we used the same data collection approach as that described in [1] for determining models of directivity. Once a sound source has been localized, the robot identifies a series of target waypoints surrounding the source at various distances and angles (Figure 2), moves to each waypoint, and collects four or more 250-msec samples of ambient noise using a 4-element microphone array mounted on its back. A sound source class is an average of those vectors created from all samples collected in the vicinity of the sound source.

Once source classes have been identified for known active sources, each new sample recorded by the robot is matched to a source using the mahalanobis distance from the new samples' feature vector to each of the class vectors. Assuming that each of the sound sources has a different sound function, this simple classification strategy works well for distinguishing between known classes of common ambient noises [5]. To allow for unknown sound sources in the environment, we also insert a set of 20 random classes into the nearest-neighbor search. Now, instead of matching to a known source, there is a good chance that a feature vector belonging to an unknown class will be closer in distance to one of these random classes.

Results and Conclusion

Manual Evaluation **-** To test the efficacy of the proposed classification algorithm, we first set up a small experiment involving two sources in a 5x5-m^2 indoor environment. An air filter with a significant directional component was placed to the right side of the room, generating noise at 50-dB. A second source, a small fountain, was then placed roughly 3-m away at the top end of the room, generating water noise at a hand measured 54-dB (Figure 3).

To build a class vector for each of the two sources, a microphone was manually moved about each of the sources to simulate the collection of samples by a moving mobile robot performing an area-coverage algorithm over the surrounding 2-m. Both sources were enabled while the samples were being collected. These samples served as the basis for a fountain class vector, and a filter class vector. An additional 20 random classes were also generated using the minimum and maximum range of the collected samples.

After source functions were approximated with a class vector, additional sampling was performed at each of 10 sample locations. Roughly the same number of samples were recorded at each location. Figure 3 shows the ratio of samples classified as belonging to the fountain vs. the filter at each sample location. Since the fountain was a significantly louder source in this test, more than half of the sample locations detected were dominated by the fountain. What Figure 3 also demonstrates, however, is the relationship between MFCC-based classification and the predicted volume of each sound source. Comparing the ratio of predicted volumes to the classification ratio revealed that at all sample positions where the expected difference in volume is greater than 2-dB the louder source dominated the classification results. For volume differences less than 2-dB, the classification result is more variable.

Robot Evaluation – After demonstrating in one environment that the predicted sound field volumes using a constant reverberant field estimate are an indicator of the classification results using sampled data, the next step was to use a mobile robot to demonstrate the same relationship in another environment. Therefore, we set up an experiment involving three sources in a larger 8x10-m^2 indoor environment. An air filter with a significant directional component was placed on one side of the room, generating noise at 62-dB. A second source, a small indoor water fountain, was then placed roughly 5-m away in the middle of the room, generating water noise at a measured 61-dB. The third source was the robot itself, a Pioneer2-dxe robot that generated approximately 47-dB of fan noise. All three sources had been localized and modeled (directivity, volume, and source class vector) earlier by the robot when only one

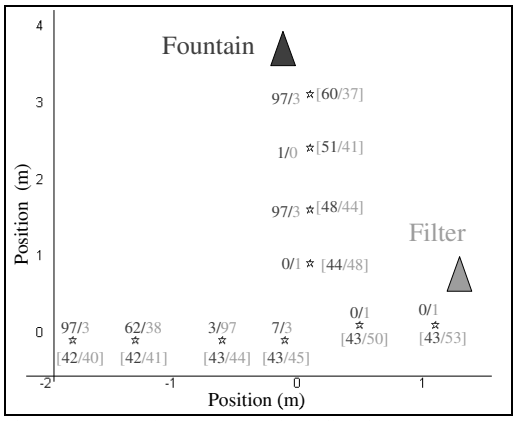

Figure 2. Hand measured classification results (top and left) vs. predicted direct field volumes (bottom and right, in brackets) for two sources, a fountain in black (nominator) and a filter in grey (denominator).

Detected Class	Predicted Loudest Source		
	Filter Loudest	Fountain Loudest	Neither
Filter	5.1%	57.1%	40.6%
Fountain	68.4%	1.6%	2.0%
Robot	19.1%	41.3%	54.3%
Other Classes	7.4%	0.0%	3.1%

Table 1 Detected class versus the class predicted to be loudest using robot measured data.

source was active. The robot class was an average of the recorded noise from an isolated part of the room. To test the classification performance of these autonomously determined classes, the robot patrolled the room five times on a circular route (Figure 2), passing each time within 1-m of each of the two sound sources while collecting audio samples (~320 samples/ patrol).

In areas of the room where no sound source was significantly louder (>1.5dB) than both other sources (1352 samples), the majority of the samples collected (54%) belonged to the robot class. The fountain, which was the next loudest source over most of the room, was detected in 39% of the samples, while the filter was detected in only 2% of the samples.

Where the sound field predictions estimated a greater than 1.5-dB difference in volume, however, the distribution changed (Table 1). In areas where the filter was louder than either the stationary fountain (136 samples), or the robot's own ego-noise, the filter was detected in 68% of the samples, as opposed to the fountain which was detected in only 5%. Where the fountain was predicted to be loudest (126 samples), the fountain was successfully detected in 57% of the samples, versus the filter in only 2% of the samples, and the robot in 41% of the samples.

These robotic results, in conjunction with the previous hand-collected results, validate the proposed method for detecting change in the environment. By first identifying regions of the room that should be dominated by known sound sources in the environment and then building MFCC's from samples collected in those areas, a robot can identify changes to each source such as being turned off, or being masked by new sound sources. Furthermore, if the change to the environment can be identified as benign, all of the data that the robot needs to readjust its predictive models of the environment can be collected autonomously without human intervention, allowing the robot to continue its surveillance of the auditory scene.

In the future, it is our intention to expand upon these initially successful results in detecting change to the auditory environment. Currently, new sounds, such as those associated with machinery breaking down, fires, intruders, etc., are only detectable if they either replace or mask an existing, known source. By building a probabilistic model, however, incorporating not only volume and classification results, but also sound source location and reverberant field models, we hope to build a general approach to detecting both large changes to existing sound sources and quiet, new sounds in the environment.

1. Martinson, E. and A. Schultz. *Robotic Discovery of the Auditory Scene*. in *IEEE Int. Conf. on Robotics and Automation*. Rome, Italy. 2007.
2. Gerkey, B., R. Vaughan, and A. Howard. *Player User Manual v2.1*. Available from: http://playerstage.sourceforge.net.[accessed September, 2005]
3. Raichel, D.R., *The Science and Applications of Acoustics*, New York, NY: Springer-Verlag. 2000.
4. Slaney, M., *"The Auditory Toolbox"*. 1994, Apple Computer Company, Apple Technical Report #45.
5. Ravindran, S., *Physiologicially Motivated Methods for Audio Pattern Classification*, Ph.D. Thesis, School of Electrical and Computer Engineering, Georgia Institute of Technology, 2006.