2011 IEEE International Conference on Robotics and Automation
Shanghai International Conference Center
May 9-13, 2011, Shanghai, China

Dynamically Reconfigurable Microphone Arrays

E. Martinson, B. Fransen

Abstract—**Robotic sound localization has traditionally been restricted to either on-robot microphone arrays or embedded microphones in aware environments, each of which have limitations due to their static configurations. This work overcomes the static configuration problems by using visual localization to track multiple wireless microphones in the environment with enough accuracy to combine their auditory streams in a traditional localization algorithm. In this manner, microphones can move or be moved about the environment, and still be combined with existing on-robot microphones to extend array baselines and effective listening ranges without having to re-measure inter-microphone distances.**

I. INTRODUCTION

The design of a microphone array has a significant impact on the mathematics of sound source localization. Arrays, for instance, are commonly designed to emphasize the region directly in front of the robot limiting noise from the sides when the signals are combined together. Microphone placements within the array can lead to: superdirective configurations, amplifying conversations at a distance; or be generic in shape, but be prepared with lots of microphones to listen to any possible approach direction [1]. They can emphasize angular measurements or be spread across a room to localize a source in 2D [2]. In general, the shape of the array should be designed for the application, so as to effectively suppress ambient noise while amplifying the target. In robotics, however, the problem with a static mounting is that the microphones do not move relative to each other. The robot base can move, but the microphones stay in a rigid, well measured mounting. Robot microphone arrays, therefore, are designed to overcome worst case configurations for a single application, or, worse yet, they are simply attached to the robot/environment wherever space permits. This is not conducive to effective noise cancellation, either in sound source localization or auditory streaming applications. Instead, the robot needs to be able to reconfigure its array configuration dynamically in response to changing tasks and environmental conditions. The goal of this work is to remove that limitation of a static array configuration, freeing up microphones to be moved about as needed either by hand or by other robots.

The reason for traditionally mounting microphones in rigid configurations, as opposed to dynamically adapting the array design, is the effects of small amounts of error on localization performance. In a 2 element array, with microphones spaced 0.3-m apart, a 1-cm error in relative microphone position can mean a 14-deg error in sound localization. With triangulation based methods for localizing sources in 2- or 3- dimensions, this angular error is then compounded by the distance from the array. A static-mounted array, therefore, is accurately measured and rigidly attached to a robot or the room to maintain measurement precision or repeatability. Whether the microphone streams are being combined through beamforming, generalized cross correlation, or independent components analysis, the precise position of each microphone relative to the others is required. This work does not violate this principal. Using a single on-robot camera, a robot localizes each microphone individually from an attached fiducial. As will be discussed in greater depth in section 3, the AR toolkit [3], when used with a modified camera image segmentation process, allows the robot to identify environmental microphone positions accurately enough to identify sound source positions in 3D. These distributed microphones can then be combined at the signal-level (i.e. using generalized cross-correlation to identify time-delay on arrival) with any on-robot microphones to change the array configuration, and improve the quality of sound localization.

II. RELATED WORK

Robot mounted microphone arrays have traditionally been subject to two classic limitations. First, on-robot microphones are located in close physical proximity to sources of robot ego-noise (e.g. motors, fans, wheels, etc.), which mask the signals of interest. Second, limited mounting space on top of or around a physical robot platform restricts the potential accuracy of sound source localization algorithms because of small inter-microphone distances. A small array, in general, can find the angle to a sound source, but not the distance [4]. Solutions to these problems tend to be application specific.

For handling robot ego-noise, as well as general environmental sources, the most common solution is targeted

Manuscript received Sept 15, 2010. This work was supported by the Office of Naval Research under job order numbers N0001408WX30007 and N0001410WX20773.

E. Martinson is a post-doctoral fellow with the U.S. Naval Research Laboratory, Washington, DC 20375. (phone: 202-404-4948; e-mail: eric.martinson.ctr@nrl.navy.mil).

B. Fransen is with Beyond Robotics, Annapolis, MD (e-mail: brf1234@gmail.com).

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
MAY 2011		**00-00-2011 to 00-00-2011**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Dynamically Reconfigurable Microphone Arrays	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) **U.S. Naval Research Laboratory,Washington,DC,20375**	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
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Presented at the IEEE International Conference on Robotics and Automation, May 9-13, 2011, Shanghai, China, Government or Federal Purpose Rights License.

14. ABSTRACT
Robotic sound localization has traditionally been restricted to either on-robot microphone arrays or embedded microphones in aware environments, each of which have limitations due to their static configurations. This work overcomes the static configuration problems by using visual localization to track multiple wireless microphones in the environment with enough accuracy to combine their auditory streams in a traditional localization algorithm. In this manner microphones can move or be moved about the environment, and still be combined with existing on-robot microphones to extend array baselines and effective listening ranges without having to re-measure inter-microphone distances.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**7**	

filtering. Speech filters, in particular, have received a lot of attention in the field of human-robot interaction for foveating a robot towards a target [5] in 1D. Valin uses specially designed filters for separating out speech signals from ambient and robot noise [1] to identify the angle (yaw, pitch) to a speech source and do auditory streaming. Also, filters modeling the precedence effect, a biological phenomena that emphasizes early sounds over reverberations, can track time-varying signals such as speech or music [6].

If a source position is necessary, however, and not just its angle, then one robotic solution is to move the physical array and sample in multiple locations. Angular estimates from each sample location, when combined together, triangulate upon one or more targets in 3D. Combining the measurements effectively becomes the difficult part, with both evidence grids [4] and the RANSAC [7] algorithm having been used for this purpose.

When the microphone array does not need to remain on the robot, then a good alternative solution in some cases is to mount microphones throughout the environment. Aware environments research has demonstrated this in home environments to keep track of their occupants [8]. Nakadai et al [2] demonstrated this successfully with 2D localization of a speaker in a single room, ultimately integrating both local measurements from an on-robot array with the room mounted array using a particle filter. The auditory streams from the robot measurements, however, were not compared directly to the room mounted microphones due to the lack of precise inter-microphone spacing. In general, however, the room mounted array solution suffers from deployment problems. Relatively rapid re-calibration techniques using high-frequency pulses [10, 11] are effective for localizing disparate microphones, but have reduced accuracy in the presence of obstacles. Furthermore, hardware setup may be prohibitive. Wires need to be strung throughout the room or built into the environment, or wireless systems need to manage batteries, base stations and interference from other wireless type devices. All of which limits deployment of large room mounted arrays to arbitrary environments.

III. Dynamic Microphone Array Design

In contrast to previous work mixing multiple microphone arrays together [2], this work integrates microphones together at the signal level. Combining data at the signal level means that sound localization probabilities are calculated for every microphone pair instead of each separate microphone array. Therefore, each detected microphone in the environment, when combined with on-robot static array, doubles the number of total sound localization estimates. Furthermore, because environmental microphones are generally located further away from the robot, the baseline of the array is extended tremendously, improving overall accuracy.

The system we have developed to test our dynamic array is shown in Figure 1. An iRobot B21r is equipped with a single Point Gray firefly camera for localizing fiducials attached to wireless microphones. The SR3000 time of flight camera seen in Figure 1 is not currently used except as a visualization aid. The wireless microphones detected by the robot are continuously streaming to 2 variable frequency base-stations mounted on the robot. The signals of both the 2 wireless base stations and 2 robot-mounted microphones are amplified with battery powered preamps and then sampled using an 8 channel PCMCIA A/D converter. Currently, all microphone streams are continuously sampled by the robot. Stream inclusion in sound localization is then guided by software. In the future, however, we hope to remove this need for a base station per wireless microphone by either synchronizing streams locally and then transmitting via traditional wireless network [10], or dynamically switching between wireless frequencies depending upon the set of visible fiducials.

Figure 1. Wireless microphones (right) are rigidly attached to fiducials. The B21r robot uses its onboard camera to localize these microphones and combine their signals with 2 overhead microphones.

The remainder of this section describes the sensing components of this system in more detail. Specifically, it covers the modifications necessary for accurate visual localization of the microphones, and the algorithms used to localize sound sources from a dynamic microphone array.

A. Visual Microphone Localization

Dynamic microphone localization is achieved through rigidly affixing fiducials to the wireless microphones. The fiducials we use are developed by the ArToolkit [3] which can recover fiducial positions in real time. When using the ArToolkit in a real time signal processing task, the system must be calibrated and parametrically optimized before accurate position estimates can be recovered. We will review calibration of ArToolkit parameters first, followed by camera calibration techniques and testing results.

The ArToolKit provides a system for generating and tracking planar markers. The system is based on the

capability of recognizing black squares printed on white backgrounds commonly known as fiducials. To recognize the fiducials, a segmentation algorithm is employed that labels black pixels as part of the square and white pixels as part of the background. For pixels containing only a white background or black square, the labeling is straightforward. Pixels on the border of the square, containing portions of both black and white colors, however, are often highly ambiguous and result in incorrect estimation of a fiducial's boundary in an image. Two factors that contribute to boundary ambiguity are the use of a Bayer pixel pattern and image blur. The Bayer pixel pattern utilizes an uneven spacing of red, green and blue pixels used in most modern cameras. The process of generating a color image requires demosaicing the pixel pattern to generate a full resolution image. The image interpolation necessary for demosaicing introduces blurring about edges of image features. In addition to blurring caused by the demosaicing process, blurring occurs due to lens/image plane effects.

The end results of image blur is the need to parametrically set the image segmentation threshold in the ArToolKit. In work presented here, the threshold boundary was modified until the segmentation boundary for our fiducial coincided with the actual boundary of the fiducial in the image. Without setting the segmentation threshold, the ArToolKit consistently underestimated fiducial size, resulting in underestimated distances between camera and fiducial.

Once the segmentation parameters were set, internal calibration of focal length and distortion was performed. To overcome image distortion, the ArToolKit's distortion calibration tools were used to solve for the center of projection and image distortion characteristics. For recovering the camera's focal length, we used the fiducials themselves. Fiducials were placed in known locations and then the focal length was modified until the distances between fiducial locations matched the actual theoretical distances between the fiducials. By using the fiducials to solve for the focal length, the internal calibration parameters of the camera were determined while further minimizing any effects of the aforementioned segmentation parameters.

Once the camera was calibrated, we tested fiducial localization accuracy by comparing distances between fiducials placed at known locations. Two tests were performed. First, we evaluated inter-fiducial location accuracy in regard to translations away from the camera with error metrics shown in Figure 2. These results show a greater error for greater distances between fiducials. This demonstrates that fiducials further away from the camera, necessary for producing larger inter-fiducial distances, caused the inter-fiducial distance to be under estimated. This deterministic underestimation can be partially attributed to inaccurate focal length calibration. With due diligence, the decrease in fiducial accuracy with respect to range could be mitigated by additional modification to the camera focal length parameters. However, the projected mean error for a 2-m inter-microphone distance was still <1-cm, which would produce a less than 5 degree error in sound localization.

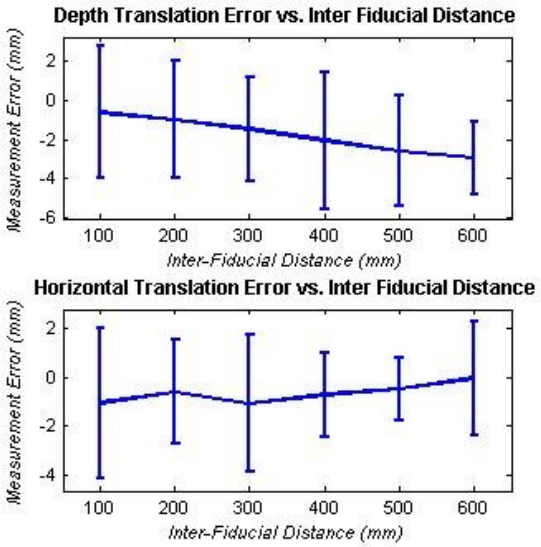

Figure 2. Measurement error vs. inter-fiducial distance for: (Top) translations away from the camera, (Bottom) horizontal translations

The second test was a horizontal translation performed at a distance of 1.5 meters. Without a scale change, the fiducial system performs more accurately in the translation test producing a low error value. Results for both horizontal and depth translation produced low error and were within the accuracy needed to perform sound localization.

B. TDOA-Based Sound Source Localization

Time difference on arrival (TDOA) based sound localization uses the fact that the speed of sound is finite and relatively slow compared to light. If the time difference between a sound arriving at each microphone is identifiable, then locations surrounding the array which might have made that sound can be extracted from the problem geometry. As mentioned previously, practical limitations such as ambient/robot noise and inter-microphone distances may restrict accuracy in all dimensions, but, particularly, range.

To evaluate our localization efforts, this work uses the generalized cross correlation algorithm to identify the actual time delay, organizing the results in a spatial likelihood [11]. Measurements are then combined over time and different physical sampling locations using an auditory evidence grid. Except for an adaptation to handle larger physical spaces, this 3D localization algorithm is used as described by Martinson and Schultz [4]. The remainder of this section describes: (1) a theoretical analysis of acceptable inter-microphone visual localization error, (2) the details of how we combined measurements together to localize a sound source together, and (3) how to reduce the computational load of 3D spatial likelihoods for real-time operation.

1) Theoretical Error Analysis

An analysis of the problem geometry allows us to identify theoretically acceptable inter-microphone localization error. Assuming that all of the error is split evenly between the two

microphones along the array axis, and that 5 degrees of sound localization error is an acceptable maximum, Figure 3 plots the maximum acceptable position error for varying angles of incidence. As the sound source angle of incidence from array normal increases, the acceptable inter-microphone position error decreases significantly. For small angles and large baselines, microphone localization error can be much worse than that of our existing visual system. But for sound sources located along the array axis (i.e. 90º), less than 1-cm of error is important for accurate localization.

2) Localizing a Sound Source in 3D

Spatial likelihoods are a sound localization approach based on maximum likelihood that uses a weighted cross-correlation algorithm to estimate the relative energy associated with all possible source locations. The idea is that the resulting cross-correlation value, adjusted for the predicted time difference on arrival, will be highest for those position/time differences corresponding most closely with the true value. The generalized cross correlation (GCC) value is determined separately for each microphone pair, and then summed across all microphone pairs for every position:

$$F_l = \sum_{a=1}^{N}\sum_{b=1}^{N}\int_{\omega} W(\omega)M_a(\omega)\overline{M_b(\omega)}e^{-j\omega(T(l,a)-T(l,b))}d\omega$$

where (M_a) is the Fourier transform of the signal received by microphone (a), $\overline{M_b}$ is the complex conjugate of (M_b), (ω) is the frequency in [rad/s], and (W) is a frequency dependant weighting function called the phase transform (PHAT):

$$W(\omega) = \frac{1}{|M_a(\omega)||M_b(\omega)|}$$

To localize a sound source, each measurement collected is scaled to [0.01,0.99] to form a likelihood. Then, each cell in a global auditory evidence grid is updated using log-odds notation to reflect this new measurement. This effectively stores measurements over time, and enables triangulation from multiple measurement positions to localize a sound source in 2 or 3 dimensions.

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

In these equations, $p(SS_{x,y}|z^t,s^t)$ 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 *inverse sensor model,* or probability that a single grid cell contains the sound source based on a single measurement.

To extract the most likely sound source position from the resulting evidence grid, cells whose value are less than 90% of maximum are discarded, and the remaining cells are clustered together. For each cluster *c*, the combined log-likelihood L_c and the weighted centroid μ_c are identified. The

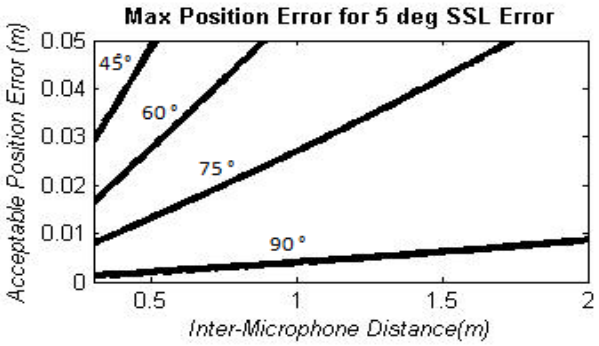

Figure 3. Maximum acceptable inter-microphone localization error given an allowable 5 degree sound localization error. The different lines represent varying angles of incidence away from the array normal.

centroid of the cluster with the greatest L_c is the most likely sound source position. Other clusters may indicate other sound sources, depending upon their combined likelihood.

3) Interpolation for Real-Time Operation

Although spatial likelihoods are determinable for large spaces in 2D at run-time (e.g. 100-m^2, 0.1-m stepSize), the expansion to 3D causes computational difficulties for real-time operation. Previous work using evidence grids avoided this limitation because all microphones were closely located in space, and, therefore, spatial likelihoods were determined over a limited region surrounding the robot (3-m from the array centroid). By allowing microphones to move as much as 3-m from the robot, however, the space over which spatial likelihoods need to be determined expands substantially.

The solution to this problem is to interpolate across calculated time-delays, rather than recalculate the GCC energy for every position in the spatial likelihood. At its core, the spatial likelihood function is a 3D representation of the underlying energy vs time-delay function. For each 3D position, a 1D time-delay is calculated for all microphone pairs, which, in turn, are used with GCC to identify a combined correlation energy. The range of potential time-delays is defined by the distance between microphones (*d)* and the speech of sound (c_s):

$$delay \in \left[-d/c_s, d/c_s\right]$$

By regularly sampling time-delay range, and calculating the corresponding energy (see Figure 4), we can then use interpolation to convert the time-delays associated with each 3D position to the corresponding energy without calculating the GCC energy for all position. For an area of 6x6x3-m^3, this means a computational reduction from 1.08x10^8 GCC

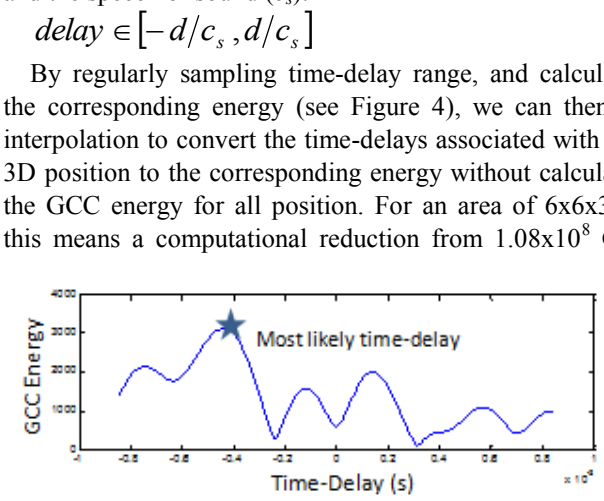

Figure 4. Cross correlation energy is dependent only on time-delay. Interpolation enables faster construction of 3D spatial likelihoods.

calculations to 2000 (the number of delays we use per sample). This makes the spatial likelihood computationally feasible for run-time calculations over larger areas and smaller grid cell sizes.

IV. System Evaluation

The dynamically reconfigurable microphone array was evaluated in 2 series of tests. In the first, 2 wireless microphones were used by themselves to localize music sources in the environment from a number of different positions. In the second tests, the 2 wireless microphones were combined with a rigid binaural array on top of the b21r robot. The robot then self-discovered wireless microphone positions with which to augment its onboard array.

The testing environment was a mobile robotics laboratory. Ventilation, transportation, and speech noise were common during testing. The sound source evaluated was a computer speaker playing an fm radio broadcast at ~65dBA. Samples were collected from both the wireless and static microphone arrays at 11025 hz in 2048 sample packets. All known microphone positions were recorded for each packet. Samples for which less than 2 microphone positions were known were discarded.

A. 2 Wireless Microphones

To evaluate the performance of the visual localization method, the first test uses just wireless microphones. The accuracy of the sound source localization method is directly related to the accuracy of the microphone localization. Therefore, in this test, the two wireless microphones were moved by hand to different pair position configurations. A total of 8 trials tracking the computer speaker were completed, with 5-9 different pair positions recorded per trial. The difference in the number of pair positions was due to the visibility of the microphones. When one microphone was poorly visible to the camera, audio was not recorded.

To analyze performance, 2-sec of audio were sampled randomly from each position and an evidence grid built. For all combinations of 3-5 positions, a combined evidence grid was constructed and a sound source identified. For [3,4,5] mic-pair positions, a total of [142,132,85] grids were constructed and sources extracted across all 8 trials. Figure 5 summarizes the combined results.

As should be expected, localization accuracy increases with the number of positions. With data from only 3 positions, a classifier with a log-likelihood threshold set for 600 would accept 60% of all sound source positions and have a mean error of 0.32-m. Using 4 positions increases performance with the same threshold to 74% and 0.26-m. 5 positions means 82% and 0.24-m. The last graph in Figure 5 shows an ROC-curve, where a positive classification is determined to be anything within 8^3-in of the target. 8" was selected because the speaker is ~8" tall. The ROC curve shows that classification performance for only 3 positions has a maximum positive classification rate of 54%. Using more positions raises that maximum to 76% and 82% for 4 and 5 positions respectively.

To compare these results better to other localization work, Table 1 compares the mean error between 2D and 3D localization for log-likelihood thresholds that accept 80% of the localized sound sources.

Table 1. Mean-error comparison between 2D and 3D localization for an 80% acceptance rate.

# of Mic Positions	Mean Error(2D)	Mean Error (3D)
3	0.26m	0.36m
4	0.20m	0.27m
5	0.18m	0.24m

B. 2 Static + 2 Wireless

Using only a standard computer sound card, a robot is limited to binaural inputs. Even when using wireless microphones, the audio signal must ultimately be sampled and recorded by the computer using an A/D converter. When the converter has more channels, however, the robot can mix onboard sensing with wireless microphones. In this second set of trials, the dynamic nature of the microphone array is demonstrated by having the robot autonomously discover microphones in the environment through robotic movement. Discovered microphones are then mixed with 2 onboard microphones to better localize a stationary sound source.

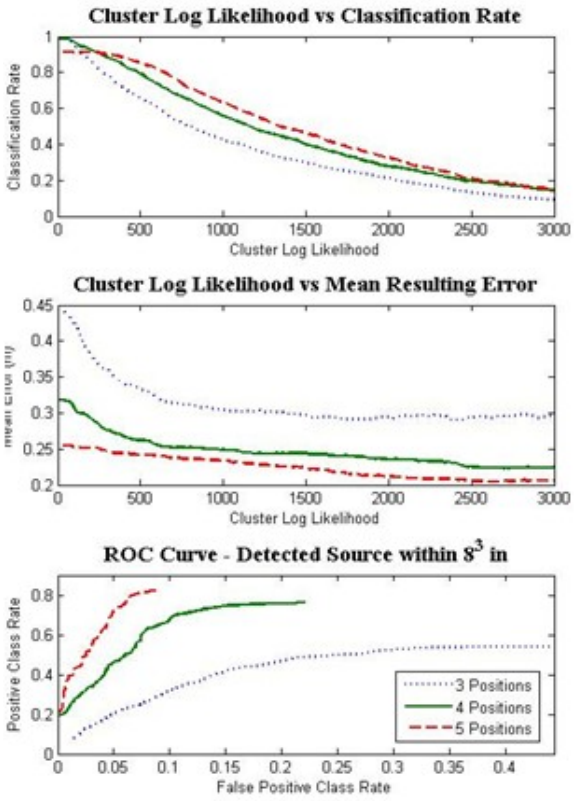

Figure 5. Localization performance increases with the number of stops. (Top) Classification rate is the percentage of samples accepted as valid sound sources. (Middle) Mean resulting error examines distance to ground truth for all sound source positions whose combined log-likelihood exceeds the threshold. (Bottom) Percentage of sources detected within 8^3-in of ground truth vs. sources outside 8^3-in.

Robot ego-noise measures 51dBA at the stationary microphones and ~49dBA ambient.

In these trials, two wireless microphones were placed within 3-m of the robot at arbitrary locations. The robot then rotated in place until it found each of the microphones. 10 samples were collected at each of 4 locations: twice without any visible microphones, and twice with at least one wireless microphone visible. A total of 10 trials were completed in this fashion. The wireless microphones were moved to new locations between each trial.

Using the self-discovered wireless microphone signals, the robot successfully localized the sound source in 3D with an average error of 0.25m. Using only the two onboard microphones, this average error increases to 0.53m. Figure 6 demonstrates why. With only the two microphones mounted on the robot, only an angle is reliably extracted by rotating the platform. By extending the array baseline, however, the sound source position becomes identifiable.

V. Conclusion

The goal of this initial work in dynamically reconfigurable microphone arrays has been to demonstrate their feasibility, and evaluate their accuracy. This we accomplished through two demonstrations, one with two wireless microphones moved by hand through an environment, and the other by having a robot discover wireless microphones and use them to augment its onboard microphone array. The accuracy of the combined sound localization system is within 0.3-m for 3D spatial coordinates, which is very good for a variety of sound localization type applications including human-robot interaction and security robotics.

The next step in this work is to apply it to robotic teams. With only two microphones, 2 or 3D source localization requires sampling from many different positions. Placing the sensors on robots and localizing them dynamically makes this strategy feasible. Unfortunately, localizing through movement is not effective for short term noise sources that disappear before robots have moved. This limitation is due to the shape of the array and the number of microphones, not the basic sound localization theory. By adding more microphones and separating them in space, localization of short-term sources becomes possible without additional sample positions. Robotic teams are interesting here, because making the individual nodes in this network mobile enables autonomous microphone repositioning for short duration source detection in response to user specified observation regions. Furthermore, it also enables dynamic reconfiguration of the network in response to changing environmental conditions. If a new ambient noise source begins to disrupt reception in one area, then one or more microphones can shift locations around the problem source to maximize signal-to-noise ratios in other areas.

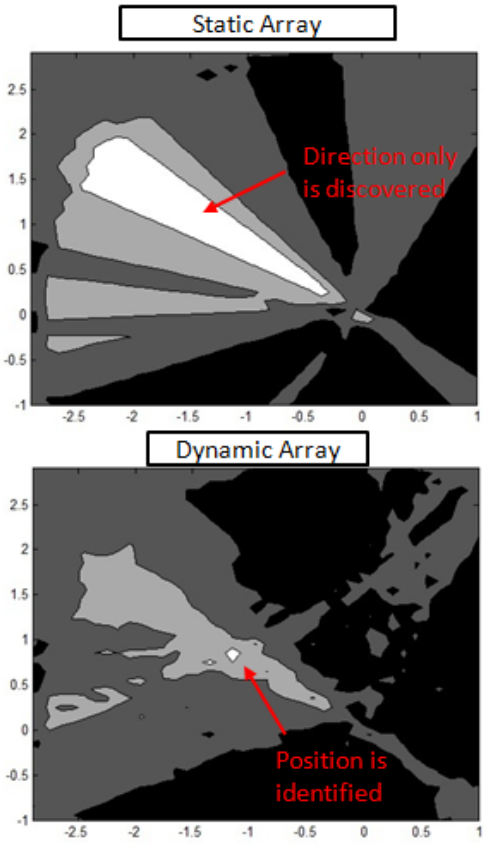

Figure 6. With only the onboard static array, rotating the robot discovers direction. Using the dynamic array with wireless microphones, a position is identified.

VI. References

[1] J.M. Valin, Auditory System for a Mobile Robot, 2005.

[2] K. Nakadai et al., "Robust Tracking of Multiple Sound Sources by Spatial Integration of Room And Robot Microphone Arrays," in *ICASSP*, Toulouse, France, 2006.

[3] H. Kato, M. Billinghurst, " Marker Tracking and HMD Calibration for a video-based Augmented Reality Conferencing System," In Proc. of the 2nd International Workshop on Augmented Reality (IWAR 99). pp. 85-94 (1999).

[4] E. Martinson and A. Schultz, "Discovery of sound sources by an autonomous mobile robot," *Autonomous Robots*, vol. 27, no. 3, pp. 221-237, 2009.

[5] E. Martinson and D. Brock, "Improving Human-Robot Interaction through Adaptation to the Auditory Scene," in *HRI*, Arlington, VA, 2007, pp. 113-120.

[6] M. Heckmann, T. Rodemann, F. Joublin, C. Goerick, and B. Scholling, "Auditory Inspired Binaural Robust Sound Source Localization in Echoic and Noisy Environments," in *IROS*, Beijing, China, 2006, pp. 368-373.

[7] Y. Sasaki, S. Kagami, and H. Mizoguchi, "Multiple Sound Source Mapping for a Mobile Robot by Self-motion Triangulation," in *IROS*, Beijing, China, 2006, pp. 380 - 385.

[8] D. Bian, G. Abowd, and J. Rehg, "Using Sound Source Localization in a Home Environment," in *Pervasive Computing*, Munich, Germany, 2005, pp. 19-36.

[9] S. Thrun, "Affine Structure from Sound," in *NIPS*, Whistler, Canada, 2005, pp. 1353-1360.

[10] L. Girod and D. Estrin, "Robust Range Estimation Using Acoustic and Multimodal Sensing." in *IROS*, Wailea, HI. 2001 p. 1312:1320

[11] B. Mungumaru and P. Aarabi, "Enhanced Sound Localization", *IEEE Trans on Systems, Man, and Cybernetics*, vol 34, no 3. 2004