

Optimizing a Reconfigurable Robotic Microphone Array

E. Martinson¹, T. Apker¹, M. Bugajska²

Abstract—Relative positioning of microphones in an array has a significant impact on how well the array can localize sound sources. Two- or three- dimensional localization accuracy of a randomly distributed array will vary widely with respect to the relative position of the sound source. Where a statically located array would be forced to work with its initial configuration, however, a reconfigurable robotic array does not have to remain in its starting position. Given a known region of interest, robots can autonomously optimize their relative configuration to improve localization accuracy where it matters the most. In this work, we propose 2 separate strategies for optimizing such a robotic array. Evaluations are completed first in simulation, and then deployed to a team of real robots.

I. INTRODUCTION

The key to building effective microphone arrays has always been to know the environment in which the array will be deployed. A small microphone array reliably estimates only direction [1], but in an environment with few ambient noise sources, localizing and amplifying in only a single dimension can be effective. In a crowded acoustical environment, however, a directional amplification pattern often picks up significant noise from the background. A distributed room-mounted microphone array, by contrast, with greater distances between individual microphones, is better suited for such conditions, as it separates sounds in multiple dimensions, localizing sources and amplifying them for signal analysis [2]. When choosing microphones for robot audition, unfortunately, there is often need for multi-dimensional separation of sound sources, but no space for a large on-robot array, and little time to setup a room a priori. In this work, we overcome these limitations by spreading the microphones out across multiple mobile robotic platforms. By utilizing robotic movement to optimize their physical array configuration for a specific area, this distributed robotic array can localize a variety of sound sources in multiple dimensions like a much larger room mounted microphone array without building a bigger physical robot.

The challenge in building distributed microphone arrays is

that randomly distributed microphones have a randomly distributed amplification pattern. That is, they will localize sound sources in some parts of room with high accuracy, and very poorly localize in other areas. To address this problem, this work takes advantage of the actuation of the individual elements in the microphone array. If a region of interest, i.e. a region in which sound activity is of greatest interest to the operator, can be defined a priori (such as a region in front of a security checkpoint, or an open area in which people congregate), then robotic microphones can redistribute themselves dynamically to improve reception and localization accuracy in that particular region. This paper demonstrates such improvements in both simulated scenarios with various array and region of interest configurations, as well as using 3 real mobile robotic platforms.

II. RELATED WORK

The multi-dimensional localization problem has been given increased attention as robot audition has matured. Initially, the focus was on combining measurements from different locations to effectively triangulate upon a sound source [1,3]. This required that sound sources remain active long enough for a robotic system to gather multiple measurements within detection range, a significant limitation when listening for speech or other short-duration noise. Murase et al [4] also took advantage of movement, but instead of the robot moving, a SLAM like approach reconstructed the path of a moving sound source. More recently, Rodemann [5] demonstrated that with biologically inspired hardware and software, movement is not always necessary for estimating both distance and angle.

There has been more effective work in tracking or localizing speech using room-mounted microphone arrays. As one or more individuals walk through a room, microphones equipped to walls and ceiling can track individuals [6], and integrate that knowledge into person identification systems [2]. Nakadai et al [7] have integrated robotic sound localization results with room mounted arrays also for tracking, but without synchronizing streams between the two. Martinson and Fransen [8] synchronized streams between robot mounted microphone arrays and scattered wireless microphones, but did not address how to improve an initially poor configuration.

In general, this work seeks to extend the advantages of these room mounted microphone arrays without manually equipping an environment a priori. Like [2,6,7], localization,

Manuscript received Mar 22, 2011. This work was supported by the Office of Naval Research under job order numbers N0001408WX30007 and 09-Y861.

¹E. Martinson and T. Apker are post-doctoral fellows at the U.S. Naval Research Laboratory, Washington, DC 20375. (phone: 202-404-4948; e-mail: {eric.martinson.ctr, thomas.apker.ctr}@nrl.navy.mil).

²M. Bugajska is with the U.S. Naval Research Laboratory, Washington, DC 20375 (magda.bugaska@nrl.navy.mil)



Fig 1. (Top) The evaluated dynamic microphone array utilizing a centrally located omni-directional camera. Microphones are rigidly attached to fiducials for visual localization. (Bottom) A robot mounted version of the omni-directional camera base-station with onboard visual processing and attached microphone array.

tracking, and ultimately audio streaming are the goals, but without the need for human array and equipment handling beyond robotic platform selection. Furthermore, a robotic instrumentation of an environment could be quickly adapted in the face of changing operational targets. This adaptive array configuration is similar to work in radio-frequency geolocation using UAV's [9]. In that domain, multiple aircraft are tasked with detecting and tracking multiple emitters using time-delay on arrival (TDOA) based algorithms. Optimal utilization of limited resources thus guides the dynamic reconfiguration of sub-arrays

III. DYNAMICALLY RECONFIGURABLE ARRAYS

A dynamically reconfigurable microphone array [8] is an array that can change its physical properties, including number and positioning of microphones, in response to external stimuli. In this work, such an array is constructed with three Pioneer3-AT robots (Figure 1, top). Two Pioneer-3AT robots are each equipped with one microphone that communicates wirelessly with a base-station. A third robot is equipped with two wired microphones and the wireless base-stations. The signals of all microphones are amplified with battery powered preamps and then sampled using an 8 channel PCMCIA A/D converter. These signals then feed continuously into the same data acquisition unit (DAQ) on the third robot, thus synchronizing the auditory streams.

A. Visual Microphone Localization

To make a dynamic array configuration useful for real time sound source localization, microphones within the array need to be localized quickly and accurately. For this purpose, we use a Point Grey ladybug-3 omni-directional camera system. The ladybug consists of six wide-angle cameras with overlapping visual spaces. The visual system localizes microphones in the environment by detecting rigidly attached 10-cm fiducials. Software constructed using the AR Toolkit [10] then identifies relative translational and rotational offset

from the fiducial to the visual frame, converting fiducial locations into a global coordinate using a pre-estimated camera transformation matrix. These same fiducials were also used in combination with robot odometry to localize individual robot team members. The localization error for such a system is generally less than 1cm, and is presented in greater detail in [8].

For this work, all microphones were localized dynamically using a centrally mounted ladybug camera (see Figure 1, top). In the future, however, we intend to move the ladybug camera onto the same robot carrying the wireless base-stations (Figure 1, bottom).

B. Sound Source Localization

With known microphone positions and synchronized auditory streams, sound source localization is completed using generalized cross correlation (GCC). Described in more detail in [5], spatial likelihoods are a sound localization approach based on maximum likelihood that uses GCC 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 GCC value is determined separately for each microphone pair, and then summed across all microphone pairs for every position:

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

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 based on an average noise sample, if known, or using the PHAT transform if not.

To localize a sound source, every half-second sample is first analyzed to determine the SNR. If the SNR in two or more microphones is >1 , then the spatial likelihood is estimated and scaled to $[0.01, 0.99]$. 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 [1].

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 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.

IV. DYNAMIC OPTIMIZATION OF A MICROPHONE ARRAY

Given a set of microphones distributed across multiple robots, how can the robots best move to minimize sound localization error across a known region of interest when using a time delay on arrival based algorithm? What motivates this problem is surveillance. With an operator to

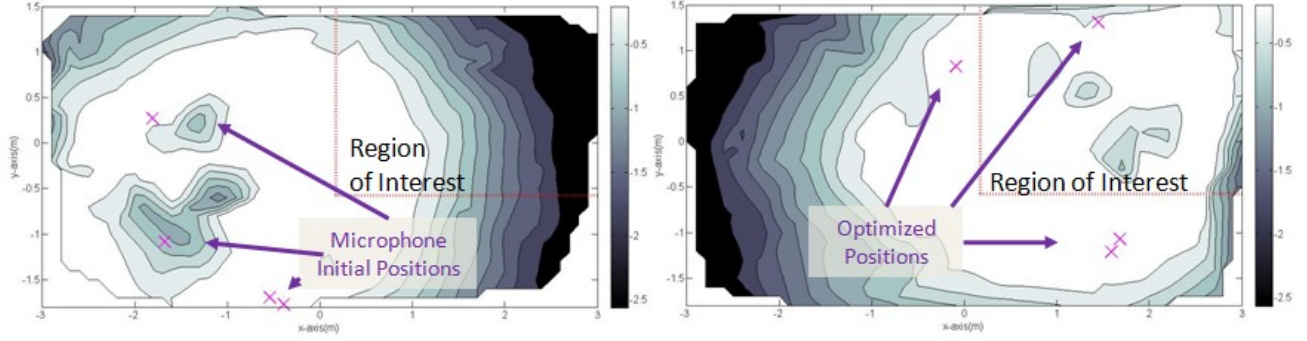


Fig 2. Simulated localization error (m) interpolated across 200 samples using (left) initial configuration and (right) optimized configuration. The dotted rectangle indicates the region of interest the microphone array was being optimized for.

identity areas of the environment in which most interesting or relevant sound sources will be located, the platforms comprising the robotic array can autonomously position themselves with greatest accuracy for listening. It is not, however, a trivial problem to select an optimal array configuration, and configuration really matters. In Figure 2, localization error is simulated for a randomly distributed set of microphones. In the region of interest, error is very high, but even within the vicinity of the microphones there are regions that are poorly localized. By changing the positions of the microphones, we can dramatically reduce error across the region of interest (Fig 2, right).

We propose two solutions to this problem. Both solutions use an iterative approach where robots move based on local decisions on how best to improve the array. The first solution is based on a cost function derived from [9], in which sensor arrays are formed based on optimal angular layout. The second solution is a simpler attraction model where robots are attracted to the region of interest while being repulsed by other robots. Both methods were evaluated in simulation. The best was also deployed on real robots.

A. Optimization Strategy 1: Spreading out the Angles

In [9] sensors in a mobile network are grouped for optimal reception based on a cost function that identified the minimum separation angle between all receivers about a suspected emitter location. For our first optimization strategy, we designed a similar cost function. Let *samples* be a selection of potential sound source positions (size = *ns*) selected at regular intervals from the region of interest. Then for each sample location *s*, identify the relative position of robot *i* to *s* in polar coordinates ($[th_{i,s}, r_{i,s}]$). Based on the empirical results discussed in section 3.B, those robots with $r_{i,s} > 3$ -m are discarded due to their distance. The remaining *N* positions are used to calculate a cost of the configuration for each sample. The total cost of any given robotic array configuration is then the mean cost for all samples.

$$th_{\min,s} = \min(th_{i,s} - th_{j,s}) \forall i \neq j$$

$$weight(s) = \begin{cases} 2\pi/3, & \text{if } N < 3 \\ 2\pi/N - th_{\min,s} \end{cases} \quad \text{Eq 2}$$

$$totalcost = \sum_s weight(s) / ns$$

Even with an established cost function, optimization remains non-trivial. In particular, our cost function does not

take into account any features of the environment such as obstacles when calculating cost over the region of interest. Even if it did, evaluating cost for every possible array would be computationally prohibitive. For this reason, we use a behavior-based approach that guides robots towards more optimal configurations in line with environmental restrictions. The cost function is incorporated into a behavior which estimates the value of potential movement directions by an individual node in the sensor array.

$$\vec{V}_{out,i} = \vec{V}_{spread,i} + \vec{V}_{cam,i} + \vec{V}_{obstacles,i} \quad \text{Eq 3}$$

The spreading vector is created as follows:

- 1) Estimate the cost of 24 hypothetical arrays in which the sensor node in question has moved 0.2-m (i.e. at 15 degree intervals)
- 2) Create a unit vector \vec{V}_{best} towards the direction that leads to the greatest reduction in cost ($\vec{V}_{best}=0$ if no improvement can be made)
- 3) Use the current array cost to weight \vec{V}_{best}

$$\vec{V}_{spread} = \frac{cost_{current}}{2\pi/3} \vec{V}_{best} \quad \text{Eq 4}$$

\vec{V}_{cam} is an attraction vector designed to keep individual nodes in range of the visual localization system. Beyond 2-m, fiducial localization is unreliable using the omnidirectional ladybug camera. Fiducials may still be detected, but they are often not uniquely identifiable. Given that $r_{cam,i}$ is the distance from node *i* to the camera:

$$\vec{V}_{cam,i} = \begin{cases} [0,0], & \text{for } r_{cam,i} < 1.6 \\ \frac{2-r_{cam,i}}{0.4r_{cam,i}} ([x_{cam}, y_{cam}] - [x_i, y_i]) \end{cases} \quad \text{Eq 5}$$

Obstacles are detected by laser and sonar readings on each robot. $\vec{V}_{obstacles}$ is a weighted repulsion vector from every obstacle point within 0.5-m of the robot.

B. Optimization Strategy 2 – Attraction/Repulsion

The principal cost estimation component of the first optimization strategy was designed to spread the robots out and surround the target. Our addition of a maximum distance requirement (e.g. robots must be within 3-m of the target to assist with cost reduction) also encourages individual nodes to move within 3-m of as many points in *samples* as possible. These basic needs, however, do not necessarily require angle computations and/or knowledge of the entire array. Instead, the problem can be approached as an area coverage scenario

in which robots are simultaneously attracted to points in the region of interest while being repelled by other robots.

$$\vec{V}_{out,i} = \vec{V}_{samples,i} + \vec{V}_{robot,i} + \vec{V}_{cam,i} + \vec{V}_{obstacles,i} \quad \text{Eq 6}$$

V_{cam} and $V_{obstacles}$ are the same as described in the previous section. V_{robot} represents the summation of individual repulsion vectors from all other robots in the array. Repulsion vectors have safety margin = 0.5m, and sphere of influence = 3.0m [11]. $V_{samples}$ is an attraction vector created from the summed attraction forces towards all points s in $samples$. Attraction forces are weighted by relative distance, as described in [11], but only assume a length of 1.0 beyond 3.0-m (e.g. sphere of influence):

$$\vec{V}_{s,i} = [x_s, y_s] - [x_i, y_i]$$

$$W_{s,i} = \begin{cases} 0, & \|\vec{V}_{s,i}\| < 1.5 \\ \|\vec{V}_{s,i}\|/1.5 - 1, & 1.5 \leq \|\vec{V}_{s,i}\| < 3.0 \\ 1.0, & 3.0 \leq \|\vec{V}_{s,i}\| \end{cases} \quad \text{Eq 7}$$

$$\vec{V}_{samples,i} = \frac{1}{10} \sum_s W_{s,i} \vec{V}_{s,i} / \|\vec{V}_{s,i}\|$$

V. SIMULATED EVALUATION

Before deployment on real robots, the optimization algorithms were first evaluated in simulation to determine general usefulness. We used simulated speech to measure localization accuracy, including masking noise, sound decay, and microphone measurement error effects.

A. Simulation Setup

Both optimization strategies were evaluated using Matlab. For this work, we assumed that the camera remained stationary at the center of the environment. Microphones were then placed into this environment at random locations within range of the camera and the array configuration performance was evaluated. Because the focus of testing was evaluating optimization algorithm performance and not obstacle avoidance, obstacles were not included in the simulation. Obstacles were, however, present in the real implementation (sec VI).

To improve array performance, each microphone calculated its V_{out} (this varies between optimization algorithms) and moved up to 0.2-m in the specified direction. The actual distance moved was directly related to the magnitude of its movement vector (e.g. $dist = |V_{out}|*0.2$). When the combined distance moved by all microphones was less than 5-cm, indicating convergence upon a stable solution, the simulation stopped and the resulting array configuration was evaluated.

B. Sound Propagation Simulation and Evaluation

Each microphone array configuration was evaluated using the same recorded 1-sec speech clip with 22050-Hz sampling rate. Given known emitter and receiver locations, an auditory stream was generated for each receiver in which the sample had been delayed and decayed as per an emission from a spherical sound source [12]. Noise was then added to each

stream assuming a constant room noise volume, with signal-to-noise ratio = 20dB at 0.3-m from the source. With constant noise, but decaying signal, the SNR at each microphone decreased with ~3dB/m from the source. The frequency characteristics of the added noise matched the spectral analysis of recorded noise in an office environment.

Sound localization accuracy was evaluated at regular intervals across the region of interest. The interval width varied with region size and shape, so as to investigate ~50 sample locations in any given region. At each location, a set of streams were generated using the simulator.

To simulate visual localization error, Gaussian noise ($\sigma=5\text{mm}$) was added at this point to all microphone locations. All streams were then analyzed as if on a real system, using generalized cross correlation to construct a spatial likelihood. Finally, the sound source position was identified using the weighted centroid of the cluster with greatest combined energy (Sec III.B).

C. Evaluation

Simulations were conducted in two stages. In the first stage, a square region of interest (6m^2) was defined with the camera at its centroid. Because all randomly selected initial positions were constrained to be within visual detection range (e.g. 2-m), most robots started within sound localization range of the entire region. 100 randomly selected initial configurations were evaluated with 4 independently moving microphones. The same initial random configuration was used for both optimization algorithms.

Table 1. Error when microphones begin in view of the region of interest.

	Mean	Var	Improvement over Random
Random	0.30	0.25	
Spreading Angles	0.11	0.07	63%
Attraction Model	0.09	0.05	70%

The purpose this first stage was to evaluate improvement when distance between sound sources was less important. In this case, both optimization strategies improved over random >60%, lowering both mean localization error and variance. Using an optimization strategy to improve localization accuracy is clearly important, even when all microphones begin within range. Under these conditions, attraction/repulsion showed some advantage over spreading angles, but the real difference was small.

The second stage of testing evaluated the more general case, in which the region of interest was randomly selected. Regions of interest were constrained to be of similar area, 6-m^2 , with no axis less than 1m in length, and be fully within the confines of the defined testing environment ($6.1 \times 3.5\text{m}$). 100 regions of interest were evaluated. 3 different array types were also evaluated for each region: 1) 3 independently mobile microphones, 2) 4 microphones, but only 3 robots (e.g. one binaural configuration, as was used with the real robots), and 3) 4 independently mobile microphones.

In this second evaluation, it is clear that using 4 microphones is better than only 3. While not surprising, it is important because it was still true when two microphones were forced to remain in close proximity as part of a binaural pair. Simply adding another microphone shows the greatest

Table 2. Array type vs optimization strategy mean error.

	3-Mics	4-Mics, 3 Robots	4-Mics
Random	0.87	0.60	0.62
Spreading Angles	0.38	0.21	0.17
Attraction Model	0.29	0.15	0.09

difference in the random distribution, an improvement of 30%, but an additional microphone in either optimized configuration improves accuracy by nearly 50%. Using a 4th independently moving microphone further improves results over the binaural configuration.

This testing stage also demonstrated the strength the attraction based optimization. Although the spreading angles approach is effective, reducing error by 73%, attraction/repulsion shows an 85% improvement. This performance improvement is most noticeable in non-square regions of interest (Figure 3). When axes are of comparable length, arrays using either optimization method demonstrate similar mean error. But as the maximum axis length increases (i.e. the region becomes less square), the attraction/repulsion model does a better job of optimizing the array. The reason for this improvement is that distance between microphones and sound sources is important to acoustic localization. With spreading angles, distance is not included in the algorithm. Robots only try to be within range. Attraction/repulsion models force robots to get as close to as much of the region as possible. As a result, the average distance using spreading angles is 1.75m, but only 1.55m using attraction/repulsion.

VI. REAL ROBOT EVALUATION

The simulation results suggest that an attraction/repulsion model is a better method for optimizing an array in response to a known region of interest. To determine whether optimizing a real robotic array was also feasible, we deployed the attraction/repulsion optimization model on a real robotic system and evaluated its performance.

A. Implementation Details

The microphones with their associated fiducials were mounted behind the SICK laser scanner on Pioneer3-AT robots with indoor wheels (Figure 1). The low-level interface to the wheels, sonar and laser sensors on the robots was provided by the Player robot server [13] version 3.2. Each robot ran a client on its embedded computer that implemented an unscented Kalman filter (UKF) to generate pose estimates based on the robots' odometry and the fiducial position estimates passed from the camera server to the robots. At approximately 2Hz, each robot reported its position to a central server that computes the next force on the robot based on its estimated position and the attraction/repulsion vector as described in Section IV.B.

Robot guidance was handled by an on-robot client and implemented a modified version of the physicomimetics framework [14] that accounted for the non-holonomic nature of the Pioneer 3AT. The UKF estimated the position of the center of the robot, where the fiducial was mounted, and the robot's heading. The guidance law applied the force generated by the potential field to a virtual particle placed

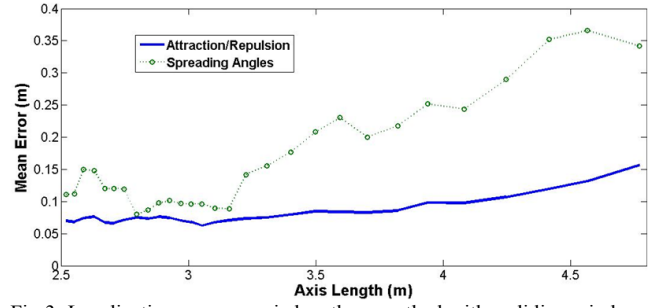


Fig 3. Localization error vs axis length, smoothed with a sliding window. Error increased faster when using spreading angles to optimize the array.

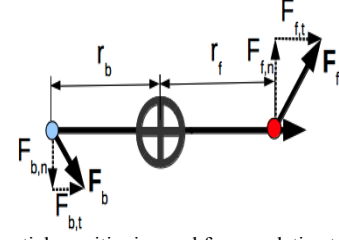


Fig 4. Virtual particle positioning and forces relative to robot centroid.

10cm in front of the center of the robot (figure 4). Obstacle avoidance in the physicomimetics framework was handled by placing an obstacle particle at the location of the nearest laser (front) or sonar (back) return, each of which applied a repulsive force to the front or back virtual particle, respectively. The component of the forces tangent to the direction of the robot, $F_{f,t}$ and $F_{b,t}$, were summed to determine the robot speed, and the normal components, $F_{f,n}$ and $F_{b,n}$, applied a torque that turned the robot toward the direction of the applied force.

B. Experimental Design

The real robot implementation of the attraction/repulsion optimization strategy was evaluated in a total of 15 robotic trials. In the first 10 trials, robots were placed initially in a line configuration within view of the centrally located camera system. The 5 remaining trials evaluated a better initial configuration roughly in the shape of a triangle. This triangular configuration is centered within the room, and so began within sound localization range of most points. It is also less symmetric than the line configuration. Figure 5 shows a map of the testing area. Besides the obstacle on which the camera was located, a wall to one side of the testing environment and a pillar on the other blocked access to some regions of the room.

Ten different regions of interest were randomly generated for these trials. Robots were informed by an operator of the region boundaries, and then autonomously optimized their configuration. Sound localization performance was measured before and after optimization, using the same set of samples for each region. 10 samples/region were randomly selected, for a total of 100 evaluated sample positions. Trials using the triangular configuration used the same region of interest set.

Sound localization performance was evaluated by manually moving a speaker to each sample position and playing each of two sounds. The first sound was a 5-sec recording of pop music that ranged that had a maximum SNR of 30 at 0.3-m from the source. The second sound was

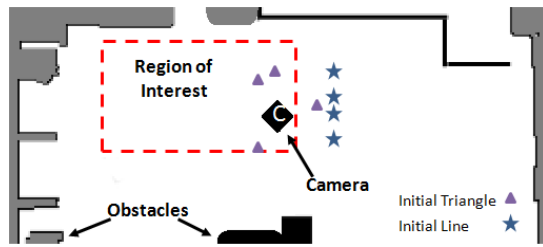


Fig 5. Experimental setup using real robots.

a recording of the spoken sentence, “Testing, testing, I am now reading a set of stories from the BBC”. The speech recording had a maximum SNR of 20. On the robots, however, the signal was masked by increased ego-noise (on-robot noise volume measured 62-dBA). As such, the maximum recorded speech SNR for any 0.5-sec sample was 3-dB, with only 30% of samples recording an SNR > 1dB.

C. Results

From an initial line configuration, there was a significant advantage, as estimated by sound localization accuracy, in optimizing the array. Table 3 summarizes the results.

Table 3. Mean localization error using real robots.

Trial	Distance Moved	Music		Speech	
		Initial(m)	Final(m)	Initial(m)	Final(m)
1	1.7m	1.36	0.46	0.86	0.24
2	10.4m	1.57	0.53	1.52*	0.52
3	2.4m	2.22	0.36	1.52	0.48
4	8.0m	1.32	0.13	1.7	0.23*
5	4.2m	1.00	0.27	0.72	0.34
6	2.7m	0.94	0.33	0.44*	0.39
7	1.8m	0.52	0.26	0.77	0.19
8	0.7m	0.52	0.47	0.43	0.50
9	7.2m	0.90	0.18	1.31	0.20
10	6.0m	1.28	0.11	1.16	0.18
Combined		1.16	0.31	1.1	0.33

** Speech SNR never passed 1dB on some sample positions (3 initial, 4 final), so no source was detected. These were discarded as anomalous.

An initial line configuration is particularly bad for estimating sound source positions. Regardless of where in the room the region of interest was located, or which sound source was being localized, the mean error for the initial line configuration was above 1m. By contrast, the optimized array configuration had a mean error of 33cm using the quieter speech source, and 31cm with the music sound source. There is a clear advantage in optimizing the array configuration when moving out of a line formation.

The triangular configuration, hypothesized to be better than the line configuration, had a mean error of 42cm, a significant improvement over the line. Optimizing this configuration further improved accuracy by 19%, or 8cm. This shows that even when the array begins in a better configuration within range of most sound sources, attraction/repulsion based optimization can still be used to enhance the sound localization capabilities of a reconfigurable robotic microphone array.

VII. CONCLUSION

Given a region of interest and the ability to move the microphone elements in an array, this work demonstrates that

not only is there a significant advantage in optimizing the relative positions of robotic microphones, but straightforward attraction/repulsion models can be used to accomplish this. In this work, we explored this problem both in simulation and on real robots using a small number of microphones to cover a 6-m² area. With an optimized configuration in place, we can determine multi-dimensional coordinates of both medium and short duration (speech in particular) noise sources in the region of interest. Simulations suggest a more than 60% improvement over random, even when all microphones begin within detection range. Real robotic results also demonstrated significantly better accuracy over common team formations such as lines and triangles. The latter optimizations using real robots were observed in the presence of obstacles. Future work on this problem will focus on generalizing the optimization by 1) making the camera base-station mobile, and 2) incorporating different environmental conditions such as terrain height, reverberation, and sound source type.

REFERENCES

- [1] E. Martinson and A. Schultz, “Discovery of sound sources by an autonomous mobile robot,” *Autonomous Robots*, vol. 27, no. 3, pp. 221-237, 2009.
- [2] A. Salah et al., “Multimodal identification and localization of users in a smart environment,” *Journal of Multimodal User Interfaces*, vol. 2, pp. 75-91, 2008.
- [3] Y. Sasaki, S. Kagami, and H. Mizoguchi, “Multiple Sound Source Mapping for a Mobile Robot by Self-motion Triangulation,” in *Proc. of Intl Conf. on Intelligent Robots and Systems*, Beijing, China, 2006, pp. 380 – 385.
- [4] M. Murase et al., “Multiple Moving Speaker Tracking by Microphone Array on Mobile Robot,” in *Interspeech*, Lisbon, Portugal, 2005.
- [5] T. Rodemann, “A study on distance estimation in binaural sound localization,” in *Proc. of the Intl. Conf. on Intelligent Robots and Systems*, Taipei, Taiwan, 2010.
- [6] Mumgumar, B. and P. Aarabi, “Enhanced Sound Localization,” *IEEE Trans on Systems, Man and Cybernetics*, Part B, 34(3), 2004
- [7] K. Nakadai et al., “Robust Tracking of Multiple Sound Sources by Spatial Integration of Room And Robot Microphone Arrays,” in *ICASSP*, Toulouse, France, 2006.
- [8] E. Martinson and B. Franssen, “Dynamically Reconfigurable Microphone Arrays,” in *Proc. of the Intl Conf on Robots and Automation*, Shanghai, China, 2011.
- [9] Marsh, L., D. Gossnik, S.P. Drake and G. Calbert, “UAV Team Formation for Emitter Geolocation,” in *Information, Decision and Control*, Adelaide, Australia 2007
- [10] H. Kato and H. Billinghurst, “Marker Tracking and HMD Calibration for a video-based Augmented Reality Conferencing System,” in *Proc. Of the 2nd International Workshop on Augmented Reality*, San Francisco, CA, 1999.
- [11] Arkin, R. *Behavior-Based Robotics*, MIT-Press, 1998
- [12] D. Raichel, *The Science and Applications of Acoustics*. New York: Springer-Verlag, 2000.
- [13] Gerkey, B. et al “Most Valuable Player: A Robot Device Server for Distributed Control.” In *Proc. of the Intl. Conf. on Intelligent Robots and Systems*. Wailea, Hawaii, 2001
- [14] Wiegand, R., M. Potter, D. Sofge, and W. Spears. “A Generalized Graph-Based Method for Engineering Swarm Solutions to Multiagent Problems”. In *Proc of 9th International Conference on Parallel Problem Solving from Nature*, 2006