An Automated Acoustic System to Monitor and Classify Birds

C. Kwan[+], K. Ho[*], G. Mei[+], Y. Li[*], Z. Ren[+], R. Xu[+], G. Zhao[+], M. Stevenson[+], V. Stanford[&], and C. Rochet[&]

[+]Intelligent Automation, Inc.
7519 Standish Place
Suite 200
Rockville, MD 20850
USA

[*]Dept. of EE
U. of Missouri at Columbia
123 Engineering Building West
Columbia, MO65211
USA

[&]NIST
Technology Building (225)
Room A216
Gaithersburg, MD 20899
USA

Abstract

Collisions between aircraft and birds have become an increasing concern for human health and safety. More than four hundred people and over four hundred aircraft have been lost globally in recent years, according to a FAA report. To minimize the number of birdstikes, microphone arrays have been used to monitor birds near the airport or some critical locations in the airspace. However, the range of existing arrays is only limited to a few hundred meters. Moreover, the identification performance in low signal-to-noise environment is not satisfactory.

Under the support of the US Air Force, Intelligent Automation, Incorporated (IAI) and the University of Missouri at Columbia, propose a novel system to improve bird monitoring and recognition system in noisy environments. First, a *microphone dish* concept (microphone array with many concentric rings) is proposed that provides very directional and long range (a few thousand meters) acquisition of bird sounds, can simultaneously pick up and *track* sound from different directions, and the cost of the dish will be a few hundred dollars. Second, an efficient recognition algorithm is proposed which uses Hidden Markov Model (HMM) and Gaussian Mixture Models (GMM). The overall system is suitable for real-time monitoring and recognition for a large number of birds.

Here we will summarize some preliminary results of our proposed method. First, we will give a brief overview of the proposed system, which consists of several major parts: microphone dish and data acquisition system, Direction of Arrival (DOA) estimation, beamformer to eliminate interferences, and bird classifier. Second, we will describe a new wideband DOA estimation algorithm and provide a comparative study between estimation results using linear array and our circular array. Third, beamforming algorithm will be introduced through a comparative study between the linear array and our circular array. A new beamforming algorithm for dish array has been developed. It was found that the dish array has several key advantages over the linear array, including less ambiguity angles, more consistent performance, etc. Fourth, bird classification results using GMM method will be presented. Fifth, the development of a prototype microphone dish will be included. A dish array consisting of 64 microphone elements has been developed

and used to collect sound data in laboratory and in an open space. Sixth, experimental results will be described to show the performance of the software and hardware.

Main Results

1. Overall System Description

Figure 1 shows the proposed system which consists a microphone dish, a data acquisition system, and software processing algorithms such as direction finder, beamformer, and bird sound classification.

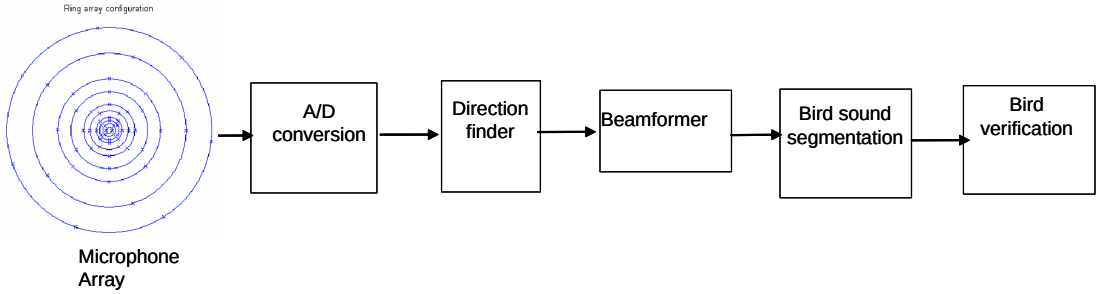


Figure 1 Proposed automated bird monitoring and recognition system.

2. DOA Estimation Algorithm

2.1 Estimation Algorithm for Circular Array

A beamformer requires the direction of arrival (DOA) of the source signal for beamforming to enhance the desired signal. The signal DOA is not known in practice and needs to be estimated. This section presents the DOA estimation of a wideband source signal, based on the MUSIC algorithm for narrowband signal.

MUltiple Signal Classification(MUSIC) algorithm is a DOA estimation algorithm for narrowband signal. For wideband signal DOA estimation, a simple technique exists by dividing the wideband signal into many narrowbands and then applying MUSIC on those narrowbands. The DOA estimation for the wideband signal is generated by combining estimated results from all the narrowband components. The process is shown in Figure 2. Other wideband DOA estimation techniques can be found in [1,2], which are computationally intensive. At present, we implemented the narrowband combining technique.

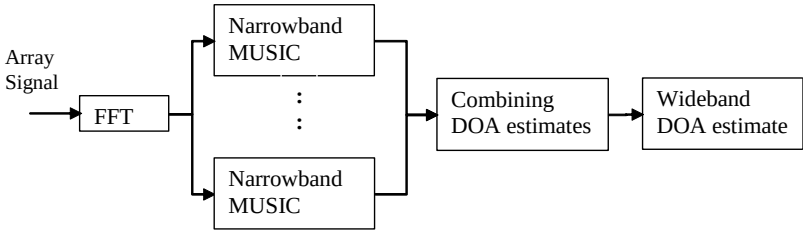


Figure 2 Block diagram of the DOA estimation of a wideband source.

As shown in Fig. 2, the DOA estimation algorithm consists of the narrowband MUSIC algorithm, which is followed by peak search technique to obtain the DOA estimate for each frequency band, and the combination of the DOAs from different frequency bands to form the final estimate.

Figure 3 shows the MUSIC spatial spectrum obtained from 4 concentric circular arrays with a total of 30 elements. The source signal used are two random amplitude narrowband signals at 500Hz, coming from $(\theta = 90^\circ, \phi = 70^\circ)$ and $(\theta = 45^\circ, \phi = 60^\circ)$ respectively, where the angle θ is with respect to the x-axis in the x-y plane and ϕ is the angle with respect to the z-axis. As shown in the figure, the MUSIC spectrum contains 2 peaks suggesting 2 DOAs.

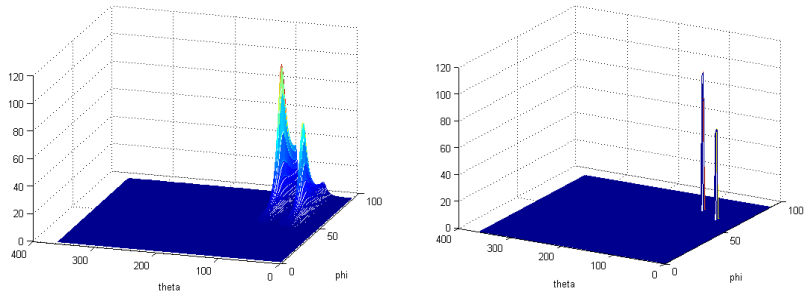


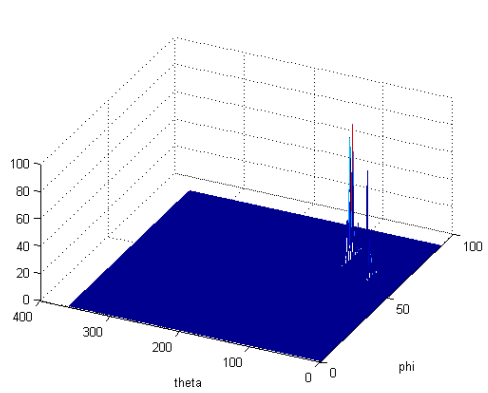
Figure 3 Narrowband MUSIC spectrum for 2 DOAs. Figure 4 Narrowband MUSIC spectrum with only local maxima.

After the MUSIC spatial spectrum is obtained, the remaining task is to identify the location of those peaks in the spectrum which correspond to the DOAs. We use the MUSIC spectrum in Figure 3 as an example to illustrate the 2-D peak searching algorithm described in [3]. Figure 4 shows such a processed MUSIC spectrum.

Finally, it should be noted that the narrowband DOA estimation results have a bias, especially in ϕ direction. We found out that when we use windowing to compute FFT of the array signal, the spectrum smearing of windowing will introduce a bias in the result. Thus longer window is preferred. And this also suggests that larger number of spectral components generally gives smaller bias in estimation result. Based on this understanding, the estimated results from narrowband MUSIC are combined in a way taking their spectrum energy into consideration.

The peak value in the MUSIC spectrum will be associated with an estimated DOA as its confidence value. A histogram is generated to combine the narrowband DOA estimates using the confidence values of the estimated narrowband DOAs as shown in Figure 5.

After obtaining the histogram of DOA estimates from different frequency components, the 2-D peak searching algorithm described in [3] is used again to yield the final wideband DOA estimate.



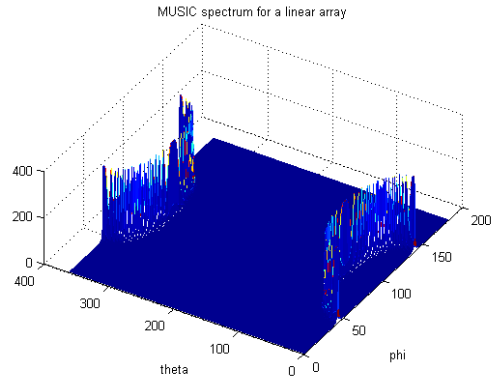


Figure 5 Combined narrowband DOA estimates. Figure 6 MUSIC spectrum for a linear array.

2.2 Comparison with DOA Estimation Using Linear Array

The DOA ambiguity set of a linear array is a cone around the linear array. Thus it cannot be used to estimate the direction of a coming signal in 3-D space. To illustrate the advantage of using a circular array instead of a linear array in DOA estimation, the MUSIC spectrum generated by an 11 element with half wavelength spacing linear array is shown in Figure 6. There is only one narrowband signal at 500 Hz coming from $(\theta = 45^{\circ}, \phi = 60^{\circ})$. The SNR is 3dB. Although there is only one signal, there are two stripes of spectrum peaks, corresponding to the ambiguity set of a cone around the linear array. It is clear that for linear array it is not possible to yield an accurate DOA estimate without ambiguity.

3. Beamforming Algorithm

The proposed beamformer design follows the technique developed by the authors [4]. We shall only summarize the array structure and some comparative results here. The details of the derivation of the array are given in [4].

The circular array has 7 rings and 102 elements. The radius of the array is about 0.5m which is very compact. Figure 7 shows the array structure. One novelty of the proposed design is that the circular array can perform wide-band beamforming, through the compound ring approach. In the compound ring structure, some rings are shared by several frequency bands and therefore resulting in savings in array elements. The proposed compound ring structure has 4 operating frequency bands as listed in the second column of Table 1. The third column in the table shows the number of rings in each band and the fourth column is the number of elements in each ring for the frequency band considered. The grouping of the rings for different bands is shown in Figure 8. The minimum separation between two array elements is

$$d = \frac{1}{4}\frac{\delta_4 \lambda_{2kHz}}{4\pi} = 0.0402m \quad (1)$$

The largest radius, and hence the size of the array, is

$$12d = 0.4827m \quad (2)$$

The details in deriving (1) and (2) is skipped here and will be provided upon request. Because of reusing array elements in different subarrays, the total number of elements is 10+14+14+18+14+18+14=102.

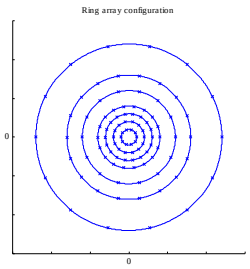


Figure 7 The proposed circular array configuration.

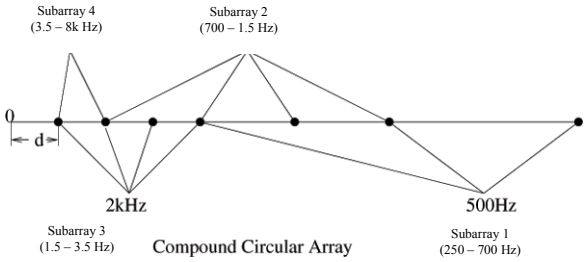


Figure 8 Grouping of the rings in the four subarrays.

Table 1: Grouping of rings into different subarrays for broad-band beamforming

	Approx. Operating Frequency Range	Number of Rings	Number of Elements in Each Ring
Subarray 1	250 Hz – 700 Hz	3	[6, 10 , 14]
Subarray 2	700 Hz – 1.5k Hz	4	[6, 10 , 14, 18]
Subarray 3	1.5k Hz – 3.5k Hz	4	[6, 10 , 14, 18]
Subarray 4	3.5 kHz – 8k Hz	2	[10, 14]

In general, the larger the number of rings in a subarray, the larger will be the attenuation in the ambient noise level. The power spectral density of birds have higher energy from 700 – 4k Hz. That is why subarrays 2 and 3 have 4 rings to provide larger attenuation to the noise.

Figure 9 shows a typical beam pattern of the proposed circular array at 1 kHz. A main advantage of the proposed design is that it provides close to a fixed level of residue side lobes.

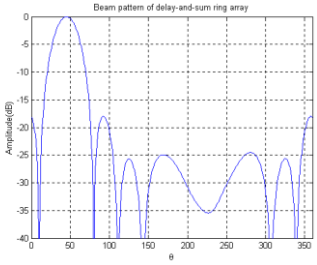


Figure 9 The beam pattern of the proposed circular array at 1 kHz.

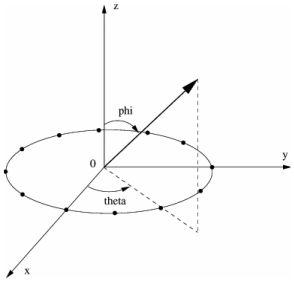


Figure 10 The coordinate system used.

The coordinate system shown below in Fig. 10, where θ corresponds to the angle with respect to the x-axis in the x-y plane, and ϕ is the angle from the z-axis with respect to the x-y plane.

For comparison purpose, a compound linear array that has the same number array elements as

the proposed circular array (102 elements) is used. The compound linear array composes of 5 subarrays operating at frequency ranges around 500Hz, 1kHz, 2kHz, 4kHz and 8kHz respectively. Each subarray contains 34 elements. Half of the elements from a subarray of higher frequency will be reused in the following lower frequency subarray. Thus total number of elements is: 34+17*4=102. Figure 11 shows a compound linear array with 5 subarrays and 4 elements within each subarray. (Subarray with as much as 34 elements are difficult to show.)

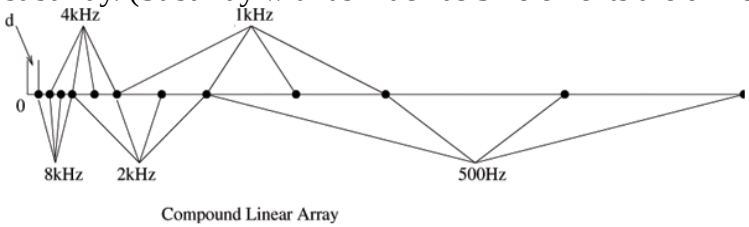


Figure 11 Configuration of compound linear array

The smallest distance between two array elements is

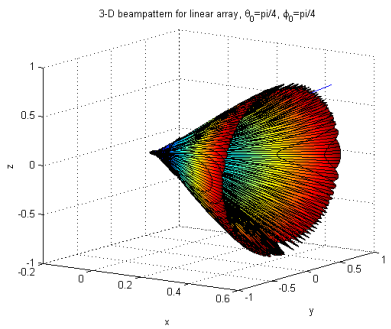
$$d = \frac{\lambda_{8kHz}}{2} = 0.0214m$$

The size of the 102 elements compound linear array is:

$$(34-1)\times\frac{\lambda_{500Hz}}{2} = 11.32m$$

which is very large.

Because of the compound array structure, the beampattern for different center frequency is same. A 3-D beampattern for one of the subarray is shown in Fig. 12, the DOA in assumed to be ($\theta = 45^\circ, \phi = 45^\circ$). A linear array has an ambiguity region that appears as a cone.



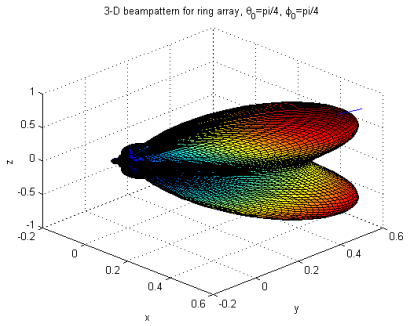


Figure 12 Three-D beampattern of compound linear array. Figure 13 Three-D beampattern for ring array

The compound ring array used is the one described earlier. It has 7 rings and contains 102 elements. The array diameter is about 1 m. The 3-D beampattern for one of the subarray is shown in Fig. 13, the DOA in assumed to be ($\theta = 45^\circ, \phi = 45^\circ$)

Besides the above beampattern comparison, a series of other comparisons between a linear array and the proposed circular array have been carried out. We have concluded that:

- Circular array has an ambiguity set of Direction of Arrival (DOA) of only 2 directions, while linear array has a larger ambiguity set of (DOA) which is cone.

- The beampattern of circular array can be rotated to arbitrary direction in the x-y plane without suffering great fluctuation. This is not the case for linear array.
- The beampattern of circular array varies with the ϕ much more than the linear array does.
- Compound linear array outperforms circular array in terms of ambient noise level most of the time but it requires very large array size.
- Compound linear array is incapable of attenuating directional interference in the DOA ambiguity set, circular array has much less ambiguity set thus it can remove the directional interference in most cases linear array fails.
- Compromise between several factors (array size, noise level, ambiguity set of DOA etc) is necessary. Overall, circular array gives better performance.

4. Bird Classification Algorithm

We have developed two different bird classification algorithms. One is based on Hidden Markov Model (HMM) and the other one is based on Gaussian Mixture Model (GMM). Based on our own evalutions [3], GMM performs better than HMM. Due to page limitations, here we only focus on GMM approach.

According to the evaluations done by National Institute of Standards and Technology (NIST) engineers, GMM has been proven to be quite useful in speaker verification applications. The birds have similar spectrum as humans. The individual component densities of a multi-modal density may model so many underlying set of acoustic classes. A linear combination of Gaussian basis functions is capable of representing a large class of sample distribution.

The bird classification consists of two major steps: 1) preprocessing the extract features; 2) applying GMM models to classify different birds.

4.1 Preprocessing to Extract Features of Birds

To identify the bird species, the algorithm we have been using is to first extract the feature vectors from the bird sound data, then match these feature vectors with Gaussian Mixture Models, each trained specifically for each bird class. The difference between the probabilities is compared to a pre-set threshold to decide if a given bird sound belongs to a specific bird class.

The feature extraction subsystem can be best described by Fig. 14.

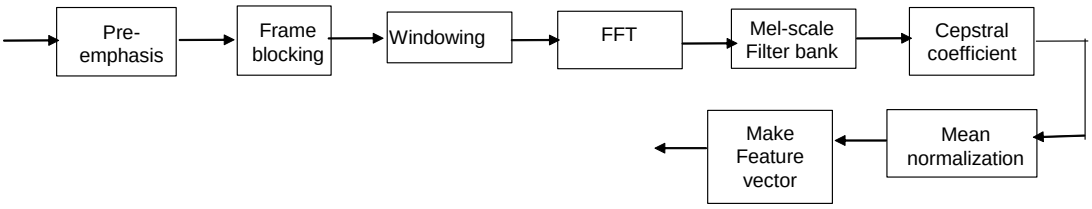


Figure 14 Preprocessing steps in the feature extraction subsystem.

The purpose of feature extraction is to convert each frame of bird sound into a sequence of feature vectors. In our system, we use cepstral coefficients derived from a Mel-frequency filter

bank to represent a short-term bird speech spectra. The digital bird sound data is first preprocessed (pre-emphasized, set to overlapped frames and windowed) and then Mel Frequency Cepstral Coefficient Analysis is applied. Typically feature extraction process compresses around 256 samples of bird sound data down to between 13 to 40 features.

4.2 Gaussian Mixture Model for Birds

The Gaussian mixture model for birds is a probabilistic model by which the distribution of data is modeled as a linear combination of several multivariate Gaussian densities. There are two motivations for using Gaussian Mixture Densities as a representation of bird identity [5]. The first is the intuitive notion that the individual components densities of a multi-modal density, like the GMM, may model some underlying set of acoustic classes. The second motivation is the empirical observation that a linear combination of Gaussian basis function is capable of representing a large class of sample distribution. The GMM is usually trained with the Expectation-Maximization (EM) algorithm to maximize the likelihood of the observation data from an individual class.

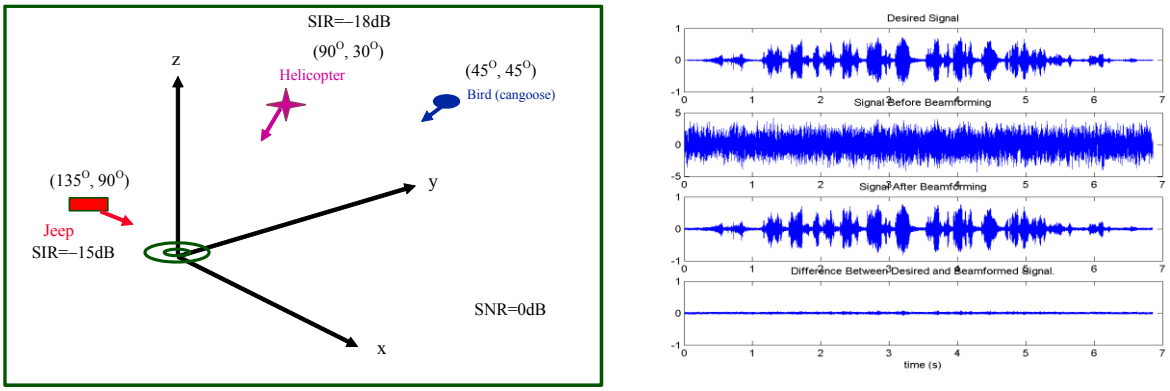


Figure 15 (a) Simulation Scenario 1; (b) Beamforming results from Scenario 1.

Our results of bird classification by using GMM performed very well [3]. Figure 15 shows one simulation scenario where the desired bird signal is CanGoose. The two interferences are helicopter noise at –18dB SIR and jeep noise at –15dB SIR. The background noise is the Hoth noise at 0 dB SNR. Hoth noise, roughly speaking, is a lowpassed Gaussian noise with spectrum similar to voice. The beamforming and classification problem is more challenging if the background noise is Hoth. This is because the noise and signal spectra overlaps extensively in the frequency domain. As a matter of fact, the problem is easier if the noise is white.

Also in Fig. 15 is the received signal before and after beamforming, and the error between the true and the beamformed signal. Before beamforming, the noise and interference dominates. The bird source signal becomes apparent after beamforming.

Table 2 shows the classification results.

Table 2: Classification Accuracy of CanGoose

	Percentage of correct classification
Before	13.11%

Beamforming	
After Beamforming	100%

Before beamforming, the classification results were not satisfactory, due to the large amount of interference and background noise. It is clear that after beamforming, the classification results improve significantly.

5. Microphone Dish Design and Hardware Prototype

5.1 Array Configuration

The compound ring array is composed of 10 rings:

Subarray 1 is composed of the 1 2 3 4th rings, with elements distributed as 4 4 6 8.
Subarray 2 is composed of the 2 4 5 6th rings, with elements distributed as (4) (8) 6 8.
Subarray 3 is composed of the 4 6 7 8th rings, with elements distributed as (8) (8) 6 8.
Subarray 4 is composed of the 6 8 9 10th rings, with elements distributed as (8) (8) 6 8.

The numbers in brackets are elements reused from a previous subarray. Total number of elements: 4+4+(6+8)*4=64.
The top view of the ring array is shown in Fig. 15.

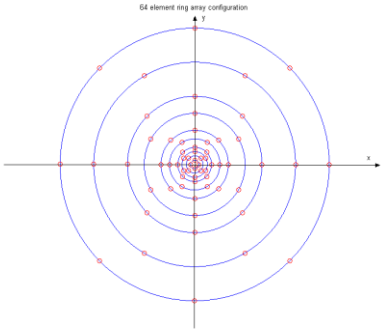


Figure 15 Circular array configuration.

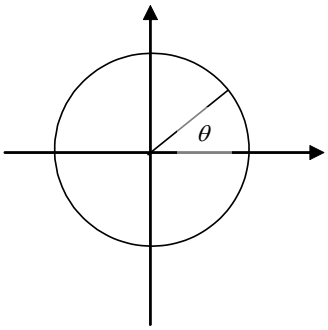


Figure 16 Definition of the angle.

Table 3 summarizes the ring radii. The largest radius is

$$R_{10} = \frac{\delta_4 \lambda_{500Hz}}{4\pi} = 0.6436178m$$

where δ_4 is the 4th root of the 0th order Bessel function of the 1st kind and $\delta_4 = 11.79$. The smallerst radius is

$$R_1 = \frac{1}{32} R_{10} = 0.02011306m .$$

Array elements on each ring are equally distrubuted. Instead of starting all the 1st element of each ring at $\theta = 0^\circ$, we use the following scheme to form a more symmetric structure for the ring array. This will reduce the fluctuation in beampattern when signal's DOA is rotated in the plane where the ring array lies.

Table 3: Summary of ring radii

Ring	1	2	3	4	5	6	7	8	9	10
Radius	R_1	$2R_1$	$3R_1$	$4R_1$	$6R_1$	$8R_1$	$12R_1$	$16R_1$	$24R_1$	$32R_1$
Number of Elements	4	4	6	8	6	8	6	8	6	8

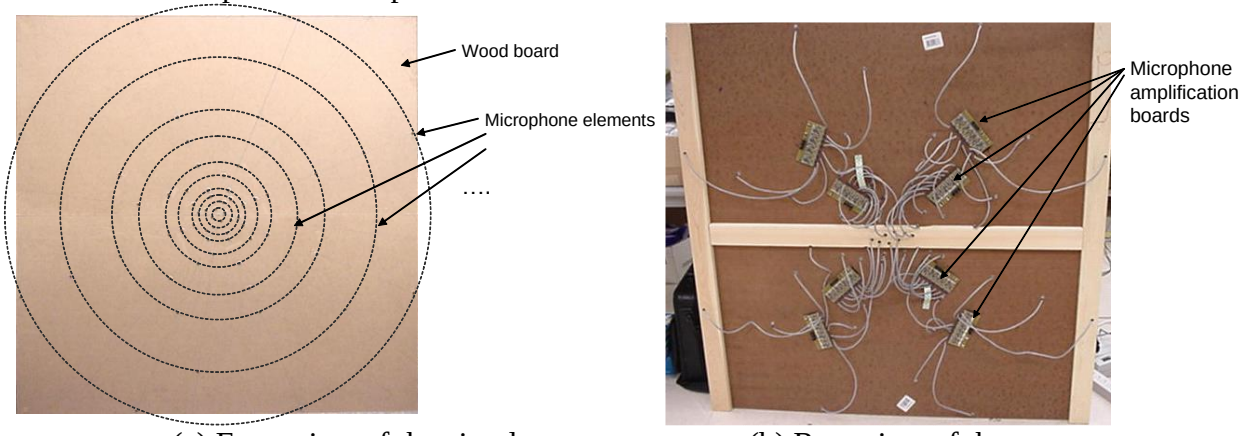
The 1st element of the 1st ring is placed at $\theta = 0^\circ$ and the remaining elements on that ring are anti-clockwise, equally placed along the circle. For the 2nd ring, its 1st element is placed at $\theta = \frac{\pi}{4}$ and the remaining elements on that ring are also anti-clockwise, equally placed along the circle. And so on for other rings. Table 4 below lists the locations (expressd by degree value of θ) of each element on the rings.

Table 4: Angular distribution of array elements

	1	2	3	4	5	6	7	8
Ring 1	0	90	180	270				
Ring 2	45	135	225	315				
Ring 3	90	150	210	270	330	30		
Ring 4	135	180	225	270	315	0	45	90
Ring 5	180	240	300	0	60	120		
Ring 6	225	270	315	0	45	90	135	180
Ring 7	270	330	30	90	150	210		
Ring 8	315	0	45	90	135	180	225	270
Ring 9	0	60	120	180	240	300		
Ring 10	45	90	135	180	225	270	315	0

5.2 Construction of the Circular Array

Based on the circular array configuration described in Section 5.1, we constructed a microphone array as shown in Fig. 17. The array board was spliced by two pieces of wood. Holes were drilled and microphones were put into them.

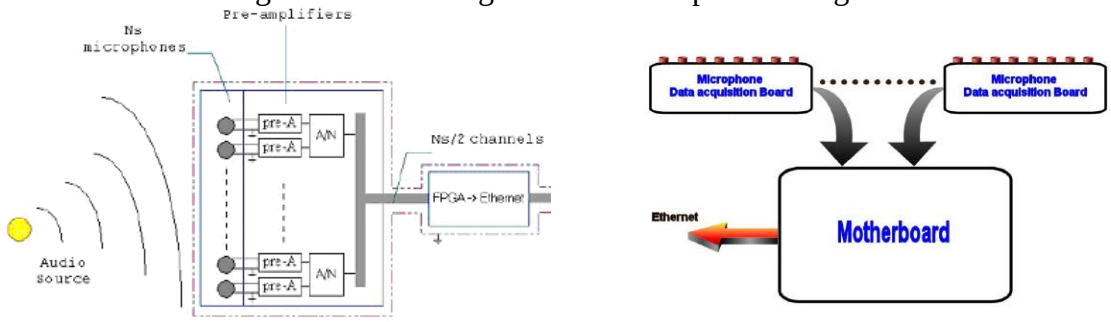


(a) Front view of the circular array; (b) Rear view of the array

Figure 17 Finished circular microphone array. It has a dimension of 1.2 meter in diameter.

5.3 Data Acquisition Electronics Associated With the Circular Array

We built data acquisition hardware for the microphone array. Figure 18 shows how we implement the data acquisition system. Basically, there are two main parts in this system. One is the microphone data acquisition board which is analog and performs A/D, the other one is the mother board which is digital and sends digital data to computer through Ethernet.



(a) Data acquisition schematics (b) Hardware relationships.

Fig. 18 Data acquisition system for the circular microphone array.

Microphone data acquisition board

Figure 19 is an overview of the microphone array analog board. There are 8 microphones on this board. We need to have at least 64 microphones in our application, so there will be 8 analog boards in this system. There are three circuit parts on this board:

- The microphone amplification stage.
- The digitalization stage.
- The motherboard connection stage.

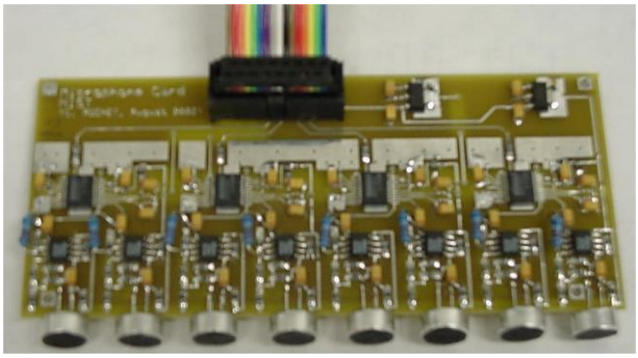


Figure 19 A finished microphone board.

Figure 20 The overview of the motherboard

Motherboard

In Fig. 20, the FPGA is the gathering part of the data acquisition system. The cables at the top of the picture are the data collection cables connected to the 8 analog boards. The red DIP is here to fix the MAC address of the microphone array. The LEDs are to give the status of the microphone array.

6. Experimental Results

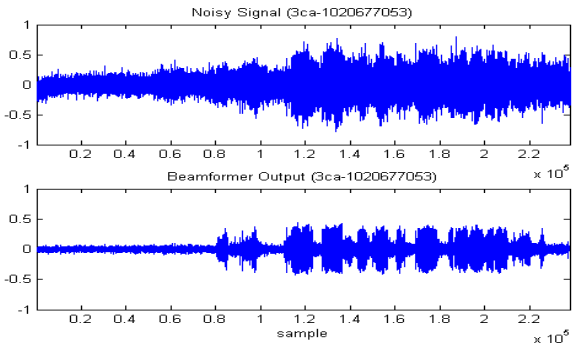
6.1 DOA Estimation Results

The microphone dish system was fully tested by taking many measurements in our laboratory as well as in an open field (parking lot). Table 4 summarizes one experiment in the open field. The experimental setup involved microphone dish and the data acquisition system to collect the data, two laptops displaying the sound sources. One source emulated the bird and the other one emulated an aircraft. The distance between the sources and the microphone dish was about 40 ft. It can be seen that there are less than 10 % of estimation error in the θ direction and 2% of error in the ϕ direction. There are three major sources of errors. One is that the distance between the sources and the microphone dish is still not long enough. Hence the microphone dish can not be treated as a point. The second reason is that distance measurements were done manually and may have some inherent errors. The third reason is that the number of elements in the dish is only 64. The angular resolution will be much better if more than 100 elements are used in the array.

Table 4: Summary of DOA estimation and the expected DOA values

	DOA Estimation		Expected DOA	
	θ	ϕ	θ	ϕ
Interference 1	171	30	180	28.52
Bird 1	329	40	360	40.06

6.2 Beamformer Outputs



Once the DOAs are estimated, beamforming algorithm eliminated the effects of interference and collected a clean bird signal. Figure 21 shows signals before and after the beamforming. It can be seen the before beamforming, the signals were very noisy. However, after beamforming, the signals were very clean. We could not hear any interference and background noise after the beamforming.

Figure 21 Signals before beamforming and after beamforming.

Conclusions and Future Research Directions

There are three major research directions:

- Reduce the beamwidth by increasing the number of microphones
  In this prototype system, we have only 64 elements. In order to reduce the beamwidth, we need to increase the number of elements to about 200. This is not easy as each microphone is digitized at a sampling rate of 22 kHz. Also the memory requirement will be huge in real-time system
- Integrate hardware and software to produce a real-time bird monitoring system

As can be seen from the previous sections, there are many software and hardware elements. The integration of these components is very challenging.

- Field tests
  Once the integration is done, we will bring the system to the airports for tests and evaluations.

Acknowledgements

This research was supported by the Air Force Office of Scientific Research under contract F49620-02-C-0044.

References

[1] H.Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources," *IEEE trans. Acoust., Speech, Signal Processing*, vol. ASSO-33, pp. 1114-1122, Oct. 1985.

[2] K M. Burckley and L. Griffiths, "Broad-band signal-subspace spatial-spectrum(BASS-ALE) estimation," *IEEE trans. Acoust., Speech, Signal Processing*, vol. 36, pp.953-964, July 1988.

[3] C. Kwan, K. Ho, et al., *Phase 1 Progress Report 3 to the Air Force*, April, 2003.

[4] Y. Li, K.C. Ho, and C. Kwan, "Design of Broad-band Circular Ring Microphone Array for Speech Acquisition in 3-D," accepted Int. Conf. on Acoustics, Speech, and Signal Processing, 2003.

[5] D. A. Reynolds and R. C. Rose, "Robust Text-Independent Speaker Verification Using Gaussian Mixture Speaker Models," *IEEE Trans. Speech and Audio Processing*, vol. 3, no. 1, 1995.