Design and Development of an Acoustic Sensor Array for Anomaly Detection

Dibyendu Roy[1], V.Ramu Reddy[1], Parijat Deshpande[1], Ranjan Dasgupta[1]
[1]TCS Research & Innovation
Email: {roy.dibyendu, ramu.vempada, parijat.deshpande, ranjan.dasgupta}@tcs.com

Abstract— **Anomaly detection is the identification of an anomaly with regards to misplaced objects or intruders in an indoor environment. It can be done by mapping a room along with its stationary objects with an ultrasonic acoustic frequency response. In this paper, we explore the feasibility of using active acoustic imagery in air and simulate phased array beam forming techniques to achieve a suitable acoustic sensor array design for a portable mobile robot which can be applied to detect the presence/absence of anomalous objects in a room. The selected room is insonified with a broadband acoustic signal and the multi-path reflection effects contained in the reflected signal time-series in enclosed rooms are analyzed to detect anomalies. The related hardware is designed with the same feasibility criterion that the developed system needs to be deployed on a portable mobile robot. There is a trade-off between image resolution and range with the array size, number of elements and the imaging frequency and has to be iteratively simulated to achieve the desired acoustic sensor array design. The designed acoustic imaging array system is targeted for use in surveillance missions for intruder alerts (anomaly detection) and imaging objects during dark and smoky scenarios where conventional optic based systems do not function very well.**

Keywords-Acoustic Imagery; Acoustic Sensor Array; Phased Array Beamforming; Anomaly Detection; Object Localization

I. INTRODUCTION

Imaging of underwater objects is primarily conducted by acoustic imagery due to the severe attenuation of electro-magnetic waves in water. Acoustic imagery underwater has varied range of significant applications such as side-scan sonar, mine hunting sonar [1], [2]. Cetaceans have perfected underwater imaging while bats employ ultrasonic imaging in air routinely while flying at night. Several man-made SONAR techniques involve collecting reflected wave data from targets and synthesize object images. However, for acoustic imaging in air, the frequency and array size i.e. aperture size need to be carefully selected to achieve desired results for a 2D planar or beam forming techniques are employed to steer beams and synthesize it. Additionally, the number of elements, inter-element spacing and their sampling frequency respecting Nyquist criterion have to be selected to realize a portable system capable of mounting on a mobile robot.

Phased array beam-forming techniques involve electronically steering a narrow-band waveform along azimuth and elevation directions. Using the two-dimensional, fully populated acoustic array, all array elements insonify simultaneously and the waveforms are transmitted using proper phase-shifts at each array element so that the combined beam is steered in both azimuth and elevation directions, covering the complete target surface. Phased array beam-forming techniques involve electronically steering a narrow-band waveform along azimuth and elevation directions. Using the two-dimensional, fully populated acoustic array, all array elements insonify simultaneously and the waveforms are transmitted using proper phase-shifts at each array element so that the combined beam is steered in both azimuth and elevation directions, covering the complete target surface.

II. STATE OF THE ART

In [3], the authors describe their approach in achieving super directivity of elements even a low acoustic frequencies. The conventional planar array beam forming can achieve limited resolution due to the beam width of the selected array as decided by the frequency and array size. In [4], Harput *et al.* have developed a real-time acoustic imaging system by adjusting the inter-element spacing for suppressing the grating lobes. Six channels for transmission and 4 channels for receiving on field programmable gate array were used for digital beam forming with dynamic focusing. To extend the Field of View (FOV) and eliminate the grating lobe caused due to more than one wavelength inter-element spacing, therefore the receiver array is formed with 3 wavelength inter-element spacing. It is observed grating lobes at transmitter & receiver array at different places due to non-identical element spacing.

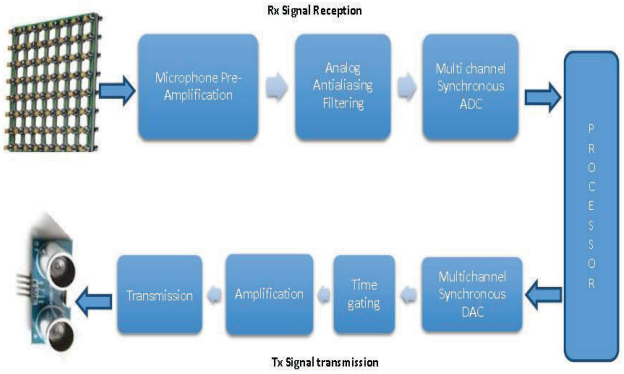
Turqueti *et al.* in [5] have presented a sonic system to locate and imaging objects in air with high quality and good resolution by using phased array techniques with 52 MEMS sensors as an acoustic array and operated at 20-100 kHz frequency range. One transducer is used to generate the scanning beam and the sensor array is employed to capture the echoes. They have used beam forming, matched filtering, spectral processing, and independent component analysis for imaging, tracking and identification of objects. On the other hand, Steckel *et al.* [6] have constructed low- cost and powerful, in-air sonar system to suit to various robotic applications. It is done by a sparse random array of receivers with the combination of broadband beam forming. Particularly sensor can produce 3-D location measurements even in the presence of multiple highly overlapping echoes.

High-resolution acoustic imaging system in air is proposed by Nihtila *et al.* in [7] by using principle of synthetic aperture with audio hardware and linear frequency modulated pulse waveform. A unique kernel function for each pixel is done for inverse filtering method. The system is demonstrated via an inverse synthetic aperture formation where steeper motor is

978-1-5090-0795-0/15/$31.00 ©2016

used to rotate the object which provides the required change in aspect angle. Measurements from conventional speaker and microphones are taken in an ordinary room where near-field distance and strong static echoes present. In our earlier works, we have performed audition based localization [8] multi-mode sense of perception using 3D information [9] using Kinect mounted on the robot. Person identification [10] [11] [12] and their activity recognitions [10] [13] were carried out using skeleton information. Anomalies like human speech, crowd noise and infant cries are also carried out using audio signals [14] [15] [16]. However, all the above works uses audio and video signals for effective localization or anomaly detection.

In this paper firstly we have explored the basic principle of phased array beam forming techniques to find the suitable array size as well as image resolution. Then the corresponding array is designed physically for mobile robot and it is used to collect some real data for identifying the presence/absence of anomalous object of a room which is term as "Anomaly Detection". Fig. 1 represents the block diagram of Acoustic Sensor array (ASA). A pulse signal is transmitted through ultrasound transmitter; the reflected signal is received by



receivers. Some engineering should be done on the received signals for detecting the anomalous object.

Fig. 1: The schematic diagram of ASA

III. ACOUSTIC ARRAY SIMULATION

Beam forming or spatial filtering is a signal processing technique, used in sensor array for directional signal transmission or reception. This is achieved by combining elements in a phased array in such a way that signals at particular angles experience constructive interference while others experience destructive interference.

A. Phased Array Beam Forming

The method of electronic beam-steering and beam-shaping is used in simulating and designing the ASA. The ASA is assumed to be in the X-Z plane while the 3-D object to be imaged is along the positive Y-direction. X-axis is the horizontal axis, while Z-axis is the vertical axis. The array is centered at the origin with N in number elements along the Z-axis and M in number elements along the X-axis; dz is the inter-element spacing along the Z-axis, whereas dx is the inter-element spacing along the X-axis. λ is the wavelength of transmission. The spatial counterpart of the temporal Nyquist criterion mandates that $dz \le \lambda/2$ and $dx \le \lambda/2$ which has been respected for suppressing the effects of side-lobes. Consider a vector $(\vec{v})$ from the origin to the point (x; y; z) in space. θ is the angle of $\vec{v}$ w.r.t. to the positive Z-direction, whereas ϕ is the angle of $\vec{v}$ w.r.t. to the positive X-direction. The magnitude of $\vec{v}$ is represented by r. Thus

$$x = r.\sin\theta.\cos\phi \quad (1)$$

$$y = r.\sin\theta.\cos\phi \quad (2)$$

$$z = r.\cos\theta \quad (3)$$

$$r = \sqrt{x^2 + y^2 + z^2} \quad (4)$$

Assume that the beam is to be steered to a particular direction given by $(\theta_T;\phi_T)$. Energy shall be received in all directions (θ,ϕ). Equation of the beam pattern is given by:

$$B(\theta:\theta_T;\phi:\phi_T) = \frac{1}{N}\frac{\sin[\frac{\pi N dz}{\lambda}(\cos\theta-\cos\theta_T)]}{\sin[\frac{\pi dz}{\lambda}(\cos\theta-\cos\theta_T)]}\times \frac{1}{M}\frac{\sin[\frac{\pi N dz}{\lambda}(\cos\phi-\cos\phi_T)]}{\sin[\frac{\pi dz}{\lambda}(\cos\phi-\cos\phi_T)]} \quad (5)$$

In this simulation, it has been considered that the individual elements are non-isotropic and having a beam pattern $B_e(\theta;\phi)$. Thus, from the Principle of Pattern Multiplication [2], the final beam pattern along a direction (θ,ϕ) for array steered to $(\theta_T;\phi_T)$ is given by

$$B_{Array}(\theta:\theta_T;\phi:\phi_T) = B(\theta:\theta_T;\phi:\phi_T).B_e(\theta;\phi) \quad (6)$$

Various combinations of parameters like number of elements and insonification frequencies are simulated for designing the ASA. An example of simulation result for B_{Array} with different elements operated at 30 kHz to 40 kHz is shown in Fig. 2. From Fig. 2, it is evident that higher number of elements cause decrease in the beam width and correspondingly smaller array increases the 3dB beam width of the main lobe.

B. Phased Array Simulation

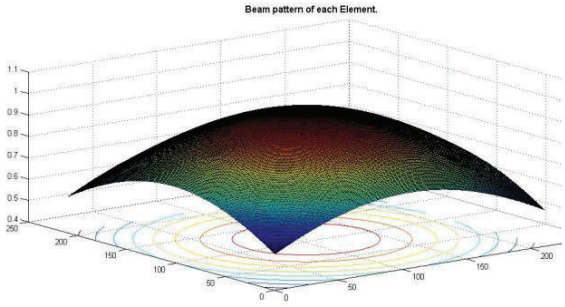
The simulation was carried out in two parts. Initially, a set of parameters was taken and simulation performed to establish the concepts. In the second phase, simulations were performed using various combinations of parameters to ascertain the most suitable combination. The details of the parameters are shown in the Table 1.The surface considered for the simulation is shown in Fig. 3. It has been assumed that the target, as in practical cases, has sufficient amount of in homogeneity on the surface as well as within the volume. Multiple reflecting points result in a composite signal being received by the array, which is the superposition in time of many signals. On receipt of the

978-1-5090-0795-0/15/$31.00 ©2016

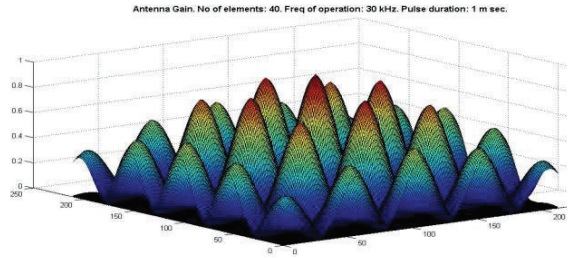
composite time signal at each scan position, correlation is performed with the transmitted pulse.

Table 1: Simulation Parameters

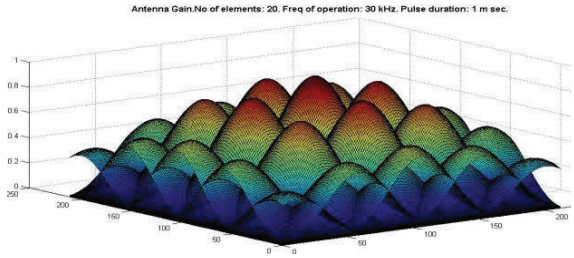
	First Phase of Simulation	*Second Phase of Simulation*
Time Duration of the pulse (msec.)	1	0.5, 1.25, 2, 2.75
Frequency	30 kHz	20, 30, 40 kHz
N, M	40, 40	20, 30, 40
dx, dz	$0.45 \times \lambda$	$0.45 \times \lambda$
Max. Horizontal steer angle	$\pm 10^{\circ}$	$\pm 10^{\circ}$
Max. Vertical steer angle	$\pm 7^{\circ}$	$\pm 7^{\circ}$
Sampling frequency of simulation	10 times of frequency	10 times of frequency



(a)



(b)



(c)

Fig 2: Representative plot of a) Beam pattern B_e (θ ; Φ) of individual non-isotropic microphone, Array beam pattern with b) 40 elements operated at 30 kHz and c) 20 elements operated at 30 kHz

Point of maximum correlation gives the distance of the point from which maximum energy has been reflected for the steering angle at that time. It is to be noted that threshold is performed after the calculation of correlation. Only, if the correlation is above a certain value (ascertained to be 0.7 after some trials), further processing is performed for that steering angle. For any scanning angle, if the value of correlation is very low, it indicates that many points are contributing to the composite signal. The estimated surface for 20 elements operated at 30 kHz is shown in Fig. 4.

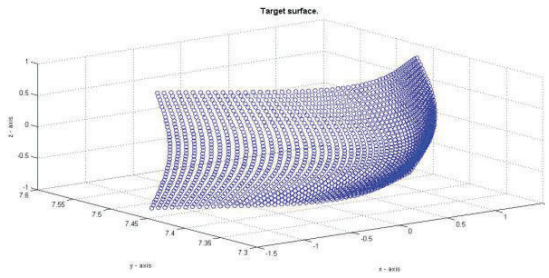


Fig 3: The target surface

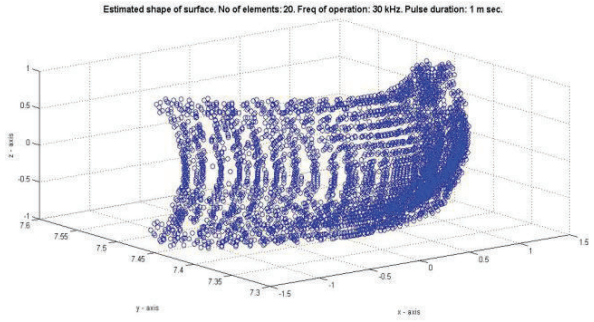


Fig 4: Estimated surface of simulation using 20 elements operated at 30 kHz

In the second phase, simulation is carried out with different combinations of frequency to ascertain the combination of parameters, which yields a viable solution for estimating the three dimensional surface. During coding of the simulation, multiple loops are coded, the "frequency of pulsed-CW signal" loop being the outermost loop and the "number of elements along one dimension of square array" loop being the innermost. The total number of combinations are $3 \times 4 \times 3 = 36$ with above mentioned parameters. For each run, using a particular combination of the above parameters, the estimated Y-coordinates of points on the surface was compared with the actual Y-coordinates of points on the surface. Absolute value of the error is then ascertained and its mean value (over number of points on the surface) is calculated. This "mean of absolute error" is considered to be the measure of performance. The same has been shown in Fig. 5 for 36 in number combinations of simulation parameters.

From the above simulation results, it is clearly seen that associated error is minimized when the frequency of transmitted pulse is 40 kHz. So to validate the above scenario, we have chosen 4*4 array design having frequency range of 40 kHz.

978-1-5090-0795-0/15/$31.00 ©2016

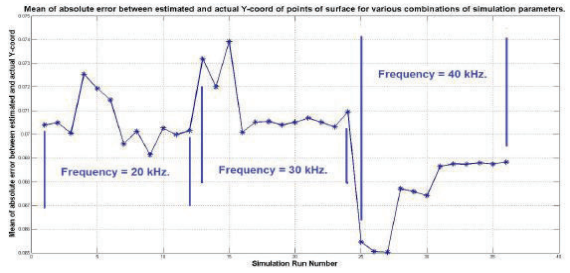


Fig 5: "Mean of absolute error" (across all points on surface) between estimated Y-coordinate and actual Y-coordinate of the surface for 36 in number combinations of simulation parameter

IV. Design of ASA

The acoustic sensor array consists of one ultrasonic module (HC-SR04) as a transmitter (receiver section disabled) and 16 (4×4) ultrasonic receivers of 40 kHz frequency mounted on a mobile robot. The spacing between the transmitter and receivers needs to be properly designed. In our case, as the operating frequency is 40 kHz and speed of sound in air is approximately 340 m/sec, the distance between two consecutive receivers is 1 cm. The electronics behind each receiver consist of a 30 kHz high pass filter and two-stage operational amplifier (Op-amp) as shown in Fig. 6. The overall gain of the amplifier is 200 (approx.).

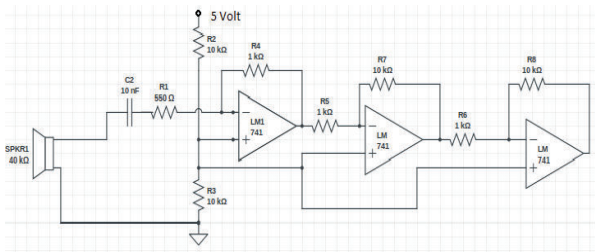


Fig 6: ASA single receiver

Fig. 7 represents the whole mechanism of the sensing array. For analyzing the receiving signals, a data acquisition system is required which has 16 analog output channels and can sample the signal at a sampling frequency of 400 kHz (10 times of the transmitted signal). Here we have used NI USB 6216 data acquisition toolbox [10]. One more advantage of this device is that it can send the data to MATLAB software directly.

V. Application of ASA

The designed ASA is deployed into a wooden box which mimics a scaled down version of a room. The experiment consists of three scenarios: empty room, with one anomaly and two anomalies (shown in Fig. 8). All data are collected through NI data acquisition toolbox to MATLAB for the diagnosis purpose.

A. Anomaly Detection

Anomaly detection denotes to the task of recognizing new or unknown patterns. Here our main focus is to detect the presence/absence of an anomalous object in a room in dark or smoky environment where optical based system fails. The disclose method enables detecting variations in a room by removing/ including objects within the room for surveillance and detection purposes/ providing intruder alerts. Fig. 9 and 11 represent the corresponding outputs of the Receiver #7 and Receiver #3 for the above three situations respectively.

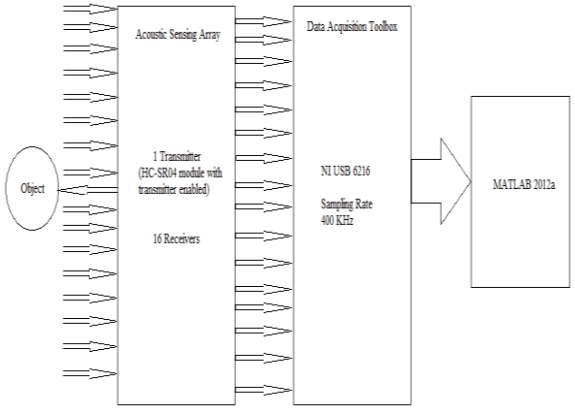


Fig 7: Data acquisition system of ASA

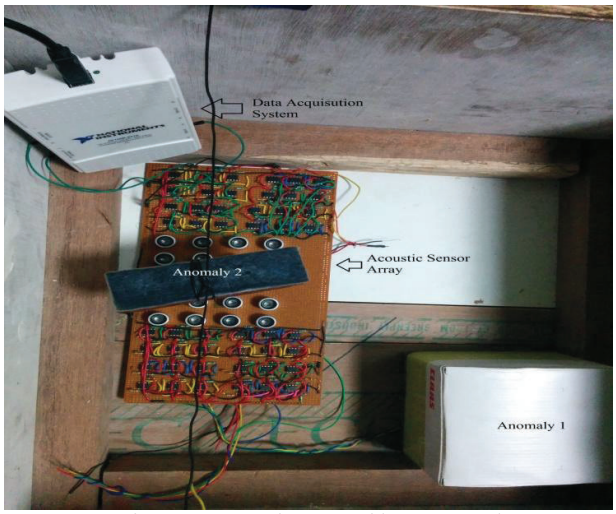
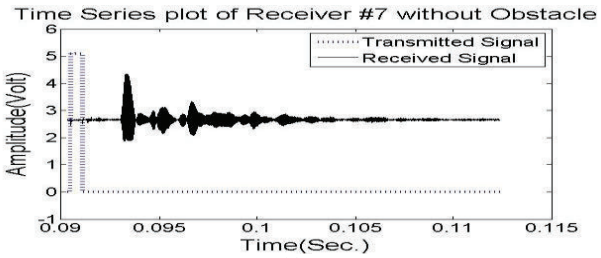


Figure 8: ASA Environment with Two Anomaly

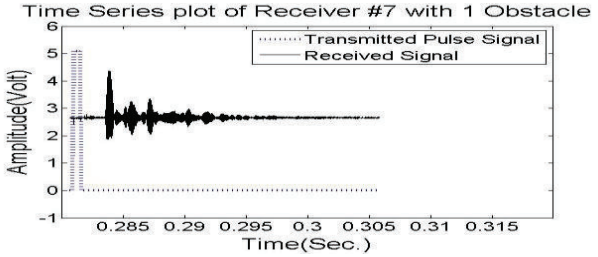


(a)

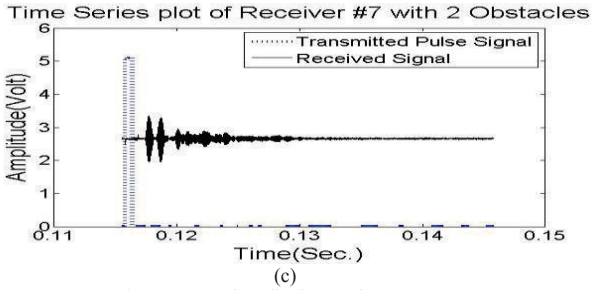
A pulse train of 1sec. duration is sending through ASA transmitter. The corresponding receiver's responses are shown in the Fig. 9 and 11. As the anomaly is far from the ASA (Fig. 8), the response variation of the receivers are less from normal condition (Fig. 9/ (b), 11/ (b)) except the Receivers #13, #12,

978-1-5090-0795-0/15/$31.00 ©2016

#5 and #4. But still anomaly can be detected easily by comparing the different Receivers responses.



(b)



(c)

Figure 9: Receiver #7 time series response

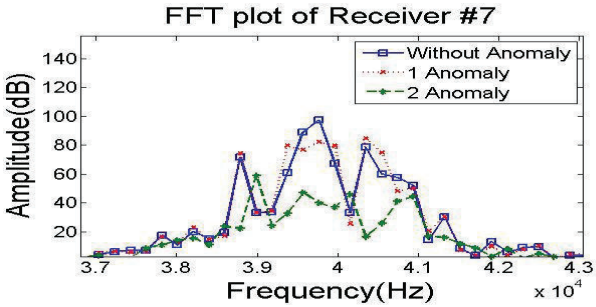
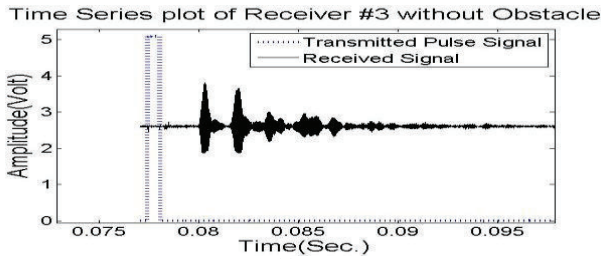


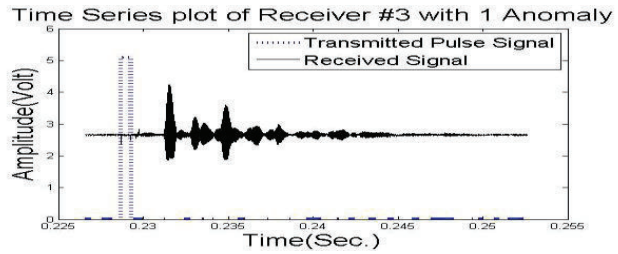
Figure 10: Frequency response of Receiver #7

For validating our discussion, we have introduced one more anomaly which is nearer to ASA (Fig. 8) along with the previous one. Fig. 9 /(c) and 11 /(c) represent the corresponding time series response of Receivers #7, #3. If all three figures of a single receiver are compared, the presence of anomalous object is clearly seen. The frequency responses of the corresponding receivers are shown in the Fig. 10 and 12. The peak variation of maximum frequency component is in the range of 75 – 120 dB when no anomaly is present. In case of single anomaly condition, the variation is nearly same (in the range of 75 – 120 dB) because the distance of the object is far from the array but the pattern provides the presence of anomaly.

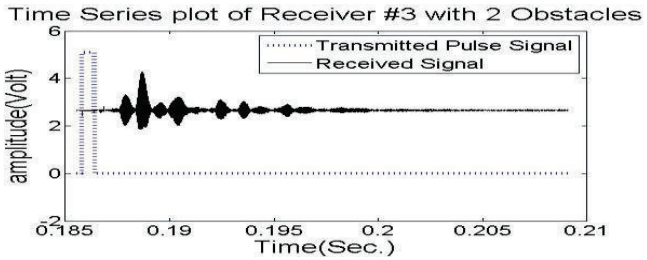
When an anomalous object is nearer to the ASA (Fig. 8), then the peak variation of maximum frequency component is changed abruptly, it is in the range of 10 – 150 dB. The peak frequency component is also changed accordingly. For no anomaly condition, peak frequency is observed at 39.8 kHz (Fig. 10) and 39.8 kHz (Fig. 12), for one anomaly condition it is at 40 kHz (Fig. 10) and 39.8 kHz (Fig. 12), whereas for two anomaly case the peak frequency is observed at 39 kHz (Fig. 10) and 39.25 kHz (Fig. 12). So it can be concluded that in presence of anomalous object nearer to the array the peak frequency is shifted.



(a)



(b)



(c)

Fig 11: Receiver #3 time series response

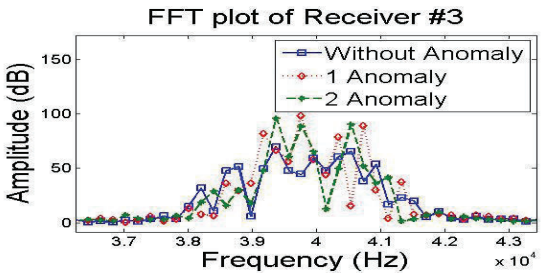


Fig 12: Frequency response of Receiver #3

B. *Object Localization*

Accurate object tracking and localization is currently a challenging problem. Tracking can be done appropriately if the location of an object at every instance of time is accurately projected. Using the echo timing of each receiver one can easily estimate the location of the anomalous object.

When the first obstacle/anomaly is placed into the box, the echo timing of all receivers is nearly same (expect Receivers

#13, #12, #5 and #4) because of distance of the object from the array. The echo timing of all receivers is 20 msec. for no anomalous condition. With one anomaly, the echo timing of Receiver #13 is 15 msec. Receiver #12 is 15.5 msec. Receiver #5 is 17 msec. Receiver #4 is 17.8 msec. The echo timing of other receivers is nearly 20 msec. So from the observation we can estimate the location of the corresponding anomaly.

Similarly for two anomaly condition, the echo timing is changed for all receivers sharply because the distance of second anomaly is nearer to the array. The echo timings of Receivers #4, #5, #8, #9 are mostly effected which determines the location of the anomaly. So using ASA we can detect the presence/absence of an anomalous object in a room as well as the location of the object. Fig. 13 represents the real-time phased shift beam-forming without obstacle condition. The beam-pattern of the corresponding signals is shown in the Fig. 14.

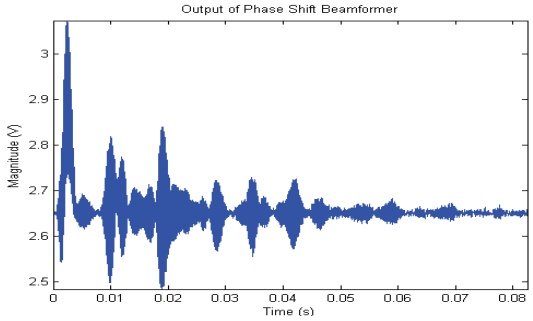


Fig 13: Phased Shift Beam-formation of ASA

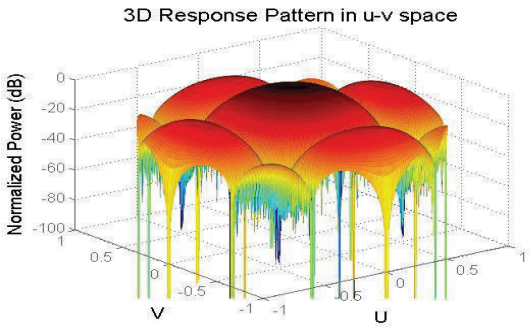


Fig 14: Beam-pattern of ASA

VI. CONCLUSION

In this paper we have simulated various configurations of planar (2D) acoustic sensor arrays and have an estimate of the beam width as well as the resolution of objects scanned / imaged. The higher number of elements results in larger array / aperture size and hence improved resolution. We have illustrated two representative simulations in this paper varying the number of elements, frequency for a fully populated planar array i.e. inter element spacing of not more than λ /2. After that we have designed the 4×4 array based on the phased array beam forming simulation. Then we have placed the array into an enclosed box (scaled down version of a room) for detecting the presence/absence of anomaly which is termed as "anomaly detection". For imaging purpose, 4 x 4 element array was simulated and found to have a broad array beam width and therefore deemed unsuitable for the current application. However, by implementing signal processing techniques for improving the directionality of the array as well as applying synthetic aperture techniques can be explored in the future to make this 4 x 4 array suitable.

REFERENCES

[1] M. Klein. "Side scan sonar. In International handbook of underwater archaeology", pages 667-678. Springer, 2002.

[2] J. L. Sutton. "Underwater acoustic imaging". Proceedings of the IEEE, 67(4):554-566, 1979.

[3] M. Crocco and A. Trucco. "Design of superdirective planar arrays with sparse aperiodic layouts for processing broadband signals via 3-d beamforming" IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP), 22(4):800-815, 2014.

[4] S. Harput and A. Bozkurt. "Ultrasonic phased array device for acoustic imaging in air". Sensors Journal, IEEE, 8(11):1755{1762, 2008.

[5] M. Turqueti, V. Kunin, B. Cardoso, J. Saniie, and E. Oruklu. "Acoustic sensor array for sonic imaging in air". In Ultrasonic Symposium (IUS), 2010 IEEE, pages 1833-1836. IEEE, 2010.

[6] J. Steckel, A. Boen, and H. Peremans. "Broadband 3-d sonar system using a sparse array for indoor navigation". Robotics, IEEE Transactions on, 29(1):161-171, 2013.

[7] T. Nihtila, J. Jylha, and A. Visa. "High-resolution acoustic imaging in air by synthetic aperture using pixel-wise matched kernels". In SENSORS, 2014 IEEE, pages 1714-1717, Nov 2014.

[8] V. R. Reddy, P. Deshpande, and R. Dasgupta, "Robotics audition using kinect," in Automation, robotics and applications (ICARA), 2015 6th international conference on. IEEE, 2015, pp. 34–41.

[9] P. Deshpande, V. R. Reddy, A. Saha, K. Vaiapury, K. Dewangan, and R. Dasgupta, "A next generation mobile robot with multi-mode sense of 3d perception," in Advanced robotics (ICAR), 2015 Int. Conf on. IEEE, 2015, pages. 382–387.

[10] V. Ramu Reddy, K. Chakravarty, T. Chattopadhyay, A. Sinha, and A. Pal,"Recognition of who is doing what from the stick model obtained from kinect," in Multimedia and Expo Workshops (ICMEW), 2014 IEEE Int. Conf. on, IEEE, 2014, pp. 1–2.

[11] V. R. Reddy, T. Chattopadhyay, K. Chakravarty, and A. Sinha, "Person identification from arbitrary position and posture using kinect," in Proceedings of the 12th ACM Conference on Embedded Network SensorSystems. ACM, 2014, pp. 350–351.

[12] V. R. Reddy, K. Chakravarty, and S. Aniruddha, "Person identification in natural static postures using kinect," in European Conference on Computer Vision. Springer International Publishing, 2014, pp. 793–808.

[13] V. R. Reddy and T. Chattopadhyay, "Human activity recognition from kinect captured data using stick model," in Human-Computer Interaction. Advanced Interaction Modalities and Techniques. Springer International Publishing, 2014, pp. 305–315.

[14] S. G. Koolagudi, K. S. Rao, R. Reddy, V. A. Kumar, and S. Chakrabarti, "Robust speaker recognition in noisy environments: Using dynamics of speaker-specific prosody," Forensic Speaker Recognition: Law Enforcement and Counter-Terrorism, p. 183, 2011.

[15] V. R. Reddy, A. Sinha, and G. Seshadri, "Fusion of spectral and time domain features for crowd noise classification system," in Intelligent Systems Design and Applications (ISDA), 2013 13th International Conference on. IEEE, 2013, pp. 1–6.

[16] R. R. Vempada, B. S. A. Kumar, and K. S. Rao, "Characterization of infant cries using spectral and prosodic features," in Communications (NCC), 2012 National Conference on. IEEE, 2012, pp. 1–5.

978-1-5090-0795-0/15/$31.00 ©2016