2019 IEEE International Conference on Signal Processing, Information, Communication & Systems(SPICSCON)
28-30 November,2019, Dhaka, Bangladesh

Direction of Arrival Estimation through Noise Supression: A Novel Approach using GSC Beamforming and Room Acoustic Simulation

Alif Bin Abdul Abdul Qayyum, Adrita Anika, Md. Messal Monem Miah,
Md. Mushfiqur Rahman, K. M. Naimul Hasan, Md. Tariqul Islam, Sheikh Asif Imran Shouborno,
Md. Farhan Shadiq. Mohammad Ariful Haque*

Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka-1205
Email: alifbinabdulqayyum.abaq@gmail.com, adritaanika23@gmail.com, *arifulhoque@eee.buet.ac.bd

***Abstract*— A novel method for localizing or estimating the direction of a sound source from the speech mixed with different levels of noise recorded by a microphone array embedded in an Unmanned Aerial Vehicle (UAV) has been proposed in this paper. Publicly available DREGON dataset (The IEEE Signal Processing CUP 2019 dataset for static task) has been used. The detail methodology of the system for localizing the sound source in static condition of the UAV is described in this paper. Generalized Sidelobe Canceller (GSC) Beamformer on the noisy audio is used to extract the noise along the rotor directions. This extracted noise is the simulated to synthesize 8 channel audio using pyroomacoustics. Finally the extracted noise is used as the reference of the wiener filter for filtering the noise in the provided noisy audios. GCC PHAT and GCC NON LIN methods are used to estimate the elevation and azimuth of the sound source. Promising results have been found using this method to localize the sound source of human speech from the audios of snr as low as -20 dB recoreded by a microphone array embedded on a UAV. If at most 10^o of error in angle is allowed, our proposed method provides an accuracy of almost 91.67%.**

Index Terms*— static, generalized sidelobe canceller beamformer, wiener filter, pyroomacoustics, azimuth, elevation.

I. Introduction

Sound Source Localization (SSL) is a very challenging problem with increasing use in modern days. In the field of robotics, automation and rescue missions SSL plays a very key part. With the advancement of robotics, Unmanned Aerial Vehicles (UAV) commonly referred as drones are becoming very popular in different applications especially when they are equipped with different forms of sensors. In cases of natural hazards such as mass fire, earthquakes or other forms of hazards where a rescue mission is required to save people's lives but there's a lack of visual feedback, UAVs equipped with microphone may come handy to locate affected people in these hazardous situations from their voice which is in fact a SSL problem. With the target speech signal, many forms of noises (rotor noises, noises due to air friction etc.) are being added as it is received by the microphone array embedded with the UAV which makes it even more challenging. The goal of this paper is to propose a state of the art method to successfully locate sound sources in the case where the UAV is in static position.

A. Description of the dataset

Publicly available DREGON dataset (The IEEE Signal Processing CUP 2019 dataset for static task) has been used.

The database contained 300 8-channel audio recordings at 44.1 kHz and of roughly 2 seconds each in the form of wav files, named 1.wav to 300.wav. Each of these were obtained by summing a clean recording of a static loudspeaker emitting speech from an unknown (azimuth, elevation) direction in the array's frame, and a recording of UAV noise of the same length in various flight conditions and using various signal-to-noise ratios. The goal is to retrieve the azimuth and elevation angles of the static speech source for each of the provided 300 recordings.

B. Related Works

Audio acquisition via any automated or semi-automated vehicle or robot faces some common challenges, including ego noise. For example, [5] deals with self-noise considerations in case of audio enhancement for acquisition of audio via a humanoid robot, for microphones close to speakers or affected by limb movements at peripheral locations. In [9], a non-parametric Bayesian model is proposed based on non-negative matrix factorization at single channel, which is independent of motion information. On the other hand, [10] uses neural networks to predict the internal noise from angular velocities at joints of Aibo robot.

Such challenges also arise in the case of source localization from received sound. A survey of such works involving sound source localization in the field of robotics has been presented at [12]. In [1], it is presented that the problem of ego noise cancellation and sound source localization in case of unmanned aerial vehicles is novel and a sector of interest in recent research works targeting applications such as localizing emergency or distress calls.

Many works targeting the aforementioned problem of source localization tackles the problem based on algorithms modifying Multiple Signal Classification (MUSIC) and Generalized Cross Correlation (GCC). Examples of modified MUSIC algorithm include incremental variations of generalized eigenvalue decomposition or generalized singular value decomposition, namely iGEVD-MUSIC as shown in the case of humanoid robots in [15] and iGSVD-MUSIC demonstrated in case of quadrotor in [3]. On the other hand, open framework ManyEars [16] uses GCC for real-time microphone array processing to track sound source. In [2], cross correlation information of four microphones mounted on a micro air vehicle has been used to localize the sound source. Use of the speed or orientation data of self-monitoring sensors embedded within the UAV has also been proposed in [7] using GEVD-MUSIC to estimate the noise correlation matrix. In [8], neural networks have been used for sound separation and identification experiments involving UAV. In [6], several UAV ego-noise reduction algorithms have been compared, including beamforming, blind source separation or time-frequency processing, but they require the information of target sound location beforehand, and hence does not involve the problem of sound source localization.

II. METHODOLOGY

The task was divided into two parts. SNR can be as low as -15 dB or less, noise was reduced from the provided audio files. After the noise is removed considerably relatively clean audios were used to estimate the source of the speech.

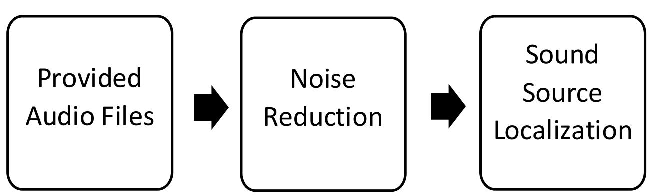

Fig 1. Flowchart of proposed methodology

For noise reduction various methods and approaches were applied. The best result were achieved by applying Generalized Sidelobe Canceller (GSC) Beamformer in the 4 rotor directions. This has been implemented using Phased Array System Toolbox of MATLAB.

This GSC Beamformer is applied on the noisy audios to extract motor noise which is later processed to use as reference signal for wiener filtering of the noisy audios. When GSC is done on a multi-channel audio, the output of the system does not remain multi-channel anymore i.e. it loses the phase information which is required for source localization. The novelty of this paper lies in applying general sidelobe cancellation on a multi-channel audio and to regain the phase information using pyroomacoustics considering the microphone array structure

A. GSC Beamformer

A GSC beamformer splits the incoming signals into two channels. One channel goes through a conventional beamformer path and the second goes into a sidelobe canceling path. The algorithm first pre-steers the array to the beamforming direction and then adaptively chooses filter weights to minimize power at the output of the sidelobe canceling path. The algorithm uses least mean squares (LMS) to compute the adaptive weights. The final beamformed signal is the difference between the outputs of the two paths.

The generalized sidelobe canceler (GSC) is an efficient implementation of a linear constraint minimum variance (LCMV) beamformer. LCMV beamforming minimizes the output power of an array while preserving the power in one or more specified directions. This type of beamformer is called a constrained beamformer. Exact weights for the constrained beamformer can be computed but the computation is costly when the number of elements is large. The computation requires the inversion of a large spatial covariance matrix. The GSC formulation converts the adaptive constrained optimization LCMV problem into an adaptive unconstrained problem, which simplifies the implementation.

In the GSC algorithm, incoming sensor data is split into two signal paths as shown in the block diagram. The upper path is a conventional beamformer. The lower path is an adaptive unconstrained beamformer whose purpose is to minimize the GSC output power.

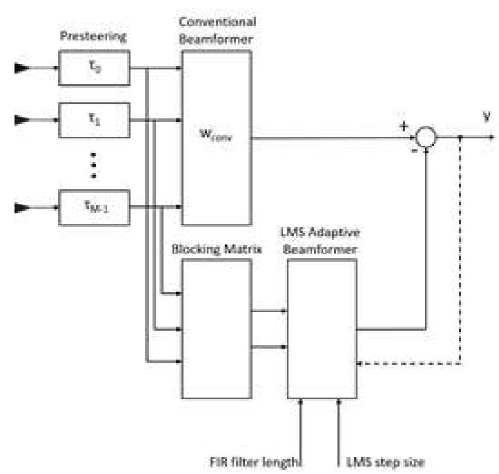

Fig 3. GSC Beamformer

The element sensor data is presteered by time-shifting the incoming signals. Presteering time-aligns all sensor element signals. The time shifts depend on the arrival angle of the signal. The presteered signals are then passed through the upper path into a conventional beamformer with fixed weights, $\mathbf{w}_{conv}$. The presteered signals are also passed through the lower path into the blocking matrix, **B**. The blocking matrix is orthogonal to the signal and removes the signal from the lower path. The lower path signals are filtered through a bank of FIR filters. The FilterLength property sets the length of the filters. The filter coefficients are the adaptive filter weights, $\mathbf{w}_{ad}$. Difference between the upper and lower signal paths are computed. This difference is the beamformed GSC output. The beamformed output is fed back into the filter. The filter adapts its weights using a least mean-square (LMS) algorithm. The actual adaptive LMS step size is equal to the value of the LMSStepSize property divided by the total signal power. Beamformer have been applied in the rotor directions.

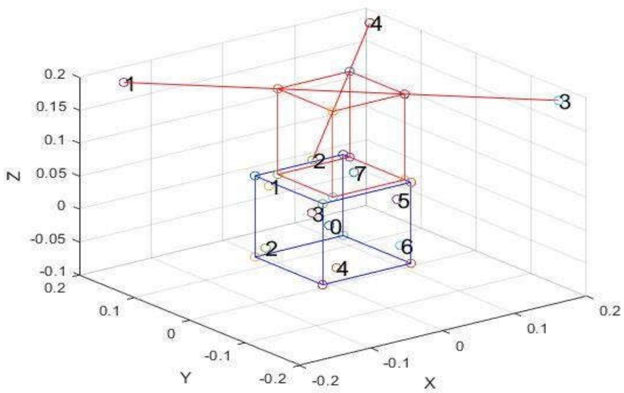

Fig 4. Plot of microphone array and rotor directions

Filter Length: 256
Sample Rate: 44100

B. Synthesizing Rotor Noise & Filtering

Thus, we have extracted signals which are used as sound sources and placed at rotor positions. Then we have simulated synthetic rotor noise by pyroomacoustics. Pyroomacoustics is a software package aimed at the rapid development and testing of audio array processing algorithms. By applying this method we have incorporated phase information and synthesized 8 channel rotor noise to be used as the reference of the wiener filter in attenuation mode to filter the noise. So, the flowchart of the noise cancelling phase is as following,

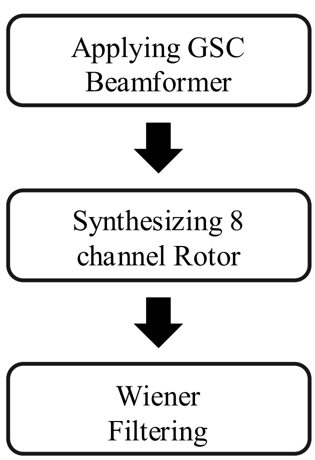

Fig 5. Flowchart of Noise Cancellation

C. Sound Source Localization

Multi-channel BSS Locate software aims at localizing audio sources in 3D audio scene based on recorded signals using an array of N microphones. Results are expressed in terms of directions (azimuth and elevation) using the centroid of this microphones array as reference.

Localization is achieved under far field assumption based on a choice between several local angular spectrum methods. Chosen method is applied to each microphone pair and resulting contributions (of all microphones pairs) are then aggregated. Available local angular spectrum methods are those used in BSS Locate Toolbox [1]: GCC-PHAT, GCC NON LIN, MVDR, MVDRW, DS, DSW and DNM. No clustering based methods is available in Multi-channel BSS Locate software.

GCC-PHAT and GCC NON LIN methods are used to obtain angular spectrum separately. Also, wiener filtering in ‘attenuation’ mode is used to attenuate rotor noises of a drone. Some special parameters which helped us to estimate better direction on given problem set are described below:

GCC-PHAT and GCC NON LIN methods are used to obtain angular spectrum separately. Wiener filtering in ‘attenuation’ mode is used to attenuate rotor noises of a drone. While designing the system, the following things are considered:

a. Number of Sources: Number of sources to be located. Though we had to identify only one source, for some cases (where extensive noise is present) our desired sound source may be 2nd or 3rd source. So, we identified more than one source and chose one among them

b. Wiener Filter in Attenuation Mode: Wiener filter aims to produce the estimation of a target random signal from an observed noisy process. It can be used in two different modes- emphasis and attenuation. Emphasis mode is used to separate the reference the signal while in attenuation mode the reference signal is to be suppressed in the output signal. In our system, reduction of rotor noises in angular spectrum is desired which requires the filter to be designed in attenuation mode.

c. Reference Signal of Wiener Filtering: We simulated individual rotor noises according to their physical position.

Wiener filtering for localization enhancement: In order to attenuate or emphasize the localization of a type of signal into a mixture, the toolbox provides some Wiener filtering tools. The typical scenario is a mixture composed of one speaker and a noise source, both located at different positions. If an excerpt of this noise is available for example, it can be used to enhance either the speaker localization or the noise localization on the mixture. This technique is only available for GCC NON LIN and GCC PHAT methods.

Averaged covariance definition:

Let us define x the multi-channel mixture (Ks samples and I channels) and xn the multi-channel signal excerpt to be attenuated or emphasized into the mixture (Kn samples and I channels). We denote as X(n,f) and $X_n(n,f)$ the corresponding I-dimensional complex valued vectors of TF coefficients.

First, let us define the averaged covariance of a signal Y(n,f)

$$R_y(f) = \frac{1}{N}\sum_{n} Y(n,f) * Y(n,f)^H$$

Attenuation mode:

In the "attenuation" mode, the signal excerpt is used to be attenuated into the mixture. The filtered version of X is computed below:

$$\tilde{X}(n,f) = \left[\{R_x(f) - \widehat{R_x}(f)\} * R_{x_n}^{-1}(f)\right] * X(n,f)$$

Thus we localize the static source.

III. Result

The results from the proposed method are quite promising. Actual azimuth and elevation angles were provided for each if the 300 audios. We have finally extracted the azimuth and elevation angles by GCC- PHAT and GCC NON LIN incorporating GSC beamforming for generating reference signal for wiener filtering. The azimuth and elevation angles are determined separately first and then angular error is calculated using following formula,

$$\Delta\sigma = arctan\frac{\sqrt{(cos\phi_2\sin(\Delta\lambda))^2 + (cos\phi_1 sin\phi_2 - sin\phi_1 cos\phi_2\sin(\Delta\lambda))^2}}{sin\phi_1 sin\phi_2 + cos\phi_1 cos\phi_2\cos(\Delta\lambda)}$$

We have compared our results with the results from GCC PHAT and GCC NONLIN method from [1]. A comparative analysis of the results are shown in the following chart,

TABLE I. Number of occurences for different angular errors

	Angular error (0^o - 5^o)	Angular error (5^o - 10^o)	Angular error (10^o - 15^o)	Angular error (15^o - 20^o)	Angular error (> 20^o)
1	235	40	3	3	19
2	242	33	6	3	16
3	180	43	5	3	69
4	184	41	4	3	68

For better understanding of the overall result we have determined some statistical parameters such as mean and deviation of the angular errors. Another chart containing mean angular error and standard deviation of the angular error of different methods,

TABLE II. Mean and standard deviation of angular error

	Mean angular error	Standard deviation
1	7.75	19.96
2	6.41	16.99
3	20.36	33.18
4	19.70	32.84

Where,
1=GCC PHAT with GSC Beamformer (Proposed)
2=CGG NON LIN with GSC Beamformer (Proposed)
3=Conventional GCC PHAT (Baseline)
4 = Conventional GCC NON LIN (Baseline)

From, the results it is apparent that, our proposed method provides significant improvement from the results of [1].

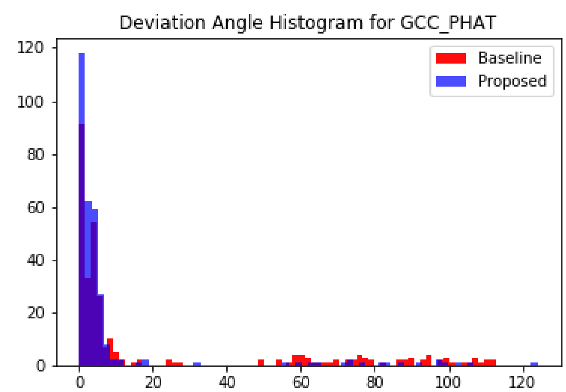

Fig. 4. Comparison of Baseline and Proposed method for GCC PHAT

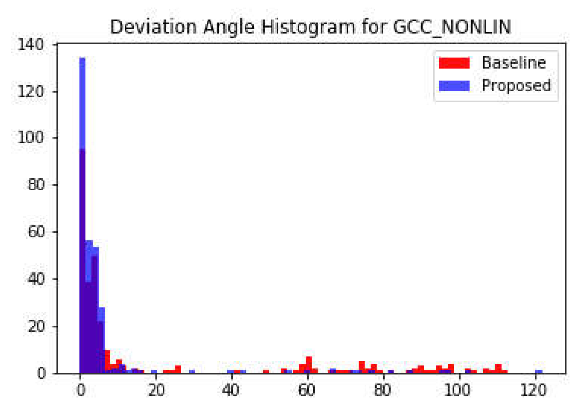

Fig. 5. Comparison of Baseline and Proposed method for GCC NON LIN

If 10^{o} of angular deviation is allowed then, the accuracy of our proposed method is 91.67% for both GCC PHAT and GCC NON LIN methods. Accuracy is increased by 17.34% by our proposed method. According to second table GCC NON LIN provides better result as mean and standard deviation of angular error both are less than that of GCC PHAT.

IV. Conclusion

In this paper, a state of the art method for determining azimuth and elevation from audio recorded by a microphone array embedded on a UAV is proposed. The results from the proposed method has a healthy accuracy with angular error less than 10 degrees. Reasonable progress has been recorded compared to similar methods which is inspiring for further research in this topic. Sound source localization by audio recorded by a particular microphone system is a relatively new area of research and our proposed method provides significant improvement in terms of accuracy. Incorporating wiener filtering with especially extracted noise signal mentioned in the methodology as reference signal of the filter is the main contribution of this method. Finally this approach can be applied to localize sound source in practical scenarios as it can locate a sound source quite accurately even from audios of -15 dB SNR. In future, we wish to work on improving the performance of our method and achieve better accuracy.

V. References

[1] Martin Strauss, Pol Mordel, Victor Miguet, Antoine Deleforge. "DREGON: Dataset and Methods for UAV-Embedded Sound Source Localization," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2018)*, Oct 2018, Madrid, Spain. pp.5735-5742

[2] M. Basiri, F. Schill, P. U. Lima, and D. Floreano, "Robust acoustic source localization of emergency signals from micro air vehicles," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012, pp. 4737–4742.

[3] T. Ohata *et al.*, "Improvement in Outdoor Sound Source Detection Using a Quadrotor-Embedded Microphone Array," in *Intelligent Robots and Systems (IROS), 2014 IEEE/RSJ International Conference on*. IEEE, 2014, pp. 1902–1907.

[4] K. Hoshiba *et al.*, "Design of UAV-Embedded Microphone Array System for Sound Source Localization in Outdoor Environments," *Sensors, 17, 2535*, pp. 1–16, 2017.

[5] H. W. Lollmann, H. Barfuss, A. Deleforge, S. Meier, and W. Keller- ¨ mann, "Challenges in acoustic signal enhancement for human-robot communication," in *Speech Communication; 11. ITG Symposium; Proceedings of*. VDE, 2014, pp. 1–4.

[6] L. Wang and A. Cavallaro, "Microphone-Array Ego-Noise Reduction Algorithms for Auditory Micro Aerial Vehicles," *IEEE SENSORS*, vol. 17, no. 8, pp. 2447–2455, 2017.

[7] K. Furukawa *et al.*, "Noise Correlation Matrix Estimation for Improving Sound Source Localization by Multirotor UAV," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*. IEEE, 2013, pp. 3943–3948.

[8] T. Morito *et al.*, "Partially Shared Deep Neural Network in Sound Source Separation and Identification Using UAV-Embedded Microphone Array," in *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*. IEEE, 2016, pp. 1299–1304.

[9] T. Tezuka, T. Yoshida, and K. Nakadai, "Ego-motion noise suppression for robots based on Semi-Blind Infinite Non-negative Matrix Factorization," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, 2014, pp. 6293–6298.

[10] A. Ito, T. Kanayama, M. Suzuki, and S. Makino, "Internal noise suppression for speech recognition by small robots." in *European Conference on Speech Communication and Technology (INTERSPEECH), 2013 IEEE/RSJ International Conference on*, 2005, pp. 2685–2688.

[11] A. Schmidt, A. Deleforge, and W. Kellermann, "Ego-Noise Reduction Using a Motor Data-Guided Multichannel Dictionary," in *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*. IEEE, 2016, pp. 1281–1286.

[12] C. Rascon and I. Meza, "Localization of sound sources in robotics: A review," *Robotics and Autonomous Systems*, vol. 96, pp. 184–210, 2017.

[13] K. Okutani *et al.*, "Outdoor Auditory Scene Analysis Using a Moving Microphone Array Embedded in a Quadcopter," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*. IEEE, 2012, pp. 3288–3293.

[14] L. Wang and A. Cavallaro, "Ear in the sky: Ego-noise reduction for auditory micro aerial vehicles," in *2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2016, pp. 152–158.

[15] K. Nakamura *et al.*, "Intelligent Sound Source Localization for Dynamic Environments," in *Intelligent Robots and Systems (IROS), 2009 IEEE/RSJ International Conference on*. IEEE, 2009, pp. 664–669.

[16] F. Grondin, D. Letourneau, F. Ferland, V. Rousseau, and F. Michaud, ´ "The ManyEars open framework," *Autonomous Robots*, vol. 34, no. 3, pp. 217–232, 2013.