3-D Localization of UAV and Detection based on Harmonics Index and Spectral Entropy Criteria

Muhammad Amjad Iqbal1, Zhao Zhao*, Xu ZhiYong, Saad Ur Rehman
School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, P. R China
engramjad@njust.edu.cn, zhaozhao@njust.edu.cn

Abstract. The popularity of drones or unmanned aerial vehicles (UAVs) will be increased enormously in future smart cities due to their usage in several domains. Significantly it’s an emerging research to localize and timely detect such kinds of objects to avoid threats. This paper considered to deploy an array of 4-microphones for acoustics localization enhancement and UAV detection optimization. For high accuracy of acoustics localization in 3-D, the cumulative algorithm of GCC-PHAT and least square (LS) based on TDOA is implemented to estimate acoustic direction of arrivals (DOAs). However, detection optimization is achieved by a spectral framework, where harmonics index estimation effectively determined sound and noise range in each frame, while spectral entropy featured noise and sound correlation, the least possible value of entropy verified the presence of UAV and optimize its detection. Moreover, experimental results confirmed the efficiency of proposed techniques, our system has robust performance for localization as it featured less than 5% errors, while detection accuracy is more than 75%.

1. Introduction
UAV is widely using in several commercial applications, on another hand it might be used for illegal purposes which violate social norms, so objective is to detect unauthorized drones. Target tracking with a camera using an image processing technique requires a high-resolution cameras which is not cost-effective [1]. Radars and millimeter waves have an advantage for surveillance of flying objects, even don’t get any affect by environmental factors [2]. Although, due to the small cross-section of UAV, radar can hardly pick up these small objects. Radio-frequency approach to detect UAV which has excellent results for long-range detection [3], but it hinders signal of other devices (Wi-Fi). Recently, machine learning approaches, linear predictive coding (LPC) and MFCC are used for speech signal and drone detection [4]. Regardless, due to misclassification and week feature identification of multipath signals, this solution is not suitable in real-time detection.

An emerging topic is a search of targets by its acoustic signature which is typically done by using a set of microphone array [5], which is cost-effective as it uses a large range of detection, correspondingly it hinder noises to maximize accuracy. For the passive method of localization, mostly two modus operandi are used in the literature, the Angle of arrival (AOA), and time difference of arrival (TDOA) [6, 7]. TDOA scheme is recommended for high accuracy localization, as it’s more efficient in both cases of 2-D and 3-D [8]. For powerful estimation of TDOA, GCC should be filtered with a known technique called phase-amplitude transform (PHAT), and hence resulting in the algorithm as GCC-
PHAT [9, 10]. This algorithm has several advantages; a well-performing attitude in reflective geography and more interesting is its sharp spectrum due to effective use of weighting function. Cleansed method is helpful to eliminate reflected waves, and to obtain those waves which are considered only direct waves and paving most of the spatial information [11]. For the sake of optimization in localization, least square (LS) method defines signal properties to find the location of target in 3-D effectively [12]. LS algorithm can do weight updates for multiple frames simultaneously, and each update does not need traditional resampling.

To detect target sound, several matrices can be selected including harmonics, and spectral entropy. Since drone sound contains harmonics features which are an integer multiple of its fundamental frequency [13]. Harmonics index estimation gives the range of target sound and noise information in each frame. The spectral entropy method [14] is used for the first time as this feature is widely helpful for detection of desired target signal. It can also de-noise the Doppler frequency signal successfully and ensure for the most part to gain coherence of information and maximum SNR.

In this paper, the design of acoustics localization and UAV detection is simple and fast in searching by using an array of 4-microphones. Array is designed to rectify the sound of 32-kHz with as low as possible mixing of interference. Section-2 presents localization and enhancement of acoustic signal. TDOA with GCC-PHAT is adapted to find time delays and the position of the sound. For peaks conversion into angles, Moore-Penrose (LS) technique is used. Localization approximation results are derived in the form of distance, elevation, and azimuth angles. The cleansing is introduced to identify direct waves, moreover, delay and sum beam-forming is applied to enhance localized signal. Section-4 elaborates the detection of UAV based on beam-formed vector, where harmonics indexes are evaluated to distinguish noise and sound range. To optimize detection of UAV, spectral entropy is implemented with different frame lengths. Conclusions with possible future advancements are briefed in section-5.

2. Localization and Enhancement of Acoustic signal in space

Microphone array elements are placed in spherical coordinates in 3-dimensional spacing whose coordinates are $M_i(x_i, y_i, z_i)$ where, $i = 1, \ldots, N$ as shown in Figure (1). GCC-PHAT and TDOA are employed to obtain geographical information, while 3-D view of acoustic emissions determined by LS method. To enhance localized signal, cleansing method with appropriate Hamming window is used, however for spatial filtration Delay and sum beamforming is used.

2.1. Generalized Cross-Correlation Phase Transformation (GCC-PHAT)

GCC-PHAT [9, 10] computes time delay $\tau$ obtained by TDOA and here function assumes that signals at microphone under observation and reference microphone are coming from the same source. GCC-PHAT finds the location of the peaks of cross-correlation between the observatory signal and reference signal. Relationship between the source signal and impulse response is represented by the following expression, $S_m(\omega) = Y(\omega) * H_m(\omega)$ while GCC of two microphones is derived as:

\[
Z(\tau) = \int_{-\pi}^{\pi} S_{m1}(\omega) * S_{m2}(\omega) e^{j\omega \tau} d\omega
\]  

(1)

\[
Z(\tau) = \int_{-\pi}^{\pi} Y_1(\omega)Y_2^*(\omega)H_{m1}(\omega)H_{m2}^*(\omega) e^{j\omega \tau} d\omega
\]

(2)

The weighting function is $H_{m1}(\omega)H_{m2}^*(\omega) = \psi_{m1m2}(\omega)$ designed to optimize performance criteria. Substituting weighting function into (2) GCC-PHAT will be of the form:

\[
Z(\tau) = \int_{-\pi}^{\pi} \psi_{m1m2}(\omega) Y_1(\omega)Y_2^*(\omega) e^{j\omega \tau} d\omega
\]

(3)

Weighting is used to emphasize the power of a directly incoming signal by suppressing noise power.

2.2. Sound Source DOA formulation based on the Least Squares Method

UAV in space is positioned at $(D \cos \theta \sin \theta, D \cos \theta \sin \phi, D \sin \phi)$ where ‘$\theta$, $\phi$, $D$’ are elevation, azimuth and distance from the origin of the reference coordinate. Apply an expression for a 4-element array, by using time difference information between $i_{th}$ and $j_{th}$ array elements, while $(r_i - r_j)$ is path distance of microphone.
\[
\begin{bmatrix}
\frac{r_2 - r_1}{v} \\
\frac{r_N - r_1}{v} \\
\frac{r_3 - r_2}{v} \\
\vdots \\
\frac{r_N - r_{N-1}}{v}
\end{bmatrix}
\begin{bmatrix}
\cos \theta \cos \phi \\
\sin \theta \cos \phi \\
\sin \phi
\end{bmatrix}
= \begin{bmatrix}
\tau_1 - \tau_2 \\
\tau_1 - \tau_N \\
\tau_2 - \tau_3 \\
\vdots \\
\tau_{N-1} - \tau_N
\end{bmatrix}
\]

(4)

Where ‘N’ represents the number of elements in an array and ‘v’ is the speed of sound. The optimal parameters $\theta$, $\phi$ can be estimated by Moore-Penrose method [15]. To solve a problem in the least square error sense, in general it is an exact solution to solve an over-determined problem.

\[
Y \cdot x = \tau
\]

(5)

‘Y’ is the structure of an array, ‘x’ is direction vector of sound and, ‘$\tau$’ is a time delay; LS method is

\[
\hat{x} = Y^+ \tau = (Y^T Y)^{-1} Y^T \tau.
\]

(6)

$Y^+$ is Moore-Penrose inverse matrix of $Y$ is given as:

\[
Y^+ = \begin{bmatrix}
u_1 \\
u_2 \\
u_3
\end{bmatrix}
\]

(7)

Thus least square estimates of azimuth and pitch angles of source in the far field are obtained as:

\[
\hat{\theta} = \tan^{-1} \left( \frac{u_1^T \tau}{u_2^T \tau} \right)
\]

(9)

\[
\hat{\phi} = \tan^{-1} \left( \frac{u_3^T \tau}{\sqrt{(u_1^T \tau)^2 + (u_3^T \tau)^2}} \right)
\]

(10)

The LS algorithm has less computational cost, and reduces the error for DOA estimation in the form of elevation and azimuth, which are obtained under real-time scenario.

2.3. DOA Improvement by Cleansing of Multi-path Effect

Peaks of GCC-PHAT function sometimes do not certainly match with accurate source location due to the multipath effect. For this concern, cleansing process is helpful to remove non-directional bins, to ensure that the microphones haven’t encountered with unnecessary information. To obtain better results of cleansing, Hamming window is employed as illustrated in (11):

\[
\mathcal{C} \tilde{Z}(\tau) = \int_{-\infty}^{+\infty} \left| w(f) \ast \tilde{Y}_{12}(f) \right| \ast e^{j2\pi ft}
\]

(11)

Here $\otimes$ refers to convolution operator and assume that additive noises are independent and also uncorrelated with source signal, while $w(f)$ is an appropriate hamming window.

2.4. Delay and Sum Beam-forming (DSB)

Beamforming is a technique to enhance a sound source in the desired direction [16]. To emphasise that to what extent the desired signal is affected by reflections and other sources. Simple delay-sum beamforming is carried out, where delays are inserted after each microphone to compensate path-delay. Later on, all samples of a relative microphone are delayed or advanced, to get a booster version of sound to form an acoustic beam in particular direction by following equation:

\[
Y(k) = \sum_{k=1}^{i} \alpha_k X_k(n - t_j)
\]

(12)

$X_k$ represents the number of microphones, $t_j$ corresponds to relative time delay and $Y(k)$ is a delay and sum beamforming output.
3. Spectral matrices for detection of UAV

To detect UAV, it’s necessary to collect useful spectral information. Several matrices can be selected, however in our case, harmonic index estimation, and spectral entropy are included.

3.1. Harmonic Frequency Index Estimation

UAV propellers generate complex sound and its acoustics feature fosters many harmonics [13], while harmonics are round about integer multiple of a fundamental frequency, which is helpful to distinguish drone sound. The time domain signal after beamforming of UAV is divided into frames. Apply FFT that data set is sparse in frequency domain especially, as energy is concentrated in harmonic structure. Find 1st fundamental in the spectrum between $F_L$ & $F_R$ and search for ‘$N$’ number of harmonic indexes as follows

$$\left[ \frac{n}{n-1} F_{n-1} * F_{\text{interval}} \right] - F_T, \left( \frac{n}{n-1} F_{n-1} * F_{\text{interval}} \right) + F_T$$

where $n=2$, and $F_T$ is threshold frequency, and Output, $F_1, \ldots, F_N$.

3.2. Spectral Entropy-based detection Optimization

Entropy is to encode signal according to its consistency as it featured noise and sound correlation [14]. The output results in the aforementioned step is under the condition of intense noise which may have an apparent error. Therefore to optimize detection, entropy-based method has been reported. The least possible value of entropy indicates the presence of UAV, smaller variations validate the stability of target sound. Calculate the section of each spectral component in the form of energy which would be lean as a signal energy probability at a certain spectrum point.

$$p_x = \frac{S(f_x)}{\sum_{k=1}^{N} S(f_k)} , x = 1, \ldots, M$$

(13)

Here $S(f_x)$ is the energy of $f_x$, and $p_x$ is related probability density, while ‘$M$’ are points of FFT frequency components. Proportionally, spectral entropy for each UAV sound frame is stated as:

$$E_x = - \sum_{k=1}^{M} p_x \log p_x$$

(14)

If a frame consists of only noise then the value of entropy will approach to 1, if it is containing noise with the target signal information, then entropy value will lead to 0.

4. Experimental Validations

Figure (1) shows experimental setup. Performance of GCC-PHAT based on TDOA was empirically estimated between reference microphone and microphone under observation. To evaluate the performance of GCC-PHAT, the Moore-Penrose LS method is utilized to estimate DOAs. According to the GCC-PHAT result, as shown in Figure (2.a), the estimation performance of DOAs exhibits some non-directional bins and it has several ups and downs. After cleansing, non-directional bins were removed, result shown in Figure (2.b) is near to ground truth which is definitive and only focuses the direct waves coming from UAV, as DOAs are not drastic and mostly it contains target sound.

![Figure 1. Field experiments using microphone-array.](image1)

![Figure 2. GCC-PHAT result & cleansing of reflections.](image2)
DOAs of sound in space of 3-D is scrutinized in Figure (3) where drone flying movement is expressed and easily examines the accuracy of DOAs. Some multi-path bins appeared which might be caused by the collision of sound to building walls. In this experiment, the localization task achieved 95% accuracy with ground reality, which in turn will have a maximum success detection rate.

Figure (4) shows a beam pattern, which has an effect of boosting the sound signal at the desired angle, while unwanted signals are combined in a quite unpredictable manner. Beamforming is applied to highlight that, to what extent target sound is deteriorated by noises. UAV is searched at an angle of 30°. The aim is to achieve the main-lobe with minimum reflections at the desired angle. It is noticed that interferences are lower, depression is quite a bit, and noise is not interfering main-lobe signal and its behaviour is stable, side-lobe regions are [10 - 25; 35 - 50] degree.

In Figure (5) index-wise estimation of harmonic frequency is presented. Background noises are mainly existing in the indexes range of 0-120, which arrives before harmonic components of UAV sound. Indexes of the sound signal obtained where peak frequencies were detected and it will arise in a harmonic correlation of fundamental frequency which is distinguishable. This feature gives a range of noise and UAV sounds in each frame to detect a target easily.

For the evaluation of spectral entropy, a specific set of isolated UAV sound is applied with three different frame lengths of 1024, 2048 and 4096. Length selection of frames is an important parameter. As it is shown in Figure (6) when we choose window size smaller it gives better results in the form of...
low entropy value. The number of samples in a frame, consequently the correlation between sound and noise is less. Slight variation of entropy in adjacent frames is a significant advantage. If the frame size is not appropriate, many samples will be lost and it will seed an error in the detection of UAV. Detection analysis for blocks of different frames is examined based on inverse RMSE relative to SNR. It is concluded that frames with less detection rate are sound of UAV having low potential, while most of the other frames exhibiting the high success detection rate, which indicates that these frames have a high potential of the target sound.

5. Conclusions
This work demonstrates UAV localization and detection, using an array of 4-microphone elements. Several problems on SSL have been addressed based on its pros and cons, we deliberated to improve localization, and summarized a novel strategy to detect UAV. GCC-PHAT with TDOA extracted the phase information to get the position of UAV and LS technique obtain DOA successfully. It has been empirically manifested that, localization has outperformance for DOA estimation, as accuracy is 95%. To detect UAV, representative algorithms of harmonics, and spectral entropy are applied. As long as simulation is quick enough to execute processing and evaluation was done passively, results of harmonics frequency index estimation and spectral entropy are realistic. Noise and sound ranges are extracted based on harmonic indexes, the gain of the UAV signal is directional for most of the frames, and spectral entropy of UAV sound is estimated with minimum entropy value, which depicts that data under observation contains target sound, as there is a very slight variation in entropy behaviour. This research provides many new openings, specifically the improvement for real-time detection pursuit. Thereafter, it will be an opportunity to distinguish if the certain UAV is viable to threat or either it is a functional flight [17]. Acoustic signature of drone which has (malicious) payload is different from a UAV that is just having a simple flight. Tracking will be plus point to extend the validity of the above work.

Acknowledgments
This work was supported by the National Natural Science Foundation of China [grant number 61401203]; the Fundamental Research Funds for the Central Universities [grant number 30918012203]; Equipment Pre-Research Program of China [grant number 41413010401].

References
[1] A. Rozantsev, V. Lepetit, and P. Fua.: ‘Detecting Flying Objects using a Single Moving Camera’, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, vol. 39, pp 879 – 892.
[2] F. Fioranelli, M. Ritchie, H. Griffiths, et al.: ‘Classification of Loaded/Unloaded MicroDrones Using Multistatic Radar’, Oct. 2015, IET Electronics Letters, vol. 51, no. 22, pp 1813–15.
[3] M. Haluza and J. Cechak.: ‘Analysis and Decoding of Radio Signals for Remote Control of Drones’, Proc. IEEE New Trends in Signal Processing, , Slovakia, Oct. 2016, pp 1–5.
[4] Muhammad Zohaib Anwar, Zeeshan Kaleem, Abbas Jamalipour.: ‘Machine Learning Inspired Sound-Based Amateur Drone Detection for Public Safety Applications’, IEEE Transactions on Vehicular Technology, 2019, Vol. 68, pp 2526 – 2534.
[5] Caleb Rascon, Ivan Meza.: ‘Localization of sound sources in robotics: A review’, Robotics and Autonomous Systems, ELSEVIER, Oct 2017, vol.96, pp.184–210.
[6] Yang, M., Jackson, D.R., Chen, J., et al.: ‘A tdoa localization method for nonlineof-sight scenarios’, IEEE Trans. Antennas Propag., 2019, 67, (4), pp. 2666-2676.
[7] Zheng, Y., Sheng, M., Liu, J., et al.: ‘Exploiting aoa estimation accuracy for indoor localization: a weighted aoa-based approach’, IEEE Wirel. Commun. Lett., 2019, (1), pp. 65-68.
[8] Xianyu Chang, Chaoqun Yang, Junfeng Wu et al.: ‘A Surveillance System for Drone Localization and Tracking Using Acoustic Arrays’, 2018 IEEE 10th (SAM) Workshop.
[9] J. Xie, H. Tao, X. Rao, and J. Su.: ‘Passive localization of mixed far field and near-field sources
without estimating the number of sources’, Research gate, Sensors, 2015, vol. 15, pp. 3834–3853.

[10]. Ali Pourmohammad, Seyed Mohammad Ahadi.: ‘Real Time High Accuracy 3-D PHAT-Based Sound Source Localization Using a Simple 4-Microphone Arrangement’, 2012, IEEE System Journal, vol.6, pp. 455 – 468.

[11]. Sangmoon Lee, Youngjin Park2 and Youn-sik Park.: ‘Cleansed PHAT GCC based Sound Source Localization’, International Conference on Control, Automation and Systems IEEE, 2010 Oct. 27-30.

[12]. Szymon Wozniak and Konrad Kowalczyk: ‘Passive Joint Localization and Synchronization of Distributed Microphone Arrays’, IEEE Signal Processing Letters, Vol. 26, no. 2, February 2019.

[13]. Hannes Pessentheiner, Martin Hagmüller, Gernot Kubin.: ‘Localization and Characterization of Multiple Harmonic Sources’, IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2016,vol. 24(8), pp. 1348-1363.

[14]. Shufen Qi, Jinghong Zhao, Xiaojie Liu.: ‘Mill Sound Detection Based On Signal Singular Entropy Noise Reduction’, Chinese Control and Decision Conference (CCDC) IEEE, Yinchuan China, may 2016, pp.3766 – 3769.

[15]. Xu Zhi Young, Zhao Zhao.: ‘Perturbation sensitivity analysis of azimuth estimation of the acoustic source for the planar microphone array’, Journal of Xidian University, 2017, doi:10.3969/j.issn.10012400.

[16]. David Kurc, Vaclav Mach, Kristian Orlovsky et al.: ‘Sound Source Localization with DAS Beamforming Method Using Small Number of Microphones’, International Conference on Telecommunications and Signal Processing (TSP), 2013 IEEE, pp. 526 – 532.

[17]. Francesco Fioraneli, Matthew Ritchie, Hugh Griffiths.: ‘Multistatic human micro-Doppler classification of armed/unarmed personnel’, IET Radar, Sonar & Navigation, 2015, Volume: 9, pp. 857 – 865.