# Efficient Phase-Based Acoustic Tracking of Drones using a Microphone Array

Paul M. Baggenstoss, Mark Springer, Marc Oispuu, Frank Kurth
Fraunhofer FKIE, Fraunhoferstrasse 20
53343 Wachtberg, Germany
Email: p.m.baggenstoss@ieee.org

***Abstract*—An efficient phase-based acoustic detection and tracking algorithm for drones is presented. The algorithm separately tracks the time difference of arrival (TDOA) of the incoming signal with respect to microphone pairs based on the phase in the discrete Fourier transform (DFT) bins. The direction of arrival (DOA) of the drone is determined by forming a solution curve corresponding to each TDOA, then clustering the curve intersections. The algorithm avoids the computationally expensive grid-search over DOA, so is significantly more efficient than beamforming. The proposed algorithm and the maximum likelihood (ML) processor (beamformer) are compared in simulated and real data scenarios. Using simulated data, the ML estimator is shown to agree with the Cramér-Rao lower bound (CRLB) and the proposed algorithm is shown to approach the performance of ML at higher SNR. In real data scenarios, the phase-based algorithm implemented with simple alpha-beta TDOA trackers consistently tracked the target through difficult maneuvers at short and long range, showing no degradation with respect to the beamformer. Other potential advantages include robustness against interference and ability to create phase-based spectrograms for classification.**

## I. Introduction

### A. Motivation and Prior Work

Due to the increasing use (and misuse) of unmanned aerial vehicles (UAV, or drones), there is increasing interest in systems that can detect, track, classify, and warn of approaching drones. Acoustic systems have advantages over infra-red, radar, and electro-optic systems for which drones present weak signatures [1]. While some specialized systems are designed for sound visualization in close-range scenarios [2], there is a need for efficient systems that can locate drones at long range in environments with interfering sources. Because distance to the drone can be estimated by triangulation of direction of arrival (DOA) estimates from multiple arrays, there is a need for inexpensive systems that measure DOA only. While most proposed systems are based on classical beamforming [1], [3], [4], there is the potential for efficient algorithms based on time difference of arrival (TDOA) estimation [5], [6], [7]. Of particular importance is the avoidance of the computationally expensive grid-search over DOA. For further increases in robustness and efficiency, time delay estimation can be based on phase difference between the signal at two sensors at a given frequency. For exponential-phase signal models, phase-based algorithms approximate the ML estimators [8] and can achieve the CR Bound at higher SNR [9]. Using phase can increase robustness for reasons that will be discused.

[1]This research was funded by German BMBF in the AMBOS project under grant 13N14269.

### B. Overview of Approach

A phase-based DOA tracking algorithm is proposed in this paper and is illustrated in Figure 1 for a simplified 3-microphone system. The microphone time-series data is segmented into overlapped time windows, then processed by discrete Fourier transform (DFT). An initial time difference of arrival (TDOA) estimate is made for each microphone pair using classical frequency-domain cross-correlation. Deviations from the initial TDOA estimate are tracked by comparing the phase difference for two microphones in a given DFT bin with the predicted phase difference to arrive at a phase error. The phase errors are combined using a novel frequency weighting into a robust estimate of TDOA error and then processed by the TDOA tracker to develop a consistent, phase locked track of the TDOA. To localize the drone, the locus of all points consistent with a measured TDOA between a pair of sensors is projected onto a circle on an imaginary distant sphere in 3-dimensional space. The intersections of these circles are clustered into a global DOA estimate (azimuth and elevation), and refined using some iterations of Newton's method. In addition to efficiency, TDOA tracking and DOA clustering both contribute to robustness for reasons to be discussed in Sections II-C and III. To the best of our knowledge the TDOA-tracking in has not been applied to the drone detection task before.

## II. Algorithm Description

### A. Maximum Likelihood (ML) Processor (Beamformer)

The ML bearing estimator was used as a performance benchmark. The beamformer arises as the implementation of the ML bearing estimator and attains the Cramér-Rao lower bound (CRLB) in theory. The assumed signal model, the ML processor, and the CRLB are detailed in appendix A.

### B. TDOA Estimation

DOA can be measured indirectly by first measuring the time delay, called time difference of arrival (TDOA), at each microphone pair (There are $\frac{M(M-1)}{2}$ different pairs in an $M$-microphone array). To initialize the TDOA estimate, we used classical frequency-domain cross-correlation, taking care to prevent circular wrap-around by zero-padding. A band-pass filter was applied before the final inverse FFT. TDOA

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

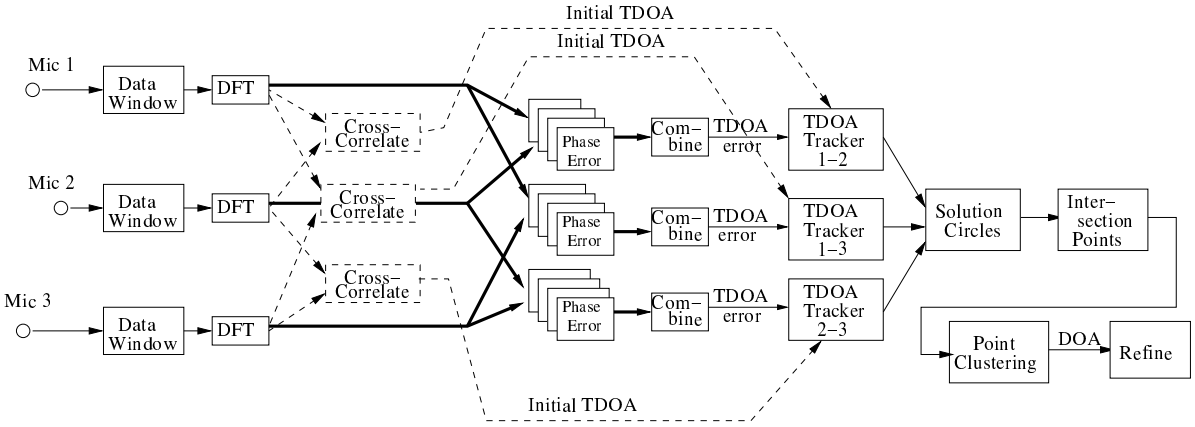


Fig. 1: System block-diagram for simplified 3-microphone scenario. Dotted lines indicate the TDOA initialization. Thick lines indicate values that exist at each frequency bin.

is obtained by peak-picking within the allowable range given the array geometry.

The initial TDOA estimate can be updated using phase measurements. Let $X_{m,i}$ be DFT bin $i$ from microphone $m$. Then, for microphones $l$ and $m$, $X_{m,i} \simeq X_{l,i}\ e^{-j2\pi i \tau_{lm} f_s N}$, where $N$ is the DFT size, $f_s$ is the sample rate, and $\tau_{lm}$ is the true time-delay between sensors $l$ and $m$. Given an initial TDOA estimate $\hat{\tau}_{lm}$, we form the following phase measurement $\phi_i^{lm} = \text{angle}(X_{l,i}\ \bar{X}_{m,i}\ e^{-j2\pi i \hat{\tau}_{lm} f_s N})$ which is related to the TDOA error $\delta^{lm} = (\tau_{lm} - \hat{\tau}_{lm})$ by

$$\phi_i^{lm} \simeq \frac{2\pi i f_s}{N}(\tau_{lm} - \hat{\tau}_{lm}). \tag{1}$$

By inverting (1), we get an ideally unbiased estimate of TDOA error for each DFT bin:

$$\delta_i^{lm} = \frac{N}{2\pi i f_s}\phi_i^{lm}. \tag{2}$$

These frequency-specific TDOA error estimates are combined into an ideally unbiased measurement of TDOA error using the formula

$$\delta^{lm} = \frac{\sum_{i=0}^{N/2} w_i \delta_i^{lm}}{\sum_{i=0}^{N/2} w_i}. \tag{3}$$

where $w_i = i^\rho\ e^{-(\phi_i^{lm})^2/(2\sigma_w^2)}$. The term $e^{-(\phi_i^{lm})^2/(2\sigma_w^2)}$, where $\sigma_w^2$ is a phase variance parameter, is used to apply a lower weight to frequencies where the phase error is high, presumably because of noise, interference, or drone dynamics. A value of $\sigma_w^2 = .25$ was used. The term $i^\rho$ is used to control the frequency weighting of the phase measurements. Notice that (2) has a frequency weighting of $1/i$, so (3) has a combined frequency weighting of $i^{\rho-1}$. If (1) is seen as a function of frequency $i$, then $(\tau_{lm} - \hat{\tau}_{lm})$ is proportional to the slope. To approximate the slope using linear regression, which also provides the approximate ML estimate of time-delay [8], an exponent of $\rho = 2$ would be required. But, the dynamics of the drone produce larger unknown phase errors at higher frequencies where sound absorption is greater (and SNR lower), suggesting a lower value of $\rho$. We found experimentally that $\rho = 0.5$ is a good compromise, but it is clearly application dependent since it is a function of signal spectrum, background noise, target dynamics, sound absorption, etc.

### C. TDOA tracking

We used the well-known alpha-beta tracker [10], [11], to track the TDOA error for microphone pair $l, m$, $\delta^{lm}$. Details are provided in appendix B. Tracking TDOA based on phase errors has some advantages. As long as the TDOA is tracked, it remains phase-locked because any bias in the TDOA estimate results in an ever-accumulating phase error. Also, track loss can be easily detected (so that the tracker can be re-intialized), by monitoring average phase error, or by detecting if there are no correlator peaks near the tracked TDOA. In the ML processor (beamformer), an interfering source can bias or even mask the local maximum due to the desired target. But, when tracking TDOA based on phase, the energy from interfering signals appearing in a DFT bin, no matter how strong, will likely produce a random (uniformly distributed) phase error because there is no correlation between the predicted phase (based on the tracked TDOA) and the interfering signal. This means that as long as the majority of the DFT bins with low phase error are due to the desired source, tracking will continue and not be interrupted by the interfering source.

## III. Time-Delay Based DOA estimation

A geometric approach is employed to estimate the DOA in form of azimuth and elevation $(\theta_1, \theta_2)$. Every sensor pair in conjunction with a TDOA estimate induces a hyperbolic surface containing all possible points which would produce the measured TDOA. If all the surfaces are intersected with a sphere of some radius $\mathcal{R}$ large enough, circles are formed on this sphere[1]. If the intersections of each pair of circles are calculated, a cloud of points that are consistent with at least two measured time delays is obtained. The core idea is

[1]Assuming the center of the two the sensors coincides with the array center.

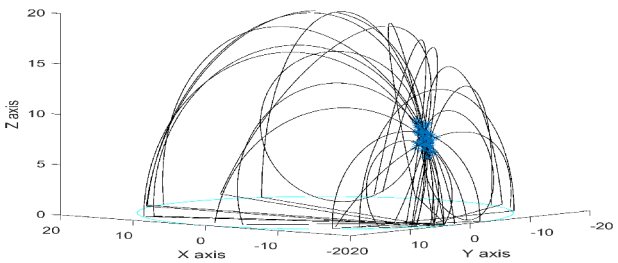


Fig. 2: Example with 28 sensor pairs, i.e. all pairs of an 8 element array.

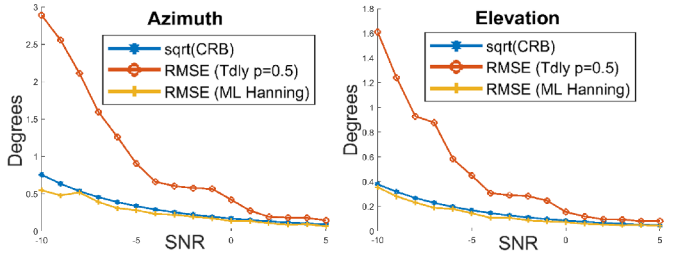


Fig. 4: Comparison of azimuth and elevation standard deviation of TDOA-Tracking to tracked Beamformer.

that these points form a cluster on the sphere somewhere in direction of the arriving sound. Outliers caused by interference that do not belong to the main cluster will not bias the main cluster center. The cluster can be extracted using conventional clustering algorithms, such as K-means or DBSCAN[12], and its geometric center be used to calculate a DOA. An example of this can be seen in Figure 2, where the circles and the intersections of the largest cluster are drawn.

## IV. Simulations

### A. Stationary Target

For comparison with the classical method and the CRLB, a stationary simulated target was employed. Data was generated in the frequency domain according to the assumed signal model (see appendix A). Figure 3 shows the standard deviation of

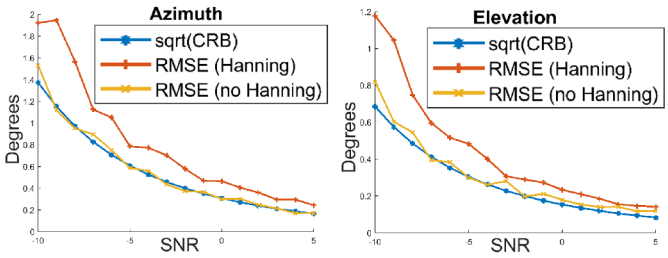


Fig. 3: Comparison of angular standard deviation to CRLB.

azimuth and elevation estimates, computed by maximizing (5) using Newton iterations. For later comparison with TDOA tracking, the standard deviation was computed once with Hanning weighting and once without. The target was simulated to appear at $\theta_1 = 50°, \theta_2 = 30°$, and the results computed using 100 independent runs. It can be concluded that the conventional processor with Newton iterations to refine DOA does indeed achieve the CRLB. Hanning weighting can be seen to reduce the effective number of samples, and therefore increases the resulting error. In the next experiment, estimation error will be reduced again by incoherent averaging over time, and used as a benchmark for evaluating the proposed approach.

The variance between the ML DOA estimator and the proposed phase-based method method is compared in Figure 4. A simple alpha-beta tracker was applied to the ML estimates so as to result in a fairer comparison with TDOA tracking. It can be seen that over the complete SNR range, we do not lose more than about 2.5 degrees in standard deviation when using time-delay tracking as opposed to the tracked maximum likelihood estimate. Furthermore at high SNR the difference in negligible. Additionally, the CRLB is plotted to confirm that through tracking we can achieve super-efficiency.

## V. Real Data Results

Data was acquired from an array consisting of $M = 10$ microphones arranged on a 35 cm diameter cylinder, with a sample rate of 25 kHz. A band pass filter of 500Hz - 12 kHz was applied in the frequency domain. Ground-truth was obtained by recording the GPS position of the drones.

Figure 5 shows the application of TDOA tracking and Beamforming/ML on real world data. The drone flew in a structured pattern, flying twice past the array (see dips in elevation at 180 and 285 seconds) out to about 100 meters. For the ML approach, the BF-function was computed on a grid of pre-formed beams spaced 2 degrees in azimuth and 1 degree in elevation, and incoherently integrated over one second. Local maxima were then detected and refined by Newton iterations. The global maximum was plotted without tracking.

On the plot can be seen the DOA estimates of the proposed method (dots), the global maximum DOA for ML (O's), and the ground truth (X's). Both the ML and the proposed method tracked the drone consistently throughout the exercise [2]. A quantitative comparison on real data could not be made because the differences between ML and TDOA tracking were smaller than errors caused by bias and delay of ground-truth data.

Note that in Figure 5, the variance in elevation is larger than the variance in azimuth. This is likely a result of array geometry that causes varying DOA accuracy and ground reflections which cause ambiguities in elevation. Also there is an offset in elevation which is mainly due to a mismatch in assumed array height and actual height of the array at the time of recording.

The computational time of the proposed method was 22 times faster than the ML processor. This primarily comes from the fact that ML requires a computationally expensive grid-search over DOA.

[2]Using detection distances provides just a rough idea of the algorithm's performance. Actual detection range is influenced by environmental factors such as background noise, the sound intensity of the drone, and microphone characteristics. To evaluate the proposed method, we compare the detection performance to the optimum (ML-) estimator.

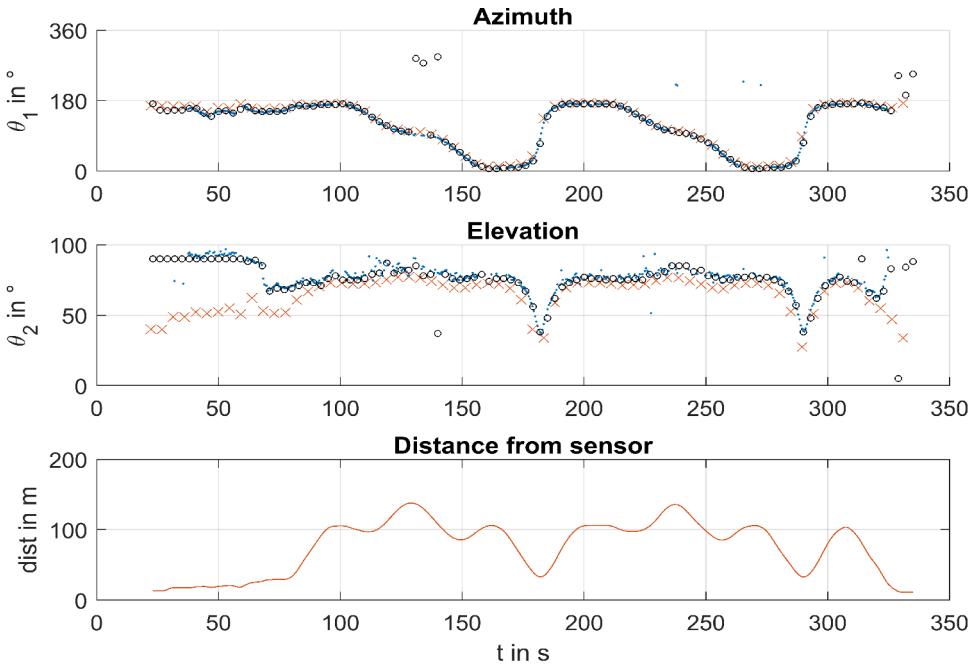


Fig. 5: Real data results. Azimuth, elevation (0 is straight up, 90 is horizon) and distance as a function of time. Key: (X) ground truth, (.) proposed phase-based algorithm, (O) ML/BF.

## VI. Analysis and Conclusions

The above results demonstrate the feasability of the proposed method in tracking drones in a real-workd scenario with no discernable loss of performance with respect to ML, but an order of magnitude reduced computational load. In addition to computational efficiency, operating using phase offers the potential for robust behavior and, as a side benefit, stable features for classification, as will be explained. Note that the ML processor is linear, making no "hard decisions" until the last step when the DOA is peak-picked and where interference can bias the location of the main peak or produce a larger interference peak[3]. On the other hand, the simpler phase-based TDOA tracker makes many hard decisions early in the processing stream, where the "safety in numbers" principle applies. Even if some microphone pairs track the wrong TDOA peak, the resulting false circle intersection points are not likely to fall in or bias the main DOA cluster. And, by estimating TDOA error using only DFT bin phase, a "hard decision", a good TDOA estimate results as long as the majority of DFT bins produce a target-derived phase error which is unbiased in the long run due to phase-locking. A side benefit of interference-rejection property of TDOA tracking using DFT bin phase (See Section II-C) is the ability to form "phase spectrograms", a potentially stable feature for classifying drones. Figure 6 illustrates such a spectrogtam created by plotting the function $e^{-\phi_i^2/\sigma_w^2}$, where $\phi_i$ is the phase error at DFT bin $i$ (see eq. 1), over six seconds of the TDOA track at one microphone pair and smoothing over time. Dark areas indicate low phase error. Clear narrow band lines from the drone's motors can be seen. Spectral lines from interfering signals will be rejected due to the lack of correlation with the target predicted phase.

[3]These errors can be handled by sophisticated tracking algorithms that are outside the scope of this paper.

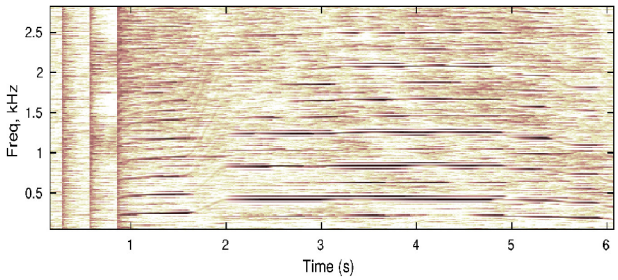


Fig. 6: Typical Phase Spectrogram.

In conclusion, we have demonstrated that phase-based TDOA tracking is sub-optimal, but loses only about two degrees or less of accuracy. DOA estimation is made significantly more computationally efficient. In real data scenarios, it is shown to track the drone consistently throughout a difficult exercise. Phase-based TDOA tracking offers the potential for interference rejection and robustness with the side benefit of an interference-free phase spectrogram for classification.

## Appendix

### A. Theoretical Analysis and CRLB

For theoretical analysis, it is assumed that there is a single target source whose signal is a sequence of independent Gaussian random variables (a white Gaussian process) with variance $\sigma_s^2$. Let there be $M$ microphones receiving this source

with an associated delay $\tau_m(\boldsymbol{\theta})$, $m = 1, \ldots, M$, where $\boldsymbol{\theta}$ is the target DOA parameter (azimuth and elevation angles), plus some independent Gaussian noise of variance $\sigma_n^2$. Let there be $N$ time samples of each microphone at sample rate $f_s$. Let these $N$ samples be transformed to the frequency domain by discrete Fourier transform (DFT). Then the frequency-domain signals at frequency bin $i$ and microphone $m$ can be written $X_{i,m} = S_i \; e^{-\frac{2\pi f_s \tau_m(\boldsymbol{\theta}) i}{N}} + u_{i,m}$, where $S_i$ is the source process at DFT bin $i$ and $u_{i,m}$ is the independent noise (independent in time and spatially across microphones). Because all quantities are real Gaussian processes, $X_{i,m}$ is a complex Gaussian random variable and the signal and noise at different frequency bins are independent and uncorrelated, but correlated across microphones at a given frequency. Let $\mathbf{X}$ stand for the collection of all frequency-domain values $\mathbf{X} = \{X_{i,m}\}$. Then $\mathbf{X}$ has distribution

$$\log p(\mathbf{X}) = -\sum_{i=0}^{N/2} \left\{ M \log(\pi) + \mathbf{X}_i^* \mathbf{R}_i^{-1} \mathbf{X}_i + \log |\mathbf{R}_i| \right\}, \tag{4}$$

where $\mathbf{X}_i$ is the $M \times 1$ vector of complex DFT bin values at frequency bin $i$ and the complex frequency-domain covariance matrix is $\mathbf{R}_i = \mathcal{E}\{\mathbf{X}_i \mathbf{X}_i^*\} = N\sigma_s^2 \mathbf{e}_i \mathbf{e}_i^* + N\sigma_n^2 \mathbf{I}$, where $\mathbf{e}_i = \left[ e^{-\frac{2\pi f_s \tau_1(\theta) i}{N}}, e^{-\frac{2\pi f_s \tau_2(\theta) i}{N}} \ldots e^{-\frac{2\pi f_s \tau_M(\theta) i}{N}} \right]'$. Note that the variance parameters can be a function of frequency bin $i$, although for notational simplicity, the index is omitted. The inverse can be expressed as $\mathbf{R}_i^{-1} = \frac{1}{N\sigma_n^2}\mathbf{I} - \frac{\sigma_s^2}{N\sigma_n^2(M\sigma_s^2+\sigma_n^2)} \mathbf{e}_i \mathbf{e}_i^*$ nd the determinant can be expressed as $|\mathbf{R}_i| = (NM\sigma_s^2 + N\sigma_n^2)(N\sigma_n^2)^{(M-1)}$. Therefore, (4) can be rewritten

$$\log p(\mathbf{X}) = C - \sum_{i=0}^{N/2} \left\{ \frac{1}{N\sigma_n^2} |\mathbf{X}_i|^2 - \frac{\sigma_s^2 |\mathbf{e}_i^*(\boldsymbol{\theta}) \; \mathbf{X}_i|^2}{N\sigma_n^2(M\sigma_s^2 + \sigma_n^2)} \right\}, \tag{5}$$

where $C$ is independent of $\mathbf{X}$ and $\boldsymbol{\theta}$.

To estimate the target position parameter $\boldsymbol{\theta}$ by maximum likelihood, one must maximize (5) over $\boldsymbol{\theta}$. The conventional beamformer arises because the maximization can be accomplished by maximizing the term $B(\boldsymbol{\theta}) = \sum_{i=0}^{N/2} \; |\mathbf{e}_i^*(\boldsymbol{\theta}) \; \mathbf{X}_i|^2$ over $\boldsymbol{\theta}$ which is tantamount to steering the beam to maximize output power.

The lower bound on the error variance for parameter $\boldsymbol{\theta}$ is given by the Cramér-Rao lower bound. The derivative of (5) with respect to a component $\theta$ of $\boldsymbol{\theta}$ is

$$\frac{\partial \log p(\mathbf{X})}{\partial \theta} = \sum_{i=0}^{N/2} \frac{2\sigma_s^2 \; \mathrm{Re}\left\{ \mathbf{e}_i^{\theta *} \; \mathbf{X}_i \; \mathbf{X}_i^* \mathbf{e}_i \right\}}{N\sigma_n^2(M\sigma_s^2 + \sigma_n^2)},$$

where $\theta$ in the exponent indicates first derivative. From this expression, the Fisher's information matrix for $\theta_1, \theta_2$ can be derived:

$$\mathbf{I}_{\theta_1,\theta_2} = \sum_{i=0}^{N/2} \frac{2\sigma_s^2 \; \mathrm{Re}\left\{ \mathbf{e}_i^{\theta_1 *} \mathbf{R}_i \mathbf{e}_i^{\theta_2} + \mathbf{e}_i^{\theta_1 \theta_2 *} \mathbf{R}_i \mathbf{e}_i \right\}}{N\sigma_n^2(M\sigma_s^2 + \sigma_n^2)}. \tag{6}$$

The CRLB is then the inverse $\mathbf{C}(\theta_1, \theta_2) = \mathbf{I}_{\theta_1,\theta_2}^{-1}$.

### B. Alpha-Beta Tracker

A separate TDOA tracker is used for each sensor pair. In the following, the tracker is described for an arbitrary sensor pair.

*1) State Variables:* State variables include the TDOA (time delay) $\tau$ and the time-delay rate of change $v$ (in seconds of delay per second) : $\mathbf{x} = [\tau, \; v]'$.

*2) Initialization:* The state variable $\tau$ is initialized to the TDOA in seconds obtained by peak-picking the cross-correlation. The velocity $v$ is set to zero.

*3) Time Update:* Let there be a state variable estimate at time update $n-1$, based on data up to and including time update $n-1$ denoted by $\hat{\mathbf{x}}_{n-1|n-1}$. To extrapolate this estimate to time update $n$, the linear time-update is $\hat{\mathbf{x}}_{n|n-1} = \mathbf{A} \; \hat{\mathbf{x}}_{n-1|n-1}$, where the time update martrix is given by $\mathbf{A} = \begin{bmatrix} 1 & K/f_s \\ 0 & 1 \end{bmatrix}$, where $f_s$ is the input data sampling rate, $K$ is the number of new input samples processed in each time update, and $K/f_s$ is the amount of time that has passed since the last time update.

*4) Alpha-Beta Measurement Update:* To update the alpha-beta filter, we use $\tau_{n|n} = \tau_{n|n-1} + \alpha\delta$, and $v_{n|n} = v_{n|n-1} + \beta\delta$, where $\delta$ is from (3) , $\alpha = .5$ and $\beta = 0.8 \; \frac{2-\alpha^2-2\sqrt{1-\alpha^2}}{\alpha^2}$. This value of $\beta$ produces critical damping [10], [11].

## REFERENCES


[1] E. E. Case, A. M. Zelnio, and B. D. Rigling, "Low-cost acoustic array for small uav detection and tracking," in *2008 IEEE National Aerospace and Electronics Conference*, pp. 110–113, July 2008.

[2] J. Busset, F. Perrodin, P. Wellig, B. Ott, K. Heutschi, T. Rhl, and T. Nussbaumer, "Detection and tracking of drones using advanced acoustic cameras," in *Unmanned/Unattended Sensors and Sensor Networks XI: and Advanced Free-Space Optical Communication Techniques and Applications* (E. M. Carapezza, P. G. Datskos, C. Tsamis, L. Laycock, and H. J. White, eds.), SPIE, oct 2015.

[3] M. Benyamin and G. H. Goldman, "Acoustic detection and tracking of a class i uas with a small tetrahedral microphone array," in *Report ARL-TR-7086*, Army Research Laboratory, Adelphi, MD 20783-1138, 2014.

[4] F. Grondin, J.-S. Lauzon, J. Bass, D. Ltourneau, A. Lussier-Desbiens, and F. Michaud, "Drone detection and localization with sound using multiple microphone arrays," in *Proceedings NSERC Canadian Field Robotics Network Symposium*, 06 2016.

[5] C. O. Tiemann and M. B. Porter, "Automated model-based localization of sperm whale clicks," in *OCEANS 2003 Proceedings*, vol. 2, pp. 821–827, Sept 2004.

[6] P. Giraudet and H. Glotin, "Real-time 3d tracking of whales by echo-robust precise tdoa estimates with a widely-spaced hydrophone array," *Applied Acoustics*, vol. 67, no. 11-12, pp. 1106 – 1117, 2006. Detection and localization of marine mammals using passive acoustics.

[7] P. M. Baggenstoss, "Processing advances for localization of beaked whales using time difference of arrival," *Journal of the Acoustical Society of America*, vol. 133, pp. 4065–4076, 2013.

[8] A. G. . Piersol, "Time delay estimation using phase data," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-29, pp. 471–477, 1981.

[9] P. M. Baggenstoss and S. Kay, "On estimating the angle parameters of an exponential signal at high snr," *Signal Processing, IEEE Transactions on*, vol. 39, pp. 1203 – 1205, 06 1991.

[10] R. Penoyer, "The alpha-beta filter," *Canadian Govt Report*, 1993.

[11] T. R. Benedict and G. Bordner, "Synthesis of an optimal set of radar track-while-scan smoothing equations," *IRE Transactions On Automatic Control*, pp. 27–32, 1962.

[12] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, 1996.