Journal of Unmanned Vehicle Systems

A Harmonic Spectral Beamformer for the Enhanced Localization of Propeller-driven Aircraft

Journal:	<i>Journal of Unmanned Vehicle Systems</i>
Manuscript ID	juvs-2018-0011.R1
Manuscript Type:	Article
Date Submitted by the Author:	13-Mar-2019
Complete List of Authors:	Harvey, Brendan; Memorial University of Newfoundland, Engineering & Applied Science O'Young, Siu; Memorial University of Newfoundland, Faculty of Engineering and Applied Science
Keyword:	Acoustic sensing, Localization, Beamforming, Collision avoidance
Is the invited manuscript for consideration in a Special Issue? :	Not applicable (regular submission)

SCHOLARONE™
Manuscripts

A Harmonic Spectral Beamformer for the Enhanced Localization of Propeller-driven Aircraft

Brendan Harvey * and Siu O'Young

Abstract: The following paper presents several array processing techniques which may be used to enhance the localization of acoustic source targets such as UAVs. A review of common methods is first provided, followed by several algorithms developed to reduce computational loads for the application of concern. A beamforming method is proposed which exploits the properties of harmonic narrowband signals such as that generated by propeller-driven aircraft to enhance direction of arrival accuracy. In addition, a regional contraction search algorithm is proposed to minimize computational loads associated with the presented localization technique. A brief comparison of the proposed method to that of standard frequency domain beamformers is also provided using both theoretical analysis and experimental data. For the case of target localization between two moving fixed-wing UAVs, it was found that the proposed Harmonic Spectral Beamforming method increased localization accuracy by 50% over the standard Steered Response Power approach.

Keywords: acoustic sensing; localization; beamforming; collision avoidance

1. Introduction

The general use of acoustic sensing to detect, localize, and track moving targets has been well studied and documented in the literature (A. M. Zelnio, E. E. Case, B. D. Rigling 2009; A. Sutin et al. 2013; E. E. Case, A. M. Zelnio, B. D. Rigling 2008; H. Salloum et al. 2015; Nielsen 2009; Zelnio 2009). However, there have been very few accounts of using this technology for Unmanned Aerial Vehicle (UAV) collision avoidance purposes. Currently, there have been no reports of air-to-air localization of another aircraft (either manned or unmanned) from a moving UAV. Ferguson utilized a small UAV (Aerosonde) fitted with two microphones to detect and localize acoustic impulses from a propane cannon located on the ground (Ferguson and Wyber 2006). Detection distances of up to 300 m with a localization bearing error of 3° was reported; although evidence for these claims were not clearly presented. Robertson also conducted experiments where a ground-based propane cannon was detected and localized from a small UAV fitted with four microphones (Robertson et al. 2007). Detection distances of up to 180 m were reported with an average localization error of 8° . Ohata used a multirotor fitted with a large number of MEMS microphones to detect and localize an acoustic source located on the ground while hovering at altitudes less than 5 m, producing localization accuracies of approximately 10° (Ohata et al. 2014; Okutani et al. 2012).

Consider the general case involving a potential mid-air collision between two aircraft as depicted below in Figure 1. The system consists of an intruding aircraft which emits some unknown acoustic signal, and a detecting aircraft fitted with a microphone array. Both aircraft are assumed to be in

* Corresponding Author: brendanb@mun.ca; Tel: +1-709-764-0636, Faculty of Engineering & Applied Science, Memorial University of Newfoundland, St. John's, NL, Canada, A1B 3X5

continuous motion with constant velocities and are separated by distances large enough such that incident waves arriving at the sensors may be treated as planar with linear fronts. Translating coordinate frames are fixed to both aircraft and give locations with respect to the GPS coordinate system. Each translating frame also contains a rotated coordinate reference which provides kinematic information relative to the respective aircraft orientation (yaw, pitch, roll). Typically, only information projected along a direct line-of-sight vector connecting the two aircraft is of concern, since this component ultimately governs characteristics of the received source signal. Relative separation distances will largely dictate source attenuation levels, with atmospheric effects also providing some typically unknown contribution. Relative motion between the aircraft will produce variations in the observed source frequency due to Doppler effects. Self-noise generated by the sensing aircraft will effectively corrupt acquired signals and thus influence the ability to perform detection and localization operations. The ultimate goal of the system is to utilize the known kinematic information of the sensing aircraft in conjunction with measured data such as acquired signal amplitude, spatial phase characteristics, and perceived source frequency, to determine information about the unknown acoustic source such as location and velocity. With respect to the above considerations, the overall system may therefore be described in terms of its acoustic properties, kinematic relationships, and signal acquisition/processing requirements.

It is expected that received signals will be harmonic and continuous in time since they are generated by propeller-driven aircraft. Due to the time sensitive nature of the operation, algorithms must be capable of operating in real-time with low latency. Due to physical limitations associated with placing a microphone array on a UAV, it is also unlikely that a uniform spacing can be achieved for all sensors. In addition, vibrations and unsteady airflow during flight operations will produce some degree of motion in the array elements effectively producing varying positional errors. This is combined with the fact that practical limitations during construction and installation of the array will also produce some degree of positional uncertainty with respect to the original intended array design. Taking the above aspects into consideration, it is apparent that a robust beamforming method is required to ensure maximum operational performance for these challenging conditions.

Thus, the purpose of the following paper is to present several signal processing methods to enhance the spatial localization of acoustic source targets under adverse conditions. That is, for situations where microphone placement may be severely limited and would generally result in poor performance (localization accuracy) for signal frequencies typically associated with small propeller-driven aircraft. To maximize localization accuracy, a beamforming method is proposed which exploits the harmonic narrowband nature of acoustic signals generated by propeller-driven aircraft. In addition, a regional contraction search algorithm is provided to minimize computational loads associated with the proposed localization technique. A brief comparison of the method to that of standard frequency domain beamformers is also provided using both theoretical analysis and experimental data.

The presented paper is an extension to work previously presented by Harvey and O'Young to evaluate the viability of acoustic sensing to establish a non-cooperative aircraft collision avoidance system (Harvey and O'Young 2015; Harvey and O'Young 2018). More specifically, localization results are provided from experiments previously conducted to determine maximum detection distances achievable between two fixed-wing UAVs (Harvey and O'Young 2018). In order to adhere to paper length requirements, many details regarding the experimental setup for the data used and prior signal processing implemented (filtering, whitening, CFAR detection, etc.) will not be directly presented, but rather the reader directed to the appropriate article where the information can be readily found.

2. Background Information

2.1 Conventional Beamforming

The most simplistic and commonly employed beamformer is the Delay-and-Sum form. Here we consider a linear array consisting of S sensors/microphones with locations given by $\vec{r}_s$ where $s \in \{1, 2, \dots, S\}$. The principle behind the approach is to simply apply an appropriate time delay such that the acquired target signals become aligned producing a coherent amplified output when summed together. Alternately, the Delay-and-Sum beamformer may be expressed in the frequency domain instead by applying the Fourier transform and realizing that a delay in the time domain equates to phase shift in the frequency domain. This form is often referred to as the Filter-and-Sum (FAS) beamformer and is given by the following expression (McCowan 2001):

$$Y(\hat{k}, f) = \sum_{s=1}^S W_s(f) \cdot X_s(f) \cdot e^{-j2\pi f \tau_s(\hat{k})} \quad (1)$$

where W_s are weighting values applied to each channel to modify directional output characteristics. The time delays τ_s are characterised by the steering direction which is given by the unit vector $\hat{k}$. The maximum output will be achieved when $\hat{k}$ aligns with the incoming wave propagation direction. For a given steering direction, the required time delay for the s^{th} microphone will be thus given by:

$$\tau_s = \frac{\hat{k} \cdot \vec{r}_s}{c} \quad (2)$$

where c is the speed of sound in air. For 3-dimensional space characterized by the azimuth and elevation angles ϑ and φ respectively as depicted above in Figure 1, the steering and position vectors will be given by the following:

$$\hat{k} = \begin{bmatrix} \cos(\vartheta) \cos(\varphi) \\ \sin(\vartheta) \cos(\varphi) \\ \sin(\varphi) \end{bmatrix} \quad (3)$$

$$\vec{r}_s = [x_s, y_s, z_s]^T \quad (4)$$

To avoid spatial aliasing in the array directivity response, the minimum sensor spacing must be less than half the incident wave length (McCowan 2001):

$$d < \frac{\lambda_{\min}}{2} \quad (5)$$

where $\lambda_{\min}$ is the wavelength for the signal of interest. Spatial aliasing reduces the ability to localize a desired source through the presence of grating or side lobes and should thus be avoided whenever possible.

2.2 Direction of Arrival Estimation

The process of spatially localizing signals is known as Direction of Arrival (DOA) estimation. There have been an abundance of DOA algorithms presented in the literature (Hero et al. 1998; Krim and Viberg 1996; Krishnaveni et al. 2013; Godara 1997). However, most can be classified into the following three categories: 1) Subspace or Beam-space, 2) Time Difference of Arrival (TDOA), and 3) Steered

Response Power (SRP). Subspace algorithms have arrived more recently in the field of beamforming, with algorithms typically offering superior localization accuracy over more conventional methods. In general, the technique uses an approach whereby the signal and noise are separated into separate component subspaces using an eigenvalue decomposition of a covariance or correlation matrix. The MUSIC, Root-MUSIC, and ESPRIT are commonly reported examples of this algorithm type (Krim and Viberg 1996). Although these methods have been successfully employed for a variety of array processing applications, they all possess certain restrictions that prevent their practical use for the application at hand. Some of these include: the requirement of a uniformly spaced array, high computational loads which often prevent real-time operation, and a high sensitivity to array positioning errors (Li and Vaccaro 1992; Krishnaveni et al. 2013; Godara 1997).

Time Difference of Arrival (TDOA) is perhaps the most commonly employed DOA method. This is largely due to its simplicity and low computational requirements which allows real-time implementation on most all digital systems (Brandstein and Silverman 1997; Chen et al. 2006; DiBiase et al. 2001; Chen et al. 2002; Krishnaveni et al. 2013). Compared to other methods such the Subspace and Steered Response Power, TDOA methods offer a significant computational advantage (Brandstein and Silverman 1997; DiBiase et al. 2001). However, the method is known suffer from poor resolution and deteriorate extensively in the presence of multiple sources, reverberation, and low SNR environments (DiBiase et al. 2001). This is due to the presence of local maxima in the cross-correlation function which may obscure the true TDOA peak and subsequently produce incorrect delay estimates. The amplitudes of these erroneous maxima depend on a number of factors such as ambient noise levels and reverberation conditions. For the application at hand, this is very problematic since the SNR values of initially detected signals will be very low and multipath conditions may be present if operating at low altitudes.

The Steered Response Power (SRP) is robust, simplistic in nature, and often performs considerably better than TDOA methods in adverse environments (Dmochowski et al. 2007). In brief, the method forms a directional “beam” which is scanned over a region of interest while calculating the power output at each location. The source location is chosen based on the direction which maximizes this output power. For the case of the Filter-and-Sum (FAS) beamformer, the SRP will be given by:

$$\text{SRP}[Y(\hat{k}, f)] = \sum_{f=f_a}^{f_b} |Y(\hat{k}, f)|^2 = \sum_{f=f_a}^{f_b} \left| \sum_{s=1}^S W_s(f) \cdot X_s(f) \cdot e^{-j2\pi f \tau_s(\hat{k})} \right|^2 \quad (6)$$

where S is the total number of acquired microphone signals, and $[f_a, f_b]$ defines the discrete frequency range of interest. The angular location estimate can then be obtained by finding the direction which maximizes the output power:

$$\langle \hat{\theta}, \hat{\phi} \rangle = \arg \max \left(\text{SRP}[Y(\hat{k}, f)] \right) \quad (7)$$

The major benefit to the SRP method is simplicity, robustness, and the ability to use an unconstrained array geometry (DiBiase et al. 2001). The major downside is increased computational load since direct closed-form solutions cannot be achieved to estimate localization values. Instead, grid-search methods are employed to scan a region of interest. If the source location is completely unknown, this may result in unacceptably large computational requirements. For example, if scanning the full spherical region surrounding an aircraft with a 1-degree resolution, a total of $360 \times 180 = 64,800$ evaluation points are required.

2.3 SRP Search Algorithms

Several SRP search algorithms have been proposed to reduce computational loads and enable real-time operation. These can be broadly categorized into three main areas: 1) Regional reduction through TDOA-based candidate location mapping (Lathoud and Magimai-Doss 2005; Dmochowski, et al. 2008, 2007; Salvati et al. 2017; Oualil, et al. 2013), 2) Iterative-based search techniques (Wax and Kailath 1983; Marti et al. 2013), and 3) Regional contraction using coarse-to-fine grid searching (Zotkin and Duraiswami 2004; Do and Silverman 2007; Peterson and Kyriakakis 2005; Nunes et al. 2014), volumetric evaluation (Cobos, et al. 2011; Lima et al. 2015), or stochastic methods (Do, et al. 2007; Do and Silverman 2010).

TDOA-based methods use time delays to determine a region consisting of potential candidate locations. Standard search methods are then applied to the reduced region to accurately determine source locations. Although shown to be effective, the downside to this approach is the possibility of poor initial region specifications due to inaccurate TDOA estimates.

Iterative-based techniques use methods such as the Steepest Decent and Newton-Raphson to obtain localization values. These methods offer the greatest computational savings but typically suffer from the possibility of false convergence due to the presence of local maxima. This is increasingly problematic for low SNR conditions, since the objective function will not have a strong global peak. For such situations, iterative methods may produce inaccurate results which are extremely sensitive to the initial search locations (DiBiase et al. 2001).

Regional contraction methods perform multiple regional searches using successively smaller grid sizes until the desired accuracy is achieved. In general, the method assigns a coarse grid to the spatial region of concern with SRP values being calculated at each point. The grid location producing the maximum power is then identified and subsequently used to establish a new search region which is defined by the neighboring grid points. The grid spacing is then reduced, assigned to the new reduced region, and the maximum power point is again located. The procedure is then repeated until the desired localization resolution is achieved. Figure 2 displayed below provides a visual depiction of the process where $\times$ indicates the maximum SRP point in the region of interest. The method is simplistic in nature and a localization convergence can always be guaranteed. However, because the SRP space will often be composed of many local maxima, there is a direct trade-off between accuracy, computational load, and convergence accuracy. It will be shown however, that such effects can be drastically reduced if the target signal frequency is narrowband and known beforehand.

3. Harmonic Spectral Beamformer

A beamforming method is now proposed which exploits the properties of harmonic signals such as that generated by propeller-driven aircraft to increase spatial localization capabilities. Termed the Harmonic Spectral Beamformer (HSB), the method combines the proposed Complex Harmonic Spectral Transform (CHST) with the standard Filter-and-Sum (FAS) beamformer to construct a processor giving a high degree of directional sensitivity. Direction of arrival is achieved through the SRP approach in conjunction with a criss-cross regional contraction search method which will also be presented.

Harmonic signals such as those generated by propeller-driven aircraft can be considered broadband with a narrowband decomposition structure. Typically, to obtain an optimal response for each harmonic

component, the signal is decomposed into multiple segments corresponding to each harmonic, and then individually processed using some beamforming algorithm (Van Veen and Buckley 1988). Although effective, this method requires a much higher computational load since several independent algorithms are required instead of just one. A potential solution to this problem is to simply transform the signal such that harmonics are now represented by a single fundamental component. This procedure was previously described by Harvey and O'Young via the Harmonic Spectral Transform (HST) defined as (Harvey and O'Young 2018):

$$\bar{H}_a[X(f)] = \left[\frac{1}{R} \sum_{r=1}^R |X(f \cdot r)|^a \right]^{1/a} \quad (8)$$

where $X(f)$ is the relevant frequency spectrum, R is the number of harmonic components considered, and a is an integer value which specifies the statistical mean form. However, the HST given above does not contain phase information, which is required to enable beamforming operations. Thus, the phase preserving Complex Harmonic Spectral Transform (CHST) is now defined according to the following:

$$j\text{-}\bar{H}_a[X(f)] = \left[\frac{1}{R} \sum_{r=1}^R |X(f \cdot r)|^a \right]^{1/a} \cdot e^{j \sum_{r=1}^R \theta(f \cdot r)} \quad (9)$$

where the j in the variable definition $j\text{-}\bar{H}_a[\]$ indicates the complex spectral transform, and the phase function θ is given by:

$$\theta(f \cdot r) = \arg[X(f \cdot r)] \quad (10)$$

Note that the phase transform operation is performed independently of the spectral peak operation. This is to avoid any possible reduction in the transformed spectral peaks, which will occur if harmonic components are out of phase with one another. In addition, the phase forms are summed using the scalar rather than vectorial method, to prevent cancelation of between different harmonics.

By combining equations and (9) (CHST and FAS), we may thus define a new Harmonic Spectral Beamformer (HSB) according to the following:

$$\text{HSB}(\hat{k}, f) = \sum_{s=1}^S \left(W_s(f) \cdot \left[\frac{1}{R} \sum_{r=1}^R |X(f \cdot r)|^a \right]^{1/a} \cdot e^{j \left[\sum_{r=1}^R \theta_s(f \cdot r) - 2\pi \cdot f \cdot \tau_s(\hat{k}) \sum_{r=1}^R (r) \right]} \right) \quad (11)$$

The rationale behind summing phase angles across harmonic components is to provide an increased array response sensitivity. This is achieved because a slight change in steering direction will now produce a much larger change in the phase function, compared to that found via the standard FAS Beamformer given by equation (1). This property is facilitated by the physical nature of the acoustic system whereby the initial phase of each harmonic component is equal at the point of generation. Although the phase values for each harmonic may be different at any given point thereafter, summation will still be maximized when the array is steered in the direction of wave propagation. To avoid spatial aliasing, the element spacing must now adhere to the following condition:

$$d < \frac{\lambda_o}{2 \sum_{r=1}^R r} = \frac{c}{2 \sum_{r=1}^R f_o \cdot r} \quad (12)$$

where λ_o and f_o is the fundamental wavelength and frequency respectively. The above relationship can arrived by considering the standard condition for spatial aliasing as given by equation (5), and the logical operation of summing each harmonic phase in a scalar manner. Since the phase of each harmonic component is combined across time, and frequency is defined as the rate of change in phase with time; the effective frequency of the harmonically transformed wave is given by the sum of each harmonic frequency. For example, if a 100 Hz and 200 Hz signal were combined by the scalar sum of their phase functions, the resulting wave would have an effective frequency of 300 Hz.

Figures 3 and 4 provide normalized directional response plots obtained via the standard FAS and HSB beamformers for a four-element Uniform Linear Array (ULA). Response values are provided over a range of spacing values and signal harmonics for a 100 Hz sinusoid. The Half Power Beam Width (HPBW) and directional gain (D_G) values are also provided in Tables 1 and 2. The precise definition of these performance metrics can be found in (Van Trees 2002). The Standard SRP using the FAS beamformer was calculated across the entire relevant bandwidth such that all harmonic components were included $[f_o, Rf_o]$. However, the SRP using the HSB was only evaluated at the fundamental frequency of the transformed spectrum since all harmonics are already represented by this single component.

From a comparison of the results displayed, it is evident that the HSB offers greater directionality and better overall performance compared to the standard FAS method. HPBW values for the HSB were considerably less for all harmonics and spacing combinations, while directivity gains were also much greater. However, it is also evident that the effect of spatial aliasing is greatly amplified for the HSB as predicted by equation (12) given above. Based on these results, one can conclude that proposed beamformer would be most effective for applications in which low frequency signals are prominent and/or array spacing is limited; such is the case for detecting small propeller-driven aircraft via UAVs. The performance of the method will be later demonstrated using experimental data.

4. Target Localization & Tracking

The following section presents a simplistic source localization method for use with the proposed Harmonic Spectral Beamformer (HSB). Using the concept of regional reduction through coarse-to-fine grid searching, a criss-cross search method is proposed which offers reduced computational loads compared to more standard techniques such as that presented in (Do and Silverman 2007).

As previously discussed, one of the issues associated with most regional reduction and iterative-based methods is the false convergence due to the presence of local maxima and a weak global peak. To combat this issue, initial grid spacing is often chosen conservatively which consequently results in unnecessary computation. At this point, it should be emphasized that all reported methods assume source frequencies are unknown. For such instances, the SRP must be calculated across a frequency band of interest or in some cases the entire signal bandwidth $[0, f_s/2]$. For narrowband signals, this will inherently reduce the effectiveness of the approach, since the addition of random noise components will reduce the dynamic range of the SRP output. Such methods will also produce high computational requirements since the steered response requires calculation for every FFT bin in the frequency band; hence the reason why SRP methods are often considered computationally expensive. For the application at hand however, this is not the case since all possible source frequencies would have been identified in the previous Constant False Alarm Rate (CFAR) detection stage (Harvey and O'Young 2018). By

exploiting this information, a much higher degree of sensitivity can be achieved in addition to great computational savings.

For a detected source signal of frequency f_o , the SRP output will be given by:

$$\text{SRP}[Y(\hat{k}, f)] = \sum_{f=f_o} \left| Y(\hat{k}, f) \right|^2 \quad (13)$$

which requires $f_b - f_a - 1$ less evaluation points, where $[f_a, f_b]$ defines the frequency band of interest. To establish a more robust form, the region closely surrounding f_o may also be evaluated to account for spectral leakage. Thus, if G guard cells taken equally about the detected frequency are also included, the SRP will now be given by:

$$\text{SRP}[Y(\hat{k}, f)] = \sum_{f=f_o-G/2}^{f_o+G/2} \left| Y(\hat{k}, f) \right|^2 \quad (14)$$

For the Harmonic Spectral Beamformer, the final SRP output can be thus obtained by applying equation (14) to the HSB previously given by equation (11):

$$\text{SRP}[\text{HSB}(\hat{k}, f)] = \sum_{f=f_o-G/2}^{f_o+G/2} \left| \sum_{s=1}^S \left(W_s(f) \cdot \left[\frac{1}{R} \sum_{r=1}^R |X(f \cdot r)|^a \right]^{1/a} \cdot e^{j \left[\sum_{r=1}^R \theta_s(f \cdot r) - \sum_{r=1}^R (2\pi \cdot f \cdot r \cdot \tau_s(\hat{k})) \right]} \right) \right|^2 \quad (15)$$

By excluding essentially all noise components contained in the frequency band of interest, the sensitivity of the method will inherently increase, which is especially true for low SNR signals. This is demonstrated in Figures 5 and 6 which display plots of the SRP as a function of angle for two SNR scenarios (-15 and 0 dB). The signal consists of a $f_o = 500$ Hz sinusoid embedded in Gaussian noise sampled at a rate of $f_s = 2000$ Hz and transformed using a 2048-point FFT. It arrives with 45-degree angle of incidence and is acquired by a ULA consisting of 4 sensors spaced 0.25 m apart. Note that -15 dB is the typical lower limit for reliable CFAR detection using the average incoherent power scheme (Harvey 2018). The isolated SRP component (P_1) was calculated using $G = 5$ Hz, while the standard broadband output (P_2) was calculated for the full spectral range $[0, f_s/2]$. From the plots, it is evident that a significant increase in localization sensitivity is obtained for low SNR values. It is also apparent that using an adaptive search approach would not fare well for the broadband case with low SNR values, since the presence of multiple peaks with approximately equal value will greatly increase the probability of false convergence. For the isolated component case however, localization peaks are significant and should lend well to regional reduction and/or iterative search techniques.

With respect to the above considerations, a regional reduction method is now proposed which utilizes a criss-cross search method rather than a uniformly spaced square grid approach. The proposed algorithm is performed as follows: A very coarse uniform 2-dimensional grid is first applied to determine the general maxima region in terms of azimuth and elevation. While holding the elevation angle constant at the median point for the range of interest, SRP values are calculated for the desired range of azimuth values. The azimuth value producing the largest SRP output is then chosen and held constant while SRP values are then calculated for the desired range of elevation angles. The grid region is then chosen about this point, contracted, and the process is repeated. This is depicted below in Figure 7, where $\odot$ indicates the first regional max value (holding φ constant), and $\times$ indicates the final regional max. It is evident that

the approach offers great computational savings since only one row and one column for each grid set is evaluated rather than all points. For example, if performing a three-level coarse-to-fine regional contraction using a 5×5 grid, a total of $5 \times 5 \times 3 = 75$ points would require evaluation. However, if using the criss-cross approach only $5 \times 5 + 2 \times 5 + 2 \times 5 = 45$ points would be required which is a significant reduction.

5. Experimental Results & Discussion

5.1 Experimental Setup

To further evaluate the performance of the proposed localization approach, the method is applied to experimental data involving two fixed-wing UAVs. In brief, the experiment involved flying two aircraft (sensing and intruding) in circuit formation with opposing flight paths to facilitate close mid-air encounters. The aircraft were assigned different altitudes and circuit radii to avoid any actual mid-air collision from taking place. The intruder was given an altitude of 150 m with a circuit radius of approximately 500 m, while the detecting aircraft was given an altitude and circuit radius of 100 m and 200 m respectively. The sensing aircraft consisted of a hobby-grade Delta X-8 powered by a single brushless DC motor. Four DPA 4053 omni-directional microphones were fitted to the aircraft via carbon fiber booms extending from the nose as depicted below in Figure 8. The microphones were secured using custom vibration isolation mounts constructed from Sorbothane material. Although the experiment was originally conducted using a total of four microphones, a malfunctioning external power supply corrupted two of the channels with excess noise rendering them useless. Thus, data processing was performed using only two of the recorded signals. Acoustic data was recorded at a sampling rate of 48 KHz using a portable Zoom H4 handheld recording unit located inside the sensing aircraft and processed offline. Figure 9 provides the array geometry and directivity response plots for the aircraft. The intruding aircraft consisted of a gasoline powered Giant Big Stik as displayed below in Figure 10. Both aircraft were fitted with ArduPilot 2.5 systems for flight control and logging of data such as speed, GPS position, altitude, and orientation. Table 3 lists the approximate sensor accuracy values for the device. To synchronize the recorded GPS data with the recorded acoustic data, the sensing aircraft was given a physical impulse (tap) prior to take off. This impulse is visible in both the recorded audio signal, and aircraft accelerometer signal, which is also synchronized with the recorded GPS signal. Signals were decimated to 2400 Hz to reduce computational requirements and transformed to the frequency domain using a 6000 pt padded FFT with a rectangular window and 50 % overlap producing a bin resolution of 0.5 Hz/bin. Further details regarding the experimental setup can be found in (Harvey and O'Young 2018).

5.4 Experimental Results

Table 4 displayed below provides the localization results obtained for the standard SRP and proposed HSB-SRP as previously given by equations (6) and (15) respectively. Figure 11 also provides plots of the two methods in which the general accuracy and detection rate associated with the two methods can be observed. These results consist of the mean and standard deviation of the angular position error at each detection point. Error values were calculated by comparing the beamforming results to aircraft GPS data provided via the two autopilot systems. For each localization method, the SRP

is calculated using both the entire noise sample band and that pertaining only to the detected frequency bin including guard cells ($f_o \pm 10$ Hz). Azimuth values were then determined via the regional contraction method previously presented. A total of six harmonics ($H = 6$) were used to enhance localization sensitivity.

The power band used for the two localization methods are different, since false alarm rates for the CFAR used is dependent on the signal bandwidth and whether any transformations such as the HST are applied. As previously mentioned, details regarding the CFAR algorithm and detection results obtained using the presented data set are currently available elsewhere and will thus not be covered here (Harvey and O'Young 2018). However, a brief summary of the detection results obtained are provided below in Table 4 for the readers convenience.

Table 5 provides the parameters values used for the regional contraction search. A total of five contractions/reductions were used following the initial spatial scan. For subsequent detections, the location from the previous evaluation point (window) is utilized as the contraction starting point instead of re-scanning the entire domain space (360°) again. From the values displayed, it is evident that the method is much more efficient than that of the single stage approach since only 35 evaluation points are required to achieve a resolution of 0.5° . In contrast, a single stage approach would require $360/0.5 = 720$ points to achieve the same resolution. Since the array effectively consisted of only two microphones (two channels were corrupted), localization of the intruding aircraft was only possible in one angular dimension (azimuth). In addition, because the two microphones are also omni-directional, the array cannot discriminate between the forward or equivalent rear positions as depicted below in Figure 12. Thus, such a setup is not truly practical for an aircraft collision avoidance system. Localization results are therefore examined simply to demonstrate the use of the Harmonic Spectral Beamformer.

Based on the results obtained, it is evident that the proposed HSB increased localization accuracy by approximately 50% compared to the standard SRP. This was expected since the method offers a much greater directional performance as previously demonstrated. This can also be observed by the directivity response plots displayed in Figure 13 for the two-element array acquiring a 135 Hz signal (typical operating frequency for the intruding aircraft). Had four microphone signals been available for processing as originally intended, it is expected the accuracy difference would have been further increased due to increased SNR values and enhanced directivity gain.

It is also evident that evaluating only the detected frequency and surrounding guard cells ($f_o \pm 10$ Hz) provides a slightly increased localization accuracy compared to evaluating the entire frequency band (10 to 492.5 Hz). Although the increase is only small, the result is significant since it indicates that the full frequency band of interest does not need to be evaluated, but only the region around the detected signal frequency instead. Such an approach offers great computational savings which is important since the SRP method is inherently computationally expensive compared to other techniques as previously discussed.

Overall, localization results do not appear particularly accurate, especially for the standard SRP case. Unfortunately, little data is available in the literature to form a direct comparison involving acoustic detection via UAVs. Of the relevant studies previously discussed in Section 1, none clearly present localization results with numerical azimuth and/or elevation error values. Reported values obtained from general experimental data using SRP methods are typically less than 5° (Zotkin and Duraiswami 2004; Dmochowski, et al. 2007; Oualil et al. 2013). However, localization accuracy is highly dependent on SNR values which are also much lower than that often reported in the literature. In addition, most experimental instances use much larger arrays typically on the order of 5 to 10 elements. It should also be noted that the flight data obtained from each aircraft has heading and positional uncertainties of

approximately 5° and 3 m respectively. For the HSB case, these values are significant since the heading uncertainty alone represents almost 50% of the total perceived localization error (source angular position is calculated relative to the detecting aircraft orientation). Thus, when considering these facts, it appears the results obtained are relatively good for the scenario at hand; at least for the HSB case. One would expect that an avoidance maneuver can be reliably performed under these circumstances by simply changing course such that the target source is now located at $\pm 90^\circ$ (head perpendicularly away from target flightpath). However, determination of the minimum allowable localization error for a given kinematic setup is complex and thus outside the scope of this paper.

5.4 Additional Considerations

As previously mentioned, data was processed offline using a desktop computer. However, to establish an effective collision avoidance system, or any aircraft detection system for that matter, algorithms must be capable of operating in real-time with low latency. Whether a particular algorithm can operate in real-time is ultimately dependent on the hardware used. For the data presented in this paper, real-time operation was simulated and achieved very easily via MATLAB using a 3.5 GHz computer with an i5 Intel core processor. As previously mentioned in Section 2.3, the SRP beamforming method has been successfully implemented in real-time on various hardware platforms. When comparing the standard SRP and proposed SRP-HSB, the number of calculations required to determine the signal power at specific spatial location are approximately equal if using the full available signal bandwidth. This is because the Complex Harmonic Spectral Transform (CHST) requires R complex summations for each frequency bin, but the resulting signal bandwidth is also reduced by a factor of R . For example, data for the presented experiment was acquired at 48 KHz and decimated to 4800 Hz, giving a signal bandwidth of 2400 Hz in the spectral domain. Since a total of $R = 6$ harmonics were considered, the resulting spectrum for the SRP-HSB case is now only 400 Hz.

The main challenge associated with using the SRP method in general is regarding the particular search algorithm used. For the proposed criss-cross search method, only $12 + 6 + 5 + 4 + 4 + 4 = 35$ SRP-HSB function evaluations are required to achieve localization upon initial detection. After detection is achieved, each tracking point would then only require $5 + 4 + 4 + 4 = 17$ evaluations if scanning an azimuth region 20° degrees about the last localization point. In contrast, using a global scan approach with a 0.5° degree azimuth resolution would require $360 / 0.5 = 720$ function evaluations upon initial detection, and $20 / 0.5 = 40$ evaluations for each subsequent tracking point.

Another aspect which may be of interest to the reader but not directly evaluated in this paper, is the ability to detect, localize, and track multiple targets. In theory, any number of targets can be detected using the distribution-free CFAR method presented in (Harvey and O'Young 2018), provided the targets do not coexist at the same frequency. Since target signals are assumed to be of unknown frequency, all possible values are tested regardless of the number of potential targets. Thus, the inclusion of multiple targets will not introduce additional computations in the detection stage. In the localization and tracking stage however, each target must be treated separately making the number of additional computations directly proportional to the number of additional targets. For the cause of multiple intruders, the proposed SRP-HSB should attain additional increased performance over the standard SRP method due to its high directivity response as previously depicted in Figure 14. However, further experimental testing and/or simulation is required to verify this.

Finally, it was mentioned at the beginning of this paper that the presented work is an extension to that previously presented by Harvey and O'Young to evaluate the viability of acoustic sensing to establish a non-cooperative aircraft collision avoidance system (Harvey and O'Young 2015; Harvey and

O'Young 2018). However, the development of such a system not only requires the successful detection, localization, and tracking of target aircraft; but also, the ability to approximate its trajectory in order to perform an appropriate avoidance maneuver. The ability to do so from a moving aircraft is very difficult since the dynamic nature of the system does not allow traditional methods which primarily use doppler shifting in combination with a closest point of approach (CPA) to determine values. It will be shown in later works, that trajectory mapping is possible by applying an iterative least squares approach to a system of equations which model acquired target signal power, perceived doppler frequency shift, and general 3D kinematics.

6. Conclusions

Based on the theoretical analysis and experimental results presented, it can be concluded that the proposed Harmonic Spectral Beamformer (HSB) is an effective means to increase direction of arrival accuracy for harmonic signals such as those generated by fixed-wing UAVs. Compared to the standard Steered Response Power, localization accuracy was increased by 50% using the proposed method. In addition, overall localization results may also be considered accurate when accounting for the measurement uncertainty associated with the data recording system (autopilot) and low number of microphones used. Had more signals been available for processing, it is expected that results would have been much more accurate. In addition to the proposed beamformer, the proposed regional contraction search method was also found to produce significant computational saving compared to a standard full grid search approach.

Although the primary purpose of the presented work is directed towards the establishment of a non-corporative aircraft collision avoidance system, developments made may form the foundation for other counter UAV operations. For example, the detection, localization, and tracking of UAVs operating near undesired areas such as airports. Such scenarios are becoming increasingly popular today and constitute a very high safety risk to the general public. Since UAVs are often very small and fly at relatively low altitudes, standard airport detection systems such as radar are not capable of efficiently or effectively offering detection and/or classification in this regard. In theory, a network of microphones could be located around the boundary of such a space to detect the presence of an intruding UAV. The microphones may act together to effectively create a giant array operating in real-time, using the currently proposed and previously presented methods to detect, localize, and track any intruder. Such a system would be extremely useful since it can operate in all weather conditions and would be very low cost compared to other technologies such as distributed optical or micro-radar systems.

7. References

- Brandstein, M.S., and Silverman, H.F. 1997. A practical methodology for speech source localization with microphone arrays. *Computer Speech and Language*. 11:91-126.
- Case, E. E., Zelnio, A. M., and Rigling, B.D. 2008. Low-cost acoustic array for small UAV detection and tracking. *IEEE national aerospace and electronics conference*. p. 110.
- Chen, J. C., Yao, K., and Hudson, R. E. 2002. Source localization and beamforming. *IEEE Signal Processing Magazine*. 19(2):30-9.

- Chen, J., Benesty, J., and Huang, Y.A. 2006. Time delay estimation in room acoustic environments: An overview. *EURASIP Journal on Applied Signal Processing*, pp. 1-19.
- Cobos, M., Marti, A., and Lopez, J. J. 2011. A modified SRP-PHAT functional for robust real-time sound source localization with scalable spatial sampling. *IEEE Signal Processing Letters*. 18(1): 71-4.
- DiBiase, J.H., Silverman, H.F., and Brandstein, M.S. 2001. *Microphone arrays - signal processing techniques and applications*. Springer-Verlag.
- Dmochowski, J. P., Benesty, J., and Affes, S. 2007. A generalized steered response power method for computationally viable source localization. *IEEE Transactions on Audio, Speech, and Language Processing*. 15(8):2510-2526.
- Dmochowski, J., Benesty, J., and Affes, S. 2008. Fast steered response power source localization using inverse mapping of relative delays. *IEEE international conference on acoustics, speech and signal processing*, p. 289.
- Dmochowski, J., Benesty, J., and Affes, S. 2007. Direction of arrival estimation using the parameterized spatial correlation matrix. *IEEE Transactions on Audio, Speech, and Language Processing*. 15(4):1327-39.
- Do, H., and Silverman, H. F. 2007. A fast microphone array SRP-PHAT source location implementation using coarse-to-fine region contraction (CFRC). *IEEE workshop on applications of signal processing to audio and acoustics*. p. 295.
- Do, H., and Silverman, H. F. 2010. SRP-PHAT methods of locating simultaneous multiple talkers using a frame of microphone array data. *IEEE international conference on acoustics, speech and signal processing*. p. 125.
- Do, H., Silverman, H. F., and Yu, Y. 2007. A real-time SRP-PHAT source location implementation using stochastic region contraction (SRC) on a large-aperture microphone array. *IEEE international conference on acoustics, speech and signal processing - ICASSP '07*. p. I-121
- Ferguson, B., and Wyber, R. 2006. Detection and localization of a ground based impulsive sound source using acoustic sensors onboard a tactical unmanned aerial vehicle. *Battlefield acoustic sensing for ISR applications*. Neuilly-sur-Seine, France. p. 16-1.
- Godara, L. C. 1997. Application of antenna arrays to mobile communications. II. beam-forming and direction-of-arrival considerations. *Proceedings of the IEEE*. 85(8):1195-245.
- Harvey, B. 2018. Signal processing methods for the detection & localization of acoustic sources via unmanned aerial vehicles (Doctoral dissertation). Memorial University of Newfoundland, St. John's, Newfoundland.
- Harvey, B., and O'Young, S. 2015. Detection of continuous ground-based acoustic sources via unmanned aerial vehicles. *Journal of Unmanned Vehicle Systems*. 4(2):83.

- Harvey, B. and O'Young, S. 2018. Acoustic detection of a fixed-wing UAV. *Drones*. 2(1):4.
- Hero, Messer, H., Goldberg, J., Thomson, D. J., Amin, M. G., Giannakis, G., Swami, A., Tugnait, J. K., Nehorai, A., Swindlehurst, A. L., et al. 1998. Highlights of statistical signal and array processing. *IEEE Signal Processing Magazine*. 15(5):21-64.
- Krim, H., and Viberg, M. 1996. Two decades of array signal processing research: The parametric approach. *IEEE Signal Processing Magazine*. 13(4):67-94.
- Krishnaveni, V., Kesavamurthy, T., and Aparna, B. 2013. Beamforming for direction-of-arrival (DOA) estimation-A survey. *International Journal of Computer Applications*. 61(11):4.
- Lathoud, G., and Magimai-Doss, M. 2005. A sector-based, frequency-domain approach to detection and localization of multiple speakers. *Proceedings. (ICASSP '05). IEEE international conference on acoustics, speech, and signal processing*. p. 265.
- Li, F., and Vaccaro, R. J. 1992. Sensitivity analysis of DOA estimation algorithms to sensor errors. *IEEE Transactions on Aerospace and Electronic Systems*. 28(3):708-17.
- Lima, M. V. S., Martins, W. A., Nunes, L. O., Biscainho, L. W. P., Ferreira, T. N., Costa, M. V. M., and Lee, B. 2015. A volumetric SRP with refinement step for sound source localization. *IEEE Signal Processing Letters*. 22(8):1098-1102.
- Marti, A., Cobos, M., Lopez, J.J., and Escolano, J. 2013. A steered response power iterative method for high-accuracy acoustic source localization. *Journal of the Acoustical Society of America*. 134(4):2627.
- McCowan, I.A. 2001. Robust speech recognition using microphone arrays (Doctoral dissertation). Queensland University of Technology, Australia.
- Nielsen, R.O. 2009. Acoustic detection of low flying aircraft. *IEEE conference on technologies for homeland security*. Boston, MA. pp. 101-106.
- Nunes, L. O., Martins, W. A., Lima, M. V. S., Biscainho, L. W. P., Costa, M. V. M., Gonçalves, F. M., Said A., and Lee, B. 2014. A steered-response power algorithm employing hierarchical search for acoustic source localization using microphone arrays. *IEEE Transactions on Signal Processing*. 62(19):5171-83.
- Ohata, T., Nakamura, K., Mizumoto, T., Taiki, T., and Nakadai, K. 2014. Improvement in outdoor sound source detection using a quadrotor-embedded microphone array. *IEEE/RSJ international conference on intelligent robots and systems (IROS)*. p. 1902.
- Okutani, K., Yoshida, T., Nakamura, K., and Nakadai, K. 2012. Outdoor auditory scene analysis using a moving microphone array embedded in a quadcopter. *IEEE/RSJ international conference on intelligent robots and systems (IROS)*. Vilamoura, Portugal. p. 3288.
- Oualil, Y., Faubel, F., and Klakow, D. 2013. A fast cumulative steered response power for multiple speaker detection and localization. *21st european signal processing conference (EUSIPCO 2013)*. p. 1.

- Peterson, J. M., and Kyriakakis, C. 2005. Analysis of fast localization algorithms for acoustical environments. Conference record of the thirty-ninth asilomar conference on signals, systems and computers. p. 1385.
- Robertson, D.N., Pham, T., Edge, H., Porter, B., Shumaker, J., and Cline, D. 2007. Acoustic sensing from small-size UAVs. Proc. SPIE 6562 unattended ground, sea, and air sensor technologies and applications IX. Orlando, USA. p. 656208
- Salloum, H., Sedunov, A., Sedunov, N., Sutin, A., and Masters, D. 2015. Acoustic system for low flying aircraft detection. IEEE international symposium on technologies for homeland security (HST). p. 1.
- Salvati, D., Drioli, C., and Foresti, G.L. 2017. Exploiting a geometrically sampled grid in the steered response power algorithm for localization improvement. Journal of the Acoustical Society of America . 141(1):586-601.
- Stevens, K.N. 1983. Acoustic properties used for the identification of speech sounds. Annals of the New York Academy of Sciences. 405(1):2-17.
- Sutin, A., Salloum, H., Sedunov A., and Sedunov, N. 2013. Acoustic detection, tracking and classification of low flying aircraft. IEEE international conference on technologies for homeland security (HST). p. 141.
- Van Trees, H. 2002. Optimum array processing. New York: Wiley.
- Van Veen, B. D., and Buckley, K. M. 1988. Beamforming: A versatile approach to spatial filtering. IEEE ASSP Magazine. 5(2):4-24.
- Wax, M., and Kailath, T. 1983. Optimum localization of multiple sources by passive arrays. IEEE Transactions on Acoustics, Speech, and Signal Processing. 31(5):1210-7.
- Zelnio, A. M., Case, E. E., and Rigling, B. D. 2009. A low-cost acoustic array for detecting and tracking small RC aircraft. IEEE 13th digital signal processing workshop and 5th IEEE signal processing education workshop. p. 121.
- Zelnio, A.M. 2009. Detection of small aircraft using an acoustic array (Master's thesis). Wright State University: Fairborn, OH, USA.
- Zotkin, D. N., and Duraiswami, R. 2004. Accelerated speech source localization via a hierarchical search of steered response power. IEEE Transactions on Speech and Audio Processing. 12(5):499-508.

Figure Captions

Figure 1: Depiction of localization scenario.

Figure 2: Illustration of coarse-to-fine grid search method.

Figure 3: Directivity response of a 0.1 m spaced ULA acquiring a 100 Hz signal with R harmonic components.

Figure 4: Directivity response of a 0.2 m spaced ULA acquiring a 100 Hz signal with R harmonic components.

Figure 5: Power spectrum and SRP output for -15 dB signal, where P_1 and P_2 indicate the isolated and broadband SRP response cases respectively.

Figure 6: Power spectrum and SRP output for 0 dB signal, where P_1 and P_2 indicate the isolated and broadband SRP response cases respectively.

Figure 7: Criss-cross regional reduction grid search.

Figure 8: Delta X-8 fitted with four DPA 4053 microphones (Harvey and O'Young 2018)

Figure 9: Delta X-8 array geometry and directional response.

Figure 10: Gasoline powered Giant Big Stik (intruder aircraft)

Figure 11: Plots of the proposed SRP-HSB and Standard SRP azimuth results.

Figure 12: Depiction of spatial ambiguity for a two-element array.

Figure 13: Directivity response for the two-element array.

Figure 1: Depiction of localization scenario.

Figure 2: Illustration of coarse-to-fine grid search method.

Figure 3: Directivity response of a 0.1 m spaced ULA acquiring a 100 Hz signal with R harmonic components.

Figure 4: Directivity response of a 0.2 m spaced ULA acquiring a 100 Hz signal with R harmonic components.

Figure 5: Power spectrum and SRP output for -15 dB signal, where P_1 and P_2 indicate the isolated and broadband SRP response cases respectively.

Figure 6: Power spectrum and SRP output for 0 dB signal, where P_1 and P_2 indicate the isolated and broadband SRP response cases respectively.

Figure 7: Criss-cross regional reduction grid search.

Figure 8: Delta X-8 fitted with four DPA 4053 microphones (Harvey and O’Young 2018).

Figure 9: Delta X-8 array geometry and directional response.

Figure 10: Gasoline powered Giant Big Stik (intruder aircraft) (Harvey and O'Young 2018).

Figure 11: Plots of the proposed SRP-HSB and Standard SRP azimuth results.

Figure 12: Depiction of spatial ambiguity for a two-element array.

Figure 13: Directivity response for the two-element array.

Table 1: Half Power Beam Width (3 dB) in degrees for various sensor spacing values (in meters).

<i>R</i>	Standard SRP					HSB SRP				
	<i>S</i> = 0.1	<i>S</i> = 0.2	<i>S</i> = 0.3	<i>S</i> = 0.4	<i>S</i> = 0.5	<i>S</i> = 0.1	<i>S</i> = 0.2	<i>S</i> = 0.3	<i>S</i> = 0.4	<i>S</i> = 0.5
1	N/A	N/A	N/A	154	102	N/A	N/A	N/A	154	102
2	N/A	N/A	122	82	62	N/A	80	51	38	30
3	N/A	160	82	58	36	82	38	25	19	15
4	N/A	104	62	46	26	46	21	15	11	5

Table 2: Directivity Gain (*D_G*) in dB for various sensor spacing values (in meters).

<i>R</i>	Standard SRP					HSB SRP				
	<i>S</i> = 0.1	<i>S</i> = 0.2	<i>S</i> = 0.3	<i>S</i> = 0.4	<i>S</i> = 0.5	<i>S</i> = 0.1	<i>S</i> = 0.2	<i>S</i> = 0.3	<i>S</i> = 0.4	<i>S</i> = 0.5
1	1.0	1.1	1.2	1.4	1.6	1.0	1.1	1.2	1.4	1.6
2	1.1	1.2	1.5	1.9	2.3	1.2	2.0	3.3	4.2	5.0
3	1.1	1.4	1.8	2.3	2.8	2.0	4.2	6.3	7.2	5.6
4	1.2	1.6	2.1	2.7	3.2	3.7	6.6	5.6	2.8	4.7

Table 3: ArduPilot sensor accuracy values

Data Type	Sensor Accuracy
GPS Position	± 1.5 m
Compass Orientation (3D)	± 2.5 °
Altitude	± 5 m
Airspeed	± 1 m/s

Table 4: Azimuth localization results.

SRP			HSB-SRB	
Power Band	10-492.5 Hz	<i>f_o</i> ±10 Hz	10-400 Hz	<i>f_o</i> ±10 Hz
Mean Error <i>ϑ</i>	24.5°	23.5°	13.0°	11.0°
STD Error <i>ϑ</i>	13.7°	13.2°	9.7°	8.8°
Detection Rate	32 %		56 %	
False Alarm Rate	1.5E-5		1.5E-5	
Mean Signal Loss Time	8.2 s		3.5 s	

Table 5: Regional contraction search parameters.

1st Stage (Initial Scan)		3rd Contraction	
Range	360°	Range	$\pm 4^\circ$
Points	12	Points	4
Spacing	30°	Spacing	2°
1st Contraction		4th Contraction	
Range	$\pm 30^\circ$	Range	$\pm 2^\circ$
Points	6	Points	4
Spacing	10°	Spacing	1°
2nd Contraction		5th Contraction	
Range	$\pm 10^\circ$	Range	$\pm 1^\circ$
Points	5	Points	4
Spacing	4°	Spacing	0.5°

Draft