Target Detection and Classification From Sub-optimal Experimental Data Using Principal Component Analysis and Feature Vector Masking

Everett R. Coots, William W. Arrasmith, Eric A. Skowbo, John V. Olson, and Barry Webster

***Abstract*—Infrasound signals of interest (SOI) have been collected from various sources including boats, aircraft and personal watercraft during a recent field exercise. The designed experiment specifically located infrasound sensors to gather the SOIs and process them using a developmental methodology. This custom pre-processing, and exploitation suite allows for the extraction of SOIs from experimental data sets that might be sub-optimal due to the presence of noise or interfering (undesired) sources. The on-going development of the infrasound sensor suite and the SOI signal processing and analysis is supported by this work.**

***Index Terms*—Feature vector, infrasound, infrasound analysis, infrasound array, infrasound classification, principal component analysis, PCA.**

I. Introduction

The gathering and processing of infrasound signals of interest (SOI) continues to evolve as increasingly more capable sensors are deployed and more sophisticated post processing algorithms are developed [1]-[3]. The scope of infrasound analysis continues to grow as well. While once highly focused on environmental sources such as earthquakes, volcanic eruptions, and bolides, the science of infrasound research continues to expand to include ever-more man-made sources such as boats, missile launches, fixed and rotary-wing aircraft, ground vehicles, machinery, and so forth [4]-[8].

While naturally occurring infrasound SOIs are continuously being gathered and analyzed, the advancement of the state of the art in processing man-made SOIs offers a wide range of opportunities. Improved algorithms for the detection, characterization and/or classification phases of the analysis process could all aid in the advancement of the field. This paper focuses on the detection and characterization phases and makes use of empirical test data gathered during a specially designed field experiment using a combination of seismic and infrasound sensors to detect a variety of man-made SOIs from a waterborne source. In conjunction with the field experiment, post-processing routines were developed to facilitate exploitation of signals of interest. The analytical methods can be validated by comparing the results with the known ground truth data of the field exercise. This paper presents a brief summary of the field measurements, the corresponding data processing algorithm and an analysis of the results of the comparison of the processed data with the ground truth.

Manuscript received March 1, 2016; revised April 30, 2016.

William. W. Arrasmith is with the Department of Engineering Systems, Florida Institute of Technology, Melbourne, FL. 32901 USA (e-mail: warrasmi@fit.edu)

Everett. R. Coots is with the Harris Corporation, Government Communication Systems Division, 1025 W. NASA Blvd. Melbourne, FL. 32904 USA (e-mail: ecoots@harris.com).

Eric. A. Skowbo is with the Northrop Grumman Corporation, Research and Technology and Military Aircraft Systems, 2000 W. NASA Blvd. Melbourne, 32904 USA (e-mail: eric.skowbo@ngc.com).

John. V. Olson is with the Geophysical Institute, University of Alaska Fairbanks, (e-mail: jvo@gi.alaska.edu).

Barry Webster is with the Department of Engineering Systems, Florida Institute of Technology, Melbourne, FL. 32901 USA (e-mail: bwebster@fit.edu).

II. Infrasound Signal Analysis

A. Overview

A field experiment was conducted in December 2010 at an ocean-side pier in Key West Florida, USA [5]. A multi-channel-sensor array was deployed for the purpose of gathering infrasound and seismic data on a variety of man-made SOIs. The collection of extensive signal information both in the form of different sources (e.g. boats, aircraft, personal watercraft) as well as varying behavioral aspects (e.g. inbound, outbound, accelerating, idling) contributed to a growing library of data being developed as part of our teams on-going research effort. Each collection represented a particular set of controlled (source and aspect) and uncontrolled (undesired signals, system noise) signal inputs which were recorded as individual experimental trials. For the subject experiment, the particular signals of interest were emanating from a pair of boats provided by the United States Coast Guard. See Fig. 1 below.

Test Boat for Experimental Signal Gathering

Fig. 1. Experimental test boat.

Once an infrasound or seismic signal of interest has been detected by our sensor suite, the actual processing of the data can begin. Pre-processing actions such as band-pass filtering may be done but is outside the scope of this paper and is not covered here. This paper will focus on the steps of the data processing flow that begin with detection and classification. For example, a sensor may detect that some infrasound signature has occurred. The location algorithm may use a

DOI: 10.7763/IJMO.2016.V6.514

method such as back azimuth projection to determine a general direction of origin and distance to the source that may be hundreds of miles over the horizon. Next, the task is to determine what caused the received sound. Here the goal is to classify the source as being perhaps a natural event (such as a volcano), or possibly an aircraft or watercraft. Then we want to further specify the source by identifying the type of aircraft (e.g. fixed wing, rotary wing, drone, jet) or the type of watercraft. Then, with even more precise analysis we desire to know information about the attitude or aspect of the source. Is the detected aircraft taking off or landing? Is the detected ship approaching or moving away? The data flow diagram for a typical infrasound signal processing system is shown in Fig. 2.

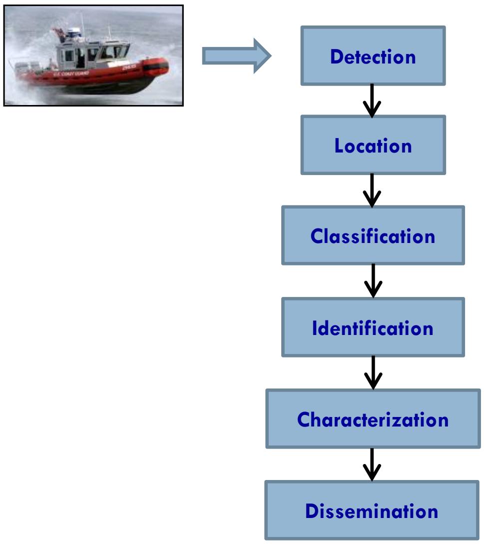


Fig. 2. Infrasound data flow diagram.

B. Classification Methodology

In order to classify a particular signal of interest, the signal must be compared to some library of known signals (or signatures) and some measure of "goodness of fit" must be established. There are numerous, well established methods available for both the signal comparison phase as well as evaluating the quality of the match [9],[10]. In this paper we will focus on the comparison of known signals and in particular, we will present a methodology for establishing a match when less than ideal experimental conditions exist.

A basic assumption when evaluating a signal of interest against a library of known signals, is that the known signals have been validated, or have some degree of trustworthiness. But where did those signals come from and how were they validated? In an ideal case the signals would be collected in a controlled experimental environment where external, interfering signal sources were eliminated or reduced to the extent possible. This would provide the cleanest (in a signal-to-noise ratio sense) signal that would then serve as the reference or library signal henceforth. But what if the only option was to attempt to create a reference signal from a less than ideal source? In this paper we will present a method for extracting key characteristics from a measured signal containing clutter and directly interfering signals and use previously collected experimental data as a test case.

The objective is to extract from the signal of interest a signature that uniquely describes the particular signal source for which the library signal is to be created. We wish to create these custom "feature vectors" in such a way that allows the derivation directly from raw measured data while simultaneously maximizing the rejection of any extraneous signals (ie: reject any content that is not due to the actual signal source). To assist us with this task we can employ the well documented data analysis method known as Principal Component Analysis (PCA) [11]. The method is particularly useful when analyzing data that is expected to be redundant (possessing duplicate data content). In this situation the size (dimension) of the data can be reduced and a simpler/smaller data set can be used to represent the signal.

Let us assume that at each point in time during the acquisition of a SOI, a measurement is logged from each of the n sensors (channels) in the network. This produces a single column vector having n rows.

$$\vec{X} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \tag{1}$$

As this is repeated across an extended acquisition of m samples, a matrix of data is produced. The first component of the Principal Component Analysis is an mxn matrix where m is the number of measurements or observations and n is the number of measurement types or channels. Each row of the data matrix is a native basis vector. We define this input data matrix X as

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix} \tag{2}$$

As our goal is to reduce and simplify our data set, we wish to find a matrix that is a linear combination of the original, native basis vectors. Given our input data matrix X, we define a desired output matrix Y and a transformation vector A. We can write the linear transformation equation

$$AX = Y \tag{3}$$

Or, in expanded general matrix form:

$$\begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} \begin{bmatrix} x_1 & \cdots & x_n \\ \vdots & \ddots & \vdots \\ x_m & \cdots & x_{mn} \end{bmatrix} = \begin{bmatrix} a_1 x_1 & \cdots & a_1 x_n \\ \vdots & \ddots & \vdots \\ a_m x_1 & \cdots & a_m x_n \end{bmatrix} \tag{4}$$

The components of vector A are actually the coefficients for the new basis vectors which allow us to execute the linear transformation from X to Y.

Now that we have a set of new bases vectors with which to define our original data, how do we select the proper subset that will allow us to redefine a new, reduced dimension set of data? Since we are comparing multiple measurement samples, and we postulate that some of the samples provide redundant data, we can calculate the covariance between pairs of measurements and then sort them to find the highest variances. In a Principal Component Analysis the coefficients with the largest variance represent the largest contributions to the signal of interest. Alternatively, measurement pairs producing low variance do not differ significantly between the data sets and are said to be redundant, and are candidates for removal. The specific method for determining precisely how many of these "principal" components to retain has been the subject of numerous publications [12]-[15].

Given two sets of data that could represent our experiment for *n* observations for two trials a and b,

$$a = \{a_1 \quad a_2 \quad \cdots \quad a_n\}\,,\; b = \{b_1 \quad b_2 \quad \cdots \quad b_n\} \tag{5}$$

the variance of the individual sets of data is calculated as shown below (set *a* for example),

$$var(a) = \sigma_a^2 = \frac{1}{n-1}\sum_{i=1}^{n}|a_i - \mu|^2 \tag{6}$$

where μ is the mean of the vector a.

If the data sets *a* and *b* are expressed as row vectors, the covariance in matrix form can then be calculated as

$$\sigma^2 = \frac{1}{n-1} ab^T \tag{7}$$

The resulting square matrix will have the variance of the two vectors along the main diagonal and the off-diagonal terms represent the covariance of all other pairwise combinations. As indicated earlier, the variances identified along the diagonal indicate the principal components. The covariance matrix is typically diagonalized to rank the components for ease of interpretation. The elements closer to the upper left corner of the ranked covariance matrix indicate the pairwise elements of the data set that have a high variance, while those towards the lower right corner of the matrix demonstrate lower variance.

The Single Value Decomposition (SVD) method is often used to calculate principal components as it is a general case of the PCA algorithm. The general form of an SVD problem begins with an eigenvector definition which can be written as

$$X\vec{v} = \lambda\vec{u} \tag{8}$$

which states that the square matrix X can be expressed in terms of the eigenvectors u and v, and the eigenvalue λ which are characteristic roots of the linear system of equations associated with the matrix X. This form is valid for representing a single trial of a given experiment. The efficiency of linear algebra allows us to write the matrix form of this equation that accounts for all trials in the data set simultaneously as shown below

$$XV = \Sigma U \tag{9}$$

SVD makes use of a linear algebra theorem that states that a rectangular matrix X, can be expressed in terms of an orthogonal matrix U, a diagonal matrix Σ, and the transpose of an orthogonal matrix V. The matrix V contains the orthonormal eigenvectors for the matrix X^TX, and the matrix U contains the orthonormal eigenvectors for the matrix XX^T.

Consider an experiment having *n* observations and *m* trials expressed in the form of Equation 9 above. Matrices U and V can be written as,

$$U = \begin{bmatrix} u_1 & \cdots & u_n \\ \vdots & \ddots & \vdots \\ u_m & \cdots & u_{mn} \end{bmatrix}, \quad and\ V = \begin{bmatrix} v_1 & \cdots & v_n \\ \vdots & \ddots & \vdots \\ v_m & \cdots & v_{mn} \end{bmatrix} \tag{10}$$

And the matrix Σ is a symmetric matrix with the rank-ordered singular values along the main diagonal.

Multiplying both sides of Equation 9 above by V^T produces the final form of the SVD shown below

$$X = U\Sigma V^T \tag{11}$$

Referring back to the earlier discussion on principal component analysis, our intention remains to establish a change of basis that allows us to restate the matrix X in a reduced dimension form by identifying the principal components of the SOI. If the data set X can be expressed in terms of the singular values in matrix Σ, a modified (alternate basis) version of the data set X can be calculated by determining a modified version of Σ. This modified matrix contains the largest (most significant in terms of variance) singular values along the main diagonal (after deciding how many to retain) while populating the rest of the matrix with zeros.

$$X_{mod} = U\Sigma_{mod}V^T \tag{12}$$

This modified version of the original data set X_{mod} is calculated by multiplying the X data matrix by the reduced/modified singular values matrix Σ_{mod}. This operation achieves one of the key benefits of principal component analysis, dimension reduction. Just as importantly, although perhaps less obviously, the elements that remain in the modified data set are those that contribute the highest variance to the data set and thus are most critical to representing the data in a reduced dimensional space. Meanwhile, the elements that have the lowest variance have been removed from the data set, as their contribution to the data set is less significant or even redundant.

III. Infrasound Field Exercise

A. Overview

In December 2010 a network of infrasound and seismic sensors was deployed to an ocean-side pier in Key West Florida in a field experiment conducted in collaboration with the U.S. Coast Guard (USCG). The focal point of the experiment was a 33 foot SPC-LE Fast Interceptor boat provided and operated by the USCG (Fig. 1). In addition, due to the public location of the test, operators were able to gather numerous infrasound and seismic signatures for a host of "targets-of-opportunity" (e.g., aircraft, pleasure boats, personal watercraft, land vehicles).

B. Sensor Configuration

The experimental sensor network consisted of a combination of seismic and infrasound sensors (4 infrasound sensors and 2 seismic sensors). While both seismic and infrasound data was collected during the experiment, only the infrasound data is used during the analysis presented here. Fig. 3 below shows a photograph of the actual sensor package as it was deployed for the field experiment.

C. Data Processing Methodology

As indicated above, the subject boat made several circuits of the experimental course to support extended data gathering.

Fig. 3. Sensor package during experiment.

This facilitated the generation of a data set that contained not only numerous samples for a given SOI (extended acquisition for a given SOI in a given aspect state), but also numerous trials of the particular SOI. For example, signal data was collected of the subject boat as it moved between the dock where the sensor platform was deployed and a buoy that was approximately 1.6 miles offshore. Having the subject boat repeat this route numerous times allowed for the collection of data for both inbound and outbound aspect states over numerous passes (trials). Fig. 4 below shows an example of the acquired temporal signature for the subject boat across 2 of the infrasound sensor channels. The vertical lines in the plots indicate the ground truth detection points within the overall acquisition. In other words, where during the acquisition did the test conductor log the closest point of approach of the subject boat? By having the ground truth data we can compare the known boat activity with the upcoming output results of the detection phase of the post processing.

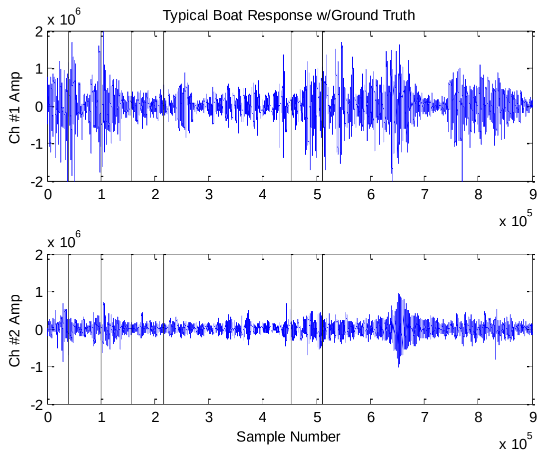


Fig. 4. Infrasound time history for boat.

Our objective is to derive a feature vector that uniquely describes the response of the subject boat. In an ideal case the boat signature would be collected in a sterile spectrum, noise-controlled environment. This would maximize the signal to noise ratio by minimizing random (white) noise as well as colored noise (signals due to other sources). Under these conditions an optimum test signature can be acquired from which the feature vector can be derived. In this experiment however, both white and colored noise are present in the recorded signals. Therefore a method was needed to overcome this empirical limitation.

First, we recognize that we only have access to signal power versus time. No phase-history information for the sampled data is available. Without a complex representation of the signal we cannot use typical signal processing methods such as the discrete Fourier transform. We can however work in the power spectrum domain. To develop the feature vector we must extract a signature from the acquired signal that uniquely describes the boat we are attempting to classify. Since we have the ground truth data we know where (when) in the acquired signals to look for the response of the boat.

We then need to separate it from the other interfering signals. If we choose to arbitrarily assign a 30 second window length to the desired feature vector we can center this around the ground truth time stamp. Fig. 5 below shows the power spectral density plot for the source boat sampled on channel #2 during passes (trials) #1-6.

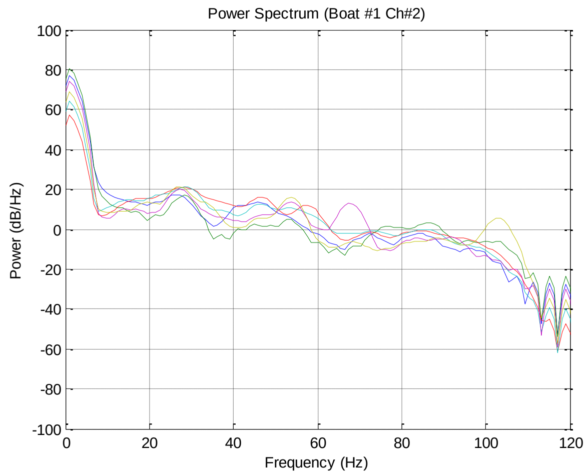


Fig. 5. Typical PSD for boat #1.

In order to build the raw data matrix X described above as the input to the single value decomposition process, we gather the power spectrum vectors for each trial as recorded by each channel. These are the x_n row vectors of Equation (2) above. After calculating the most dominant singular values (principal components) and using them to calculate the X_{mod} data array discussed above a simplified (reduced order) version of the data can be observed. As the left-hand panel of Fig. 6 below indicates, one of the trials has significantly more variance than the others. Additionally, the right-hand panel shows the Pareto plot of the singular values and indicates that the first 4 singular values should be retained to capture 95% of the variance in the data set, with the data in the remaining two trials being discarded as redundant, or having insignificant variance.

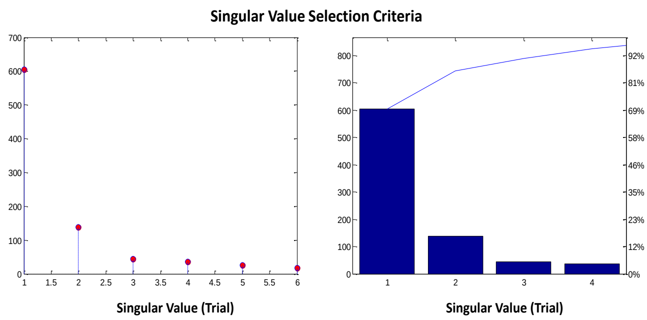


Fig. 6. Selecting singular values to retain.

While our goal is to develop a single feature vector that describes the subject boat, we must do so for each unique measurement channel due to the unique characteristics of each sensor (spatial, temporal, spectral). We do this by averaging each channel's set of feature vectors across all trials to yield a single vector for each channel. As Fig. 7 below shows, the power spectrum of each unique channel indicates that the subject boat has characteristic frequency behaviors at approximately 25 and 50 Hz.

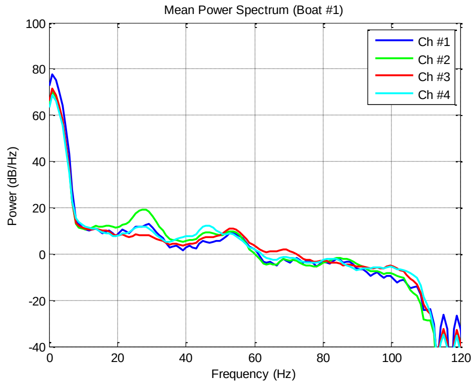


Fig. 7. Modified and mean spectra for boat #1.

To extend the validation of the conceptual method, the above process was also applied to signal data collected on a second day of the experiment where a completely different boat was provided by the U.S. Coast Guard to serve as a test subject. In this second case, a smaller boat (28ft long) with smaller (less horsepower) engines was used. The mean power spectrum (shown in Fig. 8 below) for boat #2 differs from that of boat #1 as would be expected. The lowest frequency resonance occurs at approximately 18 Hz for boat #2, vs. 25Hz for boat #1. This may help us successfully differentiate between the two boats during the classification phase of the processing.

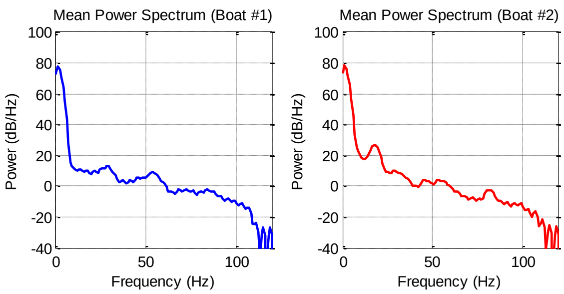


Fig. 8. Comparison of boat power spectra.

D. Data Processing Methodology Validation

Now that we have derived a representative feature vector for a boat by averaging the power spectra per channel over all trials, we wish to validate the accuracy of this approach by using our feature vector to detect & classify the boat signature from among unknown input sample data. We can go back to our original experimental data set in which our infrasound array collected power data across multiple individual trials. Recall that this data includes not just our boat signal of interest but all of the in-band signatures from a variety of other sources. A software post-processing routine was created that takes the 30 second-long feature vector and steps it in 1-second increments across the entire 2-hour long data set. For each 30 second "window" the full set of input data (grouped by channel) is processed as described above to derive the mean power spectrum for that window. That vector is then compared in a least-squared-error sense to the boat feature vector to determine how alike the two are. (Note: Even though the boat feature vector was derived from the experimental data, it has been through several processing steps. The feature vector is only a representation of the boat signature, but we hope a reasonable one. As such we would not expect to find the exact feature vector within the experimental data, but we would hope to find reasonably close matches between the two that were coincident with our known ground truth data).

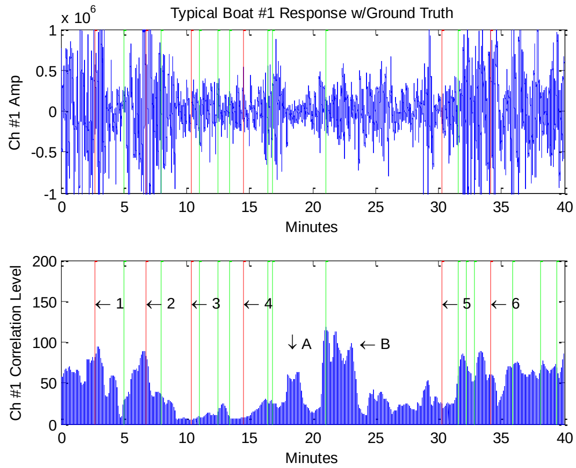


Fig. 9. Boat #1 detection example.

Fig. 9 above shows the raw data for channel #1 along with the ground truth overlaid with the red vertical lines. This represents where the test operator recorded an actual pass by our subject boat (signal of interest). The green vertical lines indicate observations of non-subject boats in the ground truth data. The lower plot in the figure shows the degree to which our analytically derived feature vector for the subject boat agrees with the data at a given point in time. Recall that the sample data was "windowed" into rolling 30-second segments for comparison. The elements numbered 1-6 in the plot represent the 6 known passes of the subject boat in outbound/inbound pairs (e.g., #1 is boat outbound, #2 is boat inbound). While the elements labeled A and B represent non-subject boat sources that registered significant levels of agreement with our feature vector. As the figure shows, the first cycle (#1 and #2) matched up very well with the ground truth indicating that the feature vector was significantly discriminated from the noise/interfering data within the sample for that point in time. Alternatively, cycle two (#3 and #4) did not register a significant detection by the algorithm. This is likely due to the low signal-to-noise ratio of the data for those samples. Comparing the raw data in the upper plot shows very little signal was recorded for cases #3 and #4 while a stronger signal level was recorded for cases #1 and #2. Additionally, the third cycle (#5 and #6) showed reasonably large agreement levels but did not align exactly with the ground truth. In the case of the outbound response at element #5, the raw data shows a large response just before the ground truth that lines up with a registered response. It is likely in this case that the ground truth recorded was slightly in error, as a review of the ground truth log shows no activity for approximately 8 minutes prior to the outbound boat signature of element #5. Two additional responses were registered between elements #5 & #6 that (according to the ground truth log) correspond to two unaffiliated boats passing through the test area. Similarly, elements A and B are attributed to boats unaffiliated with the test passing through the test area. In the case of element B, a large boat was reported in the area with loud engines.

As an additional validation case we can examine the results of a similar experiment using the separately derived feature vector for boat #2 which was collected the day after boat #1. Recall that boat #2 was slightly smaller in size than boat #1 and had a different engine configuration. Fig. 10 below shows the raw data for channel #1 along with the ground truth overlaid with the red vertical lines for boat #2 from the experiment.

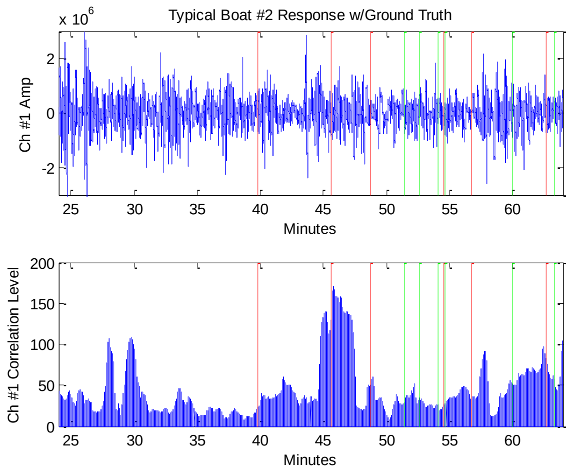


Fig. 10. Boat #2 detection example.

As with the example presented earlier for subject boat #1 the red vertical lines in the plot represent the entries in the ground truth log for the subject boat while the green lines represent other boats in the test area. While the data indicates two large responses were recorded at 27 and 30 minutes, the first pass of the subject boat (and thus the first ground truth log entry) was not made until approximately 40 minutes into the acquisition period. Only a very small response was registered for the outbound portion of the first cycle at 40 minutes while a very large response was recorded for the inbound portion of the cycle at 45 minutes. Cycle #2 at 49 and 55 minutes did not yield significant responses for either the outbound or inbound cases. For cycle #3, the outbound case was logged in the ground truth at 57 minutes but a review of the raw data suggests the ground truth may be slightly in error, with the actual boat response appearing at 58 minutes. The inbound response at 63 minutes was of similar size to the outbound.

As an additional validation exercise, we generated a purely analytical version of an infrasound sensor output channel. This enables us to evaluate the capability of the detection algorithm without the difficulty/ambiguity introduced by the background noise and interfering signals from other signals “similar” to the subject boat. To generate this test signal, we first generated a random signal composed of 3 complex sinusoids of selectable amplitude and frequency (including a small random noise component). We next overlaid a 30 second portion of this test signal with the subject boats feature vector developed in the earlier phase of the experiment. For clarity, the feature vector was inserted at the 3-minute point in the data file which corresponds with the first pass of the subject boat according to the ground truth data. The test signal was evaluated as before using a sliding 30-second window. At each window the spectrum of the test signal was compared to a mask derived from the boat feature vector. The algorithm then counts every frequency in which the test signal falls within the feature vector mask. The more frequencies in which the test signal lies between the upper and lower mask limits, the higher the correlation percentage. Fig. 11 below indicates that in an ideal situation the detection routine can identify the test boat with 100% certainty.

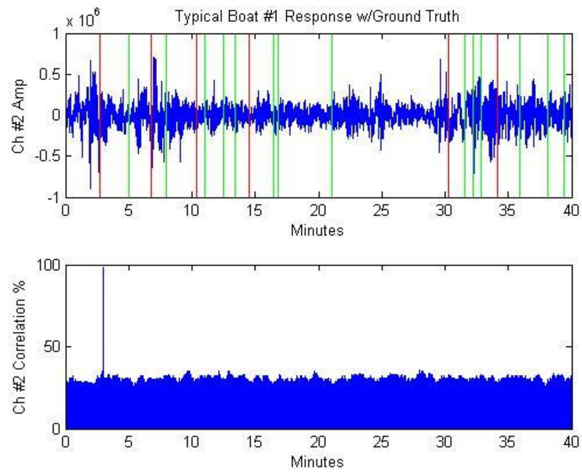


Fig. 11. Detection algorithm validation example.

Fig. 12. Feature vector detection “mask”.

In this example the "noise floor" of the data appears to be as high as 35% due to the fact that the test sample was not power-weighted versus frequency like the subject boat's feature vector. As a result, the zero-mean overlay of the two signatures creates an erroneous lower bound for the correlation percentage.

As Fig. 12 above shows, the detection mask captures the boat response fairly well when the feature vector is closely aligned with the known ground truth data and boat sample.

E. Alternate Detection/classification Methodology

Having developed and validated the feature vector detection mask discussed above, we next created a detection method that was simpler than the mean squared error based method used in the initial experiment. In the simplified method we count the percentage of each channel's unique feature vector that falls within the upper and lower mask limits as the vectors are windowed across the entire data set. This method also allows us to select a detection threshold to vary the processing sensitivity. Fig. 13 below shows the results for subject boat #1 with a 75% detection threshold. As the plot indicates, the method yielded significant detections at 3, 11 & 15 minutes and less significant detections at 6 and 30 minutes. No detection was observed at 34 minutes. (Recall that the vertical red lines indicate the known ground truth for the subject boat). The 5 of 6 detection result in this data set would yield an 83% classification success rate. By contrast, obvious incorrect detections can be seen at 5, 9, 24, 26, 28 & 29 minutes. In some cases, such as at 5 minutes the error corresponds to a known interfering boat signal. This case would be classified as a detection and classification error (false alarm). By contrast, there were numerous cases where the software correctly discriminated against interfering sources known to be in the data (green vertical lines) but did not register a detection (32-33, 36 & 39 minutes).

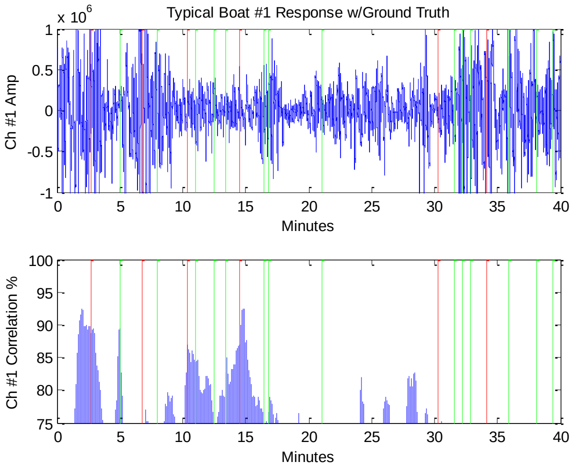


Fig. 13. Variable "mask" detection method.

As a final check of the processing algorithm, a feature vector was derived for an airplane that was recorded in the area of the sensor array during the experiment. This would allow us to contrast the boat data with a signature that should be "significantly" different. Fig. 14 below shows a comparison of the feature vector for the subject boat #1 and the airplane. As can be seen from the figure, while the boat has unique frequency characteristics at 25 & 50 Hz, the airplane has features at approximately 25, 50 and 75 Hz.

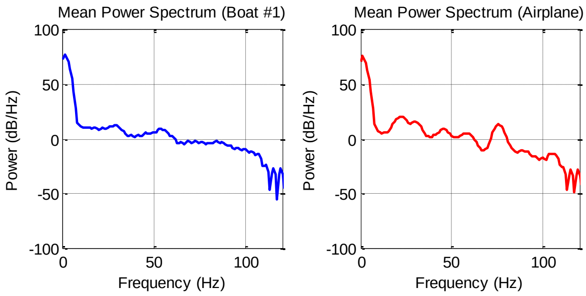


Fig. 14. Feature vector comparison.

We used the feature vector derived for the airplane as the mask and evaluated the data set as before. To the extent that the boat and airplane feature vectors are different, we would expect to see a reduced detection level for an unchanged 75% detection threshold used in the processing routine. Fig. 15 below shows that the detection rate is significantly reduced (as compared to the result shown in Fig. 13 above) using the airplane feature vector as would be expected.

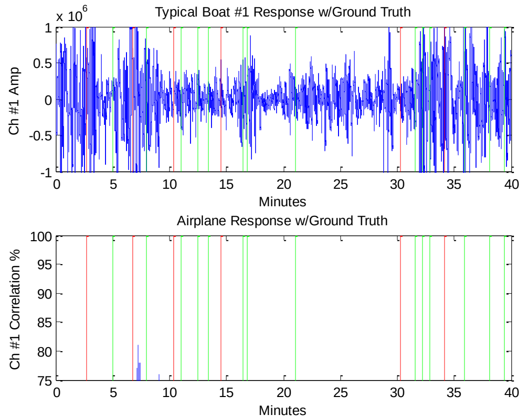


Fig. 15. Detection level w/airplane feature vector.

F. Discussion of Results

Evaluation of the level of detection for the two different boats for which two unique feature vectors were developed indicates that there is not a clear discriminator between the two boat types for the method used. Additionally, the observations indicated that significant detection levels were observed when the subject boat was not within the experimental test area. A possible explanation for this finding is that the common infrasound signature between the two subject boats was not sufficiently discriminated. Similarly, since the two subject boats used a similar engine configuration to many personal pleasure craft, it is reasonable that many of the boats observed in the test area during the experiment were similar enough in their infrasound signature as to be in-discriminable using the method described herein. From a "detection" point of view, the method is robust at detecting a variety of boats and identifying them as such. However, as a classifier a higher fidelity discriminator is needed to discern one type of boat from another.

IV. Future Work

While this field experiment included SOIs from various sources (aircraft, boats, personal watercraft) only the infrasound and seismic data from boats was analyzed for this paper. Additional work to characterize the SOIs of the other signal sources can be undertaken. Additionally, methods to improve the fidelity of the feature vectors used for classification can be investigated. For example, neural network based radial basis vectors may be able to more precisely model the boat response as compared to the power spectrum mean that was used herein.

V. Conclusion

A field experiment has been conducted in which infrasound signals of interest were collected by an array of sensors. These data were then analyzed to determine a unique feature vector that described the boats characteristic signature. The use of single value decomposition to determine the signals principal components allowed us to determine a characteristic response. This was then used to validate the method by attempting to detect boat signatures in data samples from the same experimental collection from alternate channels, sensors and days. The detection aspect of the process worked very well as numerous boat signatures were observed and validated against known ground truth data. However the very presence of so many boats and interfering signals having signatures similar to our test boats made specific classification of one boat relative to another unsuccessful. Further refinement of the feature vector and filtering methods may improve the classification results.

Acknowledgment

The authors wish to thank all those who participated in the field exercise and helped to make it a successful endeavor.

References

[1] D. N. Green and D. Bowers, "Estimating the detection capability of the International Monitoring System infrasound network", *J. Geophys. Res*., vol. 115, 2010.

[2] J. V. Olson and C. A. L. Szuberla, "Processing infrasonic array data", *Handbook of Signal Processing in Acoustics*, V2, Havelock *et al*., eds, Springer Verlag, 2009

[3] J. C. Samson and J. V. Olson, "Data adaptive polarization filters for multi-channel geophysical data," *Geophysics*, vol. 46, 1423, 1981.

[4] W. Arrasmith, E. Skowbo, and, J. Olson, "An overview of the detection and characterization of man-made signals-of-interest using an infrasound array," *INTECH 2013*, Budapest, HU, 10-13 September, 2013.

[5] W. Arrasmith, E. Coots, E. Skowbo, and J. Olson, "Systems engineering modeling and comparative analysis of various infrasound signals of interest," *International Journal of Modeling and Optimization*, vol. 5, no. 3, pp. 221-227, June 2015.

[6] D. Fee, S. R. McNutt, T. M. Lopez, K. M. Arnoult, C. A. L. Szuberla, and J. V. Olson, "Combining local and remote infrasound recordings from the 2009 Redoubt Volcano eruption", *J. of Volcanology and Geothermal Research*, vol. 259, 2013, pp. 100-114

[7] K. L. McLaughlin, A. Gault, and D. J. Brown, "Infrasound detection of rocket launches," *Science Applications International Corp (SAIC)* Arlington VA., 2000

[8] R. W. Whitaker, "Infrasound detection of rocket launches," in *Proc. the 17th Annual Seismic Research Symposium*, Sept. 12-15, 1995, pp. 997-1000

[9] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *IEEE Transactions on Neural Networks*, vol. 5, no. 4, 1994

[10] H. Wu, M. Siegel, and P. Khosla, "Vehicle sound signature recognition by frequency vector principal component analysis," in *Proc. the 15th Annual IEEE Instrumentation and Measurement Technology Conference*, St. Paul Minnesota, USA, May, 1998

[11] J. E. Jackson, *A Users Guide to Principal Component Analysis*, John Wiley & Sons, NY., 1991

[12] A. Averbuch, E. Hulata, and V. Zheludev, "Identification of acoustic signatures for vehicles via reduction of dimensionality," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 2, no. 1, pp1-22, 2004

[13] A. Blum and P. Langley, "Selection of relevant features and examples in machine learning," *Artificial Intelligence*, vol. 97, no. 1-2, pp. 245-271, 1997

[14] S .Dray, "On the number of principal components: A test of dimensionality based on measurements of similarity between matrices," *Computational Statistics and Data Analysis*, vol. 52, issue 4, 2008, pp. 2228-2237

[15] P. Peres-Neto, D. Jackson, and K. Somers, "How many principal components? Stopping rules for determining the number of non-trivial axes revisited," *Computational Statistics and Data Analysis*, vol. 49, 2005, pp. 974-997

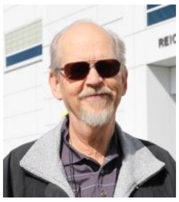
William W. Arrasmith was born in Bad Aibling, Germany on 7 January, 1961. He received his PhD in engineering physics from the Air Force Institute of Technology (AFIT) in Dayton, Ohio in 1995. He earned a MS degree in electrical engineering from the University of New Mexico in Albuquerque New Mexico in 1991. He obtained a BS in electrical engineering from Virginia Tech in Blacksburg, Virginia in 1983.

In his current position, he is a professor of engineering systems at the Florida Institute of Technology (FIT) in Melbourne, Florida, USA. Prior to FIT, he served in the United States Air Force for over twenty years culminating with a rank of Lt Colonel. During his time in the Air Force, he held several positions including chief, advanced science and technology division, applied technology directorate at the Air Force Technical Applications Center; Assistant Professor, Weapons and Systems Engineering Department, United States Naval Academy; Program Manager, Physics and Electronics Directorate, Air Force Office of Scientific Research; Director, Flood Beam Experiment, Air Force Research Laboratory (Kirtland Air Force Base); and Project Engineer, Teal Ruby Systems Program Office, Space Division. Recent related publications include William W. Arrasmith, E. Skowbo, and J. Olson, "An Overview of the Detection and Characterization of Man-Made Signals-of-Interest Using an Infrasound Array," Budapest, Hungary: INTECH 2013 Conference Proceedings, 2013; and Barry Webster, and William W. Arrasmith, "Optimal Systems Engineering Driven Search and Scan Pattern Determination for Detecting Non-cooperative Moving Ground Targets Using Micro-UAV "Swarm" Concept and Game Theory," Budapest, Hungary: INTECH 2013 Conference proceedings, 2013 (Best Paper). His research interests include applied systems engineering, advanced sensing/detection techniques, and imaging through atmospheric turbulence methods.

Dr. Arrasmith is a member of Phi Kappa Phi, Tau Beta Pi, and the American Society of Engineering Education (ASEE) and has two national and one international patent pending. He received the President's Award for Service at Florida Tech in 2013 and the Walter Nunn Excellence in Teaching Award in the College of Engineering at FIT in 2010.

Everett R. Coots was born in 1968, who was raised in Montville, CT, and graduated from Florida Institute of Technology, Melbourne, FL in 1993 with a bachelor's degree in electrical engineering, in addition to obtaining a masters-of-science in systems engineering from Florida Institute of Technology in 2007 and a Masters-of-Science in systems engineering management from Florida Institute of Technology in 2008. He is currently a doctoral student in systems engineering at Florida Institute of Technology.

He is employed by the Harris Corporation located in Melbourne Florida, as a senior antenna systems engineer in the government communications systems division. Everett has over 20-years of experience in the development of advanced antenna systems including phased arrays, deployable reflectors, and various classical aperture antennas for sea, air, land and space-based applications. Experience includes all levels of the system life-cycle: from preliminary design, to cost estimation and bid/proposal generation, to system integration and test, thru requirements verification and system delivery, to launch and initialization. Mr. Coots' current field of study is in infrasound sensor systems, infrasound signal processing and target classification algorithms.



John V. Olson is a professor of physics at the University of Alaska Fairbanks. He obtained a Ph.D. in Physics from UCLA in 1970 studying electromagnetic wave propagation in collisionless plasmas.

He worked at the University of Alberta in Edmonton, Canada for nine years as a post-doctoral fellow and research associate. In 1979 he joined the faculty of the University of Alaska Fairbanks (UAF) with a part time appointment at the Geophysical Institute at UAF. He has continued his studies of plasma waves and ionospheric disturbances while at UAF. In 1998 he initiated a program in infrasonic studies as part of the US effort in support of the Comprehensive Nuclear Test-Ban Treaty (CTBT) and has developed an active research group in infrasound studies.

Dr. Olson is a member of the American Geophysical Union (AGU), the Acoustical Society of America (ASA), and the IEEE Signal Processing Society.

Eric A. Skowbo was born in 1958, who was raised in Santa Barbara, CA, and graduated from California State Polytechnic University, Pomona, CA in 1987 with a bachelor's degree in electrical engineering, in addition to obtaining a masters-of-science in systems engineering from Florida Institute of Technology, Melbourne, FL in 2007.

He is a retired United States Air Force acquisition officer, who led missile development teams for Peacekeeper, Small Missile, and Peacekeeper Rail Garrison programs in addition to completing development and fielding the Air Force Tactical Applications Center Distributed Subsurface Network – which was the core of the United States Atomic Energy Detection System. He currently works for Northrop Grumman, located in Melbourne Florida, as Program Manager/Lead Systems Engineer for the Infrasound M2UGS R&D Group, and leads development of the Infrasound Mobile MASINT Unattended Ground Sensor (M2UGS) System. Eric has over 20 years of experience in the development of MASINT UGS systems, and led the research team that confirmed the hypothesis that UGF can be detected and characterized using unintended infrasonic emissions well outside the "fence-line". An article, Infrasound Sensor System Detection of Atlas V Rocket Launch, was published in the Northrop Grumman Technology Review Journal (Reston, VA: Northrop Grumman Corporation, 2009). Current research emphasis is on infrasound standoff detection and characterization of ground vehicles, rockets, hypersonic test vehicles, underground facilities, vehicle-borne IEDs, unmanned aerial systems, and watercraft using infrasound.

Mr. Skowbo's memberships include the Acoustical Society of AMERICA, and the Institute of Electrical and Electronics Engineers.

Barry Webster was born in Elmira, NY in 1963. He earned a B.S. in computer science from the Pennsylvania State University in University Park, PA in 1984. He went on to earn an M.S. in computer science in 1995, a Ph.D. in computer science in 2004, and an M.S. in systems engineering in 2005, all from the Florida Institute of Technology in Melbourne, FL.

He began his career in 1985 working as a systems engineer for Grumman Aerospace Corporation on Long Island, NY. He spent a total of 21 years with the company (which eventually became Northrop Grumman Corporation), holding positions in systems, software, and support engineering departments, ultimately becoming the senior database administrator for the engineering directorate. In this capacity, he was responsible for managing the operation of over 50 databases and associated applications and tools, used at installations around the world. After relocating to Melbourne, FL in 1987 he became associated with the Florida Institute of Technology, first as a student, earning three graduate degrees while working full time, then as a researcher on a sponsored program, and then as an adjunct professor. In 2007 he made the transition to Florida Tech as a full-time member of the faculty within the Department of Engineering Systems (though still consulting for Northrop Grumman part-time for another three years), where he remains to the present day. As part of his Ph.D. studies, he was also accepted for a doctoral internship at NASA's Ames Research Center in Mountain View, CA, where he worked in the Autonomous Systems Department on automated telescope control systems and algorithms for solving NP-Hard telescope scheduling problems. His research interests include artificial intelligence, decision theory, and game theory.

Dr. Webster is a member of the Association for Computing Machinery (ACM), the Institute for Operations Research and the Management Sciences (INFORMS), and the Game Theory Society (GTS). He has chaired conference sessions in database management and artificial intelligence, received several citations for his work in database management, and two Best Paper/Best Presentation awards for works on game theory.