PennState
Applied Research Laboratory

Spatial Coherence of Seabed Volume Reverberation

Final Report

Daniel C. Brown

File No.: TR-22-005
1 September 2022

Applied Research Laboratory
P.O. Box 30
State College, PA 16804-0030

Sponsored by: U.S. Office of Naval Research
Grant No.: N00014-18-1-2820

Applied Research Laboratory
State College, PA 16804

TR-22-005
1 September 2022

Sponsored by: U.S. Office of Naval Research

Spatial Coherence of Seabed Volume Reverberation

Final Report

Daniel C. Brown

1 Project Goals and Objectives

At the outset of this project the ONR Program Officer and the PI made the decision to shift the focus of this effort from the spatial coherence of volume reverberation to development of physics-based models for time series interpretation. The following report details those efforts.

Recent research and engineering have provided the Synthetic Aperture Sonar (SAS) community with a robust set of tools for acquiring high-quality data from a range of UUV-borne sensors. The overarching goal of this research program is a deeper investigation of the raw time-series data collected by these sensors. Frequently, SAS sensor data is processed to generate high-resolution sonar imagery. Can improved signal models improve the reconstructed imagery? Can improved signal models improve the interpretation of the sensed raw data? To address these questions, this program focused on developing improved physics-based models for and analysis of time-series sensed by synthetic aperture sonar systems. These physics-based models were utilized to address two areas:

1. the spatial coherence of signals measured from a bistatic, long-range SAS,
2. the sparse representation of signals projected onto a physics-based basis.

The development of coherence models supports a better understanding of how the sensor's design and employment impact the second cross-moment of the scattered field. This model has an immediate application for SAS motion estimation and longer-term benefits for interpreting signal coherence to detect and characterize the imaged scene. The development of sparse signal representations provides a method for separating time-series components for subsequent analysis. Developing these bases using physics-based methods allows subsequent signal processing and analysis to be tailored to specific physical responses.

2 Accomplishments Towards Achieving Goals

2.1 Spatial Coherence

The standard signal processing algorithms applied to SAS data generate high-resolution imagery. Historically, the process of reconstructing imagery from these sensors has rested on a number of approximations. One key assumption is known as the phase center approximation or the principle of waveform invariance. This approximation assumes the sensed scene is in the far field of the Vernier array. For many sensor geometries, this is a valid approximation; however, it is critical to determine under what conditions this approximation breaks down. This breakdown has been studied through the development of a physics-based model.

This program addressed the development of a coherence model through the van Cittert-Zernike theorem (vCZT) [1,2]. This formulation has been applied to the problem of modeling the spatial coherence of phase centers in a synthetic aperture array [3,4]. This prior modeling approach utilized the phase center approximation. Under this effort, the prior model has been expanded to remove the phase center approximation. The result is an expression for the covariance of the field measured between the reception of two pings of an active sonar. This

is the first documented model that correctly addresses the fully bistatic spatial coherence observed from sequential transmissions of a synthetic aperture sonar. A key result showed that the spatial coherence of the scattered field can be sensitive to array lengths used for existing commercial long-range SAS systems. The details of this development are captured in a journal article that is being submitted to the forthcoming Institute of Engineering and Technology's Electronics Letters special issue on "Recent Advances in Synthetic Aperture Sonar Technology". A preprint of this manuscript is appended to this report.

2.2 Sparse Signal Representation

Traditionally, the raw data collected by SAS sensors is processed to form high-resolution imagery. This is because SAS sensors provide improved image resolution over that generated by real aperture sonars, and they accomplish this with a similar form factor. These parallels have led to their use as "improved" sidelook image generation sensors. The acoustic image is a favored data representation that humans easily interpret. Consequently, this has been the primary data representation for MCM research. The generation of an image for visual interpretation is a lossy process. Additionally, this data representation may not be the most compact form for certain types of features that may be informative.

One data representation that has recently received attention is known as "acoustic color" [6, 7]. This non-image representation has proven useful for target characterization [8]. Additionally, this representation is useful because it is easily related to target scattering physics. This type of representation readily admits the design of physics-based features for target classification. An example of this is shown by Brown [9]. There, relatively simple models were combined with an acoustic-color based data representations to estimate target properties such as length and curvature directly.

To extend the development of physics-based non-image representations, this program has laid a foundation for physics-informed, model-based representation of acoustic data by developing a novel class of frames – *Enveloped Sinusoid Parseval* (ESP) Frames in collaboration with Dr. Geoff Goehle. In addition to establishing the theoretical framework for ESP frames, computational algorithms were developed to produce regularized and unregularized representations of acoustic data. Preliminary analysis was performed using experimental and synthetic time series and the performance of ESP frame-based representations for denoising and parameter estimation was compared to the STFT and Prony's Method, respectively.

The ESP frames approach was applied to AirSAS data collected against a set of objects procured by Dr. J. Daniel Park under his "Alternative Representations of Information for Acoustics" ONR research grant (N00014-19-1-2221). This dataset consists of circular SAS scans of a set of solid cylinders, hollow cylinders, and pipes. The objects all have the same external geometry with a length of 8 inches and a diameter of 2 inches, but their internal geometry varies. The time-series from these objects were separated using windows to capture specular and resonant scattering. SAS imagery was reconstructed from the separated signals and the specular, and resonant responses were clearly segregated. The details of the development of ESP Frames and their application to synthetic and measured data are detailed in a journal article that has been submitted to the Journal of the Acoustical Society of

America and a conference paper that has been submitted to the IEEE OCEANS conference at Hampton Roads. The titles for these documents are included in the following section and the pre-prints are included at the end of this report.

3 Dissemination of results

This work has produced a conference paper and two refereed journal articles. Pre-prints of these documents are included after the bibliography.

- "Spatial coherence considerations for the phase center approximation"
  - Submitted to Institute of Engineering and Technology's Electronics Letters for the special issue on "Recent Advances in Synthetic Aperture Sonar Technology"
- "Approximate Extraction of Late-Time Returns via Morphological Component Analysis"
  - Submitted to Journal of the Acoustical Society of America
- "Enveloped Sinusoid Parseval Frames"
  - Submitted to IEEE Oceans - Hampton Roads

4 Technology Transfer

We have held numerous discussions with Dr. John DiCecco at the Naval Undersea Warfare - Newport regarding the coherence modeling and the signal separation methods.

References

[1] R. Mallart and M. Fink, "The van Cittert-Zernike theorem in pulse echo measurements," *J. Acoust. Soc. Am.*, vol. 90, no. 5, pp. 2718–2727, 1991.

[2] D. C. Brown, "Modeling and measurement of spatial coherence for normal incidence seafloor scattering," Ph.D. dissertation, The Pennsylvania State University, 2017. [Online]. Available: http://etda.libraries.psu.edu/catalog/13741dcb19

[3] D. C. Brown, I. D. Gerg, and T. E. Blanford, "Interpolation kernels for synthetic aperture sonar along-track motion estimation," *IEEE J. Oceanic Eng.*, vol. 45, no. 4, pp. 1497–1505, 2020.

[4] B. Thomas and A. Hunter, "Coherence-induced bias reduction in synthetic aperture sonar along-track micronavigation," *IEEE J. Oceanic Eng.*, vol. 47, no. 1, pp. 162–178, 2022.

[5] iXblue Inc., *SAMS-50 Datasheet*, 2022. [Online]. Available: https://www.ixblue.com/wp-content/uploads/2021/12/sams-50-deep-tow.pdf

[6] S. G. Kargl, K. L. Williams, T. M. Marston, J. L. Kennedy, and J. L. Lopes, "Acoustic response of unexploded ordnance (UXO) and cylindrical targets," in *MTS/IEEE OCEANS Conf.*, Sep. 2010, pp. 1–5.

[7] D. S. Plotnick, P. L. Marston, K. L. Williams, and A. L. España, "High frequency backscattering by a solid cylinder with axis tilted relative to a nearby horizontal surface," *J. Acoust. Soc. Am.*, vol. 137, no. 1, pp. 470–480, 2015.

[8] D. P. Williams, "Acoustic-color-based convolutional neural networks for UXO classification with low-frequency sonar," in *Underwater Acoustics Conference and Exhibition*, 2019.

[9] D. C. Brown, "(U) Target length estimation from coherent response," *Journal of Undersea Research*, vol. 68, no. 2, pp. 214–234, 2018.

Pre-prints of Prepared Manuscripts

SPATIAL COHERENCE CONSIDERATIONS FOR THE PHASE CENTER APPROXIMATION

DANIEL C. BROWN AND THOMAS E. BLANFORD

Abstract. Synthetic aperture sonar image reconstruction relies on the coherence of overlapping phase centers to provide accurate micronavigation for a sensed scene. It is shown that phase centers lose coherence for near-range scattering from large SAS arrays due to the fundamentally bistatic nature of these sensors. This effect is modeled using the van Cittert-Zernike theorem and a point-based sonar scattering model. Reduction of the window length used in the delay estimation process can partially mitigate the loss of coherence at the expense of increased variance in the resulting delay estimates.

1. Background

Synthetic aperture sonar (SAS) sensors create high-resolution imagery of the seafloor through the coherent combination of transmissions from a moving array[1, 2]. Displaced phase center micronavigation is a key technique that permits robust image formation from mobile platforms (e.g. tow sleds or unmanned underwater vehicles) [3, 4]. Many studies and algorithms have been designed leveraging this technique for SAS motion estimation [5–11]. The displaced phase center technique relies on comparisons between pairs of time series recorded on successive pings of a SAS. Under the phase center approximation, a pair of signals is fully coherent when their phase centers are co-located. Signals with non-overlapping phase centers may be partially coherent due to the spatial coherence of seafloor scattering. Fundamentally, the displaced phase center method searches for the temporal and spatial offsets maximizing the coherence between signal pairs. These offset estimates are then used to determine the synthetic aperture geometry created by the sensing system.

The phase center approximation assumes the sensed scene is in the far-field of the Vernier array. Many early research SAS sensors deployed small receive arrays [12–14], and the phase

Date: September 12, 2022.

center approximation holds across the entire imaging swath for these sensors. More recently developed commercial systems are deploying large receive arrays; and in some cases, these arrays are up to three meters in length [15]. The large separation between the projector and the distal hydrophone on these arrays challenges the validity of the phase center approximation.

This paper focuses on studying the impact of the breakdown of the phase center approximation for a near-range scattered signal for SAS sensors with receive arrays spanning a variety of lengths. First, the phase center approximation itself is reviewed. A model based on the van Cittert-Zernike theorem (vCZT) is utilized to predict the population spatial coherence between signal pairs recorded on successive pings of a bistatic sensor. The predicted population coherence is then compared to a numerical estimate generated through analysis of an ensemble of pings generated from a point-based scattering model.

2. Phase Center Approximation

To review the phase center approximation, begin by considering a sonar system consisting of a single transmitter, TX, and receiver, RX. At some instant in time, sensed field is given by the coherent integration of returns from the scatterers located on an isochronous ellipsoid of revolution. Figure 1 shows the ellipse formed by the intersection of the ellipsoid and the x-y plane. The ellipse's foci are located at the transmitter and receiver positions. Except for reciprocity, unique transmitter and receiver positions produce unique ellipsoids and, therefore, unique sensed fields.

An approximation to the ellipsoidal form can be made for the case where the distance to the ellipsoid's surface is much greater than the transmitter and receiver separation. Assuming that the transmitter and receiver lie along the x-axis, that they are separated by $2d$, and the origin is the midpoint of the line joining the transmitter and receiver, the ellipsoid of

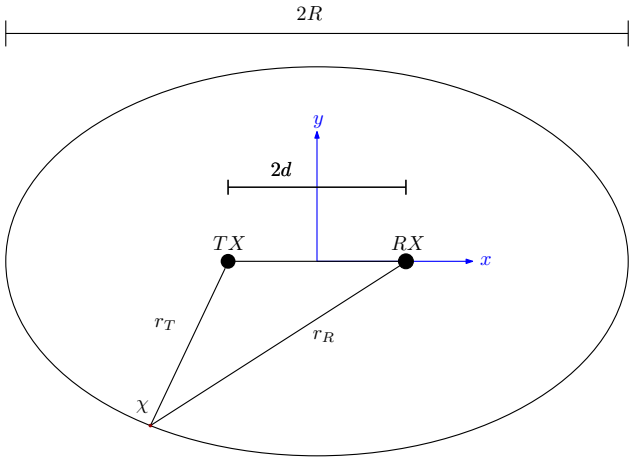


FIGURE 1. For a bistatic transmitter (TX) and receiver (RX), the field received at some instant in time is due to scattering from an isochronous ellipsoid of revolution. The intersection of this ellipsoid and the x-y plane is shown as an ellipse. For $R \gg d$ this ellipsoid is well approximated as a sphere whose surface depends only on the midpoint, which is known as the phase center.

revolution is given by

$$\frac{x^2}{R^2} + \frac{y^2}{R^2 - d^2} + \frac{z^2}{R^2} = 1 \tag{1}$$

$$x^2 + \frac{y^2}{1 - \frac{d^2}{R^2}} + z^2 = R^2. \tag{2}$$

In the far field, $R \gg d$ and the ellipsoid is well approximated as a sphere of radius R. Under this approximation, the scattered returns of any bistatic sensor pair can be approximated by a monostatic sensor located at the midpoint between the transmitter and receiver. This midpoint is known as the phase center. The sensed field is due to the integration of scatterers distributed over this spherical surface, and any pair of transmitters and receivers with a common phase center will produce an identical signal. For this reason, the phase center approximation is also described as the "principle of waveform invariance" [16, 17].

The phase center approximation holds when R is larger than the Fraunhofer distance $R \gg 2d^2/\lambda$, where λ is the wavelength [5]. As the range decreases and begins to approach the Fraunhofer distance, the effect of the approximation's breakdown is the introduction of a delay between signal observed from the a monostatic phase center and the signal observed from the bistatic sensor. Correction of this delay is required to achieve high-resolution

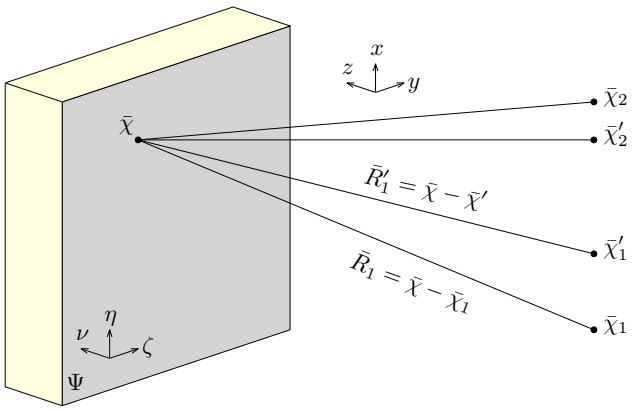


FIGURE 2. The vCZT is applied in Equation 3 using a geometry where a bistatic sonar transmits a pair of pings from $\bar{\chi}'_1$ and $\bar{\chi}'_2$. After scattering from the surface Ψ, the field is received at $\bar{\chi}_1$ and $\bar{\chi}_2$.

imagery for most (if not all) practical SAS sensors [18]. When R is much less than the Fraunhofer distance, the monostatic and bistatic signals begin to lose coherence with each other. This effect is similar to the baseline decorrelation of interferometric sonars and radars [19].

3. Analytic Model for Ping-to-Ping Coherence

The van Cittert-Zernike theorem was developed in the field of statistical optics, and it relates the spatial distribution of the intensity of a incoherent radiator to the coherence of the radiated field measured between two points [20]. This theorem has been adapted to address the spatial coherence of scattered acoustic fields [21, 22]. It has recently been extended to consider the impact of temporal windowing for seabed scattering of pulsed active sonar systems [23]. vCZT-based models for the spatial coherence of the scattered field have motivated the development of algorithms for SAS along-track motion estimation [24, 25].

The vCZT model provided by Brown, Gerg, and Blanford expresses the covariance of a pair of receivers observing temporally-windowed returns of the scattered field generated from a single transmission [24]. This single ping model is applies only to the multi-ping along-track estimation problem when the phase center approximation is valid. To directly evaluate the impact of the bistatic geometry for large synthetic aperture sonar arrays requires explicit consideration of the transmitter and receiver positions on a pair of successive pings.

The vCZT model can be directly expanded to consider sequential pings of a bistatic sonar. The covariance, less constant scaling factors, of the field measured on the temporal interval $[t_1\ t_2]$ is

$$\gamma_{12}(t_1,t_2) \propto \int_{\psi} \frac{b_1'(\bar{\chi})b_1(\bar{\chi})}{R_1'R_1}\Lambda_1(\bar{\chi},\frac{t_2+t_1}{2},t_2-t_1) \frac{b_2'(\bar{\chi})b_2(\bar{\chi})}{R_2'R_2}\Lambda_2(\bar{\chi},\frac{t_2+t_1}{2},t_2-t_1) \sigma(\bar{\chi})e^{ik(R_1'+R_1-R_2'-R_2)}\mathrm{d}\bar{\chi}. \tag{3}$$

The geometry for this equation is provided in Figure 2. The sonar transmits at the positions $\bar{\chi}_1'$ and $\bar{\chi}_2'$ and receives the scattered field at $\bar{\chi}_1$ and $\bar{\chi}_2$, where numeric subscripts indicate ping number. For a high-frequency SAS system, the scattered signal is dominated by the scattering from the surface Ψ. The transmit and receive directivity functions are b' and b respectively. $\sigma(\bar{\chi})$ is the interface scattering strength and k is the acoustic wavenumber. The temporal interval $[t_1\ t_2]$ creates a spatial "masking" of the seafloor so only a finite region contributes to the field observed at any instant in time. This effect is captured in the masking function, Λ_m, which is defined for ping $m \in [1,2]$ as

$$\Lambda_m(\bar{\chi},t,\tau) = \begin{cases} 0 & \text{if } \frac{1}{\tau}\left|t-\frac{1}{c}|\bar{\chi}-\bar{\chi}_m|-\frac{1}{c}|\bar{\chi}-\bar{\chi}_m'|\right| > \frac{1}{2} \\ \frac{1}{2} & \text{if } \frac{1}{\tau}\left|t-\frac{1}{c}|\bar{\chi}-\bar{\chi}_m|-\frac{1}{c}|\bar{\chi}-\bar{\chi}_m'|\right| = \frac{1}{2} \\ 1 & \text{if } \frac{1}{\tau}\left|t-\frac{1}{c}|\bar{\chi}-\bar{\chi}_m|-\frac{1}{c}|\bar{\chi}-\bar{\chi}_m'|\right| < \frac{1}{2}, \end{cases} \tag{4}$$

where c is the sound speed, and τ is the pulse length. Finally, many of the terms in the integrand are shown as dependent on $\bar{\chi}$ to emphasize their role in determining the spatial distribution of intensity creating the scattered field. By using this expression for the covariance of the field, the correlation coefficient can be calculated as

$$\rho_{12}(t_1,t_2) = \frac{\gamma_{11}(t_1,t_2)}{\sqrt{\gamma_{12}(t_1,t_2)\gamma_{22}(t_1,t_2)}}. \tag{5}$$

Evaluation of Equations 3 and 5 gives the population covariance of the scattered field for a given sensor, geometry, and environment. The integral in Equation 3 is generally evaluated numerically with a model for the seafloor scattering strength and measurements of the sensor directivity functions.

4. Simulation and Analysis of Ping-to-Ping Coherence

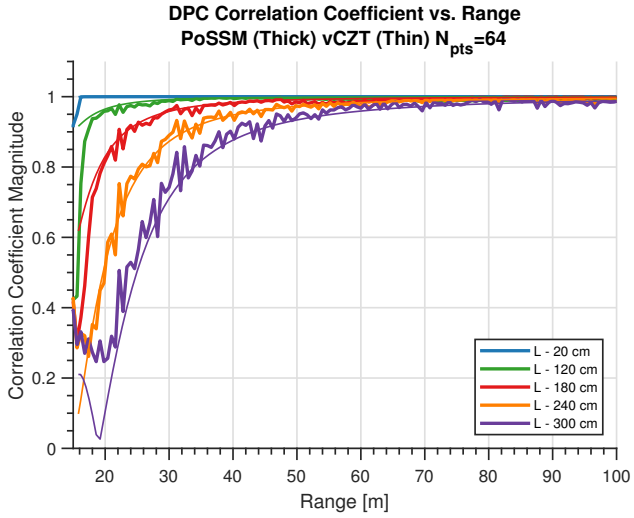
The analytic model developed above predicts the population coherence of signals received in sequential pings of a SAS system. In this section, the time series measured by a sensor is directly simulated using a point-based sonar scattering model (PoSSM) [26]. The simulated sensor operates with a center frequency of 200 kHz and a bandwidth of 30 kHz at an altitude of 15 m. The transmitters and receivers are modeled with omnidirectional vertical beam patterns and sinc patterns corresponding to transmit and receive element widths of 4 cm. The SAS is simulated with a single channel of overlap from ping to ping, and the transmitter and leading receiver channel are assumed to be co-located. Receive array lengths of 0.2 m, 1.2 m, 1.8 m, 2.4 m and 3.0 m are simulated. The phase center coherence is estimated by the peak magnitude of the complex correlation coefficient calculated from pairs of sliding-window short-time Fourier transforms (STFT). The STFTs are calculated with either 64-point or 128-point windows using rectangular weighting. Sequential short-time windows have 50% overlap.

The vCZT-based population correlation coefficient predicted by Equation 5 is compared to that calculated from the PoSSM simulations in Figure 3. The correlation coefficient was estimated using the two STFT lengths shown in Figure 3a and 3b, respectively. The model-model agreement is good over most of the simulated sensor ranges. This demonstrates that a long-range SAS sensor's loss of coherence observed at near operating ranges is well captured using a vCZT model. Also, this pair of figures shows that the near-range coherence can be partially recovered by reducing the window length used in the coherence estimation process. There is a disagreement between the models for weakly correlated signals at near ranges. The

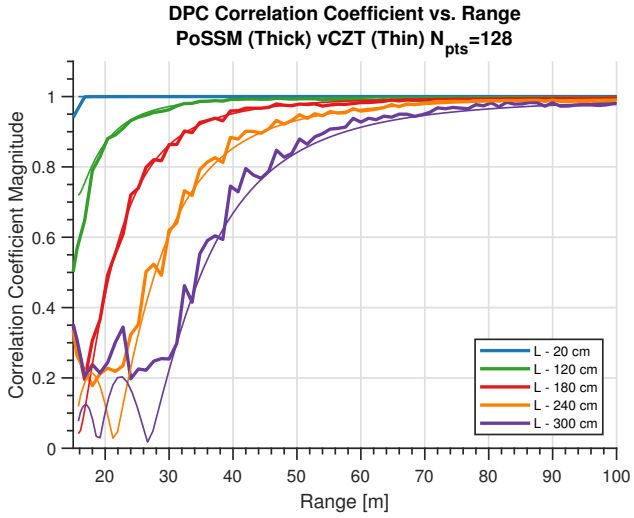
PoSSM-based simulations exhibit window length dependent asymptotic behavior between 0.2 and 0.3, while the vCZT models predict correlation coefficients that extend to zero.

The vCZT-based approach provides a model predicting the true population coherence, while the PoSSM-based approach estimates the coherence from the simulated time series. Therefore, the PoSSM-based approach is subject to the bias error associated with estimation of the magnitude of the complex correlation coefficient. The bias associated with the estimation of the magnitude of the complex correlation coefficient has been widely discussed [27–29]. This bias is further exacerbated by the fact that the coherence estimates are found by calculating the correlation coefficient across a set of temporal lags.

For large receive array lengths, the population coherence does not monotonically increase with increasing range. Instead, $|\rho|$ oscillates between approximately 0 and 0.2. This effect is observed in Figure 3a for the 300 cm array and in Figure 3b for the 240 cm and 300 cm arrays. This ripple phenomenon is a function of both the bistatic separation and the number of points in the STFT window. The complex covariance of the scattered field (mutual intensity in the statistical optics literature) originates at the ensonified region of the seafloor and propagates to the sensor according to a pair of second-order differential equations. Mathematically, this is similar to the propagation of acoustic pressure [30]. Given the mathematical similarity in their wave equations, there is a close analogy between the spatial coherence in this example and the acoustic pressure radiated from a piston transducer. In the near field of a piston, the acoustic field has a fine structure with peaks and nulls dependent upon position. In the far-field, however, the acoustic pressure changes slowly with position. The bistatic spatial coherence behaves similarly: the population coherence is very sensitive to position at near range and insensitive to position at far ranges. In Eq. 3, the range of phase angles in the complex exponential term in the integrand is small when the bistatic separation is small compared to the range. At nearer ranges this term spans a much wider range of angles and the complex exponential term oscillates over the integral. These oscillations cause the value of the integral to be very sensitive to positions.



(A)



(B)

FIGURE 3. The population coherence predicted by the vCZT model is compared to the mean sample coherence estimated from PoSSM simulated time-series using STFT windows of (a) 64 points and (b) 128 points. Mismatch for low coherence values is due to bias in estimation of the magnitude of the complex correlation coefficient from the time-series data.

This near-range ripple effect, however, may be difficult to observe in experimental data. Statistical estimation error introduces bias and variance to the magnitude of complex correlation coefficient, both of which increase with decreasing population coherence [29]. This effect of estimation bias has been observed in SAS systems for speckle coherence [31] and for DPC micronavigation [25]. The challenges with observing this ripple effect are apparent

in the PoSSM simulated time-series data. Statistical estimation error causes the correlation coefficient magnitude estimates to plateau at short range to values greater than the ripple amplitude. The increased variance in these estimates (compared to the regions of higher coherence in the plot) further obscures the ripples.

Figure 3 exhibits a trend that is an important consideration for motion estimation design: at near ranges, there is a loss of coherence that depends both on the window length and the bistatic separation. Figure 4 shows the slant range at which the population coherence exceeds 0.99 ($|\rho_{12} > 0.99|$) as a function of the bistatic separation length for several STFT window lengths. At this value of population coherence, the breakdown of the phase center approximation has essentially no effect on motion estimation. This loss of coherence is the same as would be experienced for perfectly coherent signals with 20 dB signal to noise ratio (SNR) above additive incoherent noise. As the bistatic separation increases, it is necessary to compare signals from further out in range to ensure they will be coherent. This minimum range, however, also increases for longer STFT windows.

As the range increases, the bistatic separation gradually becomes small compared to the range to the ensonified region of the seafloor. This relationship explains some of the recovery in coherence. This trend of the bistatic separation becoming small compared to the range eventually results in the validity of the phase center approximation. Note, however, that the Fraunhofer distance at center frequency for the 300 cm array is 2.4 km, assuming $c = 1500\,\mathrm{m\,s^{-1}}$. The $|\rho_{12} > 0.99|$ threshold occurs at ranges much shorter than the Fraunhofer distance for all of the bistatic separations considered here. The STFT window length is proportional to the size of the ensonified region of the seafloor, and this dimension also is a factor. (When the range exceeds the Fraunhofer distance for the array, the coherence is no longer dependent on the size of the ensonified patch of seafloor.) At long ranges, however, SNR may be less than 20 dB and cause the coherence to degrade below this threshold. Lower thresholds may be appropriate for very long range sensors.

These effects have implications for designing long range SAS arrays and motion estimation algorithms. Using short windows can recover some coherence loss, but it comes at the expense

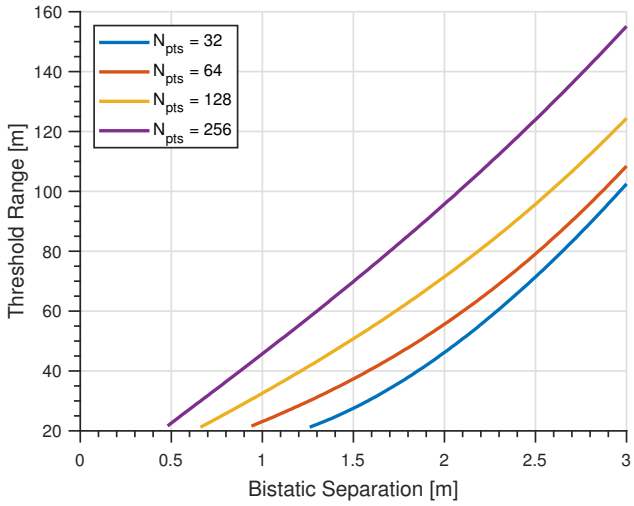


FIGURE 4. Loss of coherence is observed at near operating ranges for increasing bistatic separation. The slant range at which the population coherence degrades to $\rho_{12} = 0.99$ is plotted as a function of the bistatic separation for STFT windows of 32, 64, 128, and 256 points. As the bistatic separation increases it becomes necessary to use shorter STFT windows in the time delay estimation process used in DPC motion estimation algorithms.

of increased statistical estimation error in the magnitude of the complex sample correlation coefficients and the estimation error of the associated time delays. The variance in the correlation coefficient estimates are inversely proportional to the length of the window used in the estimate. For the time delays, the variance in the estimates is inversely proportional to the square root of the length of the window. Greater variance in the sample correlation coefficient magnitudes and the estimated time delays will lead to greater ping-to-ping errors in the sensor's along-track and cross-track motion estimates, respectively.

5. Conclusion

This paper has investigated the breakdown of the phase center approximation for near-range scattering observed by SAS sensors. This was accomplished through modeling with the vCZT to predict the population coherence and PoSSM to generate an ensemble of pings from which the sample coherence can be estimated. The vCZT method was extended to address fully bistatic scattering from sequential pings of a SAS. The vCZT and PoSSM models agree well, showing that near-range coherence is impacted by the bistatic collection geometry of large arrays. With increasing bistatic separation, a ripple effect is observed in the

population coherence at near range; however, this phenomenon was found to be challenging to observe in data as it is obscured by statistical estimation error. The bistatic coherence was found to depend on the window length used to estimate the correlation coefficient, which provides a method to mitigate coherence loss at near ranges. These effects, however, may have important implications for the design of long-range SAS arrays and motion estimation algorithms. Most long-range SAS surveys will operate at altitudes that mitigate these issues. For short-range imaging with a long-range sensor, the degradation of coherence by the breakdown of the phase center approximation and additive noise, when compounded by statistical estimation error, may place practical limits on how the SAS sensor is employed and the design of the associated motion estimation algorithms.

Acknowledgements

This work was supported by the Office of Naval Research under grant N00014-18-1-2820.

References

[1] Hayes, M.P., Gough, P.T.: Synthetic aperture sonar: A review of current status. IEEE J. Oceanic Eng. 34(3), 207–224 (2009). doi:10.1109/JOE.2009.2020853

[2] Hansen, R.E.: Introduction to synthetic aperture sonar. InTechOpen (2011)

[3] Raven, R.S., inventor; Westinghouse Electric Corporation, assignee. Electronic stabilization for displaced phase center systems. 4244036 (1981)

[4] Sheriff, R.W.: Synthetic aperture beamforming with automatic phase compensation for high-frequency sonar. In: IEEE Proc. 1992 Symposium on Autonomous Underwater Vehicle Tech., pp. 236–245. Washington, DC (1992)

[5] Bellettini, A., Pinto, M.A.: Theoretical accuracy of synthetic aperture sonar micronavigation using a displaced phase centre antenna. IEEE J. Oceanic Eng. 27(4), 780–789 (2002)

[6] Cook, D.A., et al.: Results from a hybrid synthetic aperture sonar motion estimation scheme. In: MTS/IEEE OCEANS Conf., vol. 2, pp. 1376–1381. Brest, France (2005)

208 [7] Oeschger, J.W.: Estimating along-track displacement using redundant phase centers.
209 In: Proceedings of the Institute of Acoustics, vol. 28, pp. 160–167. (2006)
210 [8] Groen, J.: Adaptive Motion Compensation in Sonar Array Processing. Delft University
211 of Technology (2006)
212 [9] Cook, D.A., Brown, D.C., Fernandez, J.E.: Synthetic aperture sonar motion estimation
213 using nonlinear least squares. In: Proc. Institute of Acoustics, vol. 28, pp. 176–182.
214 (2006)
215 [10] Cook, D.A.: Synthetic Aperture Sonar Motion Estimation and Compensation. Georgia
216 Institute of Technology (2007)
217 [11] Hunter, A.J., Dugelay, S., Fox, W.L.J.: Repeat-pass synthetic aperture sonar micron-
218 avigation using redundant phase center arrays. IEEE J. Oceanic Eng. 41(4), 820–830
219 (2016). doi:10.1109/JOE.2016.2524498
220 [12] Fernandez, J.E., et al.: Synthetic aperture sonar development for autonomous under-
221 water vehicles. In: MTS/IEEE OCEANS Conf., vol. 4, pp. 1927–1933. Kobe, Japan
222 (2004)
223 [13] Matthews, A.D., et al.: 12.75" synthetic aperture sonar SAS, high resolution and auto-
224 matic target recognition. In: MTS/IEEE OCEANS Conf., pp. 1–7. Boston, MA (2006)
225 [14] Brown, D.C., Cook, D.A., Fernandez, J.E.: Results from a small synthetic aperture
226 sonar. In: MTS/IEEE OCEANS Conf., pp. 1–6. Boston, MA (2006)
227 [15] iXblue Inc.: Sams-50 datasheet. `https://www.ixblue.com/wp-content/uploads/`
228 `2021/12/sams-50-deep-tow.pdf` (2022)
229 [16] Dickey, F.R., Edward, J.A.: Velocity measurement using correlation sonar. In: Pro-
230 ceedings IEEE Position Location and Navigation Symposium, pp. 255–264. San Diego,
231 CA (1978)
232 [17] Bradley, S.E., Deines, K.L., Rowe, F.D.: Acoustic correlation current profiler. IEEE J.
233 Oceanic Eng. 16(4), 408–414 (1991). doi:10.1109/48.90906
234 [18] Bonifant, W.W.: Interferometric synthetic aperture sonar processing. Georgia Institute
235 of Technology (1999)

[19] Zebker, H.A., Villasenor, J.: Decorrelation in interferometric radar echoes. IEEE Trans. Geosci. Remote Sensing 30(5), 950–959 (1992)

[20] Born, M., Wolf, E.: Principles of Optics. 7th ed. New York: Cambridge University Press (1999)

[21] Mallart, R., Fink, M.: The van Cittert-Zernike theorem in pulse echo measurements. J. Acoust. Soc. Am. 90(5), 2718–2727 (1991). doi:10.1121/1.401867

[22] Dahl, P.H.: Observations and modeling of angular compression and vertical spatial coherence in sea surface forward scattering. J. Acoust. Soc. Am. 127(1), 96–103 (2010). doi:10.1121/1.3268594

[23] Brown, D.C.: Modeling and Measurement of Spatial Coherence for Normal Incidence Seafloor Scattering. The Pennsylvania State University (2017). `http://etda.libraries.psu.edu/catalog/13741dcb19`

[24] Brown, D.C., Gerg, I.D., Blanford, T.E.: Interpolation kernels for synthetic aperture sonar along-track motion estimation. IEEE J. Oceanic Eng. 45(4), 1497–1505 (2020). doi:10.1109/JOE.2019.2921510

[25] Thomas, B., Hunter, A.: Coherence-induced bias reduction in synthetic aperture sonar along-track micronavigation. IEEE J. Oceanic Eng. 47(1), 162–178 (2022). doi:10.1109/JOE.2021.3103264

[26] Brown, D.C., Johnson, S.F., Olson, D.R.: A point-based scattering model for the incoherent component of the scattered field. J. Acoust. Soc. Am. 141(3), EL210–EL215 (2017). doi:10.1121/1.4976584

[27] Goodman, N.R.. On the joint estimation of the spectra, cospectrum, and quadrature spectrum of a two-dimensional stationary Gaussian process. University Heights, NY: New York University - Engineering Statistics Laboratory (1957). 10

[28] Carter, G.C., Nuttall, A.H.: Statistics of the estimate of coherence. Proc. IEEE 60(4), 465–466 (1972). doi:10.1109/PROC.1972.8671

[29] Carter, G.C., Knapp, C.H., Nuttall, A.H.: Statistics of the estimate of the magnitute-coherence function. IEEE Trans. Audio Electroacoust. 21(4), 388–389 (1973)

[30] Goodman, J.W.: Statistical Optics. New York, NY: John Wiley and Sons (1985)

[31] Synnes, S.A.V., Hansen, R.E., Sæbø, T.O.: Spatial coherence of speckle for repeat-pass synthetic aperture sonar micronavigation. IEEE J. Oceanic Eng. 46(4), 1330–1345 (2021). doi:10.1109/JOE.2021.3060812

The Applied Research Laboratory at The Pennsylvania State University

Email address: dcb19@arl.psu.edu

Approximate Extraction of Late-Time Returns via Morphological Component Analysis

Geoff Goehle, Benjamin Cowen, Thomas E. Blanford, Daniel C. Brown, and J. Daniel Park

Applied Research Laboratory, Pennsylvania State University, State College, PA, 16802, USA[a]

A fundamental challenge in acoustic data processing is to separate a measured time series into relevant phenomenological components. A given measurement is typically assumed to be an additive mixture of myriad signals plus noise whose separation forms an ill-posed inverse problem. In the setting of sensing elastic objects using active sonar, we wish to separate the early-time return (e.g., returns from the object's exterior geometry) from late-time returns caused by elastic coupling or geometric resonances.

Under the framework of Morphological Component Analysis (MCA), we compare two separation models using the short-duration and long-duration responses as a proxy for early-time and late-time returns. Results are computed for Stanton's elastic cylinder model as well as on experimental data taken from an in-Air circular Synthetic Aperture Sonar (AirSAS) system, whose separated time series are formed into imagery. We find that MCA can be used to separate early and late-time responses in both the analytic and experimental cases without the use of time-gating. The separation process is demonstrated to be robust to noise and compatible with AirSAS image reconstruction. The best separation results are obtained with a flexible, but computationally intensive, frame based signal model, while a faster Fourier Transform based method is shown to have competitive performance.

[a]goehle@psu.edu

ACKNOWLEDGMENTS

This work was sponsored in part by the Department of the Navy, Office of Naval Research under ONR award numbers N00014-18-1-2820 and N00014-19-1-2221.

I. INTRODUCTION

Underwater remote sensing using active sonar is typically performed by ensonifying the seafloor and processing the echoes to characterize the response from the objects and the environment. Synthetic aperture sonar processing is one of the primary methods used to generate imagery of the scattering intensity of the ensonified scene, and utilizes acoustic scattering phenomena that is akin to geometric optics. As such, the image formation algorithm only accounts for the early-time response of an object and tightly couples the arrival time of acoustic energy with its spatial location. However, the overall response from acoustically interrogated scene, especially the objects, supports additional responses including elastic scattering as well structural resonances. This late-time energy does not conform to the image formation model, is improperly associated with pixels during the image reconstruction process, and the energy appears in the image as smearing or blurring (Plotnick and Marston, 2016), see the top right subplot of Figure 10. Additional artifacts arise due to the fact that the late-time signal structure has been spectrally modified by the acoustic coupling, structural vibration, and re-radiation back to the receiver. Differences in the signal structure elicit alternative processing approaches.

As a first step, we are motivated to separate the overall response into multiple components, each of which share a common property. In terms of signal decomposition, we hypothesize the overall response is a superposition of multiple components as well as noise at an unknown level. Decomposing this additive mixture into their components is an ill-posed inverse problem. Separation of the early-time and late-time returns from non-homogeneous field of scatterers is a particularly challenging problem due to the diversity of acoustic effects (Pareige *et al.*, 1989). While high-Q elastic responses such as whispering gallery modes produce long-duration ringing, low-Q modes such as surface wavepackets produce short-duration ringing which can either arrive coincidentally with the geometrically scattered return or later in time (Kargl and Marston, 1989).

Various methods of separating early-time and late-time returns exist, from simple time gating to subtracting off the response of a rigid object from an elastic one with an identical geometry. A fundamental challenge with time-gating is that the early-time and late-time responses from a field of scatterers will overlap in time, preventing a clean separation of components. Subtracting the responses of objects with the same geometry but different material properties is an important tool, but limited to analytic or laboratory settings. In this paper, we approach this problem using a relatively recent technique in convex optimization known as Morphological Component Analysis (MCA) (Selesnick, 2014; Starck *et al.*, 2004, 2005). MCA is an optimization framework where each component of an additive mixture is identified by its ability to be sparsely represented by a unique linear operator, such as a frame or a dictionary. In this paper, we present two sparse representation frameworks for discriminating acoustic phenomena and compare their performance.

While we are motivated by the separation of the early-time and late-time responses, the paper will focus on the related problem of separating the short-duration and long-duration components of a time series. In this context, a short duration component is any signal component which has a short time duration, regardless of the physical source. This will include both the initial geometrically scattered return from the object as well as any late-time wavepackets resulting from surface coupling. Long duration components will generally include long tailed exponential decays caused by high-Q resonance modes. The reason we focus on short-duration/long-duration separation versus early-time/late-time is that we are motivated by the application to sonar imaging, where time series feature multiple superimposed returns with start times that are not known *a priori*. As such none of the separation techniques presented here will rely on time gating.

Section II describes the MCA framework and the two sparsification transforms featured in this paper. In Section III we use MCA to separate the short-duration and long-duration components of an analytic time series produced by Stanton's elastic cylinder model (Stanton, 1988). In Section IV we apply the same MCA techniques to experimental data collected using an in-Air circular Synthetic Aperture Sonar (AirSAS) (Blanford *et al.*, 2019) and demonstrate short-duration/long-duration separation on AirSAS imagery. Section V concludes with a discussion of MCA as applied to acoustic time series.

Notation

Brackets are used to denote the scalar elements of vectors, e.g. for $\mathbf{y} \in \mathbb{C}^N$ we have

$$\mathbf{y} = [y[0], y[1], ..., y[N-1]] \,.$$

Subscripts are used to differentiate the vectors, matrices, and parameters associated with the distinct components of the signal that we wish to separate. For example, $\mathbf{y} = \mathbf{y}_1 + \mathbf{y}_2$ denotes a vector signal $\mathbf{y}$ composed of two vector components $\mathbf{y}_i$, each of which may have an associated scalar parameter λ_i and matrix parameter $\mathbf{A}_i$ for $i = 1, 2$.

II. MORPHOLOGICAL COMPONENT ANALYSIS

For the following analysis we assume that our measured acoustic data $\mathbf{y}$ is an additive mixture of D morphologically distinct components:

$$\mathbf{y} = \sum_{i=1}^{D} \mathbf{y}_i \,. \tag{1}$$

The recovery of these components is an ill-posed inverse problem because there are infinitely many trivial solutions. The MCA framework (Starck *et al.*, 2004) addresses this issue by requiring each component $\mathbf{y}_i$ to admit sparse representation $\mathbf{x}_i$ in a corresponding transformed space given by a linear operator $\mathbf{A}_i : \mathbb{C}^{M_i} \to \mathbb{C}^N$. In this context each $\mathbf{A}_i$ is called a synthesis operator because it synthesizes coefficients $\mathbf{x}_i$ into the signal domain. We can write the MCA signal model as

$$\mathbf{y} = \sum_{i=1}^{D} \mathbf{A}_i \, \mathbf{x}_i, \tag{2}$$

under which the problem of signal separation becomes a convex optimization problem: we want to find sparse encodings $\mathbf{x}_i$ such that the original data is preserved. For $D = 2$, the optimization over sparse coefficients $\mathbf{x}_i$ is written as

$$\begin{aligned} &\arg\min_{\mathbf{x}_1, \mathbf{x}_2} \; \lambda_1 \| \, \mathbf{x}_1 \, \|_1 + \lambda_2 \| \, \mathbf{x}_2 \, \|_1 \\ &\text{s.t.} \;\; \mathbf{y} = \mathbf{A}_1 \, \mathbf{x}_1 + \mathbf{A}_2 \, \mathbf{x}_2 \,. \end{aligned} \tag{3}$$

This minimization problem identifies a sparse set of coefficients $\mathbf{x}_i$ so that the original signal $\mathbf{y}$ can be exactly reconstructed. The ℓ_1-norm, $\| \mathbf{x} \|_1 = \sum_n |x(n)|$, enforces sparsity in a minimization context by penalizing all non-zero components and by introducing a thresholding operation in the optimization algorithms discussed herein that sets small values to zero when possible. The $\lambda_i \in \mathbb{R}_+$ are tunable parameters that affect the severity of penalizing non-zero coefficients and can be used to prioritize one representation over the other. This problem is also referred to as Dual Basis Pursuit (BP) (Chen *et al.*, 2001; Selesnick, 2014). The equality constraint can be relaxed to perform denoising, a problem known as Dual Basis Pursuit Denoising (BPD), and in this case is represented by

$$\arg\min_{\mathbf{x}_1, \mathbf{x}_2} \begin{array}{l} \lambda_1 \| \mathbf{x}_1 \|_1 + \lambda_2 \| \mathbf{x}_2 \|_1 + \\ \frac{1}{2} \| \mathbf{y} - \mathbf{A}_1 \mathbf{x}_1 - \mathbf{A}_2 \mathbf{x}_2 \|^2. \end{array} \tag{4}$$

For BPD the λ_i parameters control both the weights applied to the encodings as well as the degree to which sparsity is prioritized over fidelity, with larger values of λ_i producing sparser, less accurate, reconstructions.

In MCA, our ability to identify a component via its sparse representation hinges on the aptness and mutual exclusivity of each linear operator. In other words, each transform $\mathbf{A}_i$ should admit sparse representation of its corresponding component signal $\mathbf{y}_i$, but should be inefficient in representing the other components. We wish to decompose $\mathbf{y}$ into a sum of short-duration components $\mathbf{y}_1$ and long-duration components $\mathbf{y}_2$ (as proxies for the early-time and late-time returns as discussed in Section I), and thus the problem at hand is to design $\mathbf{A}_1$ and $\mathbf{A}_2$ to describe those respective phenomena. We focus on over-complete tight-

frame operators $\mathbf{A}_i$ (Han *et al.*, 2007) which, by definition, satisfy $\mathbf{A}_i \mathbf{A}_i^* = p_i \mathbf{I}$ for $p_i > 0$. The subsequent sections discuss specific, promising selections of $\mathbf{A}_i$ for our application.

Problems 3 and 4 are convex and hence have unique, global solutions (Boyd *et al.*, 2004; Selesnick, 2014), but the solutions do not have a closed-form expression due to the non-smooth ℓ_1-norm. We can use the Alternating Direction Method of Multipliers (ADMM) to formulate these problems as a sequence of easier subproblems, whose iterative solution is guaranteed to converge to the global minimum (Eckstein and Bertsekas, 1992). The resulting algorithm is called the Split Augmented Lagrangian Shrinkage Algorithm (SALSA) (Afonso *et al.*, 2010; Selesnick, 2014). SALSA as applied to our MCA BP and BPD problems is written in Algorithm 1, where soft represents the soft-thresholding function

$$\text{soft}(x, T) = \begin{cases} \frac{|x|-T}{|x|} x & |x| > T \\ 0 & |x| \leq T. \end{cases}$$

Note that the difference between BP and BPD in Algorithm 1 is a single constant. While $\mathbf{x}_i$ converges to the solution, in the BP case it is not particularly sparse at any given iteration. As an alternative $\mathbf{u}_i = \mathbf{v}_i + \mathbf{d}_i$ also converges to the solution while being sparser at each iteration.

The λ_i scalars act as a weighting factor influencing how energy is prioritized between the $\mathbf{x}_i$ and, in the case of BPD, how much sparsity is prioritized over reconstruction fidelity. While in practice it is often useful to tune the λ_i to achieve a desired separation, in order to support comparative analysis for this paper we will use a common value $\lambda = \lambda_1 = \lambda_2$. The choice of common λ does not effect the solution for BP, but is important for BPD. When

Algorithm 1: MCA BP/BPD

Require: $\mathbf{y}$, $\mathbf{A}_i$, λ_i, μ
  initialize $\mathbf{x}_i = \mathbf{A}_i^* \mathbf{y}$, $\mathbf{d}_i = 0$, $i = 1, 2$
  if performing BP **then**
    $\alpha = \frac{1}{p_1 + p_2}$
  else
    $\alpha = \frac{1}{\mu + p_1 + p_2}$
  end if
  repeat
    $\mathbf{v}_i \leftarrow \text{soft}(\mathbf{x}_i + \mathbf{d}_i, \lambda_i/\mu) - \mathbf{d}_i$, $i = 1, 2$
    $\mathbf{c} \leftarrow \mathbf{y} - \mathbf{A}_1 \mathbf{v}_1 - \mathbf{A}_2 \mathbf{v}_2$
    $\mathbf{d}_i \leftarrow \alpha \mathbf{A}_i^* \mathbf{c}$, $i = 1, 2$
    $\mathbf{x}_i \leftarrow \mathbf{d}_i + \mathbf{v}_i$, $i = 1, 2$
  until stopping criteria met
  $\mathbf{y}_i \leftarrow \mathbf{A}_i \mathbf{x}_i$, $i = 1, 2$

performing BPD there is a maximum effective λ-value given by

$$\lambda_{\max} = \max(\| \mathbf{A}_1^* \mathbf{y} \|_\infty, \| \mathbf{A}_2^* \mathbf{y} \|_\infty)$$

such that for all $\lambda \geq \lambda_{\max}$ the solution is zero (Xenaki and Pailhas, 2019, Section V.B). We will generally choose λ as a percentage of $\lambda_{\max}$. Additionally, while it is possible to vectorize λ to achieve even finer grained control over the separation weights we will not do so here.

A. FFT MCA

A particularly simple, yet effective, form of MCA is to let the first representation be the identity, $\mathbf{A}_1 = \mathbf{I}$, and the second be the unitary Discrete Fourier Transform, $\mathbf{A}_2 = \mathbf{F}$. We refer to this as FFT MCA and in this case the solution to (3) or (4) splits a signal into two components, with the former sparse in time and the latter sparse in the frequency domain. A consequence of Fourier duality is the $\mathbf{y}_1$ component tends to be made up of broadband, short-duration elements while $\mathbf{y}_2$ tends to be made up of long-duration elements with a narrower spectrum.

A classic example is to consider the superposition of a spike on a sinusoid. Suppose we have $N = 1000$ samples with $f_s = 10\text{kHz}$ and define $\mathbf{y}$ to be

$$y[n] = \delta_{50}[n] + \sin\left(2\pi\frac{1000}{f_s}n\right)$$

where $\boldsymbol{\delta}_{50}$ is a one-hot vector at index 50. If BPD is applied to $\mathbf{y}$ using FFT MCA then Algorithm 1 will converge to

$$y_1[n] = \delta_{50}[n], \qquad y_2[n] = \sin(2\pi 1000 n/f_s). \tag{5}$$

In this case MCA separates $\mathbf{y}$ into its components exactly. This would not be true if, for example, noise were added to $\mathbf{y}$ or if $\mathbf{y}$ was the superposition of a sinusoid and a rectangular window. That is because noise and/or rectangular windows are not sparse with respect to either $\mathbf{I}$ or $\mathbf{F}$. For a noisy signal the correct approach would be to use BPD to reconstruct the signal without the noise component. For the rectangular window exact separation is easier using a different set of representations. As FFT MCA is signal agnostic and has no parameters, there isn't any way to alter the representations to fit a particular signal model.

B. ESP MCA

A more flexible set of representations are given by Enveloped Sinusoid Parseval (ESP) frames (Goehle *et al.*, 2022), a class of representations formed from enveloped and shifted sinusoids. Since ESP frames can be made using nearly arbitrary envelopes, a wide range of functions can be sparsely represented including exponentially decaying sinusoids, sinusoids with traditional windows, or modulated complex signals. Formally, given a set of non-zero

(but potentially complex) envelopes $\{\mathbf{e}_l\}_{l=0}^{L-1} \subset \mathbb{C}^N$ the vectors $\{\mathbf{a}_{l,k,m}\}$ defined by

$$a_{l,k,m}[n] = e_l[n - m \bmod N] \exp(2\pi jk(n-m)/N)$$

for $l = 0, \ldots, L-1$ and $k, m, n = 0, \ldots, N-1$ form a tight frame. Here l is the envelope index, k is the frequency index, and m is the time shift index. The synthesis operator $\mathbf{A}$ in this case is given by $A[n, l, k, m] = a_{l,k,m}[n]$. Since there are N^2L frame vectors, ESP frames are massively overdetermined. However, one of the advantages of SALSA is that it is not necessary to work with the synthesis and analysis matrices directly. Instead (Goehle *et al.*, 2022) describes FFT diagonalization, as well as additional acceleration techniques, which can be used to speed up Algorithm 1.

The goal when applying ESP frames to MCA is to find two sets of envelopes $\mathbf{e}_l^i$, where the superscript indicates the component index and the subscript the envelope index, such that the signal components $\mathbf{y}_i$ are sparsely represented by one set of frame vectors but not the other. In the ideal case $\mathbf{y}_i$ is actually equal to a frame vector for $\mathbf{A}_i$. The specific choice of envelope is often informed by the physics associated to the signal in question. In this case we wish to separate the long-duration high-Q signal components from the short-duration acoustic response of an elastic object. As such we will use decaying exponentials as envelopes for $\mathbf{A}_2$ since exponentially decaying sinusoids are an excellent signal model for long-duration ring down (Hambric, 2006). For $\mathbf{A}_1$ we will use extremely short rectangular windows since they flexibly capture short-duration signals.

As an aside, if $\mathbf{A}_1$ is generated using a single one-hot vector as an envelope, while $\mathbf{A}_2$ is generated using a single constant function as an envelope, then the resulting representations are extremely similar to the representations used in FFT MCA. This mode of ESP MCA

effectively generalizes FFT MCA, albeit not in strict mathematical terms. For example, if this degenerate ESP frame and FFT MCA are both applied to the signal described in (5) using BP with 1000 iterations, then the relative difference in the resulting $\mathbf{y}_1$ and $\mathbf{y}_2$ components is 0.077% and 0.879%, respectively.

ESP MCA Example

In order to illustrate the link between ESP frame envelopes and the underlying signal morphology we will perform ESP MCA separation using the following driven simple harmonic oscillator

$$\ddot{y} + \frac{2}{\tau}\dot{y} + 4\pi^2 f_0^2 y = \alpha \sin(2\pi f t) \tag{6}$$

with $\tau = 2\text{ms}$, $f = 15\text{kHz}$, $f_0 = 20\text{kHz}$, and $\alpha = 10^{10}$. Let y be the zero state solution ($y(0) = \dot{y}(0) = 0$). We generate $\mathbf{y}$ using $N = 1000$ samples of y with a sampling frequency of $f_s = 100\text{kHz}$. Based on what we know of the underlying dynamics we can choose envelopes which will allow us to exactly capture the decomposition of $\mathbf{y}$ into homogeneous and particular components. Since the homogenous solution to (6) must consist of exponentially decaying sinusoids we can let $\mathbf{A}_1$ be the ESP frame associated to a single exponentially decaying envelope $e_1^1[n] = \exp(-n/(\tau f_s))$ and expect $\mathbf{A}_1$ to capture the homogeneous part of $\mathbf{y}$. Similarly the particular solution to (6) must be a sinusoid so we define $\mathbf{A}_2$ to be the ESP frame associated to the single constant envelope $\mathbf{e}_1^2 = 1$ and expect $\mathbf{A}_2$ to capture the particular component. If we then perform MCA BP with 1000 iterations we get the separated signals shown in Figure 1. As we used knowledge of the ODE dynamics to define ESP frames such that $\mathbf{y}_i$ is sparsely represented by $\mathbf{A}_i$, the BP algorithm converges to the exact

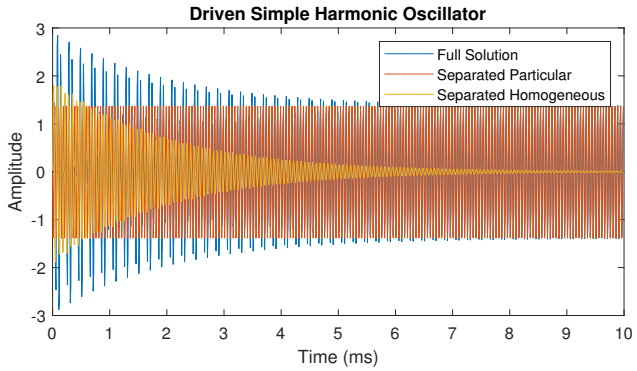


FIG. 1. ESP MCA separated homogeneous and particular components for the zero state solution to (6) using BP with 1000 iterations. The relative error for the particular solution is 0.05% and the homogeneous solution is 0.12%.

separation of $\mathbf{y}$ into its homogeneous and particular parts with the relative error in this case equal to 0.12% for the homogeneous solution $\mathbf{y}_1$ and 0.05% for the particular solution $\mathbf{y}_2$. Notably the FFT MCA approach would not be able to achieve the exact separation presented here because the homogeneous solution is not sparsely representable using the FFT.

C. MCA of Acoustic Signals

We are generally interested in separating out the short-duration returns of an elastic object from the long-duration ones. From a physical perspective the short-duration returns include the initial return of the ping reflecting off the rigid geometry of the object, as well as low-Q elastic effects and additional short-duration late-time phenomenon. The long-duration returns primarily include the high-Q resonance modes of the object. From a signal analysis perspective, however, the specific form of what constitutes a short-duration or long-duration

component is ultimately defined by the MCA representations. For FFT MCA the short-duration returns are represented using one-hot vectors, since $\mathbf{A}_1 = \mathbf{I}$, while the long-duration returns are represented using sinusoids, since $\mathbf{A}_1 = \mathbf{F}$. ESP MCA will use a frame built from short rectangular windows to capture the short-duration components (representing them as very short windowed sinusoids) and a frame built from exponentially decaying envelopes to capture the long-duration components (representing them as exponentially decaying sinusoids).

III. ANALYTIC SIGNAL SEPARATION

In this section we will demonstrate the MCA approaches presented in Section II on an analytic acoustic signal produced by the Stanton elastic cylinder model (Stanton, 1988, Section B). Section III A demonstrates separation applied to a clean impulse response from a Stanton elastic cylinder. In Section III B we demonstrate the same separation on a noisy LFM response while in Section III C we analyze the performance of the MCA techniques over a range of noise levels.

A. Impulse Response Separation

The Stanton model parameters used in this paper were chosen to represent a solid aluminum cylinder in water with a diameter of 15.25cm and a length of 30.5cm. The receiver is located 2m from the cylinder with a centered broadside orientation, and the signal is sampled at $f_s = 300$kHz. Stanton's model provides the frequency representation of the cylinder's impulse response. However, in order to minimize non-causal effects we apply a Butterworth

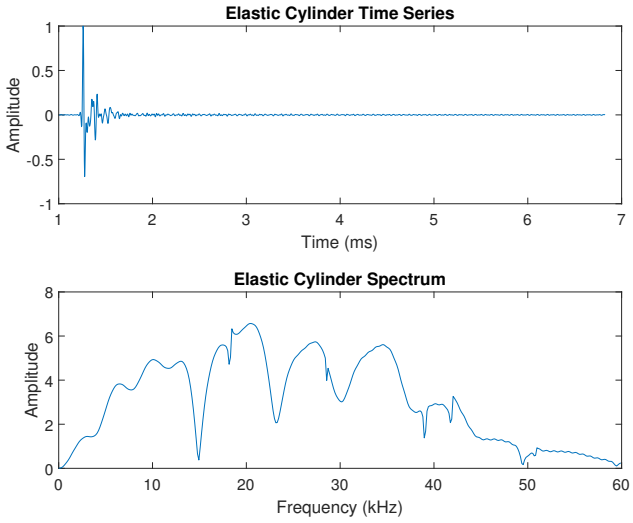


FIG. 2. Impulse response for the Stanton elastic cylinder (top) and corresponding spectral power (bottom).

filter of order 3 and threshold 0.25 to the synthesized time series. This helps to remove spectral discontinuities and produces a more natural impulse. The resulting time series, and corresponding spectrum, are shown in Figure 2. The deep nulls at 15kHz, 23kHz and 30kHz are likely caused by low-Q surface wave elastic effects. These effects are short duration, and may add constructively or destructively to the geometric scattering response. The intention is for these effects to be included in the short-duration component. The sharper, shallower nulls at 18kHz, 28kHz, 39kHz and 42kHz correspond to high-Q geometric resonance modes. These are long duration signals and are one of our primary targets for the long-duration component.

Importantly, since this model features a single return with no noise, the short/long duration separation problem is largely equivalent to early/late time separation. Consequently will use the relative error between the original time series and the separated short-duration

component in the early time, as well as the relative error between the time series and the long-duration component in the late time, as performance metrics for the MCA techniques. Specifically we fix an early-time interval I_1 lasting from 1ms to 2ms which includes the initial signal return, and a late time interval I_2 lasting from 2ms to 6ms which only includes the late signal return. Our metrics are then defined in terms of the standard ℓ_2-norm by

$$m_1 = \frac{\|\mathbf{y}\,|_{I_1} - \mathbf{y}_1\,|_{I_1}\|}{\|\mathbf{y}\,|_{I_1}\|}, \qquad m_2 = \frac{\|\mathbf{y}\,|_{I_2} - \mathbf{y}_2\,|_{I_2}\|}{\|\mathbf{y}\,|_{I_2}\|} \tag{7}$$

where $\mathbf{y}\,|_I$ indicates the restriction of $\mathbf{y}$ to the interval I. In the case where the separation is exact, we would expect m_2 to be zero, since the late-time signal contains only long-duration components, but m_1 to still be nonzero since the early-time signal does contain some long-duration energy. This effect will be minor since the signals presented in this section are dominated by short-duration energy.

FFT MCA

We begin by applying FFT MCA to the Stanton signal using BP with 1000 iterations. After performing the separation, we get the results shown in Figure 3. There we have plotted the original time series, the short-duration component and the long-duration component in the early time, in the late time, and the frequency domain. The results are quite good. FFT MCA correctly separates the loud initial response into the short-duration component while the signal tail is entirely separated into the long-duration component. Quantitatively, the short-duration component has an early-time error of 4.44% while the long-duration component has a late-time error 3.70%. The behavior of the spectrum is particularly interesting.

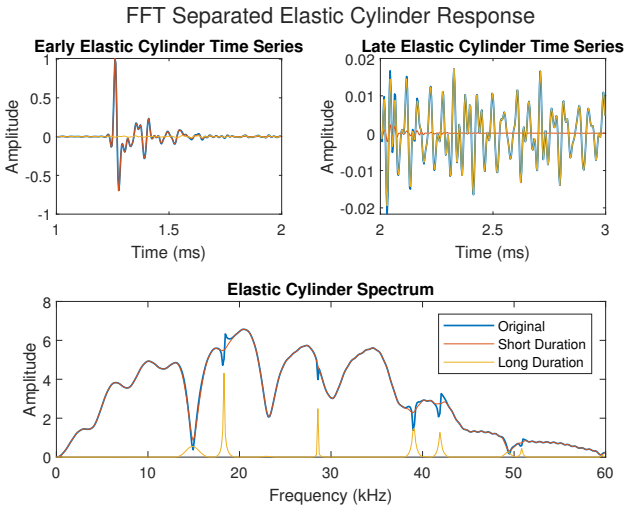


FIG. 3. Impulse response as well as the FFT separated short and long duration components for the elastic cylinder (top) and corresponding spectral power (bottom). The short-duration component has an early-time relative error of 4.44% and the long-duration component has a late-time relative error of 3.70%.

The bulk of the spectral power for the impulse response is separated into the short-duration signal, including the wide nulls caused by the low-Q elastic responses. The sharp short high-Q nulls however have been turned into distinct spikes in the spectrum of the long duration component. This has implications for feature detection as narrow peaks in the frequency domain are easier to detect and more resilient to noise than narrow nulls.

ESP MCA

As discussed in Section II B, for ESP separation we are using two categories of envelopes. In this section $\mathbf{A}_1$ will be formed using rectangular windows

$$e_l^1(t) = \begin{cases} 1 & 0 < t < T_l \\ 0 & \text{otherwise.} \end{cases}$$

The representation $\mathbf{A}_2$ will be formed using exponentially decaying envelopes

$$e_l^2(t) = \exp(-t/\tau_l).$$

We will use window lengths and time constants given by

$$T_l = 0.27, 0.54, 0.1\text{ms}, \tag{8}$$

$$\tau_l = 1.78, 3.16, 5.62, 10.00, 17.78, 31.62\text{ms}$$

Several different factors were considered in the selection of these parameters:

- The longest window length, 0.1ms, is significantly shorter than the shortest time constant, 1.78ms. This ensures the atoms are morphologically distinct and encourages better separation.
- The shortest window length, 0.27ms, is long enough to support a significant number of oscillations in the frequency ranges of interest.
- The largest time constant, 31.62ms, is big enough to support envelopes which decay very little over the length of the signal.
- The shortest time constant, 1.78ms, still produces atoms which would be considered long-duration.

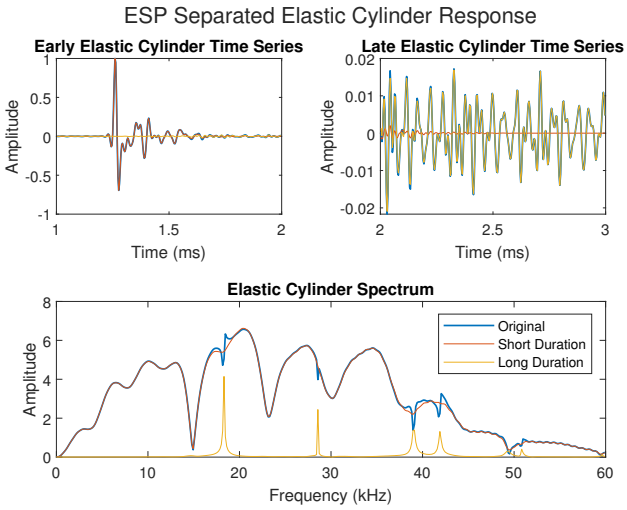


FIG. 4. Impulse response as well as the ESP separated short and long duration components for the elastic cylinder (top) and corresponding spectral power (bottom). The short-duration component has an early-time relative error of 4.04% and the long-duration component has a late-time relative error of 3.67%.

The results of ESP MCA utilizing 1000 iterations of BP are shown in Figure 4. For this time series, ESP MCA closely mirrors FFT MCA, and is nearly as effective at separating the signal as in the toy ODE case. We see that the short duration component captures almost all of the initial response, and most of the spectral power, with an early-time error of 4.04%. The long-duration component on the other hand captures the entire late-time response, with a late-time error of 3.67%, and as a result has some nice clear peaks at the high-Q null locations. Comparing the performance of both methods in Table I we see that FFT and ESP MCA perform about the same in terms of relative error, which is confirmed by a visual inspection of the separated components.

Method	m_1	m_2
FFT MCA	4.44%	3.70%
ESP MCA	**4.04%**	**3.67%**

TABLE I. Short-duration early-time relative error m_1 and long-duration late-time relative error m_2 for FFT MCA and ESP MCA for the Stanton elastic cylinder.

B. Noisy LFM Response Separation

In order to understand the impact of noise on our MCA techniques, and to represent more realistic signal processing, we will now use an LFM response from the same Stanton cylinder model to produce a noisy, matched-filtered, time series. The clean LFM scattering response was generated by convolving an LFM which sweeps from 15kHz to 45kHz over 1ms with the Stanton model impulse response presented in Section III A, again using a 300kHz sampling frequency. We also use the same Butterworth filter as the previous section, but because the LFM already effectively band pass filters the signal the effect is minimal. We generate a noisy LFM return by adding white Gaussian noise at a 10dB SNR measured against the average signal power. Lastly, matched filtering is applied to produce the final clean and noisy time series shown in Figure 5. The added noise is more apparent in the spectrum where it mostly obscures the sharp high-Q resonance nulls at 18kHz, 28kHz, 39kHz and 42kHz (18kHz and 28kHz especially).

In this section we will again use the metrics defined in 7, but measured against the clean signal. Notably the early-time portion of the LFM return is approximately 20dB louder than

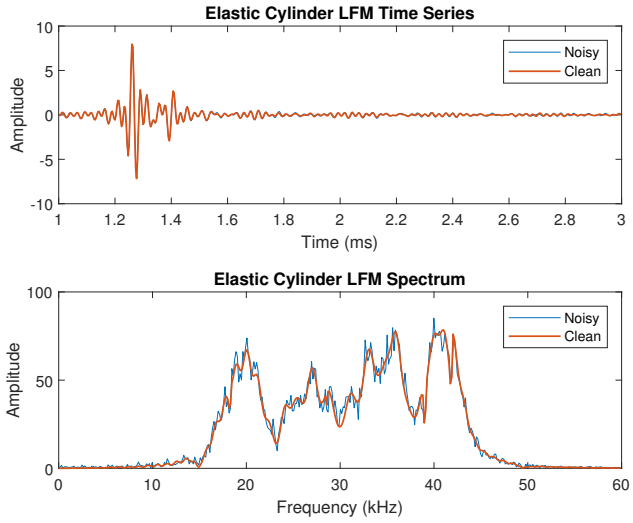


FIG. 5. Noisy and clean matched filtered LFM response for the elastic cylinder (top) and corresponding spectral power (bottom).

the late-time portion. As the SNR of the noisy signal is measured relative to the average signal power, which is dominated by the high power in the early return, the noise is much louder when compared to the quieter late-time return. Specifically, the post-matched filtered SNR of the noisy time series in the early-time is 25dB (where it is benefiting the most from filtering) while the post-matched filtered SNR of the noisy time series in the late-time is -1dB.

FFT MCA

As our signal contains broadband noise we will utilize MCA BPD, which allows for some amount of reconstruction error in order to reduce the amount of noise in the separated signals. Using FFT MCA and BPD with 1000 iterations and $\lambda = 0.01\lambda_{\max}$ we have the results in shown Figure 6. The top plots of Figure 6 compare the noisy signal, the clean

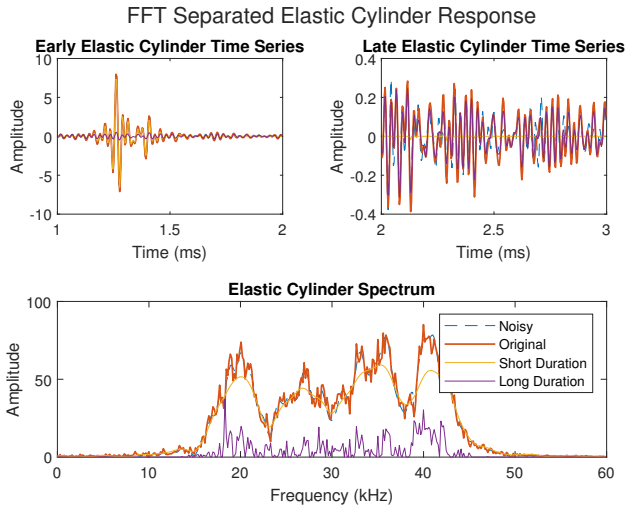


FIG. 6. Clean matched filtered LFM response, noisy matched filtered LFM response and the FFT separated short and long duration components for the elastic cylinder (top) and corresponding spectral power (bottom). Noise was added at a 10dB SNR level. The short-duration component has an early-time relative error of 17.6% and the long-duration component has a late-time relative error of 75.4%.

signal, and the separated short-duration and long-duration components in both the early-time and the late-time while the bottom plot shows the associated spectra. The addition of noise, and the use of an LFM, has had a large impact on the quality of the MCA separation. The early-time error for the short-duration component is 17.6%, and the late-time error for the long-duration component is 75.4%. While the short-duration component looks good in the time domain it is clear from the frequency domain that a significant amount of spectral information has been lost. The long-duration spectrum is also significantly muddled. While the 14kHz peak is still visible, most of the other expected resonance peaks are lost in the noise.

ESP MCA

Next we separate the noisy LFM signal using ESP MCA. While we still use rectangular and exponentially decaying envelopes we will use window lengths and time constants of

$$T_l = 0.07, 0.27, 0.54\text{ms},$$

$$\tau_l = 3.16, 5.62, 10.00, 17.78, 31.62\text{ms}.$$

The main difference between the envelope parameters for this section and Section III A is that we have removed the 0.1ms rectangular window, added a 0.07s window, and dropped the 1.78s time constant. This change is a heuristic choice driven by different characteristics of the LFM signal. Increasing the gap between the longest rectangular window and the smallest exponential time constant decreases the "overlap" between the two ESP frames and tends to make the separation more robust to noise at the cost of performance in the noise-free case. While this illustrates the flexibility of ESP frames, establishing a formal approach for determining optimal envelope parameters is an open question.

If we perform ESP MCA using BPD with 1000 iterations and $\lambda = 0.01\lambda_{\max}$ we get the results in Figure 7. We can see visually that while the separation is still largely successful, particularly in the time domain, the noise and LFM have impacted the clarity of the spectral peaks in the resonant response. The short-duration early-time error is 11.1%. The spectrum of the short-duration component is a better fit to the LFM spectrum than the FFT MCA short-duration component, but is still lacking definition. The late-time long-duration error is still a relatively high 57.3%, but represents a significant improvement over the FFT MCA.

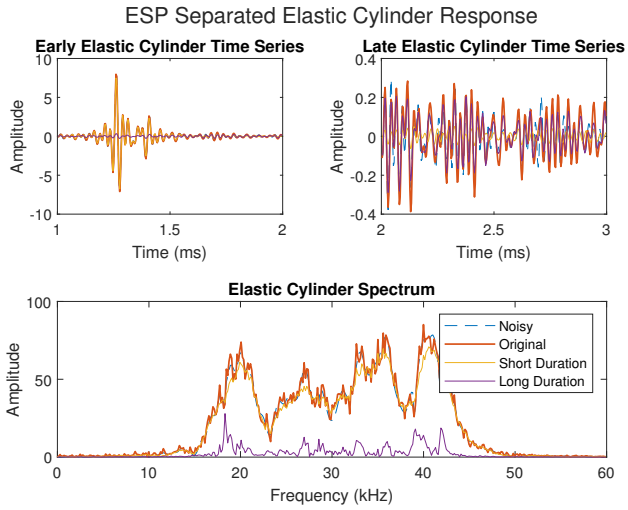


FIG. 7. Clean matched filtered LFM response, noisy matched filtered LFM response and the ESP separated short and long duration components for the elastic cylinder (top) and corresponding spectral power (bottom). Noise was added at a 10dB SNR level. The short-duration component has an early-time relative error of 11.1% and the long-duration component has a late-time relative error of 57.3%.

Additionally, the important resonant peaks are clearer with the ESP approach, especially the ones at 18kHz and 42kHz.

Table II summarizes the results of this section. We see that both methods have a large late-time long-duration component error. This is because the added noise was much louder compared to the quiet ring down, which impacts the performance of BPD. It is also worth nothing that one side effect of BPD is the overall reduction of signal power and most of the signal power lives in the early time. This is an important contributor to the early-time short-duration error. ESP MCA is the best performer of the methods, with significantly lower error rates, particularly in the late-time. This is not unexpected since ESP MCA was

Method	m_1	m_2
FFT MCA	17.6%	75.4%
ESP MCA	**11.1%**	**57.3%**

TABLE II. Short-duration early-time relative error m_1 and long-duration late-time relative error m_2 for FFT MCA and ESP MCA for the noisy elastic cylinder.

designed to fit the underlying signal while FFT MCA is signal agnostic. Of course, these results are for a particular noise level and common choice of λ. The overall performance of an individual approach can be maximized by tuning λ as is done in the next section.

C. Noise Analysis

In order to compare optimal MCA separation between the two techniques, experiments were performed using a range of noise levels and λ-values. Gaussian noise was added to the LFM signal at SNR ranging from -10dB to 30dB with the same signal processing as described in Section III B. For each SNR, 1000 noise realizations were instantiated and each was separated using FFT and ESP MCA with 1000 iterations of BPD for each of the following λ-values

$$\lambda_j = \lambda_{\max} 10^{-3+0.25j} \text{ for } j = 0, \ldots, 11. \tag{9}$$

The early-time relative error in the short-duration component and the late-time relative error in the long-duration component was computed for each combination of SNR, noise

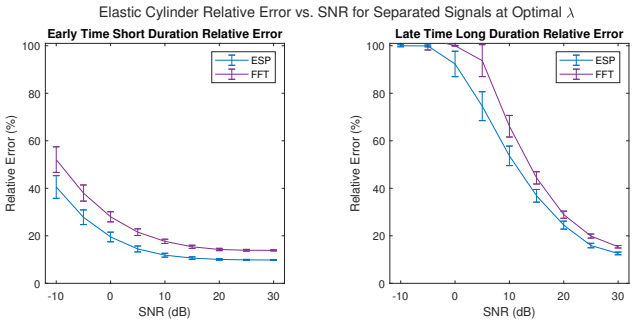


FIG. 8. Early-time short-duration component relative error (left) and late-time long-duration component relative error (right) for the elastic cylinder time series. Data points represents mean error averaged over 1000 noise realizations and the bars represent the standard deviation. The optimal λ was chosen from logarithmically selected λ ranging over $[0.001\lambda_{\max}, 0.5623\lambda_{\max}]$ using 1000 iterations of BPD.

realization, and λ-value. The means and standard deviations of those errors are plotted, for those λ_j which give the minimum error, in Figure 8 as a function of SNR.

The ESP based separation performs better than the FFT based separation. The short-duration errors in the left plot of Figure 8 display typical behavior with larger relative errors and standard deviations for lower SNR. None of the relative errors approach the 100% mark, which is an indication that even at the highest noise levels there is some amount of signal in the separated component. The short-duration component error is smallest for ESP MCA over all SNR levels.

For the long-duration component, at low SNR the relative error is either above 100% with a large standard deviation, indicating that the long-duration component is picking up loud noise, or is very near 100% with little deviation, indicating that the long-duration component is near zero. In either case, any relative errors at or above 100% should be considered

degenerate. We see more reasonable behavior above 5dB SNR with ESP MCA slightly outperforming FFT MCA. Overall ESP MCA does a slightly better job of separation in the face of noise. The FFT MCA performs reasonably well, considering it is signal agnostic, and is computationally less intensive.

Overall this section demonstrates that MCA can be utilized to separate short-duration and long-duration signal components, at least in the case of analytic signals. While the separation was clearest in the case of a clean impulse response, reasonable results were obtained for noisy LFM signals as well. For both scenarios, sharp narrow nulls in the signal spectrum were transformed into peaks in the spectrum of the late-time component, providing a more robust feature for detection and classification. Successful separation was performed with SNR as low as 10dB (-10dB as measured against the late time component), a noise regime which is achievable given current sensors.

IV. AIRSAS SIGNAL SEPARATION

In this section we will apply the MCA techniques presented in Section II to experimentally generated AirSAS time series (Blanford *et al.*, 2019). Experimental AirSAS data was collected using an in-air circular synthetic aperture experiment on two targets: an 8-inch long, 2-inch diameter copper pipe with 0.032-inch thick walls, and an 8-inch long, 2-inch diameter air-filled, hollow copper cylinder with 0.032-inch thick walls and end caps. The targets were centered on a turntable and rotated in 1 degree increments relative to a transducer array consisting of loudspeaker tweeter (Peerless OX20SC02-04) and a microphone (GRAS 46AM), see Figure 9. The tweeter transmits a 1ms 30kHz to 10kHz linear frequency

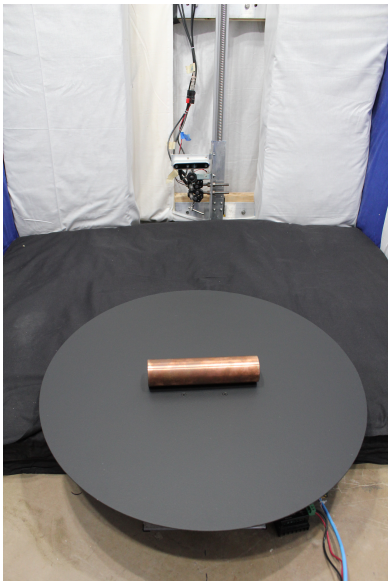


FIG. 9. Photograph of the AirSAS experimental setup.

modulated chirp and the microphone receives the signals backscattered from the target. Motion, timing, signal generation and capture is controlled from a National Instruments data acquisition platform. The recorded signals are first match filtered with the transmitted waveform. Next, the mean of the acoustic signals over all pings is subtracted from each record in order to remove the portions of the signal scattered from stationary objects in the room (stands, supports, lights, etc.). For this paper we only utilize the 3ms to 8ms portion of each time series.

We apply FFT and ESP MCA to the resulting dataset for the 0.032-inch hollow copper cylinder object in Section IV A with a noise analysis on the same dataset in Section IV B. We finish with an application of FFT and ESP MCA to the more complicated time series collected from the 0.032-inch copper pipe in Section IV C.

Despite the fact that the AirSAS data is much more complex than the analytic signal, with multiple returns arriving at different times, for this section we will continue use the

performance metrics from (7). However for the AirSAS cylinder data we use an early time-interval I_1 from 4ms to 6ms and a late-time interval I_2 from 6ms to 8ms. Since each AirSAS object scan includes 360 different time series, we will report the relative error averaged over all aspect angles, which introduces some variation into the error. While we will still view these metrics as a measure of separation performance, the fact that we expect there to be late-time short-duration energy (particularly in Section IV C) means that even in the case of perfect separation we would not expect either m_1 or m_2 to be zero. More broadly these metrics provide only a rough indication of overall performance.

A. 0.032-inch Hollow Copper Cylinder

The first dataset we will consider are the AirSAS time series collected from the 0.032-inch hollow copper cylinder. In many respects, these time series are a reasonable analogue to the Stanton model used in Section III, since at most aspect angles there is a single bright initial return potentially followed by a long-duration low power component. Notably thin walled pipes support a wider class of non-rigid phenomenon than Stanton's model. Various representations of the copper hollow 0.032-inch experimental data are shown in Figure 10. The top left subplot is a logarithmically scaled color plot of the time series amplitude. The bottom left subplot shows the associated normalized target strength, and is the spectra of each time series normalized across all aspect angles. The top right subplot is a logarithmically scaled color plot of the Polar Format Algorithm (PFA) generated image magnitude (Doerry, 2012). The bottom right subplot shows the object's k-space representation, which is the magnitude of the two-dimensional Fourier Transform of the complex PFA image. The

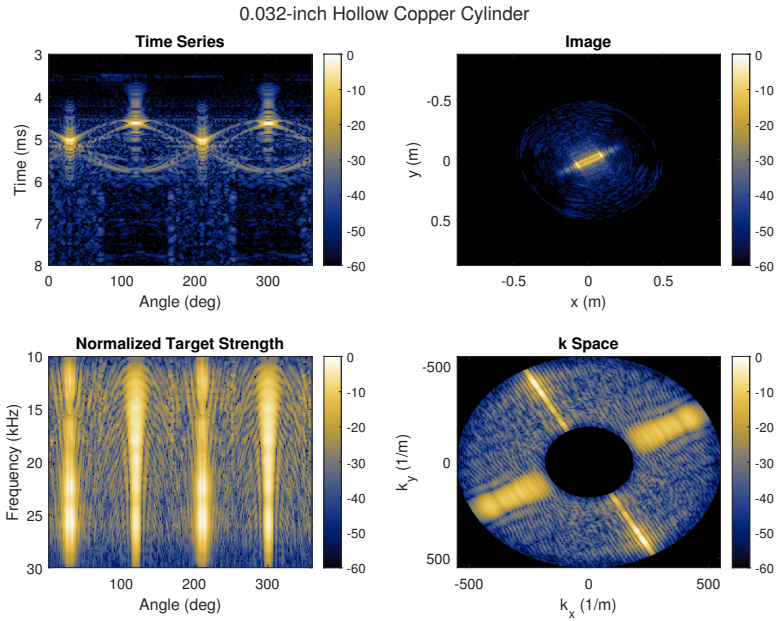


FIG. 10. Time series (top left), PFA image (top right), normalized target strength (bottom left), and k-space (bottom right) for the 0.032-inch hollow copper cylinder. All plots logarithmically scaled.

long-duration signal is clearly present in the time series representation, in bands from -10 degrees to 90 degrees and 180 degrees to 280 degrees. This late time energy is also apparent in the PFA image. Not readily apparent in either of the spectral representations is a faint signature corresponding to this late-time energy.

MCA is broadly compatible with the signal processing and image reconstruction algorithms used with the AirSAS data. For this section, we will apply MCA to the match filtered AirSAS time series individually, splitting each into short-duration and long-duration components. We then apply PFA to the separated time series to reconstruct a pair of images, one corresponding to the short-duration components and the other to the long-duration

components. We produce normalized target strength representations corresponding to the short-duration and long-duration components as well.

FFT MCA

To start with, we will apply FFT MCA using 1000 iterations of BP to the 0.032-inch hollow copper cylinder as described above. Since we are utilizing BP, the separated time series as well as the corresponding PFA images will add up exactly to the original dataset from Figure 10. After image formation, the PFA images associated to the separated short-duration and long-duration components are shown in Figure 11, along with their normalized target strength representations, on a common color scale. The separation looks fairly clean in the time domain image, with the extended ringing response from the cylinder principally in the long-duration image while the brighter geometric scattering response appears in the short-duration image. There does appear to be some bleed through of the object into the long-duration image. The average short-duration early-time error is 44.9% while the average long-duration late-time error is 23.7%. One particularly interesting set of features are the hyperbolic signatures present at 20 degrees and 210 degrees in the long-duration normalized target strength plot, since these signatures were masked by the much brighter short-duration response in the bottom-left plot of Figure 10.

ESP MCA

For ESP MCA we will use the same process as above, applying BP to individual time series and producing a pair of PFA images associated to each component. We continue to

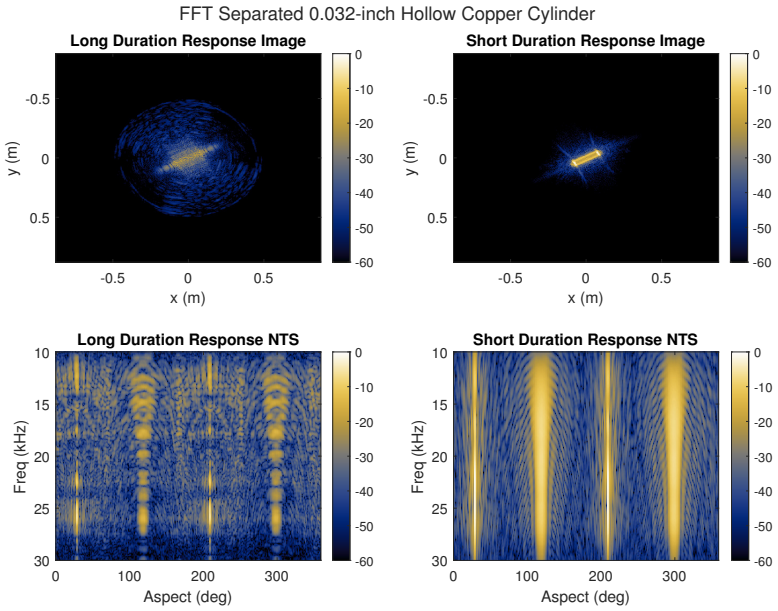


FIG. 11. PFA image (top) and normalized target strength (bottom) for the FFT separated long-duration component (left) and short-duration component (right) of the hollow copper cylinder object. The plots all share a common color scale. The short-duration component has an average early-time relative error of 44.9% and the long-duration component has an average late-time relative error of 23.7%.

use rectangular windows and decaying exponentials as our envelopes, with the same window lengths and time constants as (8). Heuristic experimentation showed these parameters produced reasonable results, although we will see that specific performance characteristics can be attained by using shorter windows and larger time constants. Using BP with 1000 iterations produces the PFA images shown in Figure 12. The separation is quite effective with the bright initial scattering almost completely contained within the short-duration response with an average early-time short-duration error of 19.0%. The long-duration component has most, but not all, of the late-time response and has an average error of 49.0%. Interestingly,

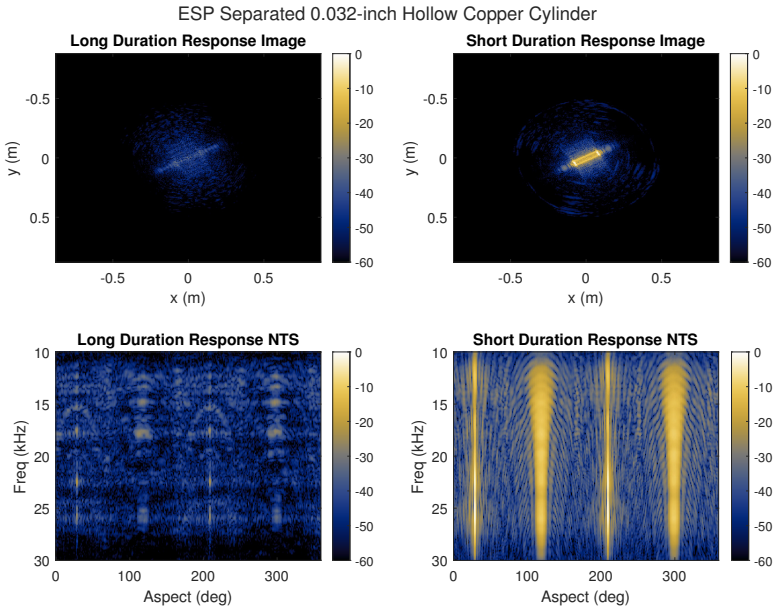


FIG. 12. PFA image (top) and normalized target strength (bottom) for the ESP separated long-duration component (left) and short-duration component (right) of the hollow copper cylinder object. The plots all share a common color scale. The short-duration component has an average early-time relative error of 19.0% and the-long duration component has an average late-time relative error of 49.0%.

there is some late-time energy in the short-duration image at the acoustic coupling angles that was not present in the FFT MCA. This energy does appear to take the form of late arriving wave packets and one theory is that these discrete returns are late arriving pulses associated to surface waves propagating on the cylinder. If this is indeed the case, then it is more appropriate for them to be part of the short-duration image. The fact that they are not present in the long-duration image negatively impacts the long-duration late-time relative error, which is consistent with our goal of separating the signal into early-time and late-time components.

Method	m_1	m_2
FFT MCA	44.9%	**23.7%**
ESP MCA	**19.0%**	49.0%

TABLE III. Short-duration early-time relative error m_1 and long-duration late-time relative error m_2 for FFT MCA and ESP MCA for the copper hollow cylinder.

Overall, compared to the FFT MCA separation we have sacrificed accuracy in the long-duration late-time for increased accuracy in the short-duration early-time, which we see in Table III. As we will demonstrate in Section IV C the ESP frame approach is flexible so this separation could be further tuned by changing the envelope parameters. (Recall that using a one-hot and constant envelope will largely reproduce the FFT result.) Moreover, there is evidence that not all of the short-duration components are early-time, which impacts the reliability of the early-time/late-time error metrics. Visually the separation appears best with the ESP MCA approach as the hyperbolic late-time features are clearer and at a higher relative power.

B. Noise Analysis

In order to understand FFT MCA and ESP MCA separation behavior over a broader range of noise levels, the analysis of Section III C has been reproduced using AirSAS data. Specifically Gaussian noise was added to the time series at each aspect angle with SNR ranging from 10dB to 40dB, all measured against a common reference power. The individual time series were then separated using FFT MCA and ESP MCA using BPD with 1000

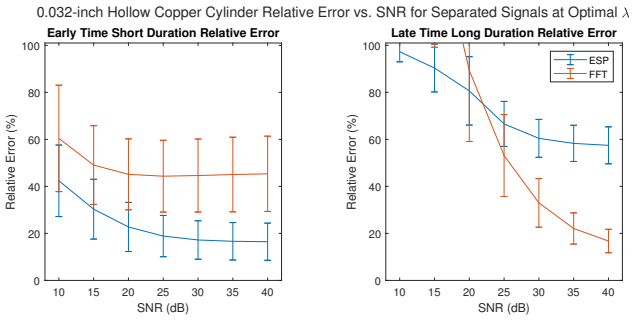


FIG. 13. Early-time short-duration component relative error (left) and late-time long-duration component relative error (right) for copper hollow cylinder AirSAS data. Data points represents mean error averaged over aspect angle and the bars represent the standard deviation. The optimal λ was chosen from logarithmically selected λ ranging over $[0.0001\lambda_{\max}, 0.5623\lambda_{\max}]$ using 1000 iterations of BPD.

iterations and λ-values given by (9). The early-time relative error in the short-duration component and the late-time relative error in the long-duration component was computed for each combination of noise level, and λ-value. We then computed the mean and standard deviation of those errors over all aspect angles. The mean and standard deviation are plotted in Figure 13 as a function of SNR for those λ_j which give the minimum error.

The results in Figure 13 are consistent with the single sample results in the previous section. First, observe that the standard deviations do not decrease as a function of SNR as they do in Section III C. As mentioned above, this is because we are averaging over different time series, not averaging over different noise realizations on the same time series. In terms of mean error, ESP MCA has a consistently lower error than FFT MCA for the early-time short-duration component. Both errors tend to decrease as a function of SNR, as expected.

For the long-duration component the late-time relative error is near or above 100% for SNR less than $\approx$ 15dB for both methods, indicating a failure of separation. There is evidence that ESP MCA starts producing viable separations at 20dB, while both methods produce viable separations by 25dB. However, the FFT MCA separation produces a lower late-time long-duration relative error for SNR greater than $\approx$ 22dB.

More broadly we have demonstrated the ability to separate SAS imagery into distinct morphological components using both FFT MCA and ESP MCA. The FFT MCA approach is superior at producing separations with lower late-time error while ESP MCA has consistently lower early-time error. Importantly both MCA methods were designed to separate short-duration and long-duration components, rather than early-time/late-time components, and will associate the late-time short-duration energy in the AirSAS data with geometric scattering. An open question is if differences in the morphology between early-time and short-duration late-time energy could be used to further the overall goal of early-time/late-time separation.

C. 0.032-inch Copper Pipe

The second data set we will consider are the AirSAS time series collected from the 0.032-inch copper pipe. This dataset is significantly more complex than the 0.032-inch hollow copper cylinder dataset, with obvious short-duration late-time energy present in the time-series. This will exercise the MCA approach to signal separation by demonstrating separation of a long duration signal superimposed on a sequence of repeated short duration signals, but will further decrease the utility of our performance metrics. Figure 10 shows logarithmically

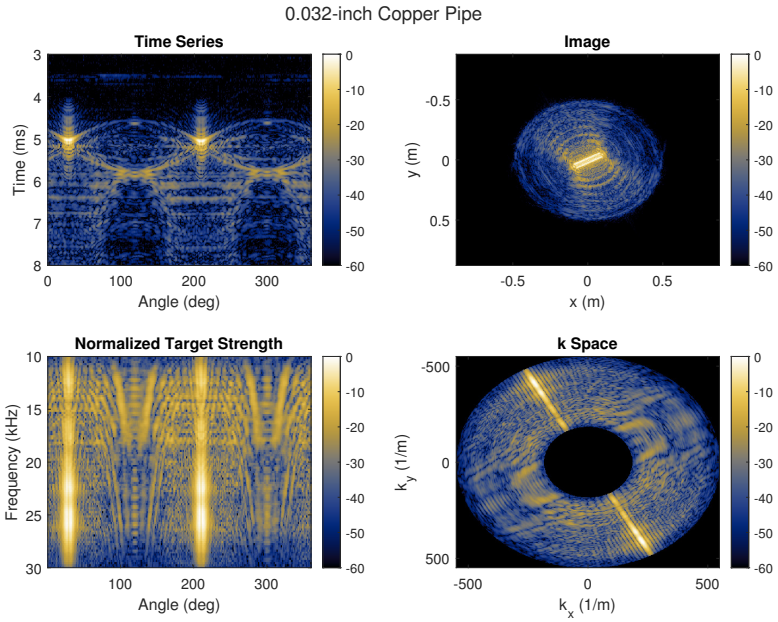


FIG. 14. Time series (top left), PFA image (top right), normalized target strength (bottom left) and k-space (bottom right) 0.032-inch copper pipe. All plots logarithmically scaled.

scaled color plots of the 0.032-inch copper pipe time series data, the associated normalized target strength, as well as the PFA image and its associated k-space representation. The aforementioned late-time short-duration energy is present in discrete "rings" around the object from -10 degrees to 90 degrees and 180 degrees to 280 degrees. We wish to understand how our MCA tools respond to this energy.

FFT MCA

Since FFT MCA is signal agnostic we apply it to the 0.032-inch copper pipe just as in Section IV A. Using 1000 iterations of FFT MCA BP we produce the separated PFA images shown in Figure 15. It is immediately apparent that most of the late-time energy, including the shorter duration late-time "rings," are contained in the long duration image. As a result

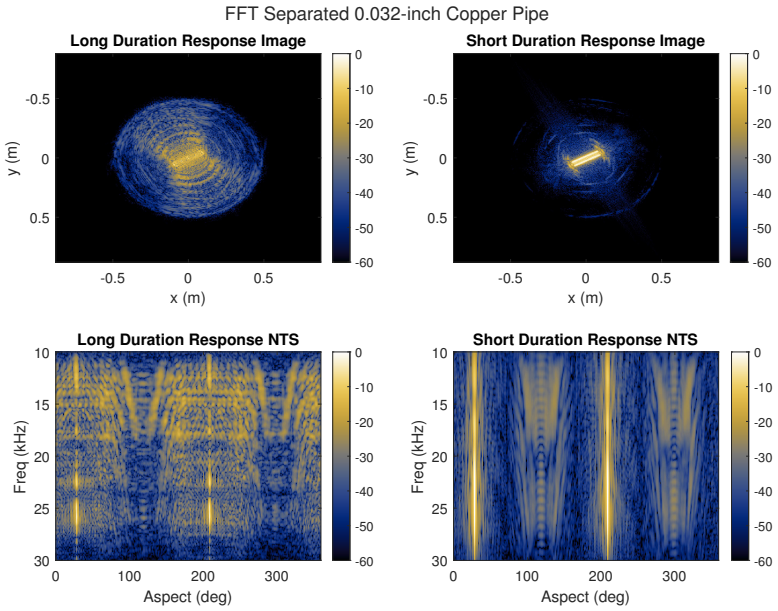


FIG. 15. PFA image (top) and normalized target strength (bottom) for the FFT separated long-duration component (left) and short-duration component (right) of the copper pipe object. The plots all share a common color scale. The short-duration component has an average early-time relative error of 64.1% and the long-duration component has an average late-time relative error of 17.7%.

the average late-time, error for the long-duration component is a relatively low 17.7%, at the cost of a higher 64.1% error for the early-time short-duration component. It is slightly unexpected that these apparently short duration signals can be more sparsely represented in the frequency domain; however, analysis of the spectrum shows that while this late time energy is apparently time limited it is nevertheless not particularly broad band. This is due to the fact that these late arriving wave packets are shaped reflections of the LFM used to ensonify the object.

ESP MCA

Next we will separate the copper pipe time series using the same ESP MCA setup as Section IV A. The short and long-duration PFA images resulting from 1000 iterations of BP are shown in Figure 16. We see that unlike the FFT MCA much of the power of the late-time short-duration wavepackets has been placed in the short-duration PFA image. The separation is a bit muddled overall, although there is clear distributed late-time energy in the long-duration plot over the expected range of angles. For ESP MCA, the long-duration error is worse than the FFT case, with an average late-time error of 62.5%, but the average short-duration early-time error is a better 27.4%. Also notable is that the normalized target strength plots in Figures 15 and 16 emphasize different features. The "V" shaped signatures between 10-20kHz around 130 degrees and 300 degrees seem to move from the short-duration component to the long-duration component.

Traditionally, one would tune the MCA λ_i parameters in order to move signal energy between the separated components to achieve some desired result. While this is often a useful practical step, the determination of λ_i is not obviously connected to the underlying signal characteristics. One of the benefits of the ESP MCA approach is that ESP windows can be tuned based off properties of the signal in question. For example, if it is desired that the long duration component include all (or more) of the late-time energy to improve early-time/late-time separation, then that can be accomplished by shortening the windows available to the short-duration component and including more quickly decaying exponentials in the long-duration representation. This makes it easier for the exponential envelopes

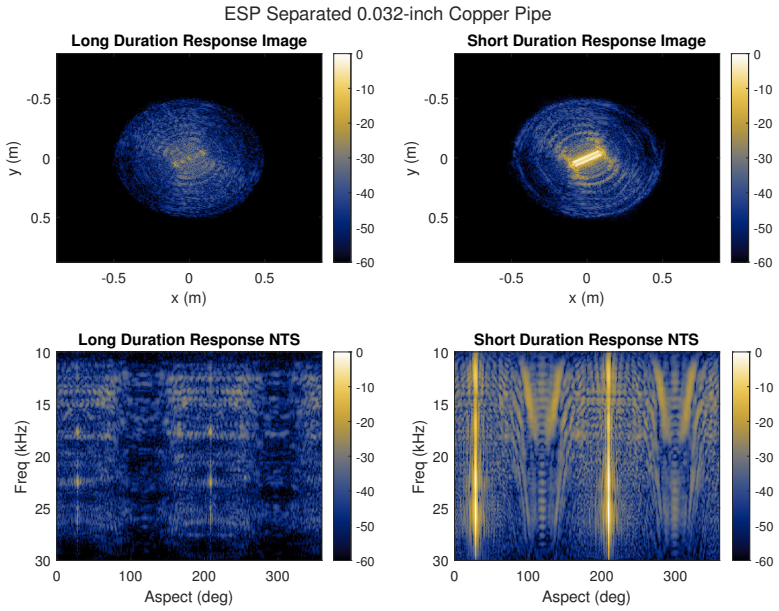


FIG. 16. PFA image (top) and normalized target strength (bottom) for the ESP separated long-duration component (left) and short-duration component (right) of the copper pipe object. The plots all share a common color scale. The short-duration component has an average early-time relative error of 27.4% and the long-duration component has an average late-time relative error of 62.5%.

to sparsely represent shorter signals and more expensive for the rectangular windows to compete for the same short signals. In this case, we shorten our rectangular window lengths considerably and add an additional time constant of 1ms, resulting in:

$$T_l = 0.01, 0.05\text{ms},$$

$$\tau_l = 1.00, 1.78, 3.16, 5.62, 10.00, 17.78, 31.62\text{ms}.$$

Performing 1000 iterations of ESP MCA BP with these parameters results in the separation shown in Figure 17.

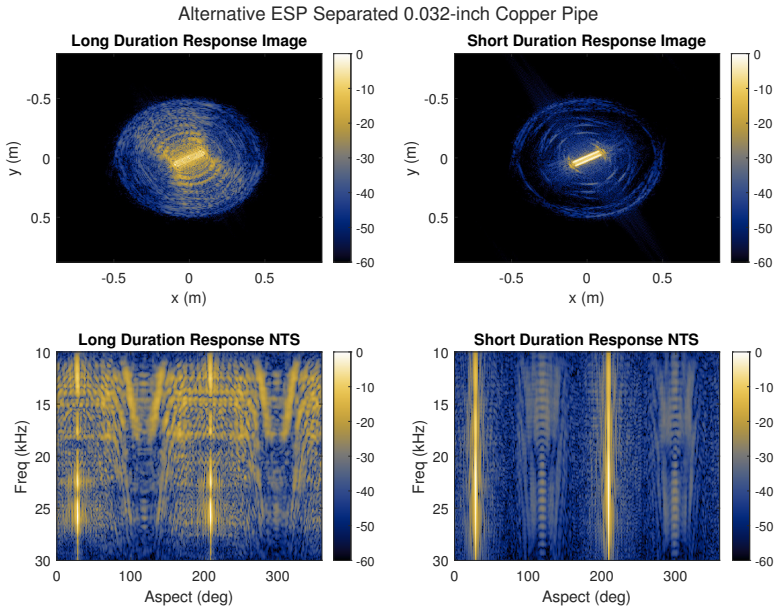


FIG. 17. PFA image (top) and normalized target strength (bottom) for the ESP separated long-duration component (left) and short-duration component (right) of the copper pipe object using an alternative set of envelopes. The plots all share a common color scale. The short-duration component has an average early-time relative error of 74.6% (3 dB) and the long-duration component has an average late-time relative error of 15.97% (16 dB).

In this case we get something that looks more like the FFT separation, with most of the late-time power (including the short-duration components) in the long-duration image. There is more late time energy in the short-duration image for this alternative ESP MCA than FFT MCA and the average short-duration early-time error is larger at 74.6%. The long-duration late-time average error 15.9%, an improvement over FFT MCA. Table IV summarizes the performance metrics for the copper pipe, with the alternative ESP MCA producing the best early time image and the ESP MCA producing the best late time image. In this case, the table does not tell the whole story. The overall performance of each method

Method	m_1	m_2
FFT MCA	64.1%	17.7%
ESP MCA	**27.4%**	62.5%
Alternative ESP MCA	74.6%	**15.9%**

TABLE IV. Short-duration early-time relative error m_1 and long-duration late-time relative error m_2 for FFT MCA and ESP MCA for the copper pipe.

depends on whether the motivating need for MCA signal separation would benefit from late-time short-duration energy grouped with the initial scattering return or with the more diffuse late-time return.

This section demonstrates some of the complexities of applying MCA to experimental data. Performance of MCA representations will depend on the signals in question and determining the correct separation parameters is not always obvious. In the case of ESP MCA, adjusting the envelope parameters can be done in a more principled fashion than adjusting the λ_i-parameters. In either case, finding a way to allow the MCA representations to be data informed, rather than completely model driven, is a potential topic for further research.

V. DISCUSSION

Motivated by the problem of separating the early-time and late-time returns from the acoustic response of an elastic object, we have presented a pair of MCA techniques which

can successfully separate the short-duration and long-duration components without the need for a reference signal or time gates. This partly isolates the late-time returns, with the geometric scattering response generally present in the short-duration component and high-Q resonances present in the long-duration components. Successful separation was achieved on both clean and noisy analytic data as well as experimentally collected in-air sonar time series.

The FFT MCA approach is signal agnostic and does a reasonable job of signal separation, performing very similarly to ESP MCA in the clean analytic signal case (see Table I). FFT MCA also has the benefit of being extremely fast, but is rigid and cannot be tuned to fit a particular signal outside of the traditional λ_i parameters. The ESP MCA application had the best metric scores in analytic time series analysis. ESP MCA has a flexible signal model which can be tuned to support a wide variety of signals as demonstrated in Section IV C. It does take orders of magnitude longer to run than FFT MCA, however.

When applied to noisy signals we found in Section III C that MCA can turn sharp nulls in the spectrum of a time series, which are easily filled in by Gaussian noise, into peaks in the spectrum of the long-duration component, which are easier to identify in the presence of noise. This is most obvious in the Stanton model impulse response (see Figures 3 and 4) but can also be seen in the ESP MCA separation of the noisy LFM (see Figure 7). We also find evidence of this behavior in experimentally collected AirSAS data with the hyperbolic signatures that appear in the long-duration normalized target strength plots for the 0.032-inch hollow copper cylinder dataset (see Figure 12) but are obscured by the

stronger short-duration geometric response in the unseparated data. The ability to utilize features derived from separated components is a on ongoing topic of research.

Additionally, Section IV demonstrates compatibility of MCA with synthetic aperture sonar image reconstruction. The ESP and FFT MCA techniques produced short-duration and long-duration PFA images which split the initial loud response of thin-walled cylindrical objects from the diffuse energy produced by long-duration ringing. This was done without time gating and in the presence of both experimental noise and overlapping returns with varying start times. Notably we have presented examples where long-duration late-time energy is separated from superimposed short-duration late-time energy (see Figure 16). While the performance metric results were less clear cut for the experimental data, ESP MCA was capable of performance similar to FFT MCA (see Table IV) while being significantly more flexible in its signal model. While both MCA approaches were designed to separate short-duration/long-duration components, the ultimate goal is to separate early-time/late-time components in order to preserve the assumptions of the image formation model. It is an open question if the spectral characteristics of early-time versus short-duration late-time responses could be used in an MCA context.

Moving forward there are quite a few potential applications of MCA to acoustic signals. The filtering process could be used in the formation of SAS imagery to either reduce late-arriving energy, allowing for a sharper representation of the object, or to highlight late-time ringing energy, identifying objects with elastic behaviors from those without. Additionally, the spectral peaks resulting from high-Q modes in the long-duration component could be used as features for classification. More broadly, MCA could be used to at least partially

separate any components of a signal which feature sparse representation, such as overlapping acoustic returns from two pings with significantly different spectra, and as such FFT and ESP MCA provide flexible tools for tackling a fundamental acoustics challenge.

Afonso, M. V., Bioucas-Dias, J. M., and Figueiredo, M. A. T. (**2010**). "Fast image recovery using variable splitting and constrained optimization," IEEE Transactions on Image Processing **19**(9), 2345–2356.

Blanford, T. E., McKay, J. D., Brown, D. C., Park, J. D., and Johnson, S. F. (**2019**). "Development of an in-air circular synthetic aperture sonar system as an educational tool," Proceedings of Meetings on Acoustics **36**(1), 070002.

Boyd, S., Boyd, S. P., and Vandenberghe, L. (**2004**). *Convex optimization* (Cambridge university press).

Chen, S. S., Donoho, D. L., and Saunders, M. A. (**2001**). "Atomic decomposition by basis pursuit," SIAM review **43**(1), 129–159.

Doerry, A. W. (**2012**). "Basics of polar-format algorithm for processing synthetic aperture radar images." https://www.osti.gov/biblio/1044949, doi: 10.2172/1044949.

Eckstein, J., and Bertsekas, D. P. (**1992**). "On the douglas—rachford splitting method and the proximal point algorithm for maximal monotone operators," Mathematical Programming **55**(1), 293–318.

Goehle, G., Cowen, B., Park, J. D., and Brown, D. C. (**2022**). "Enveloped sinusoid parseval frames" doi: 10.48550/ARXIV.2204.08418.

615 Hambric, S. A. (**2006**). "Structural acoustics tutorial—part 1: vibrations in structures,"
616 Acoustics Today **2**(4), 21–33.
617 Han, D., Kornelson, K., Larson, D., and Weber, E. (**2007**). *Frames for Undergraduates*
618 (American Mathemnatical Society, Providence, Rhode Island).
619 Kargl, S. G., and Marston, P. L. (**1989**). "Observations and modeling of the backscattering
620 of short tone bursts from a spherical shell: Lamb wave echoes, glory, and axial reverbera-
621 tions," Acoustical Society of America Journal **85**(3), 1014–1028, doi: 10.1121/1.397485.
622 Pareige, P., Maze, G., Izbicki, J., Ripoche, J., and Rousselot, J. (**1989**). "Internal acoustical
623 excitation of shells: Scholte and whispering gallery-type waves," Journal of applied physics
624 **65**(7), 2636–2644.
625 Plotnick, D. S., and Marston, P. L. (**2016**). "Multiple scattering, layer penetration, and
626 elastic contributions to SAS images using fast reversible processing methods," in *European*
627 *Conference on Synthetic Aperture Radar*, pp. 1–3.
628 Selesnick, I. (**2014**). "L1-norm penalized least squares with salsa" http://cnx.org/
629 content/m48933/, connexions.
630 Stanton, T. K. (**1988**). "Sound scattering by cylinders of finite length. ii. elastic cylinders,"
631 The Journal of the Acoustical Society of America **83**(1), 64–67, doi: 10.1121/1.396185.
632 Starck, J.-L., Donoho, D., and Elad, M. (**2004**). "Redundant multiscale transforms and
633 their application for morphological component separation," Technical Report.
634 Starck, J.-L., Elad, M., and Donoho, D. L. (**2005**). "Image decomposition via the combi-
635 nation of sparse representations and a variational approach," IEEE transactions on image
636 processing **14**(10), 1570–1582.

637 Xenaki, A., and Pailhas, Y. (**2019**). "Compressive synthetic aperture sonar imaging with
638 distributed optimization," Journal of the Acoustical Society of America **146**(3), 1839–1850.

Enveloped Sinusoid Parseval Frames

Geoff Goehle*, Benjamin Cowen†, J. Daniel Park‡, and Daniel C. Brown§
Applied Research Laboratory,
Pennsylvania State University,
State College, PA
Email: *goehle@psu.edu, †bmc6220@psu.edu, ‡jdp971@psu.edu, §dcb19@psu.edu

***Abstract*—This paper presents a method of constructing Parseval frames from any collection of complex envelopes. The resulting Enveloped Sinusoid Parseval (ESP) frames can represent a wide variety of signal types as specified by their physical morphology. Since the ESP frame retains its Parseval property even when generated from a variety of envelopes, it is compatible with large scale and iterative optimization algorithms. The ESP construction provides an analysis and synthesis transform pair that can be tuned according to the characteristics of the signal to which it is applied. This work provides examples of ESP frame generation for both synthetic and experimentally measured signals. Furthermore, the frame's compatibility with distributed sparse optimization frameworks is demonstrated. Numerical experiments on acoustics data reveal that the flexibility of this method allows it to be simultaneously competitive with a signal agnostic short-time Fourier Transform in time-frequency processing and also with Prony's Method for time-constant parameter estimation, surpassing the shortcomings of each individual technique.**

I. Introduction

The problem of decomposing a digital signal into oscillating components captures a broad spectrum of applications. Basic signal processing tools such as the Fourier Transform, the short-time Fourier Transform (STFT), wavelets, empirical mode decomposition, and others, generally seek to rewrite the signal as a linear combination of elementary signals or *atoms* [1], [2], [3]. Examination of the signal in the alternative representation can reveal important properties [4], enable specialized denoising [5], [6], improve detection [7] performance, and make parameter estimation easier [8].

However, physical signals often violate the assumptions underlying these techniques, making the resulting decompositions difficult to interpret. Researchers have developed a wide variety of approaches for representing complicated and nuanced signals in terms of elementary components, ranging from common, generic textures [9], [10] to tunable wavelets [2], to fully data-driven methods [11], [12]. In some work a superset of multiple transforms is employed to capture temporally overlapping components of the signal [13], [14].

The modern extension of this approach is to fix an overcomplete basis (or frame) of elementary atoms, then employ convex optimization to infer a weighting of the atoms that best fits the data [15]. Regularization plays an important role in characterizing the solution to overdetermined systems with a common choice being sparsity. Sparse optimization can produce coefficients that are mostly zero, so that only a small subset of available atoms are actually used to compose the signal's representation. Sparse representations are both efficient and interpretable [16].

A different type of approach is to approximate the signal with a physics-based model whose parameters are inferred. These parameters may indirectly define a superposition of atoms, but atom coefficients are not optimized directly. Prony's Method is a classical example that seeks to compute the poles of an infinite impulse response filter whose impulse response best matches the given data [1], indirectly representing the data as a superposition of exponentially decaying sinusoids. This technique is still employed in practice [17] because its model is explainable in terms of the signal morphology. However, the underlying assumption that the signal has a decaying exponential envelope and identification of the signal start time are critical to its performance.

This paper presents a procedure for generating a STFT frame using any number of complex signal envelopes, called the Enveloped Sinusoid Parseval (ESP) frame. The ESP frame uses the same underlying structure as a traditional STFT, but while the usual STFT is largely signal agnostic, an ESP frame can be easily tuned to represent a wide variety of signals by supplying relevant envelopes. It is compatible with modern convex analysis techniques, and because it retains the Parseval property, it is efficient and practical to deploy in large-scale and distributed iterative optimization algorithms.

Section II presents the ESP approach, including formulas for accelerated computation, and implementation details for incorporating it into regularized least squares problems. Section III provides an ESP frame construction for theoretical and experimental time series, and demonstrates ESP as a nonlinear denoising filter in comparison with an STFT frame. Section IV demonstrates the ESP frame's utility in parameter estimation in comparison with Prony's Method.

II. Enveloped Sinusoid Parseval Frames

Section II-A defines the ESP frame for a general set of complex envelopes, and provides formulas for accelerated computation. Optimization algorithms for L_1-regularized least-squares based coefficient inference are presented in Section II-B.

A. Definition

Consider the complex finite-dimensional Hilbert space $\mathbb{C}^N$. All norms ($\|\cdot\|$) are computed in the ℓ_2 sense unless otherwise stated. A *tight frame* is a collection of vectors $\{\mathbf{a}_k\}_{k=0}^{K-1}$ in $\mathbb{C}^N$ and $\alpha > 0$ such that

$$\| \mathbf{w} \|^2 = \alpha \sum_{k=0}^{K-1} |\langle \mathbf{w}, \mathbf{a}_k \rangle|^2 \text{ for all } \mathbf{w} \in \mathbb{C}^N. \tag{1}$$

A *Parseval frame* is a tight frame with $\alpha = 1$. Given a tight frame we define the *analysis operator* to be the linear map from vectors $\mathbf{w} \in \mathbb{C}^N$ to frame coefficients $\mathbf{A}\,\mathbf{w} \in \mathbb{C}^K$ such that

$$Aw[k] = \langle \mathbf{w}, \mathbf{a}_k \rangle. \tag{2}$$

Elements of $\mathbf{A}$ are given by $A[m,n] = \overline{a_m[n]}$. The defining characteristic of tight frames is that the original vector can be reconstructed from the frame coefficients via the formula [18, Prop. 3.11]

$$\mathbf{w} = \frac{1}{\alpha}\,\mathbf{A}^*\,\mathbf{A}\,\mathbf{w} = \frac{1}{\alpha} \sum_{k=0}^{K-1} \langle \mathbf{w}, \mathbf{a}_k \rangle\, \mathbf{a}_k\,. \tag{3}$$

The *synthesis operator* $\mathbf{A}^*$ is the conjugate transpose of the analysis operator, and maps frame coefficients back to the signal space. In the case that $\mathbf{A}$ is a Parseval frame, $\mathbf{A}^*$ serves as the frame's left-inverse.

We define the class of *Enveloped Sinusoid Parseval Frames* by applying enveloping functions to the (non-unitary) Discrete Fourier Transform (DFT) basis. First, specify the envelopes as a collection of vectors $\{\mathbf{e}_l\}_{l=0}^{L-1}$ which are not identically zero. We then have the following from [19].

Theorem 1. *Given a set of nonzero N-dimensional vectors $\{\mathbf{e}_l\}_{l=0}^{L-1}$, the vectors $\{\mathbf{a}_{l,k,m}\}$ defined by*

$$a_{l,k,m}[n] = e_l[n - m \bmod N] \exp(2\pi j k(n-m)/N) \tag{4}$$

for $l = 0, \ldots, L-1$ and $k, m, n = 0, \ldots, N-1$ form a tight frame with $\alpha = N \sum_l \| \mathbf{e}_l \|^2$.

Notably Theorem 1 also follows from viewing the frame as a multi-window STFT. The frame vectors $\mathbf{a}_{l,k,m}$ have indices corresponding to the envelope l, the modulation frequency k, and the circular shift m.

For ESP frames $\mathbf{A}$ and $\mathbf{A}^*$ are very large, so direct computation of matrix products is expensive. However both analysis and synthesis can be sped up significantly using the Fast Fourier Transform (FFT). If we define the vector $\mathbf{c}_{k,l}$ as $c_{k,l}[m] = A^* w[k,l,m]$ one can show, after some matrix algebra, that

$$\mathbf{c}_{k,l} = \mathbf{F}^*\,\mathbf{D}(\mathbf{S}^k\,\mathbf{F}\,\mathbf{H}(\mathbf{e}_l))\,\mathbf{F}\,\mathbf{w}, \tag{5}$$

$$\mathbf{w} = \frac{1}{\alpha}\,\mathbf{F}^* \sum_{k,l} \mathbf{D}(\mathbf{S}^{-k}\,\mathbf{F}\,\mathbf{e}_l)\,\mathbf{F}\,\mathbf{c}_{k,l}, \tag{6}$$

where $\mathbf{F}$ represents the DFT operator, $\mathbf{S}$ right circular shifting, $\mathbf{D}(\mathbf{v})$ the diagonal matrix with entries given by $\mathbf{v}$, and $\mathbf{H}$ the (conjugate linear) operator $Hw[n] = \overline{w[N-n \bmod N]}$. In practical terms, the formulations in (5) and (6) admit parallelization of the time dimensions, and the underlying computations can be sped up using GPU parallelization.

B. Coefficient Inference

Frames, by their nature, do not uniquely represent vectors and for any given vector $\mathbf{w}$ there will be a linear subspace of frame coefficients $\mathbf{c}$ such that $\frac{1}{\alpha}\,\mathbf{A}^*\,\mathbf{c} = \mathbf{w}$. Sparse frame representations are important in the ESP setting because the canonical frame coefficients $\mathbf{A}\,\mathbf{w}$ are computed independently, and thus for a general set of envelopes are expected to contain redundant information. Through application of L_1-regularization, atoms can be made to "compete" with each other and a sparse coefficient vector, with much less redundancy, can be computed that maintains exact reconstruction of the input signal (or in the case of noisy data, maintains some allowable error). In the ideal scenario, regularization can be used to identify superimposed components of a signal formed from a linear combination of frame vectors. This is most effective with highly distinct envelopes.

In formal terms, L_1-regularization entails finding frame coefficients that solve either the Basis Pursuit (BP) problem

$$\arg\min_{\mathbf{c}} \| \mathbf{D}(\boldsymbol{\lambda})\,\mathbf{c} \|_1 \text{ such that } \frac{1}{\alpha}\,\mathbf{A}^*\,\mathbf{c} = \mathbf{w} \tag{7}$$

or the Basis Pursuit Denoising (BPD) problem

$$\arg\min_{\mathbf{c}} \| \mathbf{D}(\boldsymbol{\lambda})\,\mathbf{c} \|_1 + \frac{1}{2} \left\| \frac{1}{\alpha}\,\mathbf{A}^*\,\mathbf{c} - \mathbf{w} \right\|^2, \tag{8}$$

where $\boldsymbol{\lambda} > 0$ is a weight vector that allows the user to control relative penalization between coefficients. Note that $\mathbf{D}(\boldsymbol{\lambda})\,\mathbf{c}$ is equivalent to element-wise multiplication. A constant parameter λ is often used in place of a vectorized $\boldsymbol{\lambda}$.

These convex optimization problems can be solved by the Split Augmented Lagrangian Shrinkage Algorithm (SALSA), which is an instance of the Alternating Direction Method of Multipliers (ADMM) for which convergence is proved [20]. The steps of SALSA [21, Algorithm 4] applied to the BP problem (7) are written in Algorithm 1, where soft$(\cdot)$ denotes the soft-thresholding function

$$\text{soft}(x,T) = \begin{cases} \frac{|x|-T}{|x|}x & |x| > T \\ 0 & |x| \le T. \end{cases}$$

Note that for a Parseval frame $\alpha = 1$ in Algorithm 1.

When the frame does not admit a sparse representation of the signal, as is the case with noisy data, BPD may be used to search for a sparse solution at the cost of reconstruction error. In this case, SALSA is also performed using Algorithm 1 with the addition of a $(1 + \mu/\alpha)^{-1}$ coefficient in the computation of $\mathbf{x}_n$. Both algorithms can be sped up by the use of (5) and (6) for analysis and synthesis, as well as the addition of the predictor-corrector-type acceleration described in [22, Algorithm 8].

For BPD, note that there exists $\lambda_{\max}$ given by

$$\lambda_{\max} = \| \mathbf{A}\,\mathbf{w} \|_\infty \tag{9}$$

Algorithm 1 Basis Pursuit / Basis Pursuit Denoising Algorithm

Initialize $\mu > 0$, $\mathbf{x}_0 = \mathbf{A}\,\mathbf{w}$, and $\mathbf{d}_0 = 0$
if performing BP **then**
  $\beta = 1$
else
  $\beta = (1 + \mu/\alpha)^{-1}$
end if
while stopping criteria not satisfied **do**
  $\mathbf{v}_n = \text{soft}(\mathbf{x}_{n-1} + \mathbf{d}_{n-1}, \boldsymbol{\lambda}/\mu) - \mathbf{d}_{n-1}$
  $\mathbf{x}_n = \mathbf{v}_n + \beta\,\mathbf{A}\left(\mathbf{w} - \frac{1}{\alpha}\,\mathbf{A}^*\,\mathbf{v}_n\right)$
  $\mathbf{d}_n = \mathbf{x}_n - \mathbf{v}_n$
end while

such that for all $\lambda \geq \lambda_{\max}$ the BPD solution is zero [23, Section V.B]. The subsequent BPD experiments set λ as a percentage of $\lambda_{\max}$ with $\lambda = 0.1\lambda_{\max}$ as a common choice.

At convergence, the choice of μ does not impact the solution for Algorithm 1, but it impacts the convergence rate. In the subsequent experiments μ is set to λ/p where p is the 99th percentile of the initial coefficient magnitudes $|x_0[k,l,m]|$. This causes the first soft threshold of either the BP or BPD algorithm to zero-out 99% of the coefficients. In the case of vector-weighted BPD $\mu = \text{mean}(\boldsymbol{\lambda})/p$ is selected.

III. Denoising

In this section the robustness of the ESP frame approach to noise is evaluated and BPD is used to filter noisy time series, both synthetic generated and experimentally collected. Notionally, if the envelopes are chosen so that the ESP frame vectors are a good model for the signal, then the L_1-regularized representation of the clean signal will be sparse. Then, when BPD is applied to the noisy signal, the sparsification of the signal will preferentially remove noise and increase SNR. The ESP frame denoising performance is compared to the same of an STFT-based frame. The selected STFT is windowed with cosine functions so that it is also a Parseval frame as described in [21]. The STFT frame uses a window length of 64, resulting in 2,432 frame coefficients. Finally, the tradeoffs between sparsity and the reconstruction error for BPD regularized ESP and STFT frame coefficients are compared.

For the following sections we will start with either a synthetically or experimentally generated signal $\mathbf{h}$ and will form a noisy signal $\mathbf{h}_N$ by adding white Gaussian noise at some specified SNR (measured against the power of $\mathbf{h}$). We do not expect BPD to exactly reconstruct the original signal, even in the noise-free case, and we track the reconstruction error using the relative error of the reconstructed signal $\mathbf{h}_R$

$$E = \frac{\|\,\mathbf{h}_R - \mathbf{h}\,\|}{\|\,\mathbf{h}\,\|}.$$

The residual $\mathbf{h}_R - \mathbf{h}$ can be used to compute the reconstructed SNR via

$$\text{SNR} = 20\log\left(\frac{\|\,\mathbf{h}\,\|}{\|\,\mathbf{h}_R - \mathbf{h}\,\|}\right).$$

We will look for this reconstructed SNR to produce a gain over the SNR of the added Gaussian noise as an indication that the BPD noise reduction process has been successful.

Section III-A presents denoising applied to a synthetically generated time series using a specially constructed ESP frame. Section III-B presents a similar denoising analysis, but applied to experimentally collected time series. Finally Section III-C describes a comparative analysis between ESP frame denoising and STFT based denoising.

A. Synthetic Data

Consider an ESP frame engineered to detect resonance frequencies, such as those from the transfer function $H : \mathbb{C} \to \mathbb{C}$ defined by

$$H(z) = \frac{z-1}{(z-\alpha)(z-\overline{\alpha})(z-\beta)(z-\overline{\beta})}, \qquad (10)$$
$$\alpha = -1/\tau_a + 2\pi j f_a, \quad \beta = -1/\tau_b + 2\pi j f_b,$$

where $f_a = 5$kHz, $\tau_a = 3$ms, $f_b = 13$kHz, and $\tau_b = 0.8$ms. The synthetic signal $\mathbf{h}$ was generated with $N = 1000$ samples and sampling frequency $f_s = 100$kHz by applying this transfer function to the Kronecker delta vector $\mathbf{v}$ where $v_{50} = 1$ and $v_i = 0$ for $i \neq 50$ (visualized in the top-left subplot of Figure 2). The impulse is chosen so that the synthetic signal starts at 0.5ms. It can be shown via partial fractions [24] that $\mathbf{h}$ is a combination of shifted exponentially decaying sinusoids. This suggests that an ESP frame constructed from exponential envelopes would be appropriate for analysis of this signal.

Given a sampling frequency f_s define the ESP envelopes $\mathbf{e}_l$ such that

$$e_l[n] = \exp\left(-n/(f_s\tau_l)\right) \qquad (11)$$

for $\tau_l > 0$. These envelopes are parameterized by the time constants τ_l and we will use them to construct an ESP frame with atoms consisting of shifted, exponentially decaying, sinusoids. If

$$\tau_l = 10^{l/5-4} \text{ for } l = 0, \ldots, 8$$

then $L = 9$ and the time constants range from 0.1ms to 10ms. From (4) we construct the ESP frame atoms

$$a_{l,k,m}[n] = c_l \exp\left(-\frac{n - m \bmod N}{f_s\tau_l} + 2\pi j\frac{k(n-m)}{N}\right) \qquad (12)$$

where the c_l are chosen so that $\|\,\mathbf{e}_l\,\| = (NL)^{-1/2}$. The vectors $\mathbf{a}_{l,k,m}$ form a Parseval frame by Theorem 1. With this configuration there are $N^2L = 11$ million frame functions parameterized by time constant (0.1ms - 10ms), modulation frequency (-50kHz - 50kHz), and circular time shift (0ms - 5ms).

Since ESP frame coefficients are three dimensional they can be difficult to visualize. In order to produce something similar to a time/frequency analysis (TFA) plot we take the Maximum Intensity Projection (MIP) along the envelope axis and plot the magnitude of the coefficients as functions of the time shift and frequency modulation parameters. As an example, the resulting image for the *unregularized* ESP frame coefficients of $\mathbf{h}$ is

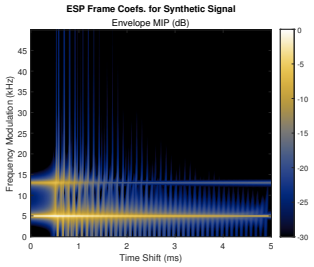


Fig. 1. Plot of the ESP frame coefficients envelope MIP for the synthetic signal $\mathbf{h}$. All intensities are plotted in dB relative to the maximum coefficient amplitude.

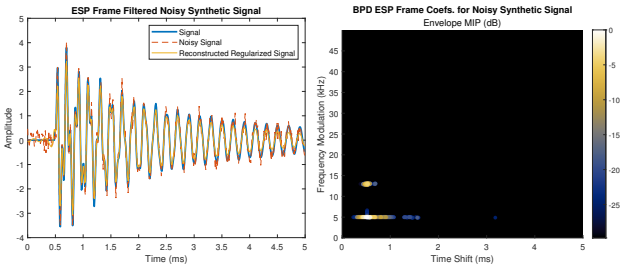


Fig. 2. Plots of the synthetic signal $\mathbf{h}$, $\mathbf{h}_N$ with 10dB SNR, and the signal reconstructed from the BPD solution (left), ESP frame coefficient envelope MIP with oversized pixels (right). The regularization is shown at iteration 1000 with $\lambda = 0.1\lambda_{\max}$. All intensities are plotted in dB relative to the maximum ESP frame coefficient amplitude.

shown in Figure 1. The most important fact about the ESP envelope MIP is that, unlike a traditional TFA, it does not show the amount of signal energy present at a particular time and frequency. Instead it shows the amplitude of a coefficient associated to a particular time shift and frequency modulation of a frame envelope. Indeed, the unregularized ESP frame coefficients behave similarly to a matched filter bank in that the coefficient power will generally have a maximum when the time shift and modulation frequency causes an envelope to correlate to a component in the underlying signal. For instance, the two main frequency peaks of $\mathbf{h}$ are clearly visible in Figure 1, while the time shift peaks are present but broader. As expected, the frequency with the smaller time constant has less power than the frequency with the longer time constant.

BPD on Synthetic Data: Moving on to sparse representations of $\mathbf{h}$, Figure 2 shows the BPD coefficient MIP and corresponding reconstruction of a noisy version $\mathbf{h}_N$ of the synthetic signal. The noisy signal was generated using an SNR of 10dB, and 1000 iterations of BPD were computed with $\lambda = 0.1\lambda_{\max}$. The computed sparse coefficient vector has 2008 nonzero ESP frame coefficients with a sparsity of 99.98%, all clustered around the correct frequencies and time shift. The relative error between the regularized reconstructed signal and the pure signal $\mathbf{h}$ is 17.4% with a reconstructed SNR of 15.2dB, a 5.2dB gain from the initial SNR. The left subplot of Figure 2 reveals that the reconstructed signal has less additive noise, particularly later on in the time series.

Using Algorithm 1 with the STFT frame in place of the ESP frame produces the coefficients shown in Figure 3. The L_1-

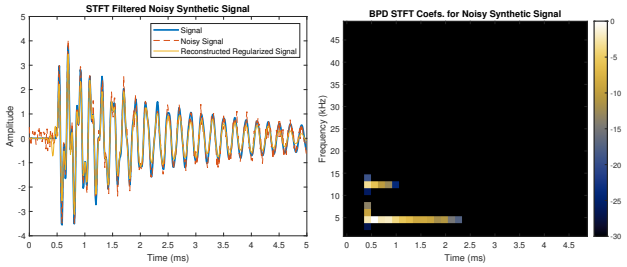


Fig. 3. Plots of the synthetic signal $\mathbf{h}$, $\mathbf{h}_N$ with 10dB SNR and the reconstructed regularized signal (left) and scaled color plot of the STFT frame coefficients for $\mathbf{h}_N$ with a window length of 64 (right). The coefficients have been regularized using BPD with $\lambda = 0.1\lambda_{\max}$ and 1000 iterations. The projected intensities are plotted in dB relative to the maximum frame coefficient amplitude.

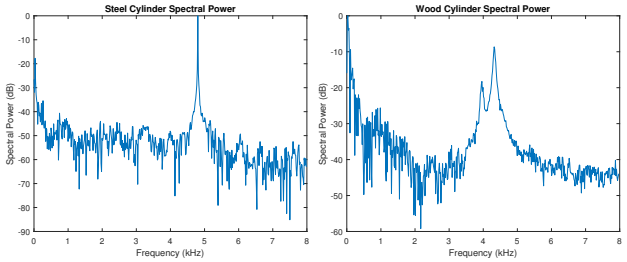


Fig. 4. Spectral power density for the steel cylinder (left) and wood cylinder (right) datasets. The primary peak for the steel cylinder dataset occurs at 4.83kHz. The primary peak for the wood cylinder dataset occurs at 4.35kHz and a secondary peak occurs at 3.94kHz.

parameter $\lambda = 0.1\lambda_{\max}$ was set using the $\lambda_{\max}$ corresponding to the STFT, and 1000 iterations were used. The solution has 58 nonzero coefficients with a sparsity of 97.41%. The relative error between the regularized reconstructed signal and the pure signal is 24.8% with a reconstructed SNR of 12.8dB. The increase in reconstructed SNR is 3dB smaller than the ESP frame case. Of course, this is for single noise level and choice of λ. A more detailed comparative analysis is presented in Section III-C

B. Experimental Data

The datasets for this section were collected by tapping an 8-inch long, 2-inch diameter steel cylinder and wooden cylinder with an impact hammer and recording the emitted sound. The recordings were taken with a sampling frequency of $f_s = 16$ kHz, after downsampling. The hammer was outfitted with a force sensor which triggered the time series to start recording at the moment the hammer impacted the cylinder. Overall 2,200 samples were taken over 0.55 seconds for each tap. The normalized spectral power densities for both cylinders are displayed in Figure 4. The densities were computed using the first 250ms of the corresponding time series.

In order to avoid any transient effects of the impact, the signal used for analysis starts at the 100th sample, with $N = 1024$ samples from the starting point. Since the signal is decaying, the same exponential envelopes $\mathbf{e}_l$ defined in (11)

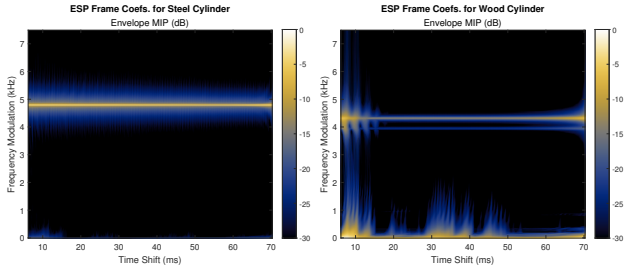


Fig. 5. Frame coefficient envelope MIP for steel cylinder (left) and wood cylinder (right) tap data. Intensities are shown on a dB scale relative to the maximum frame coefficient amplitude.

are appropriate for this analysis, with time constants

$$\tau_l = 10^{l/4-3} \text{ for } l = 0, \ldots, 10,$$

so $L = 11$ and the τ_l range from 1ms to 316ms. About 11.5 million ESP frame vectors follow via (12), and are parameterized by time constant (1ms - 316ms), modulation frequency ($-$8kHz - 8kHz), and time shift (0ms - 70.25ms).

The unregularized steel and wood cylinder ESP coefficient envelope MIPs are displayed in Figure 5. The main response in the 4-5kHz frequency range is clearly visible in both sets of coefficients, as is the secondary frequency response in the wood cylinder time series. However, the coefficients are spread across the entire time shift axis, which is in conflict with the physical setup of the experiment.

BPD on Experimental Data: Next we apply regularization to the experimental time series. Similar to the synthetic case, noisy signals $\mathbf{h}_N$ were generated by adding white Gaussian noise at 10dB SNR. Figure 6 shows the resulting BPD regularized ESP coefficients, using 1000 iterations and $\lambda = 0.1\lambda_{\max}$ as before. The resulting coefficients for the steel cylinder data have a relative reconstruction error of 17.6% with a reconstructed SNR of 15.1dB, a 5.1dB gain. The regularized coefficients have a sparsity of 99.95% with 5745 nonzero coefficients. The coefficients for the wood cylinder data have a relative reconstruction error of 27.6% with a reconstructed SNR of 11.2dB. The regularized coefficients have a sparsity of 99.97% with 2998 nonzero coefficients. The primary frequency responses for both objects are visible, as is the secondary frequency response for the wood cylinder. For the steel cylinder data the frame coefficients are spread across the entire time shift axis, which is somewhat inconsistent with our expectations for the physics of the experiment and may represent intrinsic (non-Gaussian) experimental noise.

Since the regularization process prioritizes sparsity, it tends to filter out both the noise and low magnitude resonant components. This can be seen for both data sets. All of the frequency components are either clustered around the main frequency peaks or are very low frequency, so that quieter frequency responses (such as the small peak at 2.43kHz in the steel cylinder data) have been filtered out. A promising observation is that the reconstructed signal for the wood cylinder data seems to have very few high frequency components after 30ms. This is consistent with the physics of the experiment

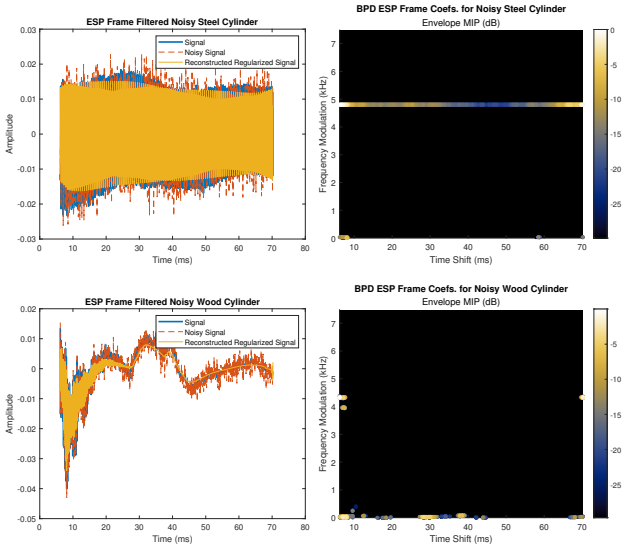


Fig. 6. Plots of $\mathbf{h}$, $\mathbf{h}_N$ with 10dB SNR and the reconstructed regularized signal (left), frame coefficient envelope MIP with oversized pixels (right) for steel cylinder (top) and wood cylinder (bottom) tap data after 1000 iterations with $\lambda = 0.1\lambda_{\max}$. Intensities are shown on a dB scale relative to the maximum frame coefficient amplitude.

since the wood cylinder resonant frequencies have short time constants.

For comparison the regularized STFT frame coefficients are displayed in Figure 7. They are computing using 1000 iterations of BPD and $\lambda = 0.1\lambda_{\max}$. The STFT coefficients exhibit very little power outside the primary frequency responses. The steel cylinder coefficients have a reconstruction error of 17.3% with a reconstructed SNR of 15.2dB and a sparsity of 97.3% with 68 nonzero coefficients. The wood cylinder data has a reconstruction error of 38.6% with a reconstructed SNR of 8.2 dB, a 1.8 dB loss, and a sparsity of 99.1% with 24 nonzero coefficients.

For the steel cylinder data, the STFT frame performance is comparable to that of the ESP frame, with the caveat that the actual number of nonzero coefficients is much smaller. This is less true for the wood cylinder data, where the reconstructed SNR goes from a small gain for the ESP frame to a small loss for the STFT. The STFT frame also visually does a poorer job of reconstructing the time series, particularly the late-time wave shape (see bottom left plot in Figure 7).

C. Denoising Analysis

In the previous sections λ and SNR were fixed. In order to illustrate a broader picture of the sparsity/reconstruction tradeoffs for ESP and STFT frames Figure 8 displays BPD results applied to a variety of initial SNRs and λs. The marker style denotes the algorithm, and colors denote a fixed SNR level ($-$30dB, $-$15dB, 0dB, 15dB, 30dB). A range of percentages of $\lambda_{\max}$ were computed for each algorithm. Each resulting coefficient vector is a point on the graph, where its position along the x-axis corresponds to its percentage of nonzero coefficients. Note $\lambda_{\max}$ is computed separately for each frame using (9), and the y-axis is the reconstructed SNR.

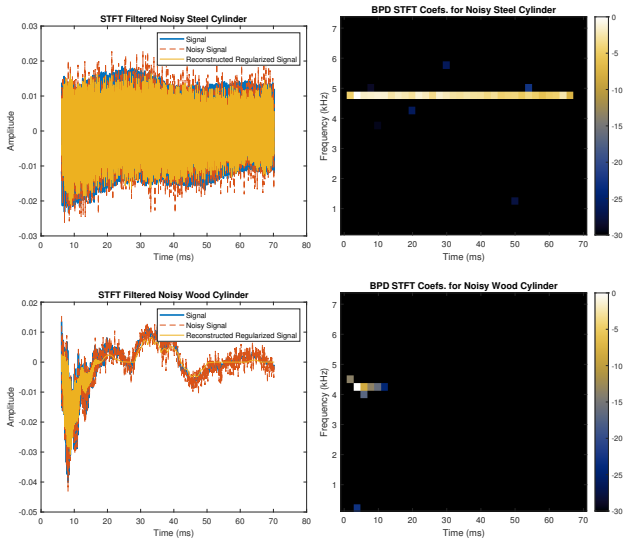


Fig. 7. Plots of $\mathbf{h}$, $\mathbf{h}_N$ with 10dB SNR and the reconstructed regularized signal (left) and scaled color plots of the STFT frame coefficients (right) for steel cylinder (top) and wood cylinder (bottom) tap data after 1000 iterations with $\lambda = 0.1\lambda_{\max}$. Intensities are shown on a dB scale relative to the maximum frame coefficient amplitude.

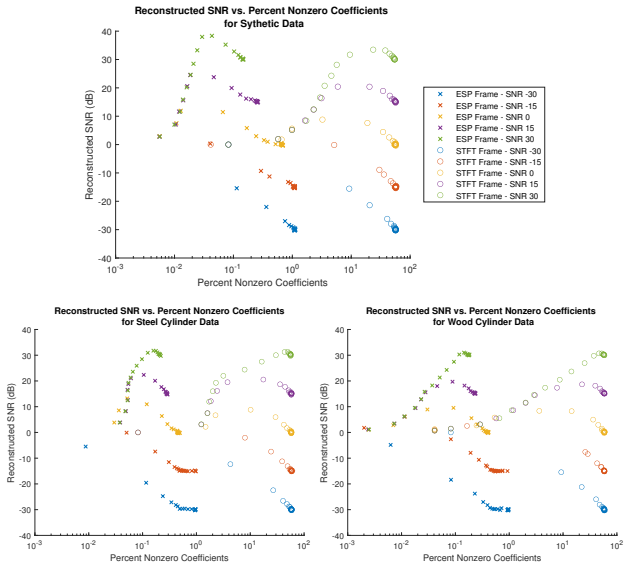


Fig. 8. Reconstructed SNR vs. percentage of nonzero coefficients for the synthetic signal (top), steel cylinder time series (bottom left) and wood cylinder time series (bottom right) using both an ESP frame and a STFT frame over SNR ranging from -30dB to 30dB. The BPD algorithm was applied to the noisy time series using 1000 iterations and λ ranging logarithmically from $0.00001\lambda_{\max}$ to $\lambda_{\max}$. Larger λ-values occur to the *left* of smaller λ-values.

	0dB SNR	15dB SNR	30dB SNR
Synthetic ESP Gain	**11.9**	**9.5**	**8.5**
Synthetic STFT Gain	8.8	5.4	3.4
Steel ESP Gain	**13.2**	**7.4**	**1.8**
Steel STFT Gain	8.8	5.5	1.3
Wood ESP Gain	**9.5**	**4.7**	**1.0**
Wood STFT Gain	8.3	3.7	0.7

Table I. Maximum Reconstructed SNR gain for signals denoised using BPD with ESP and STFT frames.

In general, the STFT frame has a greater percentage of nonzero coefficients than the corresponding ESP frame coefficients, generally falling in the 50-60% range. This is expected because the ESP frame is highly overdetermined and therefore has many more coefficients overall. For small λ the percentage of nonzero ESP frame coefficients ranges from approximately 15% for large SNR to approximately 30% for low SNR.

For both frames, when the SNR of the noisy signal is less than 0dB the reconstructed SNR increases (and the number of nonzero coefficients decreases) as λ grows until the BPD solution becomes zero at $\lambda_{\max}$. This indicates the denoising process was unsuccessful. Conversely, when the SNR of the noisy signal is greater than or equal to 0dB the reconstructed SNR increases as λ grows until it hits some maximum, and then decreases until it reaches zero. In this case, the regularization process is initially removing more noise power than signal power, resulting in an SNR gain, until it hits some optimal λ that depends on the frame and the signal. After this optimal value, further increases in λ cause the regularization process to reduce both noise and signal power until the BPD solution becomes zero at $\lambda_{\max}$.

The gain of the optimal reconstructed SNR over the SNR of the noisy signal is indicative of the frame's ability to denoise and is shown in Table I. The key takeaway is that the optimal SNR gain is consistently larger for the ESP frame than it is for the STFT frame. This is true across all three signals for the range of noise SNR's which produced successful denoising. This maximum SNR gain is greater than indicated by the examples in Sections III-A and III-B, particularly in the case of the wood cylinder data.

Overall we find that the ESP and STFT frames have similar performance with regards to denoising. However, the ESP frame based denoising produces gains which range from 0.3dB to 5.1dB higher than the STFT frame. The ESP frame gain is better for the synthetic signal than for the experimental signals as the ESP frame vectors are more closely aligned to the time series in the synthetic case.

IV. Parameter Estimation

In addition to denoising, another important application of ESP frames is parameter estimation. Since the unregularized ESP frame coefficients are correlation based we expect the coefficients to have peaks when the frame vector is well matched with the signal in question. Our approach will be to use the parameter values associated to these peaks as estimates for feature parameters in the underlying signal. While this

produces an unbiased estimation in the case of a single atom by the Cauchy-Schwartz inequality, in the case of signals with multiple components this estimate is not necessarily unbiased. If the frame is constructed so that the signal is known to have a sparse frame representation then this issue can be addressed via L_1-regularization. Importantly, if the signal does *not* have a sparse frame representation then the L_1-regularization process can introduce significant bias into the parameter estimation process, as we see below.

For this section's analysis we will focus on the identification of resonances in the signals presented in Section III. We use the frames discussed in that section to estimate resonance frequencies and time constants with both unregularized and sparse coefficients. Since the resolution on the time constant axis is poor we will interpolate to get more precise estimates. As the time constants are sampled on a logarithmic scale we use a *geometric* average weighted by the coefficient amplitudes

$$\tilde{\tau} = \left(\tau_{l-1}^{|c_{k,l-1,m}|}\tau_{l}^{|c_{k,l,m}|}\tau_{l+1}^{|c_{k,l+1,m}|}\right)^{\beta} \text{ s.t.} \qquad (13)$$
$$\beta = (|c_{k,l-1,m}| + |c_{k,l,m}| + |c_{k,l+1,m}|)^{-1}.$$

We compare the performance of our ESP frame based estimates to Prony's Method [1], a least squares regression based approach for estimating decaying resonances. Prony's Method assumes the signal can be modeled by a sum of exponentially decaying sinusoids which start at time zero. When this assumption holds, and the order of the least squares regression matches the number of poles in the signal, Prony's Method is capable of producing extremely accurate estimates. However, it is also known that noise and late starting signals can adversely affect Prony's Method. To address these issues, in the case of noise we will be utilizing the SVD-based noise reduction techniques described in [1, Section 11.9]. For late starting signals we will utilize a time shift to ensure the exponential decay starts at time zero. Unlike the ESP frame approach this requires us to know, or estimate, the number of poles and the start time of the exponential decay. It is not expected that ESP frames will outperform Prony's Method in terms of accuracy in optimum conditions. Instead however, the intention is to utilize ESP frames on signals where the number of poles or the start of the resonance component is not known *a priori*.

Section IV-A presents examples of estimating resonance parameters using synthetic time series, as well as a comparison between the parameter estimation performance of ESP frames and Prony's Method. Section IV-B presents a similar analysis using experimental time series and also discusses the use of weighted BPD to encode prior knowledge when generating sparse coefficients.

A. Synthetic Time Series

For the synthetic time series presented in Section III-A we know that $\mathbf{h}$ contains two resonances with frequencies and time constants of 5kHz and 3ms, and 13kHz and 0.8ms that both start at 0.5ms. Using Prony's Method on a shifted version of the clean signal with 4 poles we can recover the parameters for each resonance exactly. However, if we apply Prony's Method directly to $\mathbf{h}$ *without* the shift we find estimates of 5.03kHz and 4.46ms, and 12.86kHz and 1.21ms. For comparison if we use the unregularized ESP frame coefficient peaks near 5kHz and 13kHz and (13) we estimate the frequency and time constants of the two resonances to be 5kHz and 2.52ms and 13kHz and 0.63ms. Here we have recovered the frequency components exactly and the error in the time constants is better than when Prony's Method is applied without a shift. That being said, for all future Prony's Method estimates we will apply any shifts necessary to ensure optimum performance.

While the estimates above were taken from a clean signal, we are generally interested in parameter estimation in the presence of noise. If we add noise at 10dB SNR, as described in Section III, and use Prony's Method (with shifting and 30 poles filtered to 4 using SVD) to estimate the resonance parameters we get 5.01kHz and 2.50ms, and 12.95kHz and 0.55ms. For comparison if we compute the BPD regularized ESP frame coefficients as in Figure 2 we obtain estimates of 5kHz and 2.48ms, and 13kHz and 0.95ms. In this case the regularized BPD approach does a slightly better job of estimating the frequency parameters and is about as accurate as Prony's Method at estimating the time constants.

Noise Analysis: While the examples above indicate that ESP frames can be reasonably utilized as a parameter estimation tool, a further comparative analysis with Prony's Method is warranted. Specifically we wish to compare the bias and variance of ESP frame and Prony's Method based parameter estimates in the presence of added noise. To this end noise was added to the synthetic time series at levels ranging from -15dB SNR, resulting in predominantly noise, to 30dB SNR, resulting in predominantly signal. At each noise level the frequency and time constant of both resonance peaks was estimated using unregularized ESP frame coefficients, sparse ESP frame coefficients, and Prony's Method. For the unregularized ESP frame coefficients the parameter estimates were generated using the coefficient peaks near 5kHz and 13kHz and (13). The BPD regularized coefficients were generated using $\lambda = 0.1\lambda_{\max}$, 1000 iterations, and the same parameter estimation process as the unregularized case. Finally for Prony's Method we use a shifted version of the noisy time series with 30 poles filtered to 4 via SVD. Each of these estimates was generated for 100 noise realizations and the resulting estimation mean and standard deviation are plotted in Figure 9 with bias and standard deviation values at the 30dB level shown in Table II.

The frequency estimates have similar performance, for both bias and standard deviation, across all three estimation techniques and produce quality estimates at or above 0dB SNR. The one exception is the Prony's Method 13kHz peak frequency estimate at 0dB SNR, which has a notably larger bias and standard deviation. There is comparatively more variability in the time constant estimates. None of the methods produces viable estimates at the -15dB SNR level. At 0dB SNR the ESP frame based estimates are significantly better

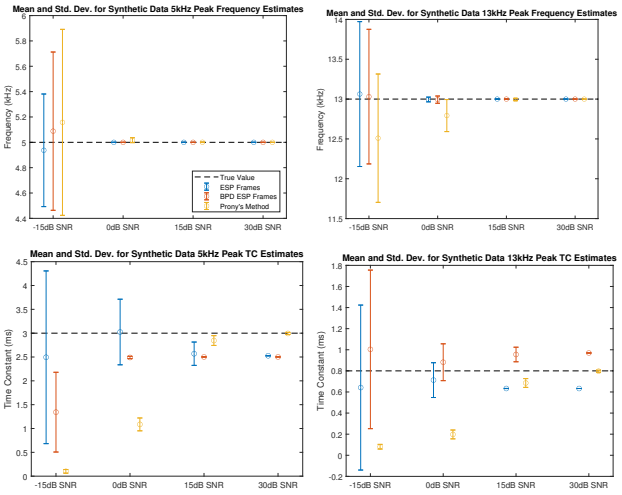


Fig. 9. Means and standard deviations for ESP frame based and Prony's Method based estimates of resonance frequency (top) and time constant (bottom) for the synthetic time series 5kHz peak (left) and 13kHz peak (right). Mean is indicated by plotted point and standard deviation by the length of the whiskers. True parameter values are indicated by the dashed line. The mean and standard deviation were computed using 100 estimates from signals with added noise at the indicated SNR.

	5kHz Res. Freq.	13kHz Res. Freq.	5kHz Res. TC	13kHz Res. TC
ESP Bias	**0.0**	**0.0**	−0.47385	−0.16711
ESP Std. Dev.	**0.0**	**0.0**	0.00018629	9.1013e−5
BPD ESP Bias	**0.0**	**0.0**	−0.49882	0.16968
BPD ESP Std. Dev.	**0.0**	**0.0**	0.00070589	0.0010405
Prony Bias	−1.9896e−5	−0.00046569	**−0.006141**	**−0.003243**
Prony Std. Dev.	0.00034626	0.0022198	**0.02189**	**0.0096209**

Table II. Bias and standard deviations for ESP frame and Prony's Method parameter estimates at 30dB SNR. Frequency values are in kHz and time constant (TC) values are in ms.

than the Prony's Method based estimate. At 15dB SNR all three methods have similar performance while the Prony's Method estimates are significantly better at the 30dB SNR level, as can be seen in Table II.

B. Experimental Time Series

While we know the true values of the resonant frequencies and time constants for the synthetic data, and were able to precisely control the amount of added noise, we do not have this luxury for the experimental time series. Instead we will use a Prony's Method based estimate created using the original experimental time series as the "true" value for the primary resonance time constant for the steel and wood cylinder data. Using Prony's Method with 16 poles filtered to 8 using SVD we estimate the steel cylinder resonance at 4.80kHz has a time constant of 166.8ms and the wood cylinder resonance at 4.32kHz has a time constant of 5.22ms. As a point of comparison, if we use the unregularized ESP frame coefficients, Figure 5, to estimate the frequency and time constant we get 4.80kHz and 177.87ms for the steel cylinder data and 4.32kHz and 5.66ms for the wood cylinder data. There is good agreement between the frequency estimates and the time constant estimates are reasonably close. This is a positive indication that we will be able to utilize the

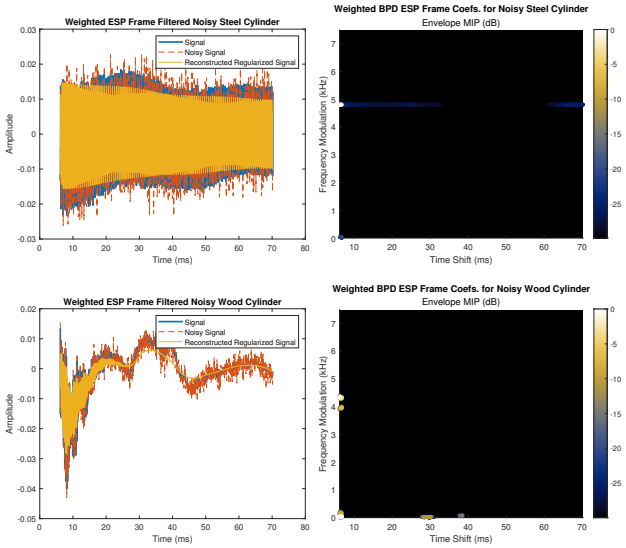


Fig. 10. Plots of $\mathbf{h}$, $\mathbf{h}_N$ with 10dB SNR and the reconstructed regularized signal (left), frame coefficient envelope MIP with oversized pixels (right) for steel cylinder (top) and wood cylinder (bottom) tap data after 1000 iterations with $\boldsymbol{\lambda} = \lambda_{\max}\mathbf{w}$. Intensities are shown on a dB scale relative to the maximum frame coefficient amplitude.

unregularized coefficients for parameter estimation. On the other hand, if we apply the same estimation process to the BPD regularized coefficients, Figure 6, we get an estimate of 4.80kHz and 54.95ms for the steel cylinder and 4.33kHz and 5.72ms for the wood cylinder. While the wood cylinder estimate is similar in both cases, the time constant for the steel cylinder estimate is significantly biased as a result of the regularization.

Weighted Basis Pursuit Denoising: One potential method for dealing with the bias introduced by regularization, and to account for the fact that the regularization does not produce coefficients consistent with our understanding of the physics of the experiment, is to utilize a vectorized $\boldsymbol{\lambda}$ in Algorithm 1. We know from the physical setup of the experiment that, outside of noise components, the signal should have a zero time shift. We can encode this prior knowledge into the BPD process by nonconstant $\boldsymbol{\lambda}$. Specifically define $\mathbf{w}$ such that

$$w_{l,k,m} = \begin{cases} 0.1 & m < 10 \\ 0.2 & m \geq 10. \end{cases}$$

This vector places extra weight on coefficients with a time shift greater than or equal to 0.156ms, increasing the cost of using these coefficients in the optimization. We allow time shifts of up to $m = 10$ so that the signal may start with varying phase. We then define $\boldsymbol{\lambda} = \lambda_{\max}\mathbf{w}$ and apply BPD. Using 1000 iterations produces the set of coefficients shown in Figure 10.

It is clear that the weighting is successful at concentrating the frame coefficients at smaller time shifts. For the wood cylinder most of the power is in the first millisecond, however there are some coefficients at 30ms and 40ms which correspond to similar late time energy seen in the unregularized STFT (not shown). The reconstruction accuracy of

Dataset	Recon. Error (%)	Res. Freq. (kHz)	Res. TC (ms)
Steel	0.0	4.80	177.9
BP Steel	17.6	4.80	**54.95**
WBP Steel	24.0	4.80	177.63
Wood	0.0	4.32	5.66
BP Wood	27.6	4.33	5.72
WBP Wood	34.7	4.33	5.84

Table III. Reconstruction error and estimated resonance frequency and time constant for unregularized, BP regularized and weighted BP regularized ESP frame coefficients for steel and wood cylinder time series data.

	Wood Res. Freq.	Steel Res. Freq.	Wood Res. TC	Steel Res. TC
ESP Bias	-0.00031189	$\mathbf{0.0031189}$	$\mathbf{-0.00039911}$	$\mathbf{-0.026285}$
ESP Std. Dev.	0.0	$\mathbf{0.0}$	$\mathbf{0.0016578}$	$\mathbf{0.004194}$
BPD ESP Bias	0.015283	0.0031189	0.1756	-0.21042
BPD ESP Std. Dev.	0.0	0.0	0.0068052	0.020472
Prony Bias	$\mathbf{-6.3537e-5}$	0.0038046	-0.56135	-17.662
Prony Std. Dev.	$\mathbf{0.00052306}$	$3.9251e-5$	0.068351	5.3159

Table IV. Bias and standard deviations for ESP frame and Prony's Method parameter estimates at 30dB SNR. Frequency values have are in kHz and time constant (TC) values are in ms.

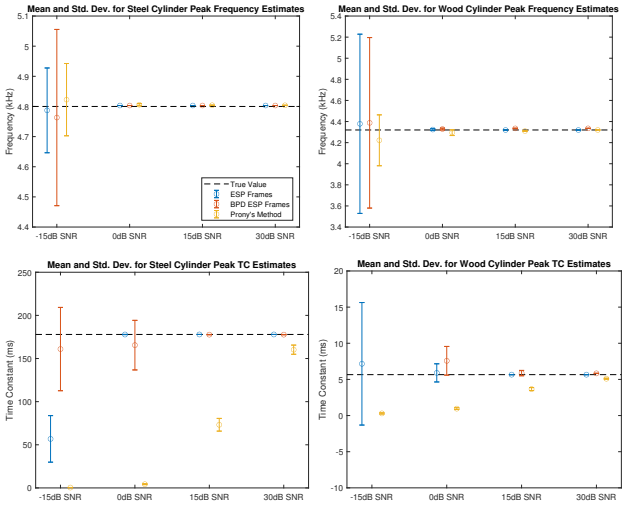


Fig. 11. Means and standard deviations for ESP frame based and Prony's Method based estimates of resonance frequency (top) and time constant (bottom) for the steel cylinder (left) and wood cylinder (right) experimental time series. Mean is indicated by plotted point and standard deviation by the length of the whiskers. The best measured value is indicated by the dashed line. The -15dB SNR BPD ESP estimate mean for the wood cylinder data is outside the scale of the plot. Mean and standard deviation were computed using 100 estimates from signals with added noise at the indicated SNR.

the weighted basis pursuit coefficients is worse than for the unweighted basis pursuit coefficients, see Table III, due to the increase in the average value of $\boldsymbol{\lambda}$. More importantly the weighting improves the time constant parameter estimation for the steel cylinder to 177.63ms, while the estimate for the wood cylinder resonance time constant is consistent across all estimation techniques.

Noise Analysis: As part our investigation of ESP frames as a parameter estimation tool we performed the same analysis presented in Section IV-A on the experimental time series. Specifically we added noise to the experimental time series at levels ranging from -15dB SNR to 30dB SNR. At each noise level the frequency and time constant of each resonance was estimated using unregularized ESP frame coefficients, *weighted* BPD regularized ESP frame coefficients, and Prony's Method. For the unregularized ESP frame coefficients the parameters were estimated using (13). The regularized coefficients were generated using $\boldsymbol{\lambda} = \lambda_{\max}\,\mathbf{w}$ and 1000 iterations, as in the previous section. Finally for Prony's Method we use 16 poles filtered to 8 via SVD. Each of these estimates was generated for 100 noise realizations and the resulting estimated means and standard deviations are plotted in Figure 11 with the bias and standard deviations at 30dB SNR shown in Table IV. Recall the "true" parameter values were generated using Prony's Method as described at the start of Section IV-B.

As in Section IV-A all three methods have similar behavior with regards to estimation of the frequency parameter, producing quality estimates at or above the 0dB SNR level. None of the methods produces a viable time constant estimate at -15dB SNR. At 0dB SNR the ESP frame based estimates are reasonably close to the true value while the Prony's Method estimates are near zero. At the 15dB and 30dB SNR levels the ESP frame based estimates have a significantly smaller bias than Prony's Method, see Table IV. Overall we find that ESP frames can be used to estimate resonance parameters and are competitive with Prony's Method, particularly in the presence of noise.

V. Discussion

This paper presented a method of constructing Parseval frames from any collection of complex envelopes. The resulting ESP frames can represent a wide variety of signal types as specified by their physical morphology. Since the ESP frame retains its Parseval property it is compatible with large scale and iterative optimization algorithms such as SALSA, and sparse sets of ESP frame coefficients can be generated using traditional convex optimization. This work presented examples of ESP frame generation, as well as L_1-regularized coefficient generation, for both synthetic and experimentally measured signals. The use of sparse coefficients for both denoising and parameter estimation was also demonstrated.

When seeking sparse sets of ESP frame coefficients we generally expect the signal will not exactly equal a small linear combination of frame vectors. This can be due to the presence of noise or because of poor resolution in the envelope parameter. While noise can be mitigated using any number of techniques, including BPD, the fact that the dimension of the ESP frame is given by N^2L, where L is the number of envelopes, means that achieving a very fine resolution along the envelope axis can be computationally infeasible. In either case, we expect that the optimal set of BP coefficients may not be particularly sparse and that high levels of reconstruction error may be needed to produce sparse coefficients using BPD. For many applications, though, it is not necessary to achieve true sparsity and instead we simply desire the ESP frame coefficients to produce discrete peaks.

With regards to denoising, we found that ESP frames are competitive with the STFT as a noise reduction tool, producing larger SNR gains over a range of noise levels. While in terms

of percentages the ESP Frame representations were sparser than the STFT representations, because the ESP frame is so large the STFT frame regularization ends up producing significantly fewer nonzero coefficients. Additionally the ESP frame approach also takes longer to converge and is more computationally intensive. In the ideal scenario we expect that ESP frame based denoising will outperform STFT based denoising since the ESP frame can be used to encode a desired signal model while the STFT is signal agnostic.

We also found that ESP frames are competitive with Prony's Method when applied to resonance parameter estimation. The unregularized ESP frame coefficients perform about as well as the regularized ESP frame based estimates across all test cases while being less computationally intensive. The ESP frame approaches produced viable estimates of the time constant parameter at a lower SNR than Prony's Method. At very high SNR Prony's Method produced better estimates in the synthetic time series case, while the ESP frame estimates were better for the experimental time series. This is consistent with the fact that Prony's Method is very accurate when its underlying signal model is a good match for the time series. It is not thought that ESP frames will outperform Prony's Method in terms of accuracy in optimal conditions. Instead however, the intention is to utilize ESP frames on signals where the number of poles or the start of the resonance component is not known *a priori*.

There are a number of possible future applications for ESP frames, ranging from Morphological Component Analysis to generating feature sets for use in signal classification [19]. Another potential avenue of investigation is to try and allow the envelope parameter to vary as part of the L_1-regularization procedure. This could enable the ESP frame envelopes to be more data informed and may further enhance sparsity. Overall, Enveloped Sinusoid Parseval frames are a flexible signal analysis tool, particularly when combined with convex optimization, and offer a wide range of applications. The ESP frame can be easily tuned to represent a wide variety of signals simply by providing their relevant envelopes. It is compatible with modern convex analysis techniques, and is efficient and practical to deploy.

ACKNOWLEDGMENT

This work was sponsored in part by the Department of the Navy, Office of Naval Research under ONR award numbers N00014-18-1-2820 and N00014-19-1-2221.

REFERENCES

[1] S. L. J. Marple, *Digital Spectral Analysis*. Mineola, New York: Dover Publications, Inc., 2019.

[2] I. W. Selesnick, "Wavelet transform with tunable Q-factor," *IEEE transactions on signal processing*, vol. 59, no. 8, pp. 3560–3575, 2011.

[3] P. Flandrin, G. Rilling, and P. Goncalves, "Empirical mode decomposition as a filter bank," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 112–114, 2004.

[4] G. Cai, I. W. Selesnick, S. Wang, W. Dai, and Z. Zhu, "Sparsity-enhanced signal decomposition via generalized minimax-concave penalty for gearbox fault diagnosis," *Journal of Sound and Vibration*, vol. 432, pp. 213–234, 2018.

[5] H. Guo, J. E. Odegard, M. Lang, R. A. Gopinath, I. W. Selesnick, and C. S. Burrus, "Wavelet based speckle reduction with application to SAR based ATD/R," in *Proceedings of 1st international conference on image processing*, vol. 1. IEEE, 1994, pp. 75–79.

[6] I. W. Selesnick and M. A. Figueiredo, "Signal restoration with overcomplete wavelet transforms: Comparison of analysis and synthesis priors," in *Wavelets XIII*, vol. 7446. International Society for Optics and Photonics, 2009, p. 74460D.

[7] A. Parekh, I. W. Selesnick, D. M. Rapoport, and I. Ayappa, "Detection of k-complexes and sleep spindles (DETOKS) using sparse optimization," *Journal of neuroscience methods*, vol. 251, pp. 37–46, 2015.

[8] P. Gerstoft, A. Xenaki, and C. F. Mecklenbräuker, "Multiple and single snapshot compressive beamforming," *The Journal of the Acoustical Society of America*, vol. 138, no. 4, pp. 2003–2014, 2015.

[9] G. Easley, D. Labate, and W.-Q. Lim, "Sparse directional image representations using the discrete shearlet transform," *Applied and Computational Harmonic Analysis*, vol. 25, no. 1, pp. 25–46, 2008.

[10] J. Ma and G. Plonka, "The curvelet transform," *IEEE signal processing magazine*, vol. 27, no. 2, pp. 118–133, 2010.

[11] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proceedings of the 26th annual international conference on machine learning*, 2009, pp. 689–696.

[12] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[13] M. Elad, J.-L. Starck, P. Querre, and D. L. Donoho, "Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)," *Applied and computational harmonic analysis*, vol. 19, no. 3, pp. 340–358, 2005.

[14] B. Singh and H. Wagatsuma, "A removal of eye movement and blink artifacts from EEG data using morphological component analysis," *Computational and mathematical methods in medicine*, vol. 2017, 2017.

[15] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, vol. 57, no. 11, pp. 1413–1457, 2004.

[16] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.

[17] N. Karpushkin, F. Guber, and A. Ivashkin, "Application of the prony least squares method for fitting signal waveforms measured by sampling ADC," *AIP Conference Proceedings*, vol. 2163, no. 030006, 2019.

[18] D. Han, K. Kornelson, D. Larson, and E. Weber, *Frames for Undergraduates*. Providence, Rhode Island: American Mathemnatical Society, 2007.

[19] G. Goehle, B. Cowen, T. E. Blanford, J. D. Park, and D. C. Brown, "Approximate extraction of late-time returns via morphological component analysis," 2022, Journal of the Acoustical Society of America, Submitted.

[20] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2345–2356, 2010.

[21] I. Selesnick, "L1-norm penalized least squares with SALSA," 2014, connexions. [Online]. Available: http://cnx.org/content/m48933/

[22] T. Goldstein, B. O'Donoghue, S. Setzer, and R. Baraniuk, "Fast alternating direction optimization methods," *SIAM Journal of Imaging Sciences*, vol. 7, no. 3, pp. 1588–1623, 2014.

[23] A. Xenaki and Y. Pailhas, "Compressive synthetic aperture sonar imaging with distributed optimization," *Journal of the Acoustical Society of America*, vol. 146, no. 3, pp. 1839–1850, 2019.

[24] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1989.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
01-09-2022	Final	Aug 01, 2018 - July 31, 2022

4. TITLE AND SUBTITLE
Spatial Coherence of Seabed Volume Reverberation

5a. CONTRACT NUMBER

5b. GRANT NUMBER
N00014-18-1-2820

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Daniel C. Brown

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The Pennsylvania State University Applied Research Labotatory
Office of Sponsored Programs
110 Technology Center Building
University Park, PA 16802-7000

8. PERFORMING ORGANIZATION REPORT NUMBER
TR-22-005

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
875 North Randolph Street
Arlington, VA 22203-1995

10. SPONSOR/MONITOR'S ACRONYM(S)
ONR

11. SPONSORING/MONITORING AGENCY REPORT NUMBER

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for Public Release; Distribution is Unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This program focused on the development of improved physics-based models for and analysis of time-series sensed by synthetic aperture sonar systems. These physics-based models were utilized to address two areas: (1) the spatial coherence of signals sensed by a bistatic, long-range SAS and (2) the sparse representation of signals defined by a physics-based basis. The development of coherence models supports a better understanding of how the sensor's design and employment impact the second cross-moment of the scattered field. This has an immediate benefit for SAS motion estimation and longer-term benefits for interpreting signal coherence to detect and characterize the imaged scene. The development of sparse signal representations provides a method for separation of time-series components for subsequent analysis. Developing the bases using physics-based methods allows subsequent signal processing and analysis to be tailored to specific physical responses.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE	SAR		Dan Brown
U	U	U			19b. TELEPONE NUMBER (*Include area code*) (814) 865-1193