

Detecting the direction of arrival and time of arrival of impulsive transient signals

Too Yuen Min, Mandar Chitre, and Venugopalan Pallayil
 Acoustic Research Laboratory, Tropical Marine Science Institute
 National University of Singapore, 12A Kent Ridge Road, Singapore 119223
 E-mail: {ymtoo, mandar, venu}@arl.nus.edu.sg

Abstract—The ability to detect the direction of arrival and time of arrival (DoA-ToA) of propagating signals generated by underwater sources using array of sensors is crucial for passive sensing. This is a challenging task especially for transient and impulsive signals which have similar acoustic signature. We consider these signals, received by each sensor in a sensor array over an observation period. We assume that only a small number of these signals exist in the observation window, and that they have distinct direction of arrival (DoA) and time of arrival (DoA). The sensor array data of these signals can be transformed into a DoA-ToA space and we should expect this to be sparse. We show that this transformation can be written as an underdetermined linear system. We suggest robust methods to recover the DoA-ToA of the signals by enforcing sparsity in the DoA-ToA space. Through receiver operating characteristic (ROC) analysis, we show in numerical simulations that our methods outperform conventional practices such as cross-correlation-based time difference of arrival (TDoA) and beamforming. We present a scatter plot of the detected DoA-ToA based on the acoustic pressure sensor array recording in Singapore waters to show that the sparse DoA-ToA technique is robust and produces imaging results that match the known underwater man-made structures at the experimental site.

I. INTRODUCTION

Collapse of the cavitation bubble generated by rapid closure of snapping shrimp claw produces a transient impulsive signal ranging from few Hertz to more than 200 kHz [1]. This episodic loud sound or “snap” can be found mostly in warm shallow waters [2] [3] [4]. The acoustic signature of a snap can be characterized by a precursor pulse triggered by the full closure of the shrimp claw, followed by the rapid growth and collapse of the cavitation bubble to create a main peak. A reverberation time ensues, allowing the oscillations of the main peak to return to rest [5]. Figure 1 illustrates a time series acoustic pressure recording of a snap in Singapore waters. Theoretical model predicts that the width of the main peak of a shrimp snap is in the order of 100 ps [1]. However, in practice, the measured peak width is probably in the order of few μs because of the low pass filtering effect of the ocean as well as the limited sampling rate [6]. Au and Banks show that the reverberation of a snap is within 100 μs while Legg et al. calculate the reverberation time of a snap as 1.2 ms [7] [8]. The difference could be attributed to the impulse response of the environment and the recording system.

Detecting the DoA-ToA of the snaps not only facilitates the study of underwater ambient noise but also allows these ambient sources to be the sources of opportunity for several

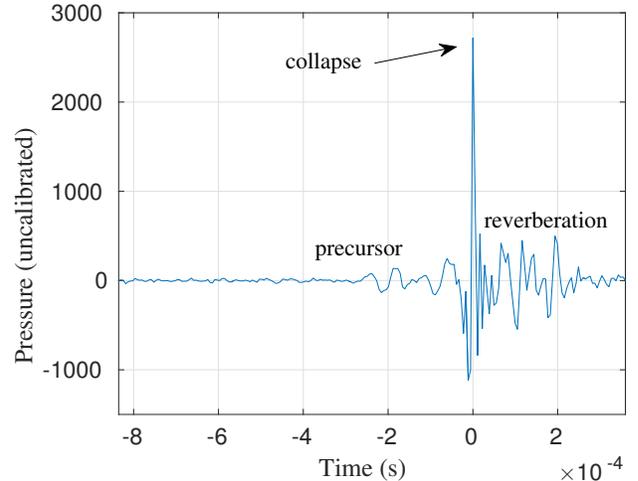


Fig. 1: An acoustic pressure recording of a snap in Singapore waters.

potential applications in underwater acoustic sensing. The depth of autonomous underwater vehicle (AUV) can be passively estimated using snapping shrimp noise [9]. Density of shrimp might indicate the ecological state of coral reefs [10] [11]. The mapping of low density clusters on known coral reefs might suggest dead reefs. Passive ranging of silent submerged objects underwater using solely the DoA-ToA of snaps has been shown to be feasible [12]. Spatial distribution of snapping shrimp can provide rough visualization on man-made structures [13]. We explore a method to detect the DoA-ToA of impulsive transient signals such as those originating from the snapping shrimp.

II. SIGNAL MODEL

To formally introduce the problem, we present a signal model to describe the acoustic pressure sensor array data of the impulsive transient signals. Let the acoustic center of an array be the origin of a Cartesian coordinate system. Let J be the number of transient impulsive signals arriving from the far-field, i.e., the signals are represented by $\{(\phi_j^*, \theta_j^*, \Gamma_j^*)\}_{j=1}^J$. Each signal is associated with a 3-tuple $(\phi_j^*, \theta_j^*, \Gamma_j^*)$, consisting of the azimuth angle, elevation angle, and the ToA. The acoustic pressure data of sensor m of the array at time

$t_n = \frac{n}{F_s}$, $n \in \mathbb{Z}_{\geq 0}$ can be written as

$$x_m(t_n) = \sum_{j=1}^J s_j(t_n - (\Gamma_j^* + \tau_m(\phi_j^*, \theta_j^*))) + v_m(t_n) \quad (1)$$

where F_s is the sampling rate and $\mathbb{Z}_{\geq 0}$ denotes non-negative integer. $s_j(t_n)$ is the acoustic pressure of signal j at time instance t_n . $\tau_m(\phi, \theta)$ is the time delay associated with the additional time needed for a source signal to reach the sensor m , and is given by $\tau_m(\phi, \theta) = \frac{\mathbf{p}_m^T \mathbf{q}(\phi, \theta)}{c}$, where $\mathbf{q}(\phi, \theta) \in \mathbb{R}^3$ is the unit vector in the signal propagating direction, $\mathbf{p}_m \in \mathbb{R}^3$ is the sensor location, c is the speed of sound in water and transpose is denoted by superscripts \top . $v_m(t_n)$ is the noise of the sensor, and is assumed to be spatially and temporally white, and uncorrelated to the signals.

The discrete Fourier transform (DFT) of the T collected snapshots of the sensor data can be written as

$$\mathbf{x}_m(f_k) = \sum_{j=1}^J \exp(-j2\pi f_k \tau_m(\phi_j^*, \theta_j^*)) \exp(-j2\pi f_k \Gamma_j^*) s_j(f_k) + \mathbf{v}_m(f_k) \quad (2)$$

for $f_k = \frac{k}{T} F_s$, $k = \{0, 1, \dots, T-1\}$ where $j = \sqrt{-1}$. Let the matrices and column vectors be represented in bold upper-case and bold lower-case letters, respectively. The DFT of M -sensor array data can be expressed as

$$\mathbf{x} = \mathbf{A}\mathbf{c} + \mathbf{v} \quad (3)$$

where $\mathbf{A} = [\mathbf{A}_1, \dots, \mathbf{A}_{|\mathcal{B}|}]$ is an overcomplete array response matrix represented as a partitioned matrix, with column sub-matrices of the form: [14]

$$\mathbf{A}_b = \begin{pmatrix} \mathbf{a}(f_0, \phi_b, \theta_b) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{a}(f_1, \phi_b, \theta_b) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{a}(f_{T-1}, \phi_b, \theta_b) \end{pmatrix} \quad (4)$$

where $\mathbf{a}(f_k, \phi_b, \theta_b) = [\exp(-j2\pi f_k \tau_1(\phi_b, \theta_b)), \dots, \exp(-j2\pi f_k \tau_M(\phi_b, \theta_b))]^\top$, and $b = 1, 2, \dots, |\mathcal{B}|$ where \mathcal{B} is the discrete space of all possible DoA and the cardinality of a set is denoted by $|\cdot|$. $\mathbf{x} = [x_1(f_1), x_2(f_1), \dots, x_M(f_1), x_1(f_2), \dots, x_M(f_{T-1})]^\top$ is the column vector of the Fourier coefficients of the array sensor data. $\mathbf{c} = [\mathbf{c}_1, \dots, \mathbf{c}_{|\mathcal{B}|}]^\top$ is the column vector with J non-zero blocks denoting the phase shifted Fourier coefficients of the arrivals. \mathbf{v} represents the DFT of sensor noise. Note that the ToA functions are non-linear. The next step is to extend the overcomplete sensing matrix to include the discrete space occupied by all possible ToAs. This extended formulation does not scale well as the size of the sensing matrix increases exponentially with the number of possible ToAs.

III. DOA-TOA POWER MAP

The discrete space of all possible DoA-ToA is denoted by \mathcal{S} . The DoA-ToA power map \mathbf{z} is a vector indexed by \mathcal{S} such that the amplitude of the element indicates the power of the

snap. Based on the aforementioned signal model, we outline the cross-correlation-based and beamforming-based method in generating DoA-ToA power map. Following that, we discuss the recent development of high resolution beamforming based on sparse DoA and further extend the idea to sparse DoA-ToA.

A. Cross-correlation-based TDoA (XCorr)

We can find the ToA of a snap at the sensor by cross-correlating the snap with the time-series sensor data. Since a snap is a random process, we do not know a priori the acoustic signal structure of the snap. In existing literature, snaps are usually approximated by segments of the time-series sensor data, which consists of distinct peaks [5] [15]. Let sensor m' be the reference sensor. In order to find the main peaks of the snaps in data of sensor m' , the data is preprocessed to form an enveloped sensor data such as

$$\tilde{x}_{m'}(t_n) = \sqrt{x_{m'}^2(t_n) + \mathbb{H}[x_{m'}(t_n)]^2} \quad (5)$$

where \mathbb{H} is the discrete Hilbert transform operator. N peaks exceeding a selected threshold, denoted by u , of the enveloped sensor output are identified as snaps with the ToA at sensor m' indicated by $\zeta_{m',i}$ for snap $i = 1, 2, \dots, N$. Let snap i be

$$s'_i(t_n) = \begin{cases} x_{m'}(t_n) & \text{if } |t_n - \zeta_{m',i}| \leq \frac{l_s}{2} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where l_s is the time length of a snap. The process of identifying snaps from the sensor array data is named as snap detection. TDoA of snap i between sensor m' and $m \neq m'$ can be obtained by solving

$$\arg \max_{-l_1 \leq t \leq l_1} \left| \sum_n x_m(t_n) s'_i(t_n - t) \right| \quad (7)$$

where the TDoA is bounded by the maximum time lag l_1 which is dependent on the size of the array. The DoA-ToA detection i , denoted by $(\phi'_i, \theta'_i, \Gamma'_i)$, maximizes the number of consistent DToAs of snap i . Hence, we have $\{(\phi'_i, \theta'_i, \Gamma'_i)\}_{i=1}^N$ and without loss of generality, we can construct the DoA-ToA power map of this method, denoted by $\mathbf{z}^{\text{XCorr}}$, such that the non-zero elements are indexed by $\{(\phi'_i, \theta'_i, \Gamma'_i)\}_{i=1}^N$ with amplitude 1 and other elements are zeros. Snap detection is sensitive to the threshold value u . A large u would result in the detector failing to pickup weak signals, whereas a small u would result in detecting too much noise. Furthermore, the use of cross-correlation for TDoA estimation is susceptible to error. This is due to the fact that multiple transient impulsive signals, which include snaps and the corresponding multipath signals, can be found within the maximum time lag of a particular snap. The acoustic signature between different arrivals cannot be easily distinguished. Interested readers can refer to [16] [17] [18] for further details regarding a variety of XCorr implementations.

B. Beamforming

1) *Delay and sum (DAS)*: The time domain DAS beamformer estimates the propagating signals using coherent and

incoherent sum [19]. Formally, the time domain DAS beamformed output can be written as

$$\alpha^{\text{DAS}}(\phi, \theta, \Gamma) = \frac{1}{M} \sum_{m=1}^M x_m(\Gamma - \tau_m(\phi, \theta)) \quad (8)$$

where $(\phi, \theta, \Gamma) \in \mathcal{S}$. The vectorization of $\alpha^{\text{DAS}}(\phi, \theta, \Gamma)$ for all the elements in \mathcal{S} is denoted by $\boldsymbol{\alpha}^{\text{DAS}}$ and the DoA-ToA power map of the DAS beamforming-based method is simply $\mathbf{z}^{\text{DAS}} = |\boldsymbol{\alpha}^{\text{DAS}}|^2$. In fact, \mathbf{z}^{DAS} is noisy due to the high sidelobe of the DAS beamformer.

2) *Hough*: To reduce the noise in the power map, the sensor array data can be preprocessed such that peaks of enveloped sensor data, exceeding the threshold u , are set to ones while others are set to zeros across all the sensors. Subsequently, the time domain beamformed output of the preprocessed sensor array data yields the power map denoted by $\mathbf{z}^{\text{Hough}}$. We use superscript Hough for the annotation because it is analogous to finding multiple planes in the preprocessed sensor array data [12]. Like the cross-correlation-based DToA method, choosing threshold u is non-trivial. This method resembles the DAS-based method for small u , while at large u , it suppresses weaker snaps. The original idea of the Hough-based method is proposed in [15] [12].

C. Sparse DoA-ToA

1) *Block-sparse (BS)*: With the assumption that only a small number of wideband signals arrive at the array in the observation window, a high resolution beamformer based on sparse DoA has been proposed [20] [14]

$$\hat{\mathbf{c}}^{\text{BS}} = \min_{\mathbf{c}} \frac{1}{2} \|\mathbf{A}\mathbf{c} - \mathbf{x}\|_2^2 + \lambda \sum_{b=1}^{|\mathcal{B}|} \|\mathbf{c}_b\|_2 \quad (9)$$

and a more intuitive form of this beamformer based on basis pursuit denoising (BPDN) is given by [21]

$$\begin{aligned} \hat{\mathbf{c}}^{\text{BS}} &= \min_{\mathbf{c}} \sum_{b=1}^{|\mathcal{B}|} \|\mathbf{c}_b\|_2 \\ \text{s.t. } &\|\mathbf{A}\mathbf{c} - \mathbf{x}\|_2 \leq \epsilon. \end{aligned} \quad (10)$$

The frequency domain beamformed output $\hat{\mathbf{c}}^{\text{BS}}$ is an estimate of \mathbf{c} in Equation 3 based on the objective function, L1/L2-norm. This mixed norm enforces block sparsity on the frequency domain beamformed output which promotes spatially sparse wideband signal reconstruction [22]. ϵ^2 denotes the upper bound on the noise power $\|\mathbf{v}\|_2^2$. λ controls the tradeoff between model misfit and block-sparsity of the frequency domain beamformed output. Let

$$\mathbf{V} = \frac{1}{\sqrt{T}} \begin{pmatrix} \exp(-j2\pi f_0 t_0) & \cdots & \exp(-j2\pi f_0 t_{T-1}) \\ \vdots & \ddots & \vdots \\ \exp(-j2\pi f_{T-1} t_0) & \cdots & \exp(-j2\pi f_{T-1} t_{T-1}) \end{pmatrix} \quad (11)$$

be the discrete Fourier transform matrix and \mathbf{W} be the block diagonal matrix of \mathbf{V} . The DoA-ToA power map of the BS beamforming-based method can be computed such as $\mathbf{z}^{\text{BS}} = |\mathbf{W}^H \hat{\mathbf{c}}^{\text{BS}}|^2$. However, wideband signal is a necessary but not a

sufficient condition for a snap. For instance, Equation 10 can be rewritten as

$$\begin{aligned} \hat{\boldsymbol{\alpha}}^{\text{BS}} &= \min_{\boldsymbol{\alpha}} \sum_{b=1}^{|\mathcal{B}|} \|\mathbf{V}\boldsymbol{\alpha}_b\|_2 \\ \text{s.t. } &\|\mathbf{A}\mathbf{W}\boldsymbol{\alpha} - \mathbf{x}\|_2 \leq \epsilon \end{aligned} \quad (12)$$

where $\mathbf{c}_b^{\text{BS}} = \mathbf{V}\boldsymbol{\alpha}_b^{\text{BS}}$ is the time domain beamformed output in direction b . Since \mathbf{V} is a unitary matrix, minimizing $\|\mathbf{V}\boldsymbol{\alpha}_b^{\text{BS}}\|_2 = \|\boldsymbol{\alpha}_b^{\text{BS}}\|_2$ does not impose sparsity on the time domain beamformed output signal in direction b . This is in contrast with the prior knowledge of the transient impulsive signal which has sparse support in the time domain.

2) *Sparse (S)*: The time domain beamformed output of the transient impulsive signal should be sparse in both DoA and ToA. We can reformulate Equation 3 as

$$\mathbf{x} = \mathbf{A}\mathbf{W}\boldsymbol{\alpha} + \mathbf{v} \quad (13)$$

and a sparse solution of $\boldsymbol{\alpha}$ can be recovered via

$$\begin{aligned} \hat{\boldsymbol{\alpha}}^{\text{S}} &= \min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_0 \\ \text{s.t. } &\|\mathbf{A}\mathbf{W}\boldsymbol{\alpha} - \mathbf{x}\|_2 \leq \epsilon. \end{aligned} \quad (14)$$

The time domain beamformed output $\hat{\boldsymbol{\alpha}}^{\text{S}}$ is an estimate of the inverse DFT of \mathbf{c} in Equation 3 based on the objective function, L0-norm, which calculates the total number of non-zero elements of a vector. Minimizing L0-norm yields the sparsest solution of Equation 13 and also the sparsest DoA-ToA power map. However, there is no efficient algorithm to solve Equation 14 to obtain the optimal solution. The closest possible convex relaxation of the problem is

$$\begin{aligned} \hat{\boldsymbol{\alpha}}^{\text{S}} &= \min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \\ \text{s.t. } &\|\mathbf{A}\mathbf{W}\boldsymbol{\alpha} - \mathbf{x}\|_2 \leq \epsilon \end{aligned} \quad (15)$$

which can be solved efficiently using either second-order cone program or first order method.

3) *Reduced-sparse (rS)*: To minimize the effect of steering and model mismatch, a very large sensing matrix \mathbf{A} has to be constructed so that all the array responses of the signals can be accurately represented by linear combination of the columns of \mathbf{A} . This leads to a computationally expensive optimization problem in Equation 14. An intuitive way to improve the efficiency of this method is to reduce the size of \mathbf{A} . Energy of the DAS beamformed output defined by

$$\mathbf{y}_b = \sum_{\Gamma} |\alpha(\phi_b, \theta_b, \Gamma)|^2 \quad (16)$$

can serve as an indicator to eliminate the obvious redundant DoAs and hence only a subset of \mathcal{B} are retained to generate a reduced-size \mathbf{A} . Formally, this subset can be written as $\tilde{\mathcal{B}} = \{(\phi_b, \theta_b) \in \mathcal{B} | \mathbf{y}_b > \mathbb{P}_p[\mathbf{y}]\}$ where $\mathbb{P}_p[\mathbf{y}]$ refers to p^{th} percentile of the value in vector \mathbf{y} . To avoid any confusion, the method of using the complete \mathbf{A} and reduced-size \mathbf{A} are coined as sparse method (S) and reduced-sparse method (rS) respectively. The sparse DoA-ToA intensity maps can be denoted as $\mathbf{z}^{\text{S}} = |\hat{\boldsymbol{\alpha}}^{\text{S}}|^2$ and $\mathbf{z}^{\text{rS}} = |\hat{\boldsymbol{\alpha}}^{\text{rS}}|^2$.

IV. DISCUSSION

A. DoA-ToA detection

We discuss a process to extract the DoA-ToA detection $\{(\phi_i, \theta_i, \Gamma_i)\}_{i=1}^N$ from the DoA-ToA power map \mathbf{z} . This is obvious for $\mathbf{z}^{\text{XCorr}}$ as the non-zero element of $\mathbf{z}^{\text{XCorr}}$ is exactly the DoA-ToA of the snaps. Other methods tend to recover the DoA-ToA power map of snaps based on different criteria. Since the acoustic signal of the snaps spans more than one sample and the array responses of the adjacent DoAs are highly correlated, $\mathbf{z}^{\text{Hough}}$, \mathbf{z}^{DAS} , \mathbf{z}^{BS} , \mathbf{z}^{S} , and \mathbf{z}^{rS} , to some extent, comprise multiple high amplitude elements (main peak of the snaps) surrounded by low amplitude elements (oscillation of the main peak and spatial leakage of the snaps, sensor noise, etc.). Locating these high amplitude elements gives the DoA-ToA detection.

Let Δ_ϕ , Δ_θ , and Δ_Γ be the threshold distances of the DoA-ToA detection. Let Δ_z be the threshold amplitude of the DoA-ToA detection. The DoA-ToA detection \mathcal{D} can be obtained using peak finding algorithm in a 3-dimensional space. This can be done firstly by setting the small value elements in \mathbf{z} to zeros based on the threshold amplitude and then selecting local maximums among the adjacent non-zero elements. Closely located peaks are eliminated based on the threshold distance because these peaks are probably due to the acoustic variability of snap or spatial leakage through highly correlated array response in DoA. We summarize the peak finding algorithm in Algorithm 1. We present a simple example in Figure 2 to further illustrate the algorithm. For convenient visualization, we display \mathbf{z} in a 1-dimensional form.

Algorithm 1 Peak finding

Require: \mathbf{z} , \mathcal{S} , $\mathcal{D} = \emptyset$, Δ_z , Δ_ϕ , Δ_θ , Δ_Γ

- 1: $\tilde{\mathcal{S}}$ contains the indexes of local minimum of the thresholded \mathbf{z} using Δ_z
- 2:

$$(\phi', \theta', \Gamma') = \arg \max_{(\phi, \theta, \Gamma)} \mathbf{z}(\phi, \theta, \Gamma) \quad \text{s.t. } (\phi, \theta, \Gamma) \in \tilde{\mathcal{S}}$$

- 3: $\mathcal{Q} = \{(\phi, \theta, \Gamma) \mid |\phi' - \phi| > \Delta_\phi, |\theta' - \theta| > \Delta_\theta, |\Gamma' - \Gamma| > \Delta_\Gamma, (\phi, \theta, \Gamma) \in \tilde{\mathcal{S}}\}$
 - 4: $\tilde{\mathcal{S}} = \tilde{\mathcal{S}} \cap \mathcal{Q}$, $\mathcal{D} = \mathcal{D} \cup (\phi', \theta', \Gamma')$
 - 5: repeat 2, 3, 4 until $\tilde{\mathcal{S}} = \emptyset$
-

B. Practical considerations

The transient signal is sparse in time since it has very few non-zero samples within T . Stable recovery from the frequency domain to time domain transient signal is possible using partial knowledge of Fourier coefficients. By randomly choosing a set of frequencies such that index of frequency $k \in \Omega \subset \{0, 1, \dots, T-1\}$ we can reduce the memory requirement for storing huge \mathbf{A} . Within the observation period, the DoA of the signals is sparse in \mathcal{B} which corresponds to a widely spread wavenumber (spatial frequency). High-resolution DoA estimation can be achieved by having a random sample of the spatial information using a sparse array.

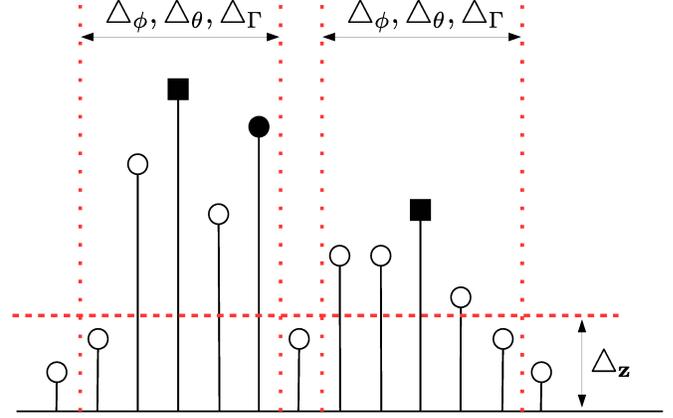


Fig. 2: The two filled squares are the DoA-ToA detections. Both filled squares and filled circle are the selected peaks. The horizontal dashed line illustrates the threshold amplitude and the vertical lines show the threshold distances.

Assuming a known $\|\mathbf{v}\|_2$ is impractical since the arrival of a snap is random. We can identify the main peaks of the snaps using simple thresholding and peak finding as described in the previous section. Then, we define the trimmed data of sensor m as

$$x_m^t(t_n) = \begin{cases} 0 & \text{if } t_n \in \{\zeta_{m,i}\}_{i=1}^{N_m} \\ x_m(t_n) & \text{if otherwise} \end{cases} \quad (17)$$

where $\zeta_{m,i}$ is a ToA of detected snap i at sensor m and N_m is the number of detected snaps at sensor m . We can estimate $\epsilon = \|\mathbf{x}^t\|_2$ where \mathbf{x}^t is the Fourier coefficient vector of the trimmed sensor array data. As snaps are highly impulsive compared to sensor noise, the bounds play an important role in eliminating the main peaks of the snaps in the sensor data.

V. NUMERICAL SIMULATIONS

We study the detection performance of the aforementioned methods based on the simulated sensor array data of the impulsive transient signals in 3 different cases as shown in Table I. The impulsive transient signal j was generated according to an exponentially damped sinusoid represented as

$$s_j(t) = \begin{cases} \beta_j \exp(-bt_n) \cos(2\pi f^s t_n), & \text{if } t_n \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (18)$$

where β_j is the amplitude of the main peak of signal j , b is the decay constant, and f^s is the frequency of the sinusoid. $b = 10000$ and $f^s = 30000$ Hz were chosen to imitate the transient and wideband behaviour of the snaps. Figure 4 shows the unit amplitude simulated signal in time domain. Two signals were generated in case 1 and case 2. The signals were purposely simulated with DoA-ToAs close to each other in case 2. Case 3 is similar to case 2 with the additional three weak signals. ROMANIS, a 2D planar array measuring 1.3 m in diameter and comprising 508 sensors, was used to record all of these signals which were contaminated by IID Gaussian noise of

zero mean and variance σ^2 , with the peak signal to noise ratio (PSNR) defined by

$$\text{PSNR} = 10 \log \left(\frac{\frac{1}{J} \sum_j \beta_j^2}{\sigma^2} \right). \quad (19)$$

Figure 3 shows the sensor placement of ROMANIS. Let $\mathcal{S} = \mathcal{S}_\phi \cup \mathcal{S}_\theta \cup \mathcal{S}_\Gamma$ where the set of azimuth angles $\mathcal{S}_\phi = \{-20^\circ, -19.5^\circ, \dots, 20^\circ\}$, the set of elevation angles $\mathcal{S}_\theta = \{-20^\circ, -19.5^\circ, \dots, 20^\circ\}$ and the set of time of arrivals $\mathcal{S}_\Gamma = \{0, 1, \dots, 499\}$. We set the threshold u to 1st percentile of the sensor data for snap detection. The typical maximum time lag of XCorr is the diameter of the receiver divided by speed of sound which gives 0.844 ms. Array response \mathbf{A} was generated by randomly selecting 64 frequency points while the reduced-size \mathbf{A} was based on 90th percentile of the energy of the DAS beamformed output. We computed $\mathbf{z}^{\text{XCorr}}$, $\mathbf{z}^{\text{Hough}}$, \mathbf{z}^{DAS} , \mathbf{z}^{BS} , \mathbf{z}^{S} , and \mathbf{z}^{rS} accordingly. Optimization toolbox, SPGL1, was used to solve for \mathbf{z}^{BS} , \mathbf{z}^{S} , and \mathbf{z}^{rS} [23] [24].

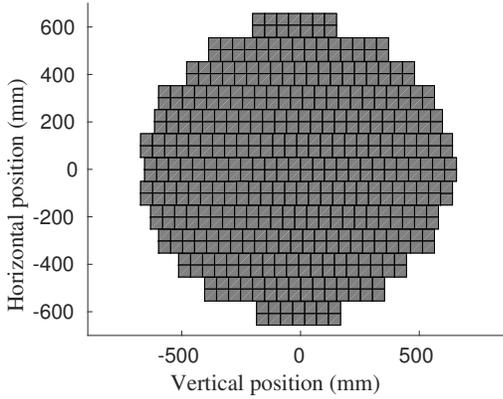


Fig. 3: Sensor placement of ROMANIS. Each gray cell represents a sensor.

TABLE I: Simulation setup.

Case	Signal	β	ϕ (degree)	θ (degree)	ΓF_s (sample)
1	s_1	1	-6.3°	-8.1°	49.2
	s_2	1	-6.1°	5.2°	249.3
2	s_1	1	-6.3°	3.4°	49.2
	s_2	1	-6.1°	5.1°	72.3
3	s_1	1	-6.3°	3.4°	49.2
	s_2	1	-6.1°	5.2°	72.3
	s_3	0.8	5.8°	4.2°	125.7
	s_4	0.8	11.1°	-3.8°	230.3
	s_5	0.8	2.4°	9.6°	400.8

Given the power maps, DoA-ToA detections were obtained based on the approximate resolution of ROMANIS $\Delta_\phi = 1^\circ$, $\Delta_\theta = 1^\circ$ while $\Delta_\Gamma = 0.510$ ms which is the reverberation time of the simulated snap. The detected DoA-ToA is true positive (TP) if it is within the threshold distance of the actual DoA-ToA. The true positive rate (TPR) is the number of TP divided by the number of actual DoA-ToA. False positive count (FPC) is defined by the number of detections which is beyond the threshold distance of the actual DoA-ToA. By

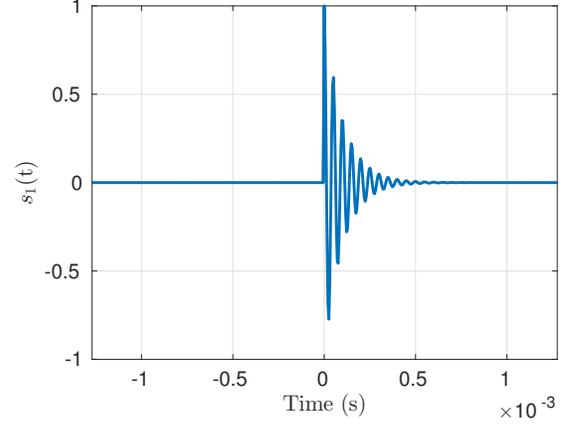


Fig. 4: Time domain simulated s_1 .

varying the threshold amplitude, the TPR vs FPC curves were plotted based on the average of 100 realizations.

As the amplitudes of the non-zero elements of $\mathbf{z}^{\text{XCorr}}$ are the same, there is only one point in all of the TPR vs FPC plots for XCorr regardless of the threshold amplitude. In case 1, DAS, S, and rS methods achieve optimal detection performance, i.e., TPR=1, FPC=0, by choosing the right threshold amplitude. Others are close to optimum with a small number of FPC. Similar results can be obtained from case 2 except that the detection performance of XCorr and Hough are observed to degrade extensively. When the DToA of two signals is less than the maximum time lag of XCorr, the method fails to identify the snap across sensors using cross-correlation. When two signals are close in DoA and ToA, Hough suffers from a large number of false positive detections. In case 3, XCorr, Hough, and BS have poor detection performance compared to the other methods.

Using the complete set of 508 sensors of ROMANIS, we noticed that DAS, S and rS are able to achieve optimal detection in all the simulated cases. To differentiate the performance among these methods, we reduced the effective diameter of ROMANIS by considering sensors within the radius of 0.3 m as shown in Figure 6 and recomputed case 3. A small aperture receiver is always preferable provided the detection performance can be maintained. However, this effectively decreases the resolution and PSNR of the receiver. The detection performance of S and rS outperform DAS when using the scaled-down array in case 3. One of the reasons is that DAS seems to be easily susceptible to false positive detection at 10 dB and 20 dB PSNR. In general, the detection performances of S and rS are fairly consistent and are reliable in a variety of circumstances such as the simulated cases. Even though rS uses the reduced-size \mathbf{A} , it does not suffer performance degradation in detecting the DoA-ToA of snaps.

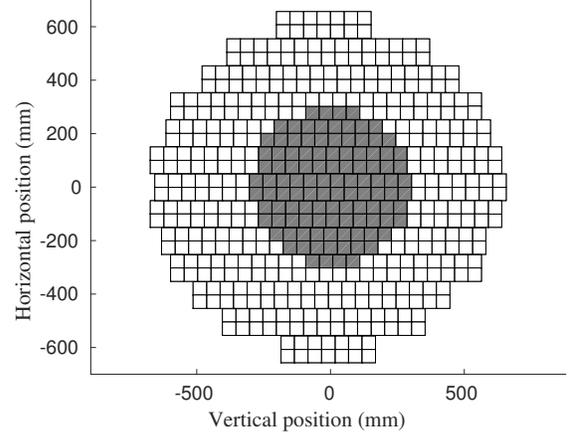
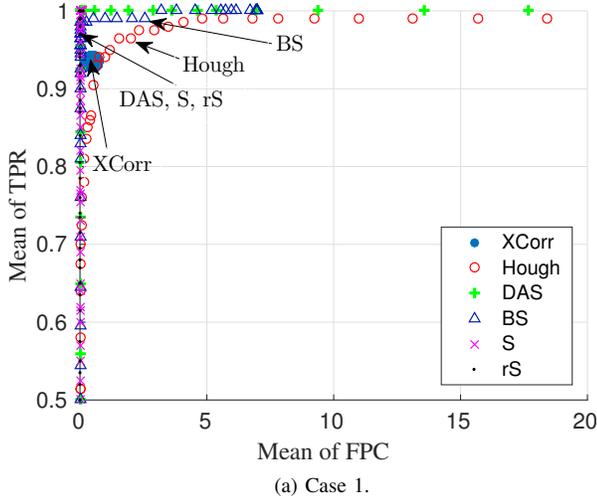


Fig. 6: Sensor placement of scaled-down ROMANIS. The gray colored cells represent the sensors used in the simulation.

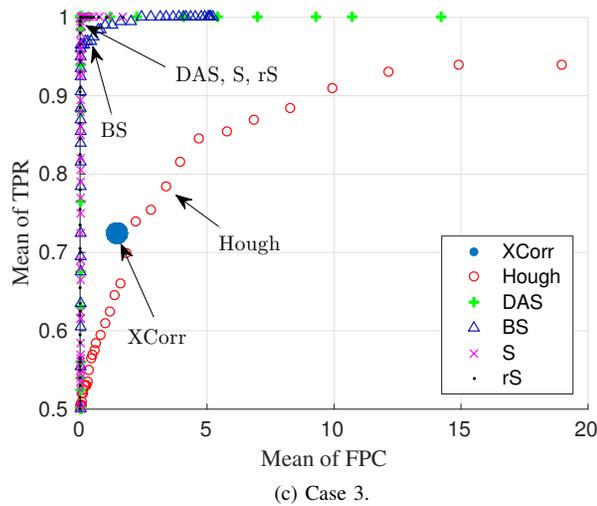
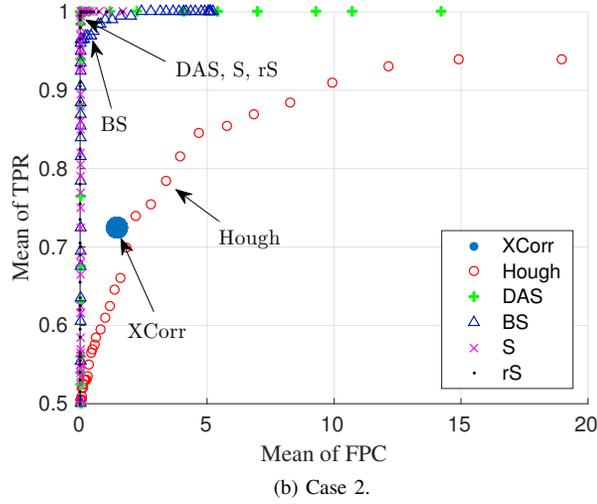


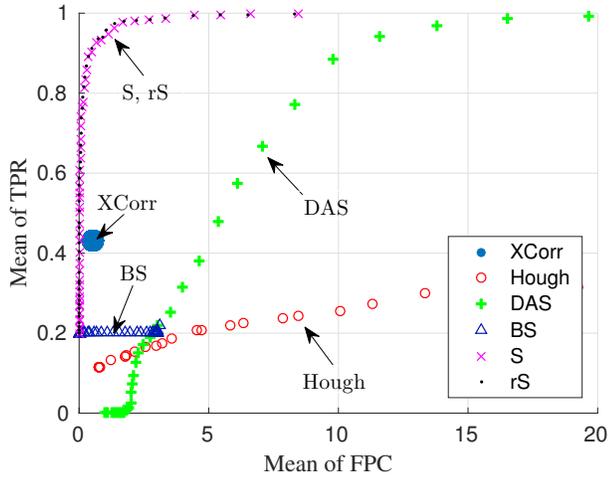
Fig. 5: Simulations at 10 dB PSNR.

VI. EXPERIMENTAL RESULTS

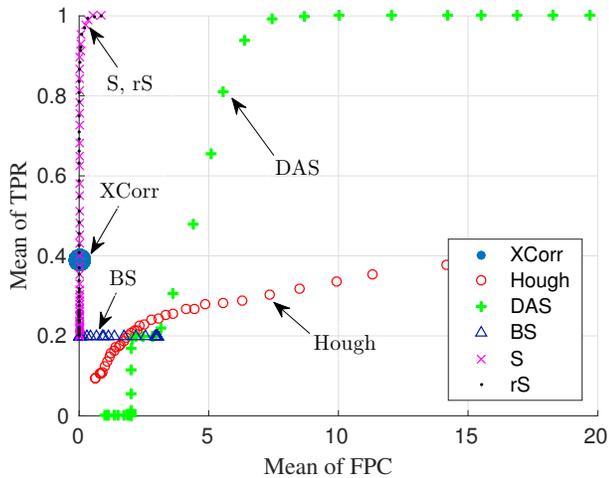
An experiment was conducted at Selat Pauh anchorage in Singapore waters. ROMANIS was deployed from a barge at $1^{\circ}12.967'N$, $103^{\circ}44.382'E$ with average water depth of 15 m. The DoA-ToA detection performance of all the methods are compared based on the 5-second acoustic pressure data collected during the experiment. There were two long term mooring buoys in the field of view of ROMANIS, buoy 1 and 2 being at a range of about 135 m and 246 m respectively from ROMANIS. These buoys present interesting targets because snapping shrimp attach themselves to the anchor lines to form colonies that are structurally different from those lodging on the seabed.

We define the set of azimuth angles $\mathcal{S}_{\phi} = \{-50^{\circ}, -49.5^{\circ}, \dots, 50^{\circ}\}$, the set of elevation angles $\mathcal{S}_{\theta} = \{-30^{\circ}, -29.5^{\circ}, \dots, 40^{\circ}\}$ and the set of time of arrivals $\mathcal{S}_{\Gamma} = \{0, 1, \dots, 3920\}$ for \mathcal{S} . We set the threshold u to 1st percentile of the sensor data for snap detection. We consider a full-sized ROMANIS (1.3 m diameter) and a scaled-down ROMANIS (0.6 m diameter). This corresponds to 0.844 ms and 0.390 ms maximum time lag for XCorr. A reduced-size array response \mathbf{A} was generated by randomly selecting 512 frequency points based on 90th percentile of the energy of the DAS beamformed output. We computed \mathbf{z}^{XCorr} , \mathbf{z}^{Hough} , \mathbf{z}^{DAS} , \mathbf{z}^{BS} , and \mathbf{z}^{rS} accordingly. For DoA-ToA detection, we set the threshold distance $\Delta_{\phi} = 1^{\circ}$, $\Delta_{\theta} = 1^{\circ}$, and $\Delta_{\Gamma} = 0.510$ ms while the threshold amplitude $\Delta_{\mathbf{z}}$ is defined by the 99.999th percentile of \mathbf{z} . The DoA-ToA detection is repeatedly computed over a 5-second ROMANIS dataset without overlapping.

Let N_a be the number of detected DoA-ToAs and ΔN_a be the change in the number of detected DoA-ToAs with respect to the full-sized ROMANIS. We verify the detection performance by plotting the detected DoA in Figure 8 and showing the changes in the number of detected DoA-ToA between full-sized and scaled-down ROMANIS in Table II.



(a) Case 3 at 10 dB



(b) Case 3 at 20 dB

Fig. 7: TPR vs FPC plot of case 3 using the scaled-down ROMANIS.

Hough generally has the worst performance in terms of the ability to compute the DoA of snapping shrimp noise. This may due to the large number of false positive detection. The DoA plot of XCorr using full-sized ROMANIS differs greatly from the one using scaled-down ROMANIS indicating an extensive performance degradation. DoA plot of DAS seems to have consistent detection performance regardless of the ROMANIS size. However, we observe that the change in the number of detected DoA-ToAs of DAS is huge, approximately 52%, from using full-sized to scaled-down ROMANIS. This shows that the method may sensitive to aperture size as well as the PSNR of sensor data. Note that BS has the largest number of detected DoA-ToA but with considerably lesser coverage in DoA. One particular DoA might contain multiple detections in the ToA-axis as the BS method only enforces sparsity in DoA space.

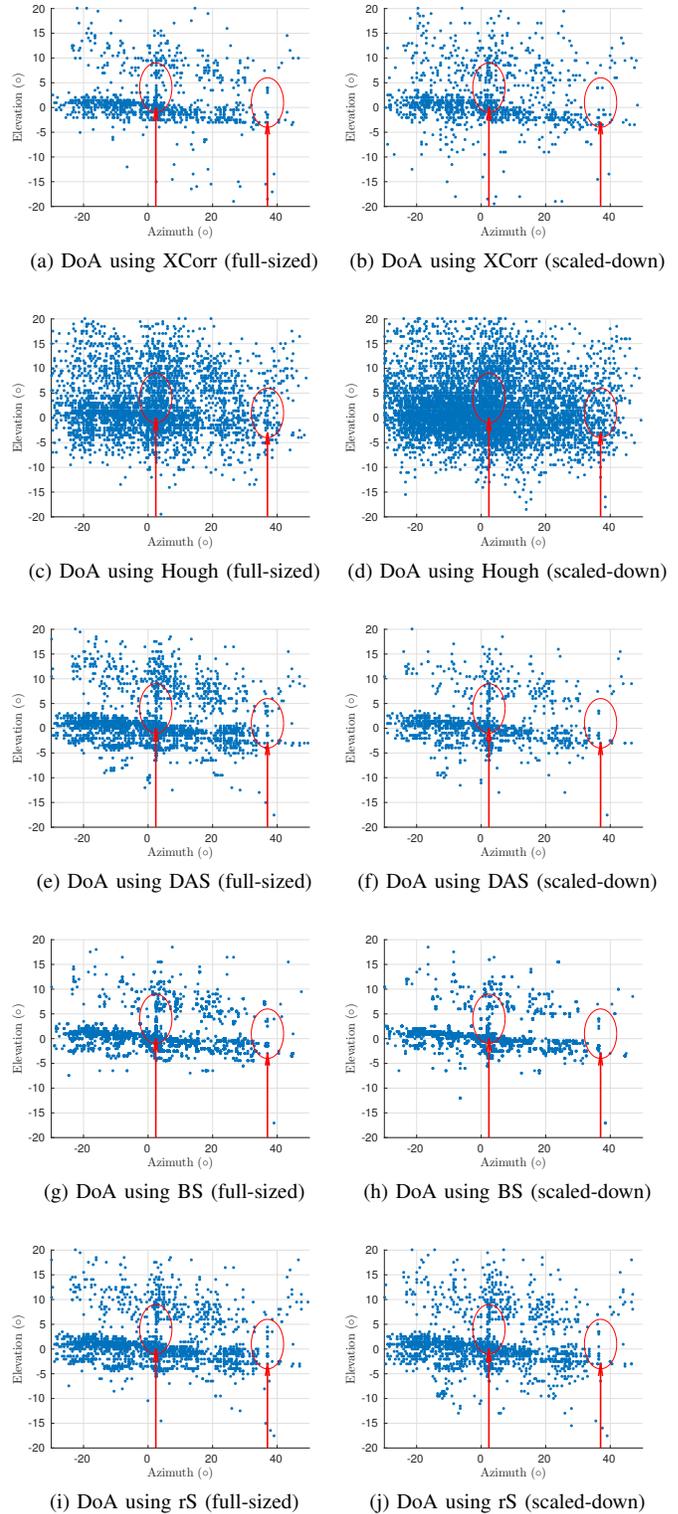


Fig. 8: The blue points in the scatter plots are the detected DoA. The circles mark buoy 1 and buoy 2 with the arrows indicating the azimuth angle of the buoys.

The DoA plot using rS mainly consists of two layers: the bottom layer, which corresponds to shrimp colonies stationed

TABLE II: Number of detected DoA-ToAs.

Method	ROMANIS	N_a	ΔN_a
XCorr	full-sized	1051	30%
	scaled-down	1371	
Hough	full-sized	3623	109%
	scaled-down	7560	
DAS	full-sized	2465	52%
	scaled-down	1177	
BS	full-sized	8993	28%
	scaled-down	6434	
rS	full-sized	2046	9.6%
	scaled-down	1850	

on the seabed, and the top layer which may be due to surface reflection of the snapping shrimp noise. There is significant amount of arrivals propagating at 2.4° azimuth angle which probably originated from snapping shrimp colonies on buoy 1. The proposed method tends to discover more arrivals from buoy 2 at roughly 37° azimuth angle. According to the GPS coordinates of ROMANIS, buoy 1, and buoy 2, the azimuth angle between a straight line from ROMANIS to buoy 1 and a straight line from ROMANIS to buoy 2 is approximately 36° . This shows that the detected DoA is close to the calculated azimuth angles of the buoys.

VII. CONCLUSION

We formulated an underdetermined linear system to model the array sensor data largely dominated by impulsive transient signals such as those generated by the snapping shrimp. We outlined some of the existing methods in detecting DoA-ToA of impulsive transient signals. We explored the option of using sparse DoA-ToA as the prior knowledge for the signal support in detecting the DoA-ToA of these impulsive signals. We demonstrated that the proposed method has several advantages over existing methods via numerical simulations. Unlike the conventional approach, it is robust to signals that are close in both DoA and ToA. The detection performance is maintained even in smaller aperture receiver with lesser number of sensors. We demonstrated that the method is able to reveal the bearings of the long term mooring bouys as far as 246 m away from the receiver. In fact, the proposed method could be applied to DoA-ToA detection problems involving any impulsive transient signals. It has been shown that optimal DoA-ToA detection of the impulsive transient signal can be acquired by randomly undersampling Fourier coefficients of the sensor array data. Future study should focus on determining minimum number of frequency points for optimal DoA-ToA detection given an array of sensors.

ACKNOWLEDGMENT

The authors would like to acknowledge colleagues at ARL for their support in the development of ROMANIS and during the field experiments. We would also like to thank Ms. Ong Lee Lin for proofreading the paper.

REFERENCES

[1] M. Versluis, B. Schmitz, A. von der Heydt, and D. Lohse, "How snapping shrimp snap: through cavitating bubbles," *Science*, vol. 289, no. 5487, pp. 2114–2117, 2000.

[2] M. W. Johnson, F. A. Everest, and R. W. Young, "The role of snapping shrimp (crangon and synalpheus) in the production of underwater noise in the sea," *Biological Bulletin*, pp. 122–138, 1947.

[3] D. H. Cato and M. J. Bell, "Ultrasonic ambient noise in australian shallow waters at frequencies up to 200 khz," DTIC Document, Tech. Rep., 1992.

[4] M. Chitre, M. Legg, and T.-B. Koay, "Snapping shrimp dominated natural soundscape in singapore waters," *Contrib Mar Sci*, vol. 2012, pp. 127–134, 2012.

[5] M. W. Legg, A. J. Duncan, A. Zaknich, and M. V. Greening, "Analysis of impulsive biological noise due to snapping shrimp as a point process in time," in *OCEANS 2007-Europe*. IEEE, 2007, pp. 1–6.

[6] M. Chitre, K. T. Beng, and J. Potter, "Origins of directionality in snapping shrimp sounds and its potential applications," in *OCEANS 2003. Proceedings*, vol. 2. IEEE, 2003, pp. 889–896.

[7] W. W. Au and K. Banks, "The acoustics of the snapping shrimp *synalpheus parneomeris* in kaneohe bay," *The Journal of the Acoustical Society of America*, vol. 103, no. 1, pp. 41–47, 1998.

[8] M. Legg, A. Duncan, A. Zaknich, and M. Greening, "An exploratory analysis of non-poisson temporal behaviour in snapping shrimp noise," in *Annual Conference of the Australian Acoustical Society: Acoustics in a Changing Environment, Acoustics 2005*, 2005. [Online]. Available: <http://researchrepository.murdoch.edu.au/14403/>

[9] M. J. Alam, E. H. Huntington, and M. R. Frater, "Passive auv tracking using sound produced by shrimps," in *Oceans-San Diego, 2013*. IEEE, 2013, pp. 1–5.

[10] S. D. Simpson, M. Meekan, J. Montgomery, R. McCauley, and A. Jeffs, "Homeward sound," *Science*, vol. 308, no. 5719, pp. 221–221, 2005.

[11] H. P. Zenil, V. G. Encomio, and R. G. Gilmore, "Passive acoustics as a monitoring tool for evaluating oyster reef restoration," *The Journal of the Acoustical Society of America*, vol. 129, no. 4, pp. 2698–2698, 2011.

[12] M. Chitre, S. Kuselan, and V. Pallayil, "Ambient noise imaging in warm shallow waters; robust statistical algorithms and range estimation," *The Journal of the Acoustical Society of America*, vol. 132, no. 2, pp. 838–847, 2012.

[13] T. Y. Min and M. Chitre, "Localization of impulsive sources in the ocean using the method of images," in *Oceans-St. John's, 2014*. IEEE, 2014, pp. 1–6.

[14] D. Malioutov, M. Çetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *Signal Processing, IEEE Transactions on*, vol. 53, no. 8, pp. 3010–3022, 2005.

[15] S. Kuselan, M. Chitre, and V. Pallayil, "Ambient noise imaging through joint source localization," in *OCEANS 2011*. IEEE, 2011, pp. 1–5.

[16] K. T. Beng, T. E. Teck, M. Chitre, and J. R. Potter, "Estimating the spatial and temporal distribution of snapping shrimp using a portable, broadband 3-dimensional acoustic array," in *OCEANS 2003. Proceedings*, vol. 5. IEEE, 2003, pp. 2706–2713.

[17] B. G. Ferguson and J. L. Cleary, "In situ source level and source position estimates of biological transient signals produced by snapping shrimp in an underwater environment," *The Journal of the Acoustical Society of America*, vol. 109, no. 6, pp. 3031–3037, 2001.

[18] S. E. Freeman, M. J. Buckingham, L. A. Freeman, M. O. Lammers, and L. Gerald, "Cross-correlation, triangulation, and curved-wavefront focusing of coral reef sound using a bi-linear hydrophone array," *The Journal of the Acoustical Society of America*, vol. 137, no. 1, pp. 30–41, 2015.

[19] D. Johnson and D. Dudgeon, *Array signal processing: concepts and techniques*. Simon & Schuster, 1992.

[20] D. Malioutov *et al.*, "A sparse signal reconstruction perspective for source localization with sensor arrays," in *MS thesis, Mass. Inst. Technol*, 2003.

[21] A. Xenaki, P. Gerstoft, and K. Mosegaard, "Compressive beamforming," *The Journal of the Acoustical Society of America*, vol. 136, no. 1, pp. 260–271, 2014.

[22] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.

[23] E. van den Berg and M. P. Friedlander, "Probing the pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2008. [Online]. Available: <http://link.aip.org/link/?SCE/31/890>

[24] —, "SPGL1: A solver for large-scale sparse reconstruction," June 2007, <http://www.cs.ubc.ca/labs/scl/spgl1>.