Localization of soniferous fish using a sparse hydrophone array and conventional steered response power method

by

Phillip Moore Deville

December, 2019

DIRECTOR OF THESIS: Mark W. Sprague, PhD.

Major Department: Physics

A sparse seven-channel hydrophone array was deployed in intertidal marsh creeks located within the North-Inlet-Winyah-Bay Nature Reserve at Baruch Marine Field Laboratory in Hobcaw Barony, Georgetown, South Carolina, USA. for the purpose of localizing the sounds produced by local fish and invertebrates. A conventional approach to localize sound sources was used, consisting of a delay-and-sum beamforming algorithm known as steered response power (SRP) applied to the seven-channel array data during post-processing. This was accomplished by delaying the signal recorded by each channel and combining the signals to "listen" to sounds propagating through the array from candidate source locations. The power of the combined signal was maximized, yielding a likely location for the source position. Two-dimensional localization of the soniferous fishes is possible, with a reasonable degree of robustness to noise and aliasing effects of spatial under-sampling. Accuracy is limited due to field measurements and the number of sound producers in the area.

Localization of soniferous fish using a sparse hydrophone array and conventional steered response power method

A Thesis

Presented To the Faculty of the Department of Physics

East Carolina University

In Partial Fulfillment of the Requirements for the Degree

Masters of Science in Physics

by

Phillip Moore Deville

December, 2019

Localization of soniferous fish using a sparse hydrophone array and conventional steered response power method

by

Phillip Moore Deville

APPROVED BY: _____

DIRECTOR OF THESIS: _____

Mark Sprague, PhD.

COMMITTEE MEMBER: _____

John Kenney, PhD.

COMMITTEE MEMBER: _____

Gregory Lapicki, PhD.

EXTERNAL COMMITTEE MEMBER: _____

Joseph Luczkovich, PhD.

CHAIR OF THE DEPARTMENT

OF PHYSICS: _____

Jefferson Shinpaugh, PhD.

DEAN OF THE GRADUATE SCHOOL: _____

Paul J. Gemperline, PhD.

Acknowledgments

I would like to thank the members of my committee for taking time out of their busy schedules to help me develop my ideas and get them written down on paper. I would like to thank Drs. Luczkovich and Sprague for getting me down to South Carolina and allowing me to participate in what was incredibly fun field research. I learned a lot down there and through the analysis of data collected there. These two members would always take time out for me with me and to discuss ideas. I would like to thank Dr. Kenney for his support throughout the thesis writing process and for many conversations particularly about sampling signals in space and time that were illuminating. I would like to thank Dr. Lapicki for helping me sharpen my technical presentation and formulae. Such detail orientation requires effort and it was very beneficial to the overall product to get a set of more experienced eyes on the paper.

I want to thank the scientists at USC Baruch Marine Field Institute: Dr. Dennis Allen, Dr. Matthew E. Kimball, and Julie Harding who assisted in the deployment of the array and who were very welcoming to us. I would also like to thank Cecilia Krahforst for her expert help and guidance during the array deployment and recovery process.

I want to thank the East Carolina University Department of Physics – for everything. The opportunity to work, learn, and grow. All the friendships and relationships that I have developed and fostered. These are some of the most caring, genuine, down-to-earth people I have ever met, and I am grateful to have experienced them.

Contents

List of Tables viii

List of Figures ix

List of Symbols xv

1 Introduction 1

1.1 Spatial filtering . . . 3

1.2 Physical assumptions . . . 4

2 Background 9

2.1 Plane waves . . . 10

2.2 Spherical waves . . . 12

2.3 Direct path propagation . . . 13

2.4 Hydrophone model . . . 15

2.5 Direction of arrival (DOA) . . . 17

2.6 Pairwise time delay . . . 18

2.6.1 Plane waves . . . 19

2.6.2 Spherical waves . . . 21

3 Apertures and array fundamentals 23

3.1 Finite continuous aperture . . . 24

3.1.1 Aperture smoothing function . . . 25

3.1.2 Resolution . . . 30

3.1.3 Sidelobe height . . . 31

3.1.4 Aperture ambiguities 31
3.1.5 Apparent velocity in a planar array 32
3.2 Spatial sampling 33
3.3 Arrays of discrete sensors 38
3.3.1 Grating lobes 40

4 Beamforming **47**
4.1 Delay-and-sum beamforming 48

5 The array **55**
5.1 The array geometry and setting for deployment 57
5.2 Calibrating the array 60

6 Steered power response method **64**
6.1 The steered-beamformer algorithm 69
6.2 Testing the algorithm 75

7 Results **82**
7.1 Array analysis 82
7.2 Calibration results 83
7.3 Steered response power results 87
7.3.1 Oyster Toadfish 87
7.3.2 Silver Perch 92
7.4 Tracking a Silver Perch 93

8 Discussion **97**

9 Conclusion **101**

References **102**

Appendix A Physical water quality data from Clam Bank monitoring site **108**

Appendix B HTI-96-MIN **111**

List of Tables

1 Specifications of the HTI-96-Min hydrophone manufactured by High Tech, inc. 55

2 Sensitivites and signal-to-noise characteristics of each hydrophone channel [26]. 56

3 Coordinates used for reference to measure distances to hydrophone position. 58

4 Coordinates for hydrophone positions based on field measurements. . 60

5 This parameter shows the arrays ability to distinguish signals propagating in at the angle of acceptance and reject waves coming in from other directions. 83

6 Field calibration data. 83

7 The calculated distances of each fish from the hydrophone array. . . . 94

List of Figures

1 The standard set of coordinates used for three-dimensional space. . . 9

2 Definition of vectors for first arrival wave propagation. 14

3 Direction of propagation vectors from a source to a discrete sensor array. 18

4 DOA with the far-field (*left*) and near-field (*right*). 19

5 Description of plane wave arrival at a microphone pair. 20

6 Description of spherical wave arrival at a microphone pair. 21

7 An aperture is only capable of measuring the wavefield to a finite extent. This is a continuous line aperture [23]. 23

8 Geometry that describes a continuous linear array. 24

9 Aperture smoothing function for a continuous line array aligned with the x-axis. L is the length of the continuous aperture. 29

10 The Rayleigh criterion [24]. 30

11 The cone of ambiguity [25]. 31

12 Representations of the spectrum of the signal that has been sampled in space. If the signal is perfectly bandlimited to $|\vec{k}| = \pi/d$ the periodic replications of the spectrum that occur due to sampling do not interfere and blend. If the space is undersampled the spectrum begins to creep into other intervals and reproducing the signal $s(m)$ is reproduced from the blended spectrum resulting in a poorer reproduction of the original signal. 37

13 In the array pattern of an aperture that spatial undersamples the field, the mainlobe is centered, the side lobes, which determine the passband of the array, are symmetric on either side for symmetric arrays. The *visible region* is the part of the array pattern that falls in between $|\vec{k}| = \pi/d$. Grating lobes appear as mainlobes. 41

14 The hypothetical discrete line array geometry plotted in two dimensions, $\theta \to \frac{\pi}{2}$. There are four equally spaced sensors, each with weight of 1. 41

15 The array pattern $|W(\vec{k})|^2$ is plotted as a function of incidence angle for several given frequencies. Notice the effect of higher frequency plane waves by the appearance of grating lobes in the array pattern. 44

16 The hypothetical array's directivity pattern. 45

17 Definition of vectors for conventional delay-and-sum beamforming analysis. 48

18 The beampattern is plotted against incidence angle for frequencies 100 Hz, 382.75 Hz, 500 Hz, 750 Hz, and 1000 Hz. 52

19 Two contour plots the first of which has been steered in direction $\alpha_x = 1$ and the second in $\alpha_x = 4$. 54

20 On the left, the main lobe for slowness $\alpha_x = 1$ has been superimposed over the wavenumber-frequency response of the array steered to look in direction α_x =4. The figure on the right shows the effect of temporal filtering. On the response of the array. 54

21 Slant forward image of the ZOOM F8 field recorder [27]. 57

22 Array plots of the hydrophone geometry at Clam Bank in the North Inlet-Winyah Bay National Estaurine Research Reserve near Georgetown, SC. The reference points are displayed as orange squares. The blue disks show the location of each hydrophone in the array. The green diamonds are the location of the calibration clanks. The reference points were very far away (*left*) and a plot excluding these points has been supplied (*right*). GPS Locations were recorded and submitted into Google Earth (*bottom*) *The plots and map are credited to M.W. Sprague.* 59

23 The red point is the center of a circle of minimal area that encloses all the intersection points for the system, and the green point is the spatial median. 62

24 Typical waveform plot of a source signal recorded on the seven hydrophone channels. 69

25 A typical grid spacing for stage 1 of the localization algorithm.The filtered spectrum is shown in red. 71

26 The brick-wall filter's response.The yellow region is the passband of the filter. The filtered spectrum is shown in red. 72

27 A sample output of the steered-beamformer algorithm after stage 1. . 75

28 A sample output of the steered-beamformer algorithm after stage 2. . 76

29 The sample sound used to text the extent of the beamformer algorithm. 76

30 Welch plot of the sample sound. 77

31 EXAMPLE: The filtered test sound without noise. 78

32 EXAMPLE: The filtered test sound with Gaussian noise. 78

33 EXAMPLE: The filtered test sound without noise. 79

34 *Top left.* Localization result of pure signal. *Top right.* Localization result with $10 \log \mathrm{SNR} = 0$ dB. *Bottom left* Localization result with $10 \log \mathrm{SNR} =$ -5 dB. *Bottom right.* Localization result with $10 \log \mathrm{SNR} =$ -13 dB. 81

35 The array pattern $|W(\vec{k})|^2$ is plotted as a function of incidence angle for several given frequencies. The array pattern is aperiodic, and increasing chaotic for higher frequencies. 84

36 The Clam Bank array's directivity pattern. 85

37 The beampattern for waves with frequency 500 Hz. The sidelobes have greater maximum here and suggest a diminishing ability to beamform accurately above these frequencies. 85

38 The results of the localization algorithm on the measured calibration tones. Hydrophone positions were adjusted until distance from true position was as minimal as possible using the spatial median. 86

39 Photo from NOAA, Credit: Andrew David, NOAA/NMFS/SEFSC Panama City; Lance Horn, UNCW/NURC – Phantom II ROV operator 88

40 SRP analysis of the first Toadfish. 89

41 Power Spectrum of the array output of the first toadfish taken with a 1024 sample width window. 90

42 Sonogram plot of the array output of the first toadfish. Taken with a 1024 sample window that overlaps by 512 samples. 90

43 SRP analysis of the second Toadfish. 91

44 A sonogram of the array's output during the time the second toadfish makes its boatwhistle taken with a 1024 sample wide window that overlaps by 512 samples. 91
45 A power spectrum of the array's output of the toadfish with the delays implemented appropriately from the SRP algorithm. The power spectrum is taken with a 1024 samples wide Hamming window. . . . 92
46 A silver perch. Brandi Noble/NOAA NMFS SEFSC Pascagoula Laboratory. 93
47 SRP analysis of the silver perch recorded at 5 s. 94
48 A power spectrum of the array's output of the silver perch with the delays implemented appropriately from the SRP algorithm. The power spectrum is taken with a 1024 samples wide Hamming window. . . . 95
49 A sonogram during the time the silver perch's drumming taken with a 1024 sample wide window that overlaps by 512 samples. 95
50 Another silver perch was localized (*top*) a 0.5 s into the recording, and then located 32 seconds (*bottom* into the recording having moved 11.7 m away. 96
51 Depth measurements [m] for three days of monitoring. 108
52 Dissolved oxygen measurements [mg/L] for three days of monitoring. 108
53 pH measurements for three days of monitoring. 109
54 Salinity measurements [g/kg] for three days of monitoring. 109
55 Temperature measurements [°C]for three days of monitoring. 110
56 The sound speed (m/s) as it varied over the 72 hours of observation. . 110

57 Mechanical outline of HTI-96-MIN hydrophones used for the seven sensor array deployed at Clam Bank [26] 111

List of Symbols

$A \equiv$ constant amplitude

$c_a \equiv$ apparent speed of sound through an array

$c \equiv$ sound speed in water

$C \equiv$ arbitrary constant

$d_{0,i}(\vec{x}) \equiv$ distance between sound source and hydrophone i in steered-response

$d \equiv$ inter-element spacing between hydrophones

$D \equiv$ ocean water depth

$D_{i,j} \equiv$ best attainable resolution in delay space

$E[n] \equiv$ energy of a discrete time signal

$E(t) \equiv$ energy of a continuous

$\mathcal{E} \equiv$ expected value

$f_{i,j}(\vec{x}) \equiv$ delay between hydrophones in steered-response

$f_s \equiv$ the sampling frequency

$f(t - \vec{\alpha} \cdot \vec{x}) + g(t + \vec{\alpha} \cdot \vec{x}) \equiv$ general form for a plane wave

$f(\vec{x}, t) \equiv$ arbitrary acoustic wavefield

$F(\vec{k}, \omega) \equiv$ spectrum of propagating wavefield

$\mathcal{F} \equiv$ Fourier transform

$\mathcal{G} = \mathcal{G}_x \times \mathcal{G}_y \times \mathcal{G}_z \equiv$ search grid

$H(\vec{k}) = W(\omega\vec{\alpha} - \vec{k}) \equiv$ wavenumber frequency response of a beamformer

$\hat{k} \equiv$ unit wavenumber vector

$\breve{k} \equiv$ discrete wavenumber

$\vec{k} \equiv$ wavenumber vector

$L \equiv$ length of a continuous line aperture

$m \equiv$ hydrophone number

$M \equiv$ total number of hydrophones

$n_m(t) \equiv$ noise recorded at each hydrophone m

$p_0 \equiv$ ambient pressure within a medium

$p' \equiv$ acoustic pressure field in a medium

$p \equiv$ pressure field within a medium

$P[n] \equiv$ power of a discrete time signal

$P(t) \equiv$ power of a continuous time signal

$\mathcal{P} \equiv$ steered response power

$r^0 \equiv$ distance between origin and the source

$r_{i,j} \equiv$ distance between calibration sound i and hydrophone j

$r_m^0 \equiv$ distance between hydrophone m and source

$r_m \equiv$ distance between origin and the m-th hydrophone

$\vec{r_m^0} \equiv$ vector pointing from hydrophone m to the source

$S(\breve{k}) \equiv$ spectrum of sampled signal in space domain

$S_c(k) \equiv$ spectrum of continuous signal in space domain

$s_c(x) \equiv$ continuous signal in space domain

$S \equiv$ ocean water salinity

$(s_{i,x}, s_{i,y}) \equiv$ position of calibration tone

$s(m) \equiv$ sampled signal in space domain

$S(\omega) \equiv$ spectrum of a time dependent signal $s(t)$

$s(x, y, z, t) \equiv$ any signal propagating in space-time

SNR $\equiv$ signal-to-noise ratios

$T \equiv$ ocean water temperature ($^\circ$ C)

$T(\mu) \equiv$ combined distance of a set of points from a spatial median

$W(\vec{k}) \equiv$ aperture smoothing function, beam pattern

$w(\vec{x}) \equiv$ weight applied to each hydrophone

$\vec{x}^0 \equiv$ position vector to a sound source

$\vec{x}_m \equiv$ position vector to m-th hydrophone

$\vec{x} = (x_1, x_2, x_3) \equiv$ arbitrary position function

$Y(\breve{k}, \omega) \equiv$ spectrum of sampled wavefield

$y_m[n] \equiv$ discrete time signal sampled by hydrophone m

$y_m(t - \Delta_m)$ time signal recorded by hydrophone m after delay

$y_m(t) \equiv$ continuous time dependent signal recorded by hydrophone m

$Z(\vec{k}, \omega) \equiv$ spectrum of arrays output

$z(\vec{x}, t) \equiv$ theoretical output of an array aperture

$\vec{\alpha} \equiv$ arbitrary slowness vector

$\vec{\alpha}_i^0 \equiv$ slowness vector of incident plane waves i

$\vec{\alpha}_j^0 \equiv$ arbitrary choice of slowness vectors j

$\Delta_m \equiv$ delays implemented by beamformer to channel m

$\vec{\zeta} \equiv$ unit vector in the direction of propagation

$\vec{\zeta}_m^0 \equiv$ unit vector describing the direction from a source to a sensor

$\hat{\mu} \equiv$ spatial median

$\rho_0 \equiv$ ambient density within a medium

$\rho' \equiv$ density field in a medium from a sound

$\rho \equiv$ density field within a medium

$\tau_{j'-j} \equiv$ the delays in arrival of calibration sound between hydrophone j' and hydrophone j

$\tau_m \equiv$ travel time for a sound from source to hydrophone m

$\tau_m(\vec{x}) \equiv$ delays calculated in the steered-response algorithm

$\omega \equiv$ angular frequency

1 Introduction

Fish bioacoustics is an interdisciplinary, specialized field of applied science focusing on sound production, reception, and communication of fish; the effect of anthoprogenic noise on fish; and active and passive acoustic technologies for fish population monitoring[1, 2]. Acoustic methods for investigating population distributions of fish are important because of the difficulties found in direct observation of species underwater. These bioacoustic studies are of direct interest to biological scientists, managers, conservationists, and commercial fishers [3], because of the significance of fish within aquatic environmental networks, the importance of fish as a food source. These studies seek to gather information about the types of fish species living in various habitats, their quantity, and their behavior, including reproduction, feeding, predator-prey relationships, and human-fish interactions. Ultimately bioacoustic studies can be used to protect these species and their natural environment.

Because most of this work approaches solutions to problems regarding the characterization of local fish populations from a conservationist angle, it is especially important that the study itself poses as small an interference with fish as possible. Passive acoustic surveys are an ideal way to investigate a habitat while not encroaching on fish behavior. Passive acoustic monitoring, in contrast to implementing an active acoustic system, depends entirely on sound produced by fish recorded using a single hydrophone or many hydrophones configured in an array. Passive acoustic methods are therefore a noninvasive way to collect data. This technique, however, is limited to soniferous species of fish only.

Often, characterizing fish populations for a given habitat involves relating the location of members of soniferous fish species within the habitat to other biological

or ecological variables, such as time of day, salinity, temperature, noise within the environment, and the proximity within the habitat to other fish species. A survey typically involves the use of one hydrophone. To locate a fish accurately within a region of space involves the use of more than one hydrophone designed to operate in concert as a sensor array. Techniques involving different configurations of sensors are useful to many disciplines of science. In this application soniferous species of fish are viewed as sources of acoustic energy.

Array signal processing concentrates on signals conveyed by propagating electromagnetic, acoustic, seismic, or gravitational wavefields. An array in this context refers to an assortment of sensors configured in some geometry at distinct locations in a domain of physical space. The sensor array transduces field energy into an electrical signal by sampling the wavefield in time at the sensor locations. Each waveform detected and output by individual sensors are combined and information about the propagating signals is calculated. There are three primary motivations to array processing: (1) to achieve a more desirable signal-to-noise ratio compared to that of a single sensor (2) to characterize the field by quantifying the energy sources [4], localizing the energy sources, and reproducing the waveforms emitted from the energy sources, (3) and to track the energy sources as they move through space [5, 6].

Much work has been done with source localization of talkers and sound source level estimation for noise [7, 8]. Increasingly, arrays have been used to study aquatic wildlife: to model acoustic wave propagations in oceans [9], to study black drum [10], plainfin midshipman fish [11], and to locate oyster toadfish [12].The use of arrays in bioacoustics to survey is still a developing field of research because of the many variables that make field research with a microphone array challenging. For one, the

array must be custom made and can be very expensive. To survey a large area, the array must be quite large, which can be cumbersome in the field to transport, deploy and recover.

The objective of this thesis is to implement such an array system to passively survey a marsh creek and to extract population information about sound producing fish species that reside or migrate through this habitat.

1.1 Spatial filtering

Spatial filtering allows for the enhancement of signals emanating from a source positioned at an arbitrary observation point, while suppressing potential signals propagating in other locations. Techniques that exploit the spatial filtering ability of an array are colloquially classified as *beamforming*, and in this category of signal processing there are distinguishing subordinate classes of specialized analyses. These techniques share the ability to locate and track positions of signal sources, and the use of adaptive algorithms to adjust the array's focus accordingly to locate and track source positions is known as *steering*. After a sensor array system has been used to remotely acquire propagating signals, the signal analysis can be implemented in real-time or as a post-process. This is achieved through the use of signal processing techniques be it electronically or computationally.

Steering as a post-process has advantages and disadvantages. It offers great flexibility in what is done to the signals for analysis as the process is the same but entirely less automated, with greater control. The steering is done with computational techniques applied to arrays and datasets by the surveyor, not electronic circuits with predefined scripts and routines. On the other hand, depending on the acquisition

time the sensor array was deployed, there can be many bytes of data, and each signal source is painstakingly analyzed on an individual basis. The usefulness of each option is determined by the application. As it will be applied to sound producing fish in estaurine environments, the post-process steering is desirable. When it comes to the bioacoustics of fish, the acoustician is listening for anomalies in noisy environments among days of audio recordings where it is "business as usual" for these animals. Interesting events can be analyzed on a singular basis using computational techniques and conclusions drawn from results in this manner. For this to be effective, some degree of automation is of course required and the algorithms must be designed and coded to accept almost any size data entry of a few types of form.

1.2 Physical assumptions

The aquatic environment that was surveyed will be considered in the model as a homogeneous quiescent medium. It is very shallow water, between 1 to 2 m, generally, even at high tide, so the pressure will vary negligibly with depth, and the water is slightly brackish but of constant density everywhere in space. The tidal patterns do indicate a flow, some small initial velocity, but take place over long time scales, so at any instant of analysis the tide's effect on the flow velocity and/or density is negligible.

For these reasons, it is ideal to assume the propagation of sound according to the linear equations of acoustics which yield a linear wave equation [13]. This idealization is adequate for the quantitative description of several acoustic phenomena. For this assumption to be valid, the ambient state of pressure and velocity fields within the medium must have certain characteristics.

Acoustic disturbances can be regarded as small-amplitude perturbations to an ambient state [13]. In a fluid, the ambient state is characterized by quantities (p_0, ρ_0, $\mathbf{v}_0$) which are the pressure, density, and fluid velocity of the medium in absence of an acoustic disturbance. The ambient-fluid variables themselves satisfy fluid-dynamic equations, and superimpose with like parameters that describe the effect an acoustic wave propagating will have on the medium:

$$p = p_0 + p' \qquad\qquad \rho = \rho_0 + \rho', \tag{1.2.1}$$

where p and ρ describe the pressure and density fields of the medium and the additional primed parameters describe the contribution from the acoustic wave to the resultant pressure and density fields within the medium. The total velocity, $\vec{v}$ is the sum of the fluid flow velocity, $\vec{v}^0$, and the particle velocity, $\vec{v}'$, of the acoustic disturbance, $\vec{v} = \vec{v}^0 + \vec{v}'$. In the homogeneous medium, the ambient field quantities are independent of position. This consequence implies that the speed of sound is constant everywhere in the medium – that the medium is non-refractive. In the quiescent medium, the ambient field variables are independent of time and the initial velocity of the medium is zero.

The assumptions just made may not apply if the speed of sound does indeed depend on the physical parameters of the medium (which are subject to change in this environment even if only over long time scales). Physical properties of the aquatic environmnet, specifically temperature, salinity and depth values will determine the speed of sound used throughout the model for a given instant of analysis. The nine-

term algorithm empirically developed by MacKenzie (1981)[14] was used:

$$\begin{aligned} c = 1448.96\ \frac{\mathrm{m}}{\mathrm{s}} + 4.519\ \frac{\mathrm{m}}{\mathrm{s}}\ {}^{\circ}\mathrm{C}^{-1}\ T - 5.304 \times 10^{-2}\ \frac{\mathrm{m}}{\mathrm{s}}\ {}^{\circ}\mathrm{C}^{-2}\ T^2 + 2.374 \times 10^{-4}\ \frac{\mathrm{m}}{\mathrm{s}}\ {}^{\circ}\mathrm{C}^{-3}\ T^3 \\ +1.340\ \frac{\mathrm{m}}{\mathrm{s}}\ \frac{\mathrm{g}}{\mathrm{kg}}^{-1}\left(S - 35\ \frac{\mathrm{g}}{\mathrm{kg}}\right) + 1.630 \times 10^{-2}\ \mathrm{s}^{-1} D + 1.675 \times 10^{-7}\ \mathrm{m}^{-1}\ \mathrm{s}^{-1} D^2 \\ -1.025 \times 10^{-2}\ \frac{\mathrm{m}}{\mathrm{s}}\ {}^{\circ}\mathrm{C}^{-1}\ \frac{\mathrm{g}}{\mathrm{kg}}^{-1}\ T\,(S - 35) - 7.139 \times 10^{-13}\ \mathrm{m}^{-2}\mathrm{s}^{-1}\ {}^{\circ}\mathrm{C}^{-1}\ T\ D^3, \end{aligned} \tag{1.2.2}$$

where c is speed of sound in sea water, T is the water temperature in ° C, S is the salinity, and D is the water depth. The formula was originally developed for the modeling of sound in oceans, but the waters in the marsh creeks surveyed has characteristics which fall within our requirements [15]: 23 to 30 °C temperature, 25 to 40 g/kg salinity, 0.6 to 2.0 m depth. This algorithm produces in the result for sound speed a standard error of 0.07 m/s. During the time the hydrophone array was recording, data was available for these parameters at the North Inlet-Winyah Bay National Estaurine Research Reserve (data from `http://www.northinlet.sc.edu/environmental-monitoring/`) Clam Bank system wide monitoring program (SWMP). This data is displayed in Appendix (A.1), and includes a plot for the variation in sound speed in Figure 55. The maximum value calculated for the sound speed was 1538 m/s and the minimum value calculated for sound speed was 1528 m/s.

Further assumptions about the medium pertain to the acoustic sources within the volume of space being surveyed. Soniferous invertebrates, such as snapping shrimp, will be regarded as point sources of acoustic energy, whose radiation pattern is described by spherical waves, regardless of the shape or size of the organism. This does

not preclude the use of far-field theory. In general, far from the source, the curvature of the wavefront decreases so that mathematically the propagating wave may be treated as a plane wave. The sources are considered stationary. The source may be in motion, but with speeds low compared to the speed of sound in water there is no significant Doppler impact on the frequency. Furthermore, in a linear model for sound propagation, it is assumed that the propagation speed does not vary with frequency, so that the medium is nondispersive. Moreover, any attenuation that occurs in wave amplitude is a consequence of the spherical wave solution to the linear wave equation – the medium is otherwise lossless. In the occurrence of several waves propagating simultaneously there is no interaction; the superposition principle applies, and the wavefield can be described as an addition of each sound wave. These considerations allow, as a consequence, an array to distinguish these sources using an array processing algorithm to spatiotemporally filter the received waveforms.

Despite banks of the channel serving as a boundary to the physical volume that was surveyed, the propagation will be treated as if it were occurring in an infinite, homogenous, isotropic medium. Sound in the free field travels in straight lines and is entirely unimpeded [16]. In these conditions, waves propagate without interference of boundaries: walls, sea-bottom, or other organisms. The environment that was surveyed is sufficiently large (≈ 3200 m^2) to suggest the use of the this model, which describes the direct path propagation from an acoustic source to an acoustic receiver, like a microphone or a hydrophone – even in reverberant environments, where acoustic reflections interfere with the direct sound. By analogy, the model could be used to study a sound source in an empty room with a microphone even with reverberant walls.

Beacuse recording of sound is nondeterministic, noise will be present in the recorded signals. One can make a reasonable assumption about the noise at each sensor: (1) that noise is independent of time and position, *i.e.* stationary [17, 18]; (2) that the noise is ergodic, the average of all the noise at each sensor will be the same of a single sensor; and (3) noise is isotropic [19]. The additive noise is considered, too, a spatiotemporally propagating wavefield, propagating in all possible directions with equal probability. Usually the noise is considered Gaussian, zero-meaned, statistically independent, and uncorrelated, with the signal [20].

In this model of sound propagation, the obvious advantage to these numerous idealizations is the inherent simplicity. Uncomplicated conditions only apply in restricted situations, and the gross simplicity will effect the efficacy of the model overall. This model, however, will be very useful in the development of a foundation for the general approach to the localization of sound-producing fish within the aquatic environments that our research team surveys.

2 Background

In array processing the signals of interest depend on both position and time. Space is considered three-dimensional within the context of the Cartesian coordinate system, and time is considered a fourth dimension. A spatiotemporal signal is written as $s(x, y, z, t)$, where x, y, z are spatial coordinates and t is a time coordinate. Alternatively, a spherical coordinate system can be used. Figure 1 shows the International Organization for Standardization (ISP) convention to three-dimensional spherical coordinates, where r is the distance from the origin located at (0, 0, 0), ϕ is the azimuthal angle measured from the positive x-axis, and θ is the altitude angle measured from the positive z-axis.

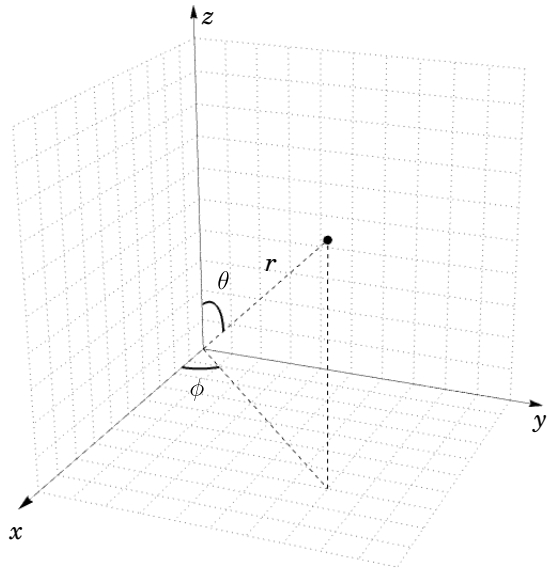


Figure 1: The standard set of coordinates used for three-dimensional space.

They are related to the rectangular coordinates through,

$$\begin{aligned} x &= r\sin\theta\cos\phi \qquad & r &= \sqrt{x^2+y^2+z^2} \\ y &= r\sin\theta\sin\phi \qquad & \theta &= \frac{z}{\sqrt{x^2+y^2+z^2}} \\ z &= r\cos\phi \qquad & \phi &= \arctan\frac{y}{x} \end{aligned} \tag{2.0.1}$$

Geometry and calculations will be treated as being in either of these physical domains, depending on the physical situation.

2.1 Plane waves

Propagating acoustic wavefields are described by the wave equation,

$$\nabla^2 p - \frac{1}{c^2}\,\frac{\partial^2 p}{\partial t^2} = 0, \tag{2.1.1}$$

where p is the pressure, t is time , and c is the sound speed. The solution to the wave equation that describes a plane wave is of the general form [13]

$$p = f(t - \vec{\alpha}\cdot\vec{x}) + g(t + \vec{\alpha}\cdot\vec{x}). \tag{2.1.2}$$

The arbitrary position is denoted by variable $\vec{x}$, and $\vec{\alpha}$ is the *slowness vector* [20]; the slowness vector is equivalent to $\vec{\alpha} = \vec{k}/\omega$, the wavenumber vector, $\vec{k}$, divided by ω the angular frequency, and $|\vec{\alpha}| = 1/c$. The slowness vector defines not only the direction of the propagating wavefield, because of its dependence on $\vec{k}$, but also its speed due to its magnitude. The functions f and g are aribtrary and are allowed to take any form; there is a sign difference in their arguements, therefore, f and g

describe waves moving in opposite directions. In many cases, there is one travelling wave in a region of space, and either f or g is accordingly set to zero – only one of these functions is required to detail the pressure in space and time. Typical solutions for a constant-frequency plane wave are represented by [13]

$$s(\vec{x}, t) = A \exp\left[i\omega(t - \vec{\alpha} \cdot \vec{x})\right]. \tag{2.1.3}$$

The symbol A, is a constant amplitude. Under a linear model for the wave equation, two solutions $s_1(\vec{x}, t)$ and $s_2(\vec{x}, t)$ involved in a linear combination is also a solution. In this way, more complicated solutions can be built as sums or integrals of complex exponentials. For example the harmonic series,

$$s(\vec{x}, t) = s(t - \vec{\alpha} \cdot \vec{x}) = \sum_{n=-\infty}^{n=\infty} S_n \exp\left[in\omega_0(t - \vec{\alpha} \cdot \vec{x})\right], \tag{2.1.4}$$

where ω_0 is the fundamental frequency of the wave. The Fourier coefficients for an arbitrary waveform $s(u)$ with period $T = 2\pi/\omega_0$, are

$$S_n = \frac{1}{T} \int_0^T s(u)\ e^{-i\ n\ \omega_0\ u}\ \mathrm{d}u. \tag{2.1.5}$$

this waveform propagates in the direction specified by $\vec{\alpha}$ with speed $c = 1/|\vec{\alpha}|$. It is allowed to contain many frequency components $\omega = n\omega_0$, but with the constraint that $\vec{k}/\omega = \vec{\alpha}$. Wave functions built in this way are periodic, due to the angular frequencies. This same approach can also be tackled in Fourier space where an integral of complex exponential can express an aperiodic waveform:

$$s(\vec{x},t) = s(t - \vec{\alpha}\cdot\vec{x}) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\omega)\exp\{i\omega(t-\vec{\alpha}\cdot\vec{x})\}\mathrm{d}\omega \tag{2.1.6}$$

$$S(\omega) = \int_{-\infty}^{\infty} s(u)e^{-i\omega u}\,\mathrm{d}u \tag{2.1.7}$$

2.2 Spherical waves

Given no ϕ or θ dependence, the general spherical wave equation can be derived,

$$\frac{1}{r}\,\frac{\partial^2}{\partial r^2}(\,r\;p) - \frac{1}{c^2}\,\frac{\partial^2 p}{\partial t} = 0. \tag{2.2.1}$$

Here, note the addition of r, which is the radial position. The solution to the wave equation that describes a spherical wave whose far-field source is at the origin takes the general form:

$$p(r,t) = r^{-1}\;f(t - |\vec{\alpha}|r) + r^{-1}\;g(t + |\vec{\alpha}|r) \tag{2.2.2}$$

Likewise, f is an arbitrary function describing a spherical wave emanating from the origin, and g is a spherical wave converging to the origin. If a source is centered at the origin, waves only move in the positive r direction away from the source and consequently $g = 0$. In this case, however, the wave shape, although similar, is altered and pressure is received differently at increasing radii with amplitude reduced by a factor of $1/r$. In addition, there is a shift in reception time of similar waveform features of $\Delta r/c$. Typical representations of single frequency spherical waves take the form:

$$s(r,t) = \frac{A}{r}\exp\{i\ \omega(t\ - |\vec{\alpha}|r\} \tag{2.2.3}$$

Because of the linearity of the wave equation, more complicated solutions can be built in a similar fashion as above:

$$s(r,t) = \frac{1}{2\pi r}\int_{-\infty}^{\infty} S(\omega)\exp\{i\omega(t - |\vec{\alpha}|r)\}\ \mathrm{d}\omega, \tag{2.2.4}$$

where $S(\omega)$ is the Fourier transform of the signal produced by the source.

2.3 Direct path propagation

In the presence of sound-reflecting surfaces, the sound waves produced by a single source propagate along acoustic multipaths. This produces the effects of reverberation; sounds reflect off objects and produce echoes. In the intertidal channel, reflections can occur off any boundary, especially the bottom and the surface in shallow water conditions. The linearity of the medium ensures treatment of combining waves through superposition, so the direct path component and the reflections are additive. Localization algorithms can separate the direct path component from the multipath components and parameterize the location of a sound producer. In any event, localization will be performed using audio trimmed to include the first arrival of a sound and some sustain, truncating before any echoes or interference. In Figure 2 is the coordinate system used to describe direct-path propagation from a sound source to a microphone. The axes are taken so that the positive z-axis is downward as is a standard in underwater acoustics.

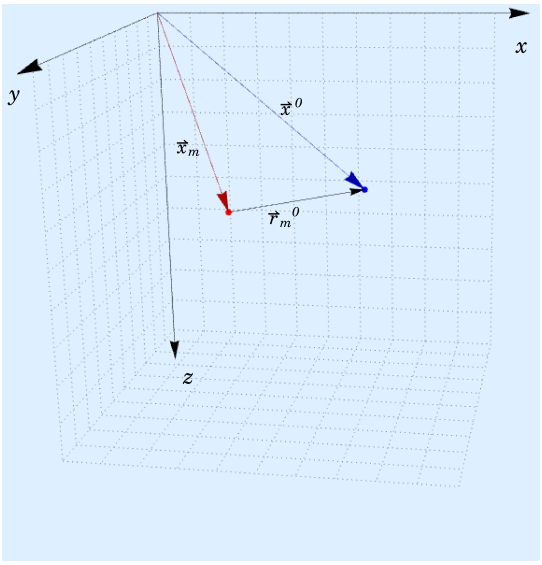


Figure 2: Definition of vectors for first arrival wave propagation.

Consider an acoustic source located at $\vec{x}_0$ with respect to an origin from which a wavefield will be propagating. A hydrophone located at $\vec{x}_m$ records the signal. The vector $\vec{r}_m^0$ is sourced at hydrophone m and points to the source location so that the distance between the source and the hydrophone is,

$$r_m^0 = |\vec{x}^0 - \vec{x}_m|. \tag{2.3.1}$$

In the treatment of soniferous fish as point sources of acoustic energy, a direct path arrival of a waveform of arbitrary shape can be modeled according to equation 2.2.3,

$$s(r,t) = A\ r^{-1}\ f(t - r/c) \tag{2.3.2}$$

Outside the region of initial excitation, the wave propagates as aforementioned, where the attenuation depends inversely on the travel distance; furthermore, delayed, under

the direct path assumption, by a time delay equal to the ratio of the travel distance to sound speed,

$$\tau_m = r_m^0/c. \tag{2.3.3}$$

The wavefield at hydrophone m for a source located at $\vec{x}_0$ is given by:

$$f(\vec{x_m}, \vec{x_0}, t) = \frac{A}{r_m^0}\ f(t - \tau_m), \tag{2.3.4}$$

where the wavefield is inversely proportional to the travel distance r_m^0. Note this is for purely a geometrical reason with the spherical coordinates in in the 3D space and has nothing to do with the attenuation which could be altogether an additional effect of absorption in murky water and/or hitting some objects such *e.g.*, other fish.

2.4 Hydrophone model

In order to transduce acoustic field energy into electrical energy, piezoelectric materials are applied within the hydrophone [21]; this type of material can change its form and helps to generate electrical potential output in response to mechanical or external pressure variations. The voltage created within the material when an acoustic disturbance perturbs its structure is proportional to the acoustic pressure, and it can be amplified, filtered, and then sampled to transform the information into a digital signal for processing. The sound field $f(\vec{x_m}, \vec{x}^0, t)$, is sampled by M sensors, denoted by index $m \in \{1, 2, ..., M\}$ located at $\{\vec{x}_m\}_1^M$ at time instants $\{t_n\}$. Each sensor m records a signal, $y_m(t) = s(t) + n_m(t)$, where $y(t)$ is the signal and $n_m(t)$ is

an additive noise term at the m-th sensor location. Compared to the analog signal whose domain is time, a digitally sampled signal has domain in samples n. In the case of periodic sampling, $t_n = nT$, where $n \in \mathbb{Z}$, where $\mathbb{Z}$ denotes the set of integer numbers, and T is the time interval between successive samples, or the sampling period. The sampling rate, f_s, is related by $1/T = f_s$ (a typical sampling rate of 44.1 kHz, would give a sampling period of 2.2676×10^{-5} s). Through a variable transformation, $t \to t_n \to nT$, and with periodic sampling T is constant the digital signal is represented:

$$y_m[n] = s[n] + n_m[n]. \tag{2.4.1}$$

The energy is given by [18]:

$$E(t) = \lim_{L\to\infty} C \int_{-L}^{L} |y(t)|^2 \, \mathrm{d}t \qquad E[n] = \lim_{N\to\infty} A \sum_{n=-N}^{N} |y[n]|^2, \tag{2.4.2}$$

where C is a constant of proportionality and subsequently the signal's power, P, can be calculated by dividing the length of each audio dataset.

$$P(t) = \lim_{L\to\infty} \frac{C}{2L} \int_{-L}^{L} |y(t)|^2 \, \mathrm{d}t \qquad P[n] = \lim_{N\to\infty} \frac{C * f_s}{2N+1} \sum_{n=-N}^{N} |y[n]|^2, \tag{2.4.3}$$

where f_s is the sampling frequency. The signal-to-noise ratio for a single hydrophone is then given by,

$$\mathrm{SNR} = \frac{P_{\mathrm{signal}}}{P_{\mathrm{noise}}} = \frac{\mathcal{E}[s^2(t)]}{\mathcal{E}[n_m^2(t)]} \Rightarrow \frac{\mathcal{E}[s^2[n]]}{\mathcal{E}[n_m^2[n]]}. \tag{2.4.4}$$

Note, assuming equal areas, the signal-to-noise ratio of powers is often expressed in decibels as the difference the level of power between the signal and the noise:

$$\text{SNR} = \beta_s - \beta_n = 10 \, \frac{\log \mathcal{E}\left[s^2[n]\right]}{\log \mathcal{E}\left[n_m^2[n]\right]} \tag{2.4.5}$$

2.5 Direction of arrival (DOA)

The incidence in time of the first maximum of a waveform, barring any echoes, is considered in this work, the time a sound-wave arrives to a microphone recording it – this says nothing about the propagation time or ultimately the moment in time when the sound was issued. In general, the exact moment a sound gets produced is unknown a priori. This first maximum occurrence corresponds to the direct-path component to the wave, representing sound that has traveled along a ray from the source to microphone m. The direction from the source to the microphone is the *direction of propagation*, and contrarily, the opposite direction is known as the *direction of arrival* (DOA); therefore, a wave traveling from North to South has a southward direction of propagation and a northward direction of arrival.

Figure 3, shows in general how the direction of arrival vectors look for direct path propagation. If $\vec{\zeta}_m^0$ is a unit vector describe the direction from a source to a sensor, then each sensor measures a different direction of arrival from the source. This is dependent, though, on the shape of the incoming wavefront. The shape of the wavefront is determined by whether the source is located in the near-field or far-field. The DOA works differently for near-field and far-field sources [22]. For sources in the near-field, this works too under the assumption of a point-source, the

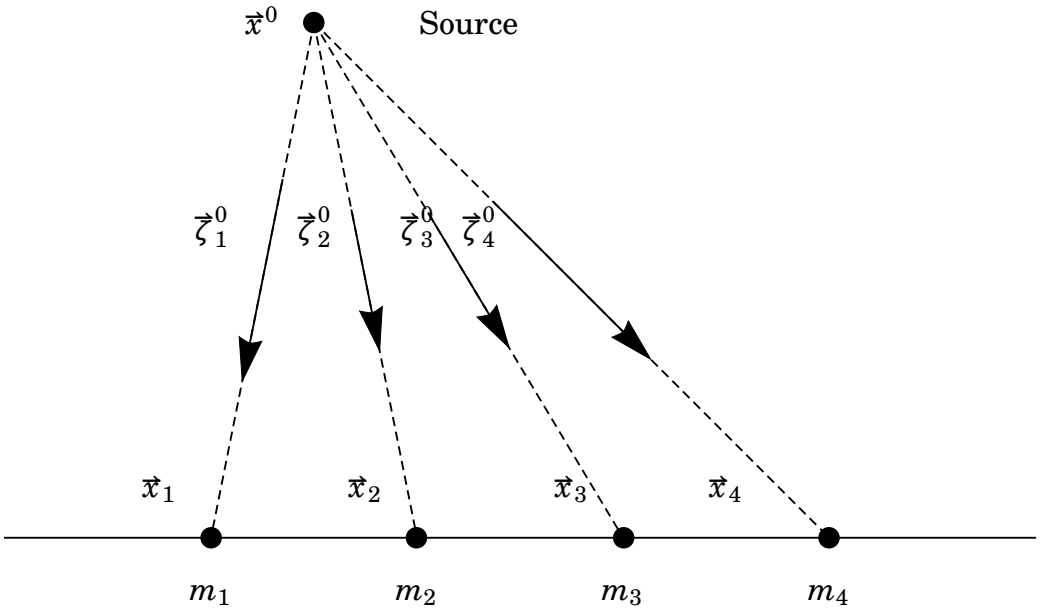


Figure 3: Direction of propagation vectors from a source to a discrete sensor array.

wavefronts arrive as curved wavefronts with spherical symmetry, and each sensor will measure DOA vectors as shown in Figure 3. In the case of wavefronts that arrive as plane waves, the same direction vector $\vec{\zeta}^0$ is measured by all sensors in an array. The effect of this is that the location can only be resolved as precise as a direction, many locations in space along the direction of that path of arrival will have the same spatial-likelihood of being the true location of the source.

2.6 Pairwise time delay

A cornerstone to beamforming/steered-response algorithms is delay analysis. Delays are calculated iteravely between possible pairs of sensors and can be implemented in such a way to achieve a position because time and position are related through the speed of sound in the medium. The ability to resolve a position or at the very

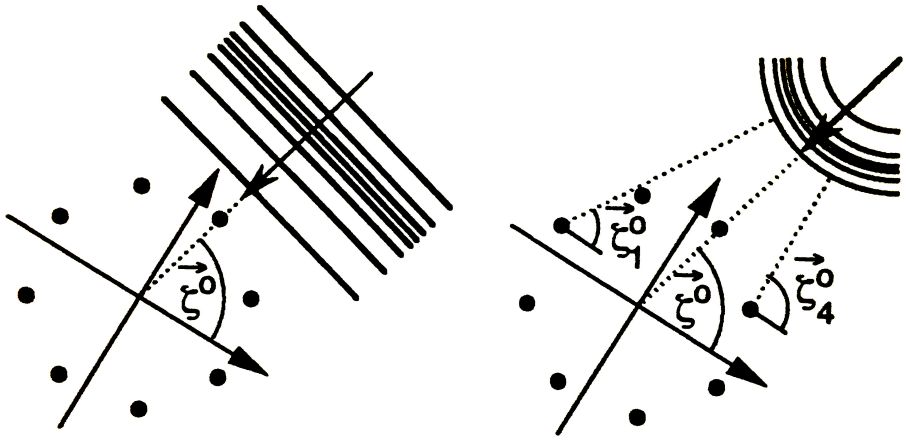


Figure 4: DOA with the far-field (*left*) and near-field (*right*).

least a direction depends on the field in which the source of the sound exists, and, the calculation can be carried out beginning with a calculation for the delay itself or choosing an estimated source location and working backwards to find the delay.

2.6.1 Plane waves

Consider a plane wave incident on a microphone pair m_1 and m_2 positioned at $\vec{x}_1$ and $\vec{x}_2$. The direction of the plane wave's arrival is denoted by the unit vector, $\hat{\zeta}^0$.

The relevant geometry is shown in Figure 5. When the wavefront reaches microphone m_1, the same pressure amplitude has yet to reach microphone m_2; the distance the wavefront must move to reach the second microphone is $\vec{d} = d\ \vec{\zeta}^0$. Using the simple model, $d = c\Delta t$, in the form presented in equation (2.3.3):

$$\tau_m = r_m/c, \tag{2.6.1}$$

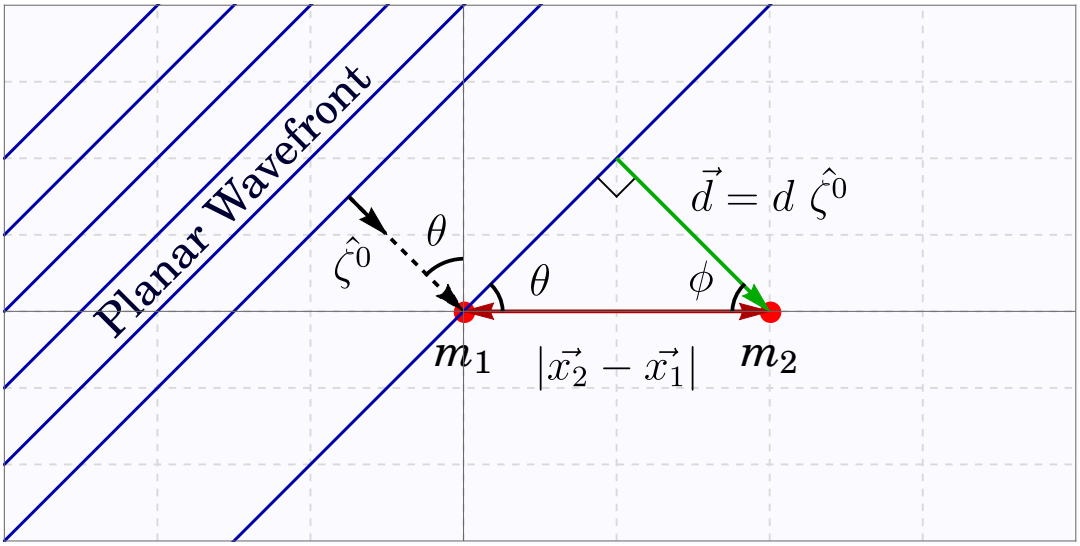


Figure 5: Description of plane wave arrival at a microphone pair.

where r_m is the distance $\vec{d}$ the wavefront must travel to reach sensor m_2 in the figure. For the one dimensional case, an array aperture aligned on the x-axis, fundamental trigonometry quickly reveals these delays can be related to the geometry through:

$$d = |\vec{x}_2 - \vec{x}_1| \sin\theta = c\ \tau_m. \tag{2.6.2}$$

An array aperture not aligned with any particular axis would represent a two-dimensional case, and for plane waves, delay calculation can be generalized to higher dimensions through the use of the dot product,

$$\tau_m = \frac{\hat{\zeta}^0 \cdot (\vec{x}_2 - \vec{x}_1)}{c} = \frac{|\vec{x}_2 - \vec{x}_1| \cos\phi}{c} \tag{2.6.3}$$

2.6.2 Spherical waves

In the near-field, and in general, the wavefronts from sources emitting acoustic energy are curved and travel as spherical waves. Moreover, nothing is lost except an opportunity for simplicity by always considering the wavefronts as curved, so an attempt will be made to be more general. For an off-axis microphone pair consisting of sensors m_1 and m_2, the vector geometry is presented in Figure 6. The source position is described by vector $\vec{x}^0$, the microphone positions are given by $\vec{x}_1$ and $\vec{x}_2$, and vectors $\vec{x}_1^0$ and $\vec{x}_2^0$ denote the distance between the source and each sensor.

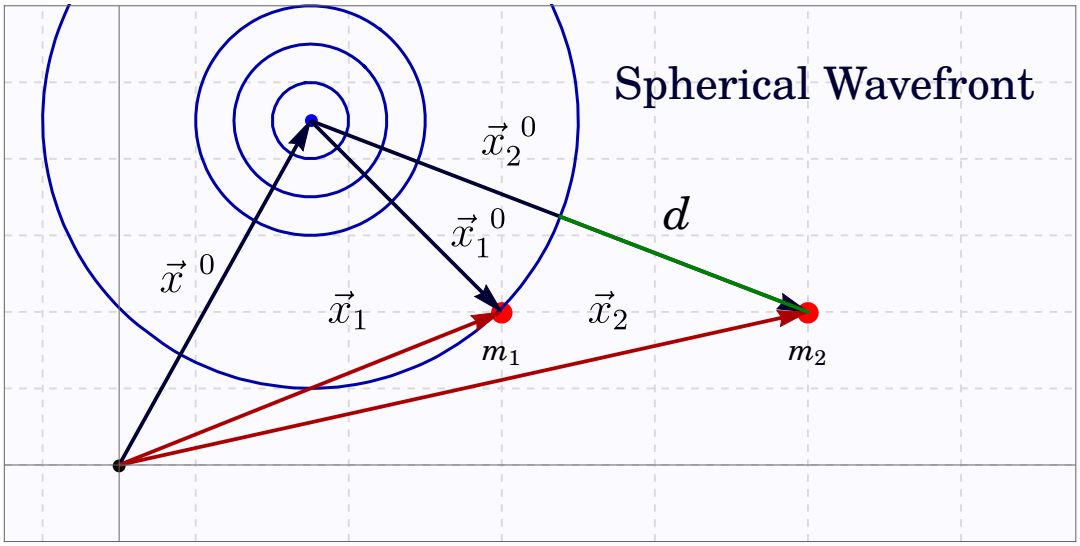


Figure 6: Description of spherical wave arrival at a microphone pair.

The distance a curved wavefront arriving at microphone m_1 must travel to arrive at m_2 is d, which is clearly the difference in the magnitudes of vectors $\vec{x}_1^0$ and $\vec{x}_2^0$. Through vector subtraction, a relationship between the time delay between each sensor in the pair, the source position, and microphone position can be established.

$$\tau_m = \frac{|\vec{x}_2^0| - |\vec{x}_1^0|}{c} \tag{2.6.4}$$

$$\tau_m = \frac{|\vec{x}_2 - \vec{x}^0| - |\vec{x}_1 - \vec{x}^0|}{c} \tag{2.6.5}$$

3 Apertures and array fundamentals

The field value at a sensor location $\vec{x}_m$ is described by some arbitrary $f(\vec{x}_m, t)$, which the sensor samples and produces an output $y_m(t)$. Ideally, the transformation is linear and of infinite bandwidth. Although recorded signals never achieve an infinite bandwidth because every acquisition time scale is finite, the hope for a linear transformation is not too misplaced. Linear devices are quite common, and, in this case, the field and the signal differ by a constant.

Spatial regions that transmits or receives field energy over a finite area are referred to as apertures. In the realm of acoustics examples for this are a microphone or speaker. Transmitting apertures are known as active and receiving apertures are known as passive. An array consists of a group of sensors combined to produce a single output; sensor arrays fall under the category of sampled apertures.

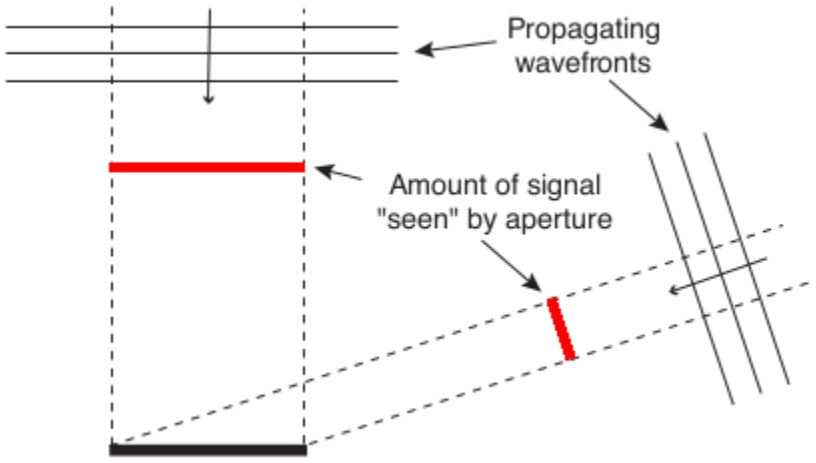


Figure 7: An aperture is only capable of measuring the wavefield to a finite extent. This is a continuous line aperture [23].

3.1 Finite continuous aperture

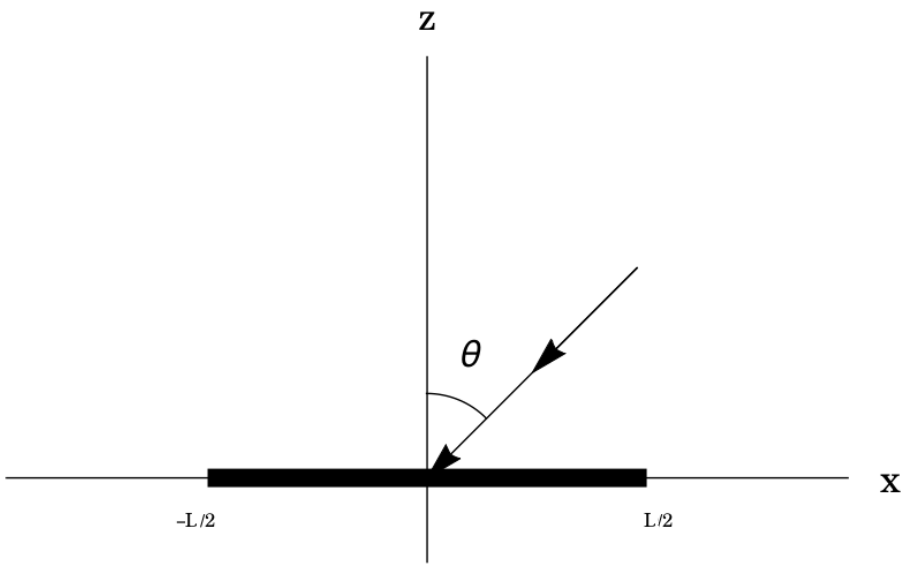


Figure 8: Geometry that describes a continuous linear array.

To describe an aperture, be it passive or active, a function, $w(\vec{x})$, called the aperture function, is defined. The spatial extent of w reflects the size and shape of the aperture. A common application of this function has w equal to 1 within the closed region where the sensors are effective at sampling the field, and zero outside this region. Fundamentally, an aperture is a window through which the field is observed. In general, aperture functions take on real values between 0 and 1 inside the aperture – values different from 1 allow for aperture weighting (shading, tapering, or apodization), which represents the relative weighting of the field within the aperture.

3.1.1 Aperture smoothing function

Treating the array aperture as a singular sensor, the output is defined as:

$$z(\vec{x}, t) = w(\vec{x})\ f(\vec{x}, t), \tag{3.1.1}$$

where f is a propagating acoustic wavefield. The *aperture smoothing function* is the Fourier transform of the sensor's weight, and to a large extent it describes how well a wavefield sampled by an array can be reproduced. To get the aperture smoothing function, calculate the space-time Fourier Transform of the sensor output, z. Real-world application of this will have a dependence on t (in samples) and will not be continuous but discrete. Because z is a multiplication of two functions that are being transformed into Fourier space, the convolution theorem is used to determine the resulting spectrum. From the convolution theorem:

$$\mathcal{F}\{f \cdot g\} = \mathcal{F}\{f\} \bigstar \mathcal{F}\{g\} = (f * g)(t) \equiv \int_{-\infty}^{\infty} f(\tau) g(t - \tau) \mathrm{d}\tau. \tag{3.1.2}$$

The three-dimension spatial to wave-number Fourier transform defined as:

$$\mathcal{F}(\vec{k}) \equiv \frac{1}{(2\pi)^3} \int_{-\infty}^{\infty} f(\vec{x}) \exp\{i\vec{k} \cdot \vec{x}\} \mathrm{d}\vec{x} \tag{3.1.3}$$

The time to angular frequency Fourier transform is defined as:

$$\mathcal{F}(\omega) \equiv \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) \exp\{-i\omega t\} \mathrm{d}t \tag{3.1.4}$$

Taking the Fourier transform each factor inside the signal function z,

$$\mathcal{F}\{w\} = W(\vec{k}) = \frac{1}{(2\pi)^3} \int_{-\infty}^{\infty} w(\vec{x}) \exp\{i\vec{k} \cdot \vec{x}\} \mathrm{d}\vec{x} \tag{3.1.5}$$

$$\mathcal{F}\{f\} = F(\vec{k}, \omega) = \frac{1}{(2\pi)^4} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\vec{x}, t) \exp\{-i(\omega t - \vec{k} \cdot \vec{x})\} \, \mathrm{d}\vec{x} \, \mathrm{d}t \tag{3.1.6}$$

The convolution will be over the variable common to both functions, $\vec{k}$, the wavenumber vector, reciprocal space-domain.

$$Z(\vec{k}, \omega) = \frac{1}{(2\pi)^3} \int_{-\infty}^{\infty} W(\vec{k} - \vec{l}) \; F(\vec{l}, \omega) \; \mathrm{d}\vec{l}, \tag{3.1.7}$$

where $\vec{l}$ is an arbitrary vector in the same units as $\vec{k}$ in reciprocal space [1/m] that represents a shift to the wavenumber vector. Z is the function z in the inverse domain. $W(\vec{k})$ is called the aperture smoothing function, and one can see, that its effect in Fourier space is that of a kernel. This is marginally analogous to blurring of images in image processing, where the "true" image will be the unadulterated spectrum of the wave-field and a filtered image post-application of a convolution matrix will represent the array output of the sampled acoustic wavefield.

Despite the blurring effect of the smoothing kernel, all information concerning the propagating signal can be recovered at the output of the array. Consider a plane wave,

$$f(\vec{x}, t) = s(t - c^{-1}\vec{\zeta^0} \cdot \vec{x}), \tag{3.1.8}$$

that is propagating in direction $\vec{\zeta^0}$. The wave-number frequency spectrum can be calculated as before through the Fourier transform. There should be one spectral line in the wavenumber spectrum for when an arbitrary wavenumber vector $\vec{k} = \vec{k^0} = \omega\vec{\zeta^0}/c$. The wavenumber-frequency spectrum of such a wavefield is

$$F(\vec{k},\omega) = S(\omega)\delta(\vec{k} - \omega\vec{\alpha^0}), \tag{3.1.9}$$

where $S(\omega)$ is the Fourier transform of $s(t)$ and $\delta(\vec{k})$ is the three-dimensional Dirac-delta function. To see the effect observing this wavefield through a finite aperture has on the spectrum of the wavefield,

$$Z(\vec{k},\omega) = \int_{-\infty}^{\infty} W(\vec{k} - \vec{l})F(\vec{l},\omega)\mathrm{d}\vec{l} \tag{3.1.10}$$

$$Z(\vec{k},\omega) = \int_{-\infty}^{\infty} W(\vec{k} - \vec{l})S(\omega)\delta(\vec{k} - \omega\vec{\alpha^0}) \tag{3.1.11}$$

$$Z(\vec{k},\omega) = S(\omega)W(\vec{k} - \omega c^{-1}\vec{\zeta^0}). \tag{3.1.12}$$

For values of the wavenumber vector that equal $\omega c^{-1}\vec{\zeta^0}$, $Z(\omega c^{-1}\vec{\zeta^0},\omega) = S(\omega)W(\vec{0})$, the output signal of the array is equal to the the propagating signal multiplied by a constant. Therefore, all of the information concerning the propagating signal can be interpreted by the output of the aperture. For other values of $\vec{k}$ the signal's frequency spectrum is multiplied by a frequency-dependent gain, $W(\vec{k} - \omega c^{-1}\vec{\zeta^0})$, that filters the relative intensities and phases of the frequency components in the signal spectrum. This result can be generalized to a superposition of traveling plane waves

with different wavenumber vectors (traveling in all directions), so that this analysis can be carried out on something mathematically similar to a wavefield it is expected to measure with an array.

$$f(\vec{x},t) = \sum_i s_i(t - \vec{\alpha_i^0} \cdot \vec{x}) \tag{3.1.13}$$

$$F(\vec{k},\omega) = \sum_i S_i(\omega)\delta(\vec{k} - \omega\vec{\alpha_i^0}) \tag{3.1.14}$$

$$Z(\vec{k},\omega) = \sum_i S_i(\omega)W(\vec{k} - \omega\vec{\alpha_i^0}) \tag{3.1.15}$$

The spectrum will be nonvanishing for any choice of $\vec{k} = \omega\vec{\alpha_i^0}$, for a particular path in wavenumber-frequency space, $\vec{k} = \omega\vec{\alpha_j^0}$, where j is an index corresponding to one of the propagating plane waves. The aperture's output will be a linear combination of the j term of the signal spectrum multiplied by a constant, $S_j(\omega)W(\vec{0})$, and the frequency components of the other planes waves in the superposition are distorted by a frequency-dependent gain factor $W(\omega\left[\vec{\alpha_j^0} - \vec{\alpha_i^0}\right])$. The output spectrum then has the form:

$$Z(\omega\vec{\alpha_j^0},\omega) = S_j(\omega)W(\omega\left[\vec{\alpha_j^0} - \vec{\alpha_i^0}\right]) + \sum_{i\neq j} S_i(\omega)W(\omega\left[\vec{\alpha_j^0} - \vec{\alpha_i^0}\right]) \tag{3.1.16}$$

The aperture smoothing window $W(\vec{k} - \omega\vec{\alpha_i^0})$ can be designed so the value $W(\omega\left[\vec{\alpha_j^0} - \vec{\alpha_i^0}\right]) << W(\vec{0})$, and in this way directionally filter by allowing the

passage of signals propagating from the desired direction, $\vec{\alpha_j^0}$, and reject signals propagating from other directions. This is the fundamental concept behind spatiotemporal filtering.

The spatial extent or size of an array aperture determines the resolution with which two plane waves can be separated. An aperture $w(\vec{x})$ with a large spatial extent has a narrow aperture smoothing function $W(\vec{k})$, which leads to minimal spectral smoothing (a closer representation to the actual acoustic field the larger an array – The perfect aperture smoothing function is infinitely narrow, $\delta(\vec{k})$; this corresponds to an aperture over all space, which is impractical.), and the more focused the aperture can be on any specific direction. The main lobe can be considered the filter's passband and the side lobes as the stop band, because the side-lobes have decreasing amplitude.

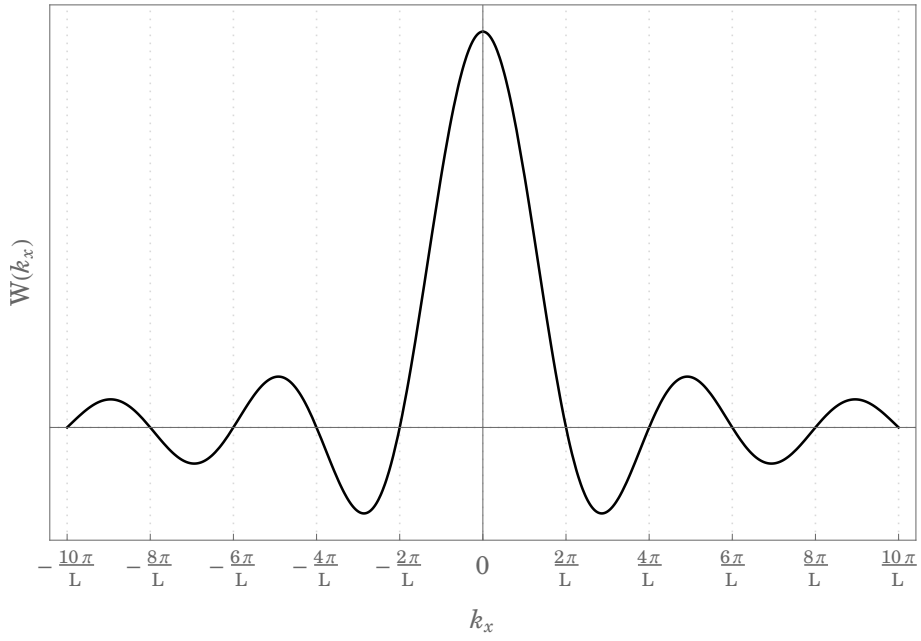


Figure 9: Aperture smoothing function for a continuous line array aligned with the x-axis. L is the length of the continuous aperture.

3.1.2 Resolution

To resolve plane waves propagating past an array, consider a wave-field of two plane waves:

$$Z(\vec{k},\omega) = S_1(\omega)W(\vec{k} - \omega\vec{\alpha_1^0}) + S_2(\omega)W(\vec{k} - \omega\vec{\alpha_2^0}) \tag{3.1.17}$$

The Rayleigh criterion is the generally accepted criterion for the minimum resolvable detail - limited when the first zero of one plane wave aperture smoothing function coincides with a peak of the other plane wave. Each plane wave causes a replica of the aperture smoothing function to appear in the wavenumber-frequency spectrum. The resolution in this context is defined as the smallest wavenumber that produces a zero in the aperture smoothing function. Resolution can be measured as the just noticeable difference in direction of propagation, as $\vec{k} = 2\pi\vec{\zeta}/\lambda$, the directional resolution is $\delta\zeta \cdot \lambda/2\pi$. In general, the Rayleigh resolution criterion requires two waves to differ by at least one period over the aperture.

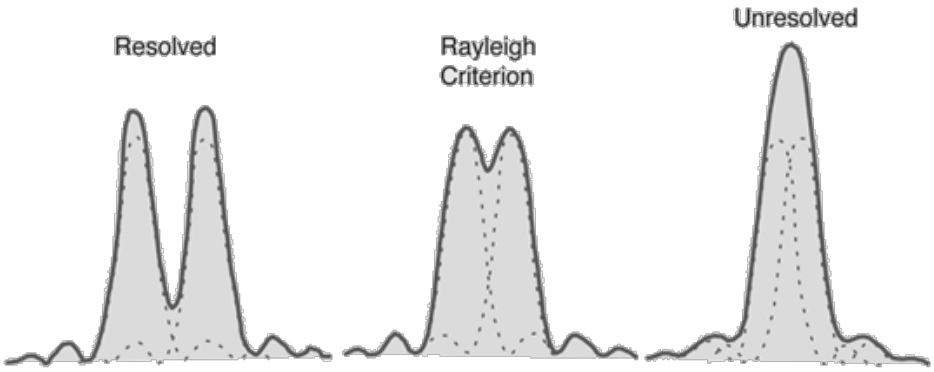


Figure 10: The Rayleigh criterion [24].

3.1.3 Sidelobe height

Height of the highest sidelobe relative to the main sidelobe measures an aperture's ability to reject unwanted noise and signals; lower sidelobes mean less unwanted frequencies in the spectrum and less noise. To get this ratio, differentiate the aperture smoothing function, and maximize to get the location and height of the highest sidelobe. Ratio is independent of length, a longer array does not give lower sidelobes.

3.1.4 Aperture ambiguities

In the case of symmetrical array apertures, waves propagating at the same speed and different directions can yield the same output – the continuous linear array is only sensitive to the x component of the wavenumber vector. A set of directions yielding identical aperture response is termed an ambiguity set, which is the set of all $\vec{\alpha_0}$ that yield the same values for $W(\vec{k} - \omega\vec{\alpha^0})$, with a constraint that $|\vec{\alpha_0}|$ is fixed – that is, the same value of magnitude, $1/c$. A linear aperture is cylindrically symmetric, the ambiguity set makes a cone around the aperture. Signals propagation from above below or to the side of the linear aperture cannot be distinguished.

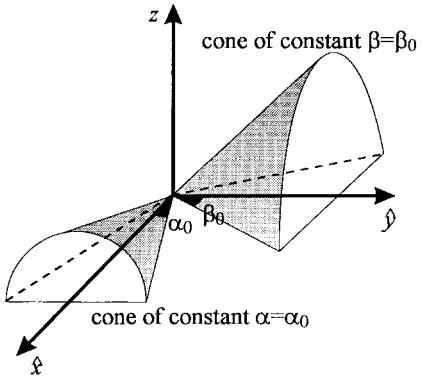


Figure 11: The cone of ambiguity [25].

3.1.5 Apparent velocity in a planar array

The velocity of the wave across the array seems different to the array than its propagation speed in reality. Illustration of this, consider the simplified case of an array that lies in the x-y plane and a monochromatic plane wave traveling past it. Consider a plane wave, $f(\vec{x},t) = \exp\{i(\omega^0 t - \vec{k^0}\cdot\vec{x}\}$, The usual parameters hold for this plane wave: $c = \omega_0/k$ with the slowness vector defined as, $\vec{\alpha^0} = \vec{k^0}/\omega^0$. The direction across the array will be defined for a plane wave traveling in the z-direction so that $\vec{k^0} =< 0, 0, -k_z >$.

Wavefronts are planes of constant phase. The points of equal field value, $f(\vec{x},t)$ always form planes like this, so that the wave's complex amplitude, which includes the phase factor $\exp\{i\phi\}$. The amplitude changes value in time simultaneously at all points in the aperture, and the wave appears to be everywhere simultaneously inside the aperture. Since the slowness vector depends on the propagation vector, the x and y components of the slowness vector are zero and therefore the speed in that direction is infinite.

Consider again, the same plane wave, except with propagation direction $+x$, $\vec{k^0} =< k_x, 0, 0 >$. The velocity in the x-direction is $c = 1/\alpha = \omega^0/k_x$, it reaches everywhere in the y in an instant ("infinite" speed), and there is no z component to the array. Another way to think about this "infinite speed" is that, the apparent magnitude of wave speed in the planar array simply won't depend on the dimension it is not in, and for propagation along the orthogonal basis vectors this term will drop out. The x-y planar array will distinguish $\alpha^0 =< \alpha_x^0, \alpha_y^0, 0 >$. In this event, the magnitude of the slowness vector can be described as, $1/c_a$, where c_a is the apparent speed of the wave perceived by the array,

$$\sqrt{(\alpha_x^0)^2 + (\alpha_y)^2} = 1/c_a \tag{3.1.17}$$

$$c_a = \frac{1}{\sqrt{(\alpha_x^0)^2 + (\alpha_y)^2}} = \frac{\omega^0}{\sqrt{(k_x^0)^2 + (k_y)^2}} \geq \frac{\omega^0}{k^0}, \tag{3.1.18}$$

which is greater than or equal to the actual speed of the wave.

3.2 Spatial sampling

In piratical application an array consists of sensors at discrete positions in space, so it cannot be considered as a continuous aperture. Essentially, these sensors, located at $\{\vec{x}_m\}$, spatially sample the wavefield at certain time instants, $\{t_n\}$. As an introduction to this type of measurement and further analysis consider two questions: "What effect does spatially sampling have on a propagating wavefield?" and "Can a true representation of the wavefield signal be reconstructed under such conditions?". Once the analysis has been carried out in one dimension it can be easily generalized to higher dimensions.

An arbitrary wavefield dependent on one spatial variable in time, $f(x,t)$ exists, propagating through space, to be recorded by a linear aperture of discrete sensors, $m \in \{1,2,3,...\}$, seperated with uniform interelement spacing, d. Temporal signals recorded at each sensor, m, will be denoted by set $\{y_m\}$. Analgous to temporal sampling, the signal recorded at each sensor is related to the wavefield by $y_m(t) = f(md,t)$, where d is called the *sampling interval*. The wavefield will be examined at time instant, t_0, an arbitrary observation time, so that the signal recorded at the m-th signal is $\{y_m(t_0)\}$; with a fixed t the only sampling being considered is

that due to the sampling in space a set of sensors positioned discretely in space does to the wavefield. There are many parallels between spatial and temporal sampling; like most discussions of sampling in time, and similarly here, ideal signals are bandlimited, with spectrum components in inverse space hopefully defined within certain limits – close fulfillment of this ideal will determine the ability of sampling to reproduce information of the continuous signal within the sampled signal. For spatial sampling, k, the wavenumber, is the spatial frequency, and inverse domain for continuous position variable x. To see these requirements met, all the wavenumber components of $f(x, t_0)$ will be located within the range $|k| \geq k_0$ and the sampling interval d is less than or equal to π/k_0. According the the sampling theorem, the continuous variable function is represented by a sum over all samples (over each sensor sampling in location space) of the sampled signal multiplied by a normalized sinc function whose argument involves a difference of the continuous variable to the discrete variable multiplied by the sampling interval that is divided by the sampling interval, this yields the Whittaker-Shannon interpolation formula [18]:

$$f(x, t_0) = \sum_{m=-\infty}^{\infty} y_m(t_0) \, \frac{\sin \pi \left(\frac{x}{d} - m \right)}{\pi \left(\frac{x}{d} - m \right)} \tag{3.2.1}$$

Most of the information of interest is located within a signal's spectrum. Can the spectrum of the continuous signal be determined from spectrum of the spatially sampled signal? For a continuous signal, $s_c(x)$, and its sampled representation, $s(m)$, the discrete-variable signal, the Fourier transform and the inverse Fourier transform are defined:

$$
\begin{aligned}
s_c(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} s_c(k)\exp\{-ikx\}\mathrm{d}k \qquad & S_c(k) &= \int_{-\infty}^{\infty} s_c(x)\exp\{ikx\}\mathrm{d}x \\
s(m) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} S(\breve{k})\exp\{i\breve{k}m\}\mathrm{d}\breve{k} \qquad & S(\breve{k}) &= \sum_{m=-\infty}^{\infty} s(m)\exp\{i\breve{k}m\}
\end{aligned}
\tag{3.2.2}
$$

Note the Fourier transform for discrete-variable signals is periodic with 2π, that is, $S(\breve{k}) = S(\breve{k} + 2\pi l)$, for any integer value of l. The continuous wavevector number, k, has units of radians/meter and equals 2π times the spatial frequency. The sampling wavenumber magnitude, $\breve{k}$, has a period of 2π, ranging from $\{-\pi, \pi\}$ typically; the units for this value are radians/sampling period.

The continuous signal s_c, a given time instant t_0, can be related to the discrete representation of the real-world signal s, through $s_c(md) = s(m)$.

$$
s_c(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} s_c(k)\exp\{-ikx\}\ \mathrm{d}k, \tag{3.2.3}
$$

Performing a change of variables, $\breve{k} = kd$, transforms the expression

$$
s(m) = s_c(md) = \frac{1}{2\pi}\int_{-\infty} \infty s_c(k)\exp\{-ikmd\}\ \mathrm{d}k \tag{3.2.4}
$$

$$
s(m) = \frac{1}{2\pi d}\int_{-\infty}^{\infty} s_c(\frac{\breve{k}}{d})\exp\{-i\breve{k}m\}\ \mathrm{d}\breve{k} \tag{3.2.5}
$$

By decomposing this integral to many integrals over intervals of 2π it can be shown that this equation fits the form of an inverse Fourier transform for discrete-variable

signals.

$$s(m) = \frac{1}{2\pi d} \sum_{-\infty}^{\infty} \int_{-\pi+2\pi p}^{\pi+2\pi p} S_c\left(\frac{\breve{k} - 2\pi}{d}\right) \exp\{-i\breve{k}m\} \exp\{i2\pi pm\} \, \mathrm{d}\breve{k} \tag{3.2.6}$$

Taking a Fourier transform of this equation gives us a relation between the spectrum of the continuous signal and the sampled signal.

$$S(\breve{k}) = \frac{1}{d} \sum_{p=-\infty}^{\infty} S_c\left(\frac{\breve{k} - 2\pi p}{d}\right) = S(kd) = \frac{1}{d} \sum_{p=-\infty}^{\infty} S_c\left(k - \frac{2\pi p}{d}\right) \tag{3.2.7}$$

Basically, the conclusion to be drawn from this is that the sampled signal's spectrum equals the sum of periodic replications of the continuous signals spectrum. If the continuous signal has no frequency components outside the domain $|k| \leq \pi/d$, there is no overlap in the periodic replications of the spectrum in the spectrum of $S(\breve{k})$. In this ideal scenario, when the continuous signal is perfectly bandlimited, there is no aliasing, and the spectrum for the continuous signal can be obtained easily from that of the recorded signal,

$$S(kd) = \frac{1}{d} \, S_c(k), \text{ for } |k| \leq \pi/d \tag{3.2.8}$$

When the continuous signal is not perfectly bandlimited, the periodic replication of the spectrum overlap and there is aliasing. In other words, one period of the sampled signal's spectrum does not equal the continuous signal's spectrum. Spectral components of the continuous signal outside of the interval $|k| \leq \pi/d$ become spectral

components of $S(\breve{k})$ inside the interval $|k| \leq \pi/d$.

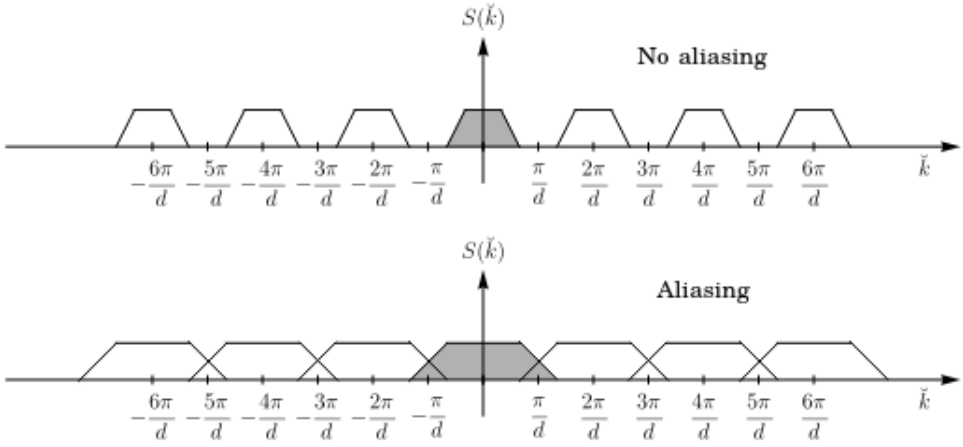


Figure 12: Representations of the spectrum of the signal that has been sampled in space. If the signal is perfectly bandlimited to $|\vec{k}| = \pi/d$ the periodic replications of the spectrum that occur due to sampling do not interfere and blend. If the space is undersampled the spectrum begins to creep into other intervals and reproducing the signal $s(m)$ is reproduced from the blended spectrum resulting in a poorer reproduction of the original signal.

This domain can be represented in terms of wavelength rather than wavenumber. In the event that $k = \omega/c \rightarrow 2\pi f/c$, the requirement to avoid aliasing becomes [23]

$$d \leq \frac{\lambda}{2}, \tag{3.2.9}$$

The interelement spacing, d, must less than or equal to half the wavelength. So, in general, there is a limit to what signals can be reconstructed perfectly by an array depending on this array parameter.

This can be generalized to higher dimensions very easily. The sampling theorem can be applied to each dimension independently. In particular for two dimensions, the interpolation formula is

$$s_c(x,y) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} s(m,n) \frac{\sin \pi \left(\frac{x}{d_x} - m \right)}{\pi \left(\frac{x}{d_x} - m \right)} \cdot \frac{\sin \pi \left(\frac{y}{d_y} - m \right)}{\pi \left(\frac{y}{d_y} - m \right)} \tag{3.2.10}$$

And the Fourier transforms, therefore, are related by:

$$S(\breve{k}_x, \breve{k}_y) = \frac{1}{d_x d_y} \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} S_c \left(\frac{\breve{k}_x - 2\pi p}{d_x}, \frac{\breve{k}_y - 2\pi q}{d_y} \right) \tag{3.2.11}$$

To represent an infinitely long signal, an infinite number of samples is required. The memory of any device used in digital signal acquisition and processing will never have an infinite capacity. Eventually the signal must be cut short. In addition the precision of the device is an issue, the signal information is represented by a finite number of bits. As a result, this condition of a perfectly bandlimited continuous signal is rarely achieved for finite time signals.

3.3 Arrays of discrete sensors

In application, spatiotemporal signals are sampled and recorded using arrays of individual sensors placed at specific locations in space. These sensors can be placed in regular or irregular patterns. Regular arrays have interelement spacings that are periodic and irregular arrays have interelement spacings that are not necessarily periodic.

The benefits of using regular arrays should be self-evident. The aperture smoothing functions of such arrays are nicely behaved and straight-forward to analyze. In this case, if designed properly, the aliasing effects that occur would be the spatial aliasing

associated with the array spacing and the time aliasing associated with sample rate.

Consider a wavefield, $f(x,t)$ with a single spatial dimension x and time t being sampled by an array with M sensors equally spaced d apart laying on the x-axis. The set of signals to be recorded by this array is defined as $\{y_m(t)\}$ – these correspond to the wavefield's values sampled every d meters: $y_m(t) = f(md,t)$. The wave-field's wavenumber-frequency response representation is given by

$$F(k,\omega) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,t)exp\{-i(\omega t - kx)\}dxdt \tag{3.3.1}$$

The spectrum of the sampled wavefield is given through

$$Y(\breve{k},\omega) = \sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} y_m[n]\exp\{-i(\omega\ n - kmd)\} \tag{3.3.2}$$

Using the relation established in the last section,

$$Y(\breve{k},\omega) = \frac{1}{d}\sum_{p=-\infty}^{\infty} F\left(k - \frac{2\pi p}{d},\omega\right) \tag{3.3.3}$$

Implementing sensor weights, the output spectrum of the array as a whole is given through

$$Z(\breve{k},\omega) = \frac{1}{2\pi}\int_{-\pi/d}^{\pi/d}\left[\sum_{p=-\infty}^{\infty} F\left(l - \frac{2\pi}{d}\right)\right] W(k-l)\mathrm{d}l \tag{3.3.4}$$

where the output of the array is a smoothed version of the source signal. The sum in brackets defines what if any aliasing occurs. If the spacing condition is met there is no spatial aliasing.

3.3.1 Grating lobes

The effect of aliasing on the beampattern is the appearance of something like main lobes appearing for different angles of arrival or wavevectors other than the desired angle or wavevector the array has been steered to identify. Signals propagating from directions corresponding to the spatial frequencies in the beampattern where the grating lobes occur are indistinguishable from signals propagating from the direction of the mainlobe. In the case of undersampled regular linear arrays and some regular multidimensional arrays, spectral peaks emerge that the mainlobe values and are periodic. In irregular arrays, grating lobes appear that do not have values equal to the mainlobe and are not periodic. It is possible, but very difficult to determine where they might occur, and do limit the spatial frequencies an array can uniquely determine, though spatiotemporal filtering can overcome this slightly. By searching for specific frequencies at specific time one can determine the distance through power optimization techniques as long as the grating lobes are indeed less than the main lobe.

An example of array analysis

A hypothetical discrete line array will be analyzed using the background information that has been presented. The proposed discrete line array will consist of four sensors with an interelemental spacing of 2 m and will be positioned at $\vec{x}_m = \{(-3,0,0),(-1,0,0),(1,0,0),(3,0,0)\}$, see Figure 1. They will be imposed with sensor weights $w_m = \{1,1,1,1\}$.

The sensors are arranged in a line and position symmetrically across the x-axis.

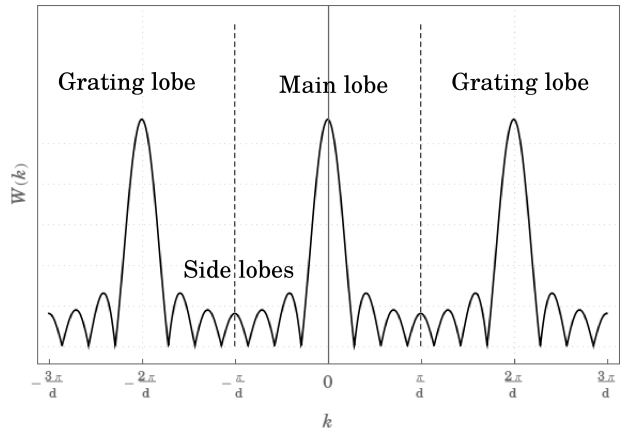


Figure 13: In the array pattern of an aperture that spatial undersamples the field, the mainlobe is centered, the side lobes, which determine the passband of the array, are symmetric on either side for symmetric arrays. The *visible region* is the part of the array pattern that falls in between $|\vec{k}| = \pi/d$. Grating lobes appear as mainlobes.

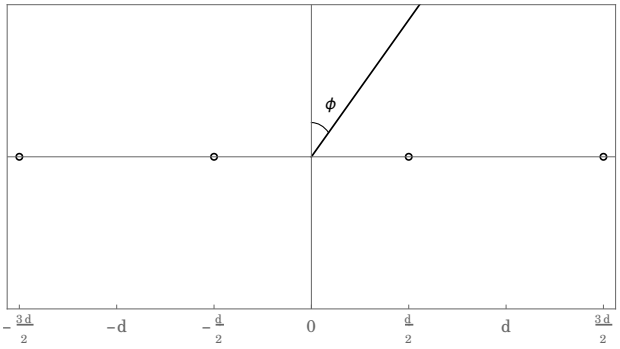


Figure 14: The hypothetical discrete line array geometry plotted in two dimensions, $\theta \to \frac{\pi}{2}$. There are four equally spaced sensors, each with weight of 1.

Throughout this analysis the azimuth angle, ϕ will be shifted by $\pi/2$ radians, as indicated by Figure 1, so that an angle of 0 radians corresponds to a direction perpendicular to the array.

Throughout this analysis the far-field assumption is made and all waves propagating in the wavefield are monochromatic plane waves of an arbitrary constant

frequency, $\omega = 2\pi f$,

$$f(\vec{x}, t) = s(t - \vec{\alpha}^0 \cdot \vec{x}). \tag{3.3.5}$$

In this representation of a monochromatic plane wave, which is dependent on position, $\vec{x}$, and time, t, we include the slowness vector $\vec{\alpha}^0$ of a wave incident on the array (any feature of an incident wave will be notated with a superscript 0). An arbitrary slowness vector, $\vec{\alpha}$ is defined as,

$$\vec{\alpha} = \vec{k}/\omega = k/\omega\hat{\zeta}. \tag{3.3.6}$$

The magnitude of the slowness vector is $|\vec{\alpha}| = 1/c$, the inverse of the wave speed, the sound speed in water – approximately $c = 1531$ m/s in this application. The wavenumber vector is $\vec{k}$, and has magnitude of $|\vec{k}| = \omega/c = 2\pi/\lambda$. The unit vector, $\hat{\zeta}$ is defined normal to the wave front, and denotes the propagation of direction for an arbitrary plane wave. The slowness vector contains 2 pieces of pertinent information simultaneously: the speed and direction of the wave.

Analysis begins usually by determining what the array pattern for given frequency of incident wave and arrival angle ϕ over several different frequencies. The array pattern is synonymous with the aperture smoothing function because it determines the amplitude and phase of the beamformed signal when the wavefield consists of a single plane wave. It is given through,

$$W(\vec{k}) = \sum_{m=0}^{M-1} w_m \exp[i\vec{k} \cdot \vec{x}_m] \tag{3.3.7}$$

There are many ways to plot such a function. To plot it in real space, one may see

$\Re[W(\vec{k})]$, $|W(\vec{k})|$, $|W(\vec{k})|^2$, $\log|W(\vec{k})|$ (dB), or $2\log|W(\vec{k})|$ (dB). One may see this given with an argument that represents an angle of arrival relative to the perpendicular bisector of the array, and it is not uncommon to see in either a Cartesian plot or a polar plot.

Figure 15 displays array patterns of a four element sensor array of spacing 2 m for frequencies 100 Hz, 382.75 HZ, 500 Hz, 750 Hz, and 1000 Hz. Central maximum in the array pattern are called main lobes and indicate a the array's sensitivity to a given direction of arrival. The smaller maxima are called sidelobes. The spatial-filtering effect of the array sampling a signal in the position domain is not perfect, and these sidelobes determine how sensitive to other angles of arrival a given array is. In fact, the ratio of the height of the sidelobe to the height of the mainlobe serves as a practical indicator of how effective an array is at discriminating directions of arrival. Consequently, a plane wave incident on an array at the angle that corresponds to the location of the main lobe will have its amplitude maximized, while plane waves arriving from other directions will have their amplitude minimized according to the array pattern. The space must be sampled sufficiently to avoid artifacts, and the inter-element spacing requirement to avoid spatial aliasing completely is:

$$d \leq \lambda/2, \tag{3.3.8}$$

and perfect replication of the wavenumber spectrum is limited to the ratio of the array spacing and the wavelength in space similar to the Nyquist sampling theorem.

For a discrete line array with 2 m interelement spacing, that means aliasing free replication of the source signal and the spectrum of the source signal for planes waves of frequency up to 382.75 Hz. Beyond this threshold, periodic replications of

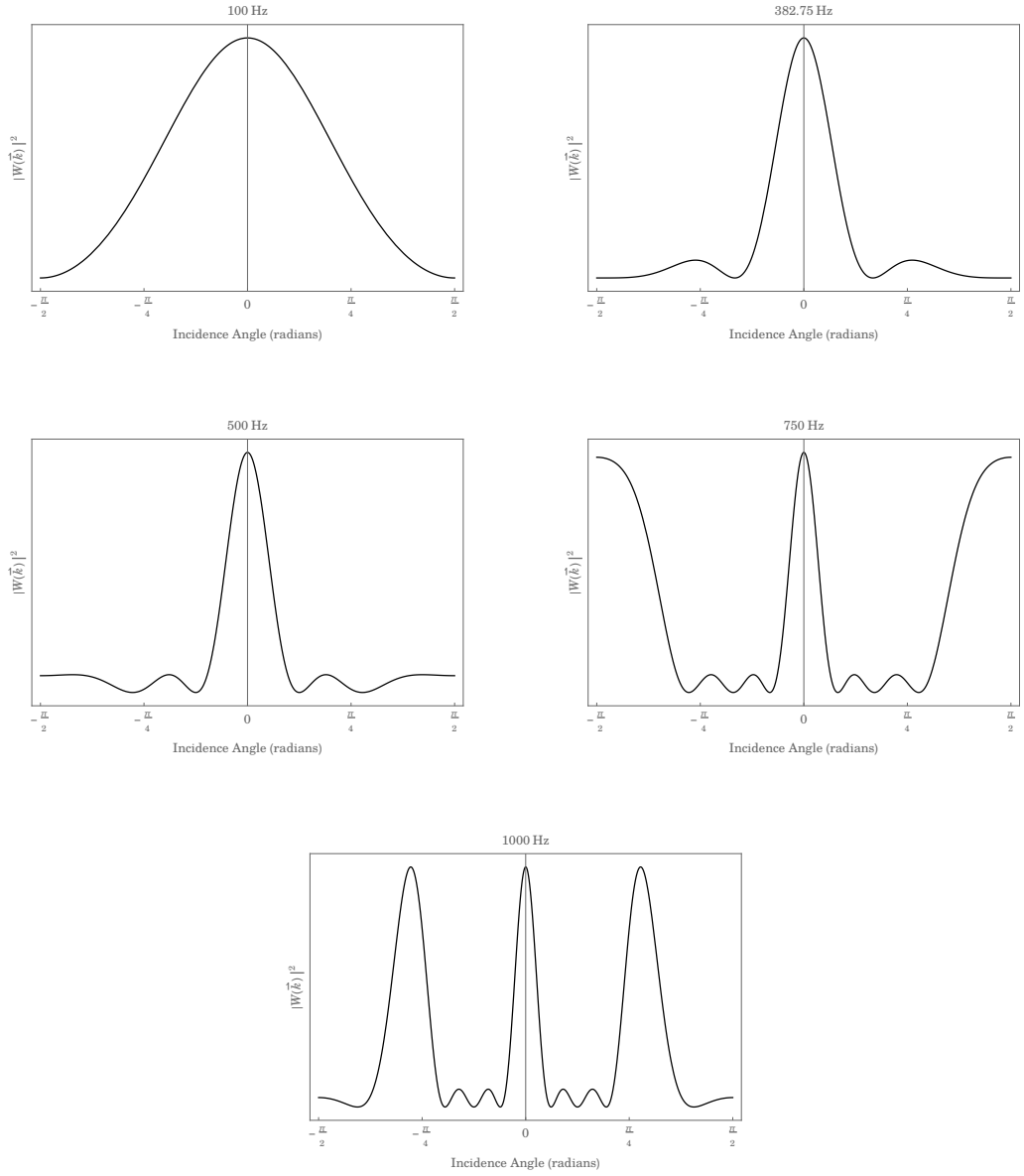


Figure 15: The array pattern $|W(\vec{k})|^2$ is plotted as a function of incidence angle for several given frequencies. Notice the effect of higher frequency plane waves by the appearance of grating lobes in the array pattern.

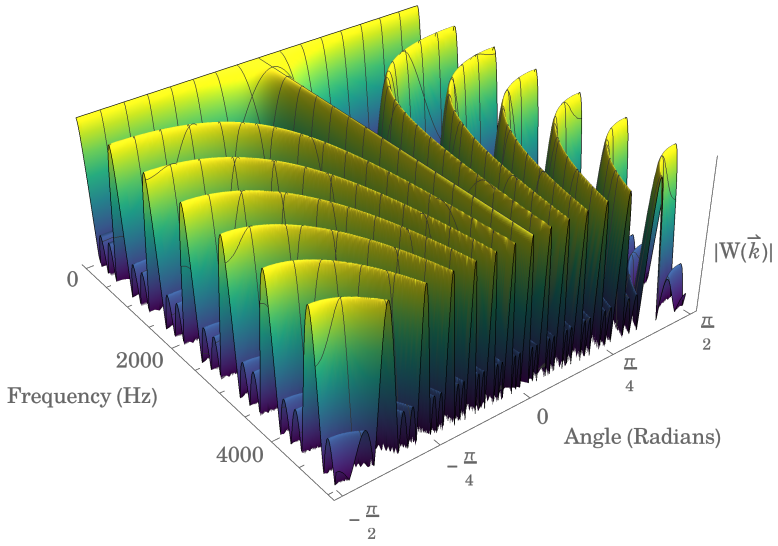


Figure 16: The hypothetical array's directivity pattern.

the source signals spectrum in Fourier space begin to blend, distorting wavenumber information. The effect of this on the array pattern is spurious mainlobes known as grating lobes in the visible region, $\phi = \{\pi/2, \pi/2\}$. The beam *directivity pattern* extends the description of array pattern over frequency space, and a three dimensional plot relating $|W(\vec{k}|^2$ to either wavenumber or incident angle and to frequency is called a directivity pattern. It can indicate the size, shape, and direction of the main beam and side lobes, and also the structure of the array pattern that pertains to emerging mainlobes for higher frequencies of propagating signals.

The directivity pattern of a four sensor array is plotted in Figure 16. The frequency range of interest is 0 Hz to 5000 Hz because soniferous fish vocalize in this region of the spectrum. The primary spectrum in the range $|\vec{k}| \leq \pi/d$ is polluted by the periodic replications and grating lobe emerge, because of the wavenumber's dependence on temporal frequency, above 382.75 Hz. The grating lobes have values that equal the

maximum value at the mainlobe, and waves of frequency higher than the threshold that are incident on the array from a direction that coincides with one of these grating lobes has its amplitude maximized in the same way as if it were incident on the array at the angle that coincides with the true mainlobe.

4 Beamforming

The objective of this research is to gather position and time information about a propagating wavefield $f(\vec{x}, t)$ by sampling the field at each sensor. This field will include noise and signals from many sources. To analyze information about a single source, such as a fish in a creek, the array must be designed to have the ability to focus on selected signals. Any attempt to focus the array to a target should at least include linear filtering through the use of bandpass filtering, and, in this way, signals occupying different frequency bands can be examined separately. Furthermore, to obtain position information about a source from any arbitrary acoustic wavefield, an array must also have characteristics that allow for spatial filtering, which is an advantage of arrays over a single omnidirectional sensor, which does not provide any spatial filtering.

An array's directional sensitivity is called its directivity pattern [17] – this is determined solely by the array's design and physical construction. To steer the sensitivity of the array the maximum of the directivity pattern must be steered so that it is directed toward a desired angle of arrival. This directivity defines the bandwidth and stopband of the spatial filter.

With large apertures, maintaining precision is difficult. Large apertures must be designed to close tolerances to not be sensitive to aberrations of the recorded signal, which was very difficult to achieve in this application. The steering, here, is achieved through signal processing techniques, and the accumulation of error will affect the resolution and results of any localization algorithm. Any type of signal processing where these techniques are relevant is colloquially called beamforming.

4.1 Delay-and-sum beamforming

This is the oldest and simplest array signal processing technique for source localization, and it is still abundantly relevant in present research. The underlying principles are fairly simple: because sound travels at a finite speed within the medium, a sound issued from a specific location at time t_0 within an aperture will be recorded at sensors m as occurring at different times $\{t_m\}$. The delays in arrival of the sound to each receiver are therefore $\{\Delta_m\} = \{t_m - t_0\}$.

If a propagating signal is present in an array aperture, the sensor outputs delayed by appropriate amounts, $\{\Delta_m\}$, and summed, reinforce the signal with respect to the noise and waves propagating in different directions. The delays that reinforce the signal are directly related to the length of time (and distance $d = ct$) of propagating between 2 sensors.

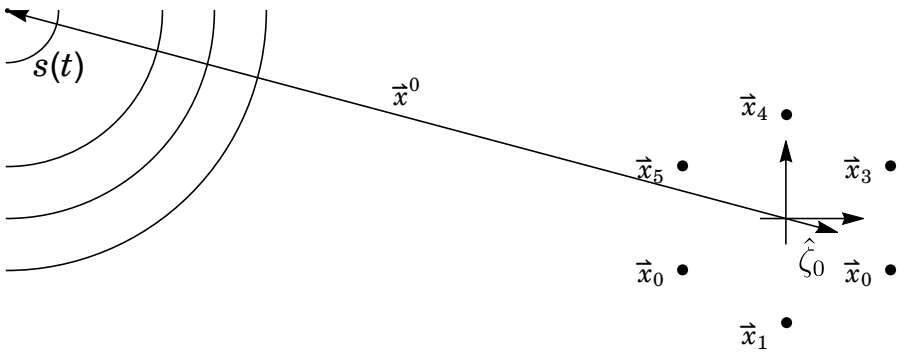


Figure 17: Definition of vectors for conventional delay-and-sum beamforming analysis.

Referring to Figure 17, $\vec{x}^0$ is a position vector pointing from the origin to the source, Let $s(t)$ be the time-dependent signal emitted at the source location, $\vec{x}^0$, that propagates through the array, $\vec{\zeta}_0$ is a unit vector in the direction of propagation (points from $\vec{x}^0$ to $\mathcal{O}$. For simplicity of notation, $|\vec{x}^0| \equiv r^0$. The overall sound field measured

by the hydrophones, $f(\vec{x}, t)$ can include the summation of contributions from any number of sources. For an arbitrary array with an arbitrary array aperture function, M sensors are located at $\{\vec{x}_m\}$, $m = 0, ..., M-1$. It is common for the array's *phase center*, defined by:

$$\sum_{m=0}^{M-1} \vec{x}_m = \vec{0}, \tag{4.1.1}$$

to be taken as the origin, although generally it does not matter. There need not be a sensor located at the phase center, and any origin may be chosen at the investigator's discretion; all sensor positions $\{\vec{x}_m\}$ are defined relatively from this position, and axes for a desired coordinate system can be chosen arbitrarily.

A delay-and-sum algorithm applies a delay Δ_m and an amplitude weight w_m to the output of each sensor so that the output of an array designed to operate as a delay-and-sum beamformer is:

$$z(t) = \sum_{m=0}^{M-1} w_m y_m(t - \Delta_m), \tag{4.1.2}$$

where the waveform recorded at each sensor, y_m, contains the signal and includes noise present in the sensor recording. The application of amplitude weights, or shading, enhances the beams shape, and reduces sidelobe levels. By allowing a summation over differing delays the beam, the maximum in an array's directivity pattern, can be steered toward a desired direction, $\vec{\zeta}_0$ or to a particular point $\vec{x}^0$. Note, two sensors are required for directional steering, 3 sensors are required for point steering in two dimensions, and four sensors are required for accurate steering in three dimensions.

Beamforming is a signal processing technique by which the main lobe of an aper-

ture's array pattern can be steered to be sensitive to other angles of arrival and maximize any signal propagating from that direction. This is achievable by including a delay term on individual sensor measurements before summing each channel together at the output of the array. The delays are given through,

$$\Delta_m = -\frac{\hat{\zeta} \cdot \vec{x}_m}{c} = -\vec{\alpha} \cdot \vec{x}_m, \tag{4.1.3}$$

where, the delays, Δ_m, depend on the direction of arrival, $\hat{\zeta}$, the position of the hydrophones, $\vec{x}_m$, and the sound speed in water, c. The delays can also be chosen to be $-\vec{\alpha}\cdot\vec{x}_m$ where the beamformer judicially discriminates incident plane waves that propagate with slowness (direction and speed), $\vec{\alpha}$. The steered array pattern introduces a delay into the domain, however this is equivalent to considering the beamformer electing to accept plane waves with slowness vector $\vec{\alpha}$. In the case of a plane wave incident on the array with slowness vector $\vec{\alpha}^0$, if the array is steered to accept plane waves such that $\vec{\alpha} = \vec{\alpha}^0$, the output signal is an undistorted replica of the source signal. The speed of sound is calculated from environmental measurements, as is typical in applications of sonar, so, comparing $\vec{\alpha}$ to $\vec{\alpha}^0$ one can conclude that the propagation direction the beamformer assumes and the propagation direction of the incident wave are equal. Conversely, if $\vec{\alpha} \neq \vec{\alpha}^0$ the array attenuates the source signal, and to assess beamformer performance one examines the wavenumber-frequency response, $H(\vec{k}, \omega) = W(\omega\vec{\alpha} - \vec{k})$. This function can describe the performance of an array set-up to operate as a beamformer in two slightly different ways. The *beampattern* describes how the array pattern varies based off different choice of frequency and wavenumber while the direction is held fixed. The *steered response* examines the array pattern for varied propagation direction while holding the frequency, ω^0, and wavenumber

vector, $\vec{k}^0$, of the incident wave fixed. These two descriptions are equivalent for a delay-and-sum beamformer. In terms of the delays imposed on the channels of the array, the wavenumber-frequency response is:

$$W(\omega^0\vec{\alpha} - \vec{k}^0) = \sum_{m=0}^{M-1} w_m \exp[-i\omega\Delta_m] \exp[-i\vec{k} \cdot \vec{x}_m]. \tag{4.1.4}$$

For the hypothetical discrete sensor array of four receivers and inter-element spacing of 2, the beam pattern is plotted for assumed directions of propagation, $\{\pi/6,\ \pi/4,\ \pi/3,\ \pi/2\}$ radians In Figure 18. The array patterns displayed are plotted versus incidence angle ϕ^0. Both angles are measured from the perpendicular to the array axis, and one can examine on each plot where $\phi = \phi^0$ the main lobe occurs meaning the array pattern is most sensitive to plane waves arriving at angles that equal the assumed direction of propagation.

As of yet, the wavefield has been considered of consisting of a single constant plane wave. In field recordings of fish this is not the case. In the case of plural waves which can have arbitrary shape, care is important. Waves propagating to the array from different directions can overlap in wavenumber-frequency space inflating the grating lobes, and discriminating direction is not always enough to reject unwanted signals. Even if the sound speed is constant and known, the various components of the wave may have different frequencies $\omega = n\omega_0$ and wavenumber vectors $\vec{k}$, but the frequencies and wavenumber vectors together must satisfy the constraint $\vec{k}/\omega = \vec{\alpha}$.

For a discrete line array like the example given, *i.e.* one that is positioned *in totem* on the x-axis, $H(\vec{k}, \omega)$ depends only on x coordinates. The previous expression for $W(\omega\vec{\alpha} - \vec{k})$ reduces to:

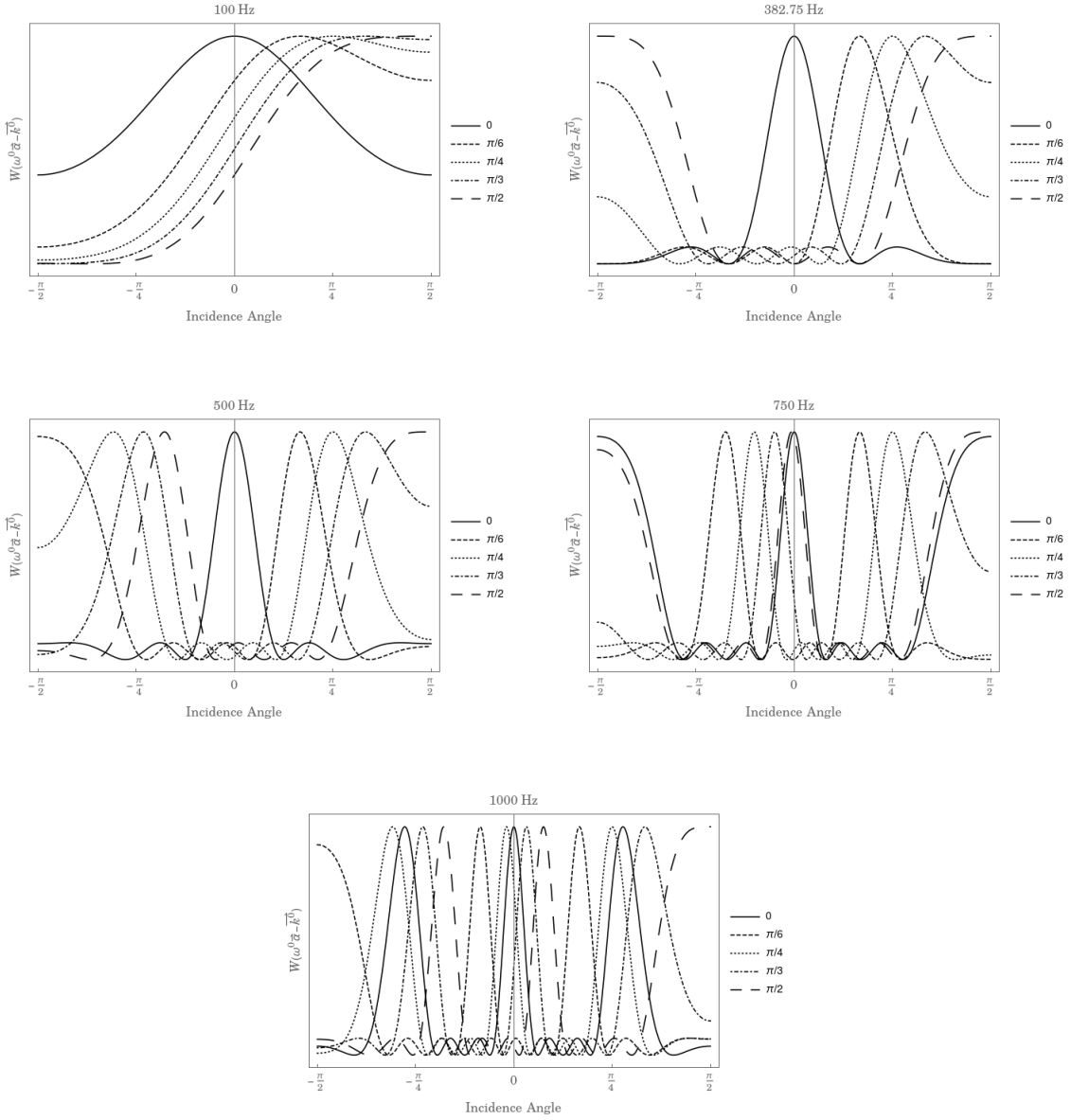


Figure 18: The beampattern is plotted against incidence angle for frequencies 100 Hz, 382.75 Hz, 500 Hz, 750 Hz, and 1000 Hz.

$$W(\omega\vec{\alpha} - \vec{k}) = \frac{\sin\frac{M}{2}(\omega\alpha_x - k_x)d}{\sin\frac{1}{2}(\omega\alpha_x - k_x)d}, \tag{4.1.5}$$

where $M = 4$ and $d = 2$. The wavenumber-frequency response can be plotted on a contour plot to examine the performance of the array and structure of the spectrum. This is done for a hypothetical discrete line array of four sensors with inter-element spacing of 2 m. The wavenumber-frequency response given in (4.4.5) is plotted for $\alpha_x = 1$ and $\alpha_x = 4$ in Figure 19. In each case, the response's main lobe extends diagonally from the origin. It is angled with respect to the axis, because it is being steered in the stated directions. Because of spatial aliasing the mainlobe is repeated as a series of grating lobes, which appear in the plot as parallel ridges with the same amplitude as the mainlobe. Based off Figure 19, the spectral response of the array is considerably different when the array is steered to look in different directions. The angle the main lobes and grating lobes make with the axis are increasingly steeper for increasing values of α_x. From Figure 20 it is obvious that while an array is steered in direction $\alpha_x = 4$ that components of a wave propagating in direction $\alpha_x = 1$ do not intersect with the main lobe of the beam pattern; however, higher frequency components of the wave in direction $\alpha_x = 1$ do overlap with grating lobes in the array's beam pattern even though it has been steered, consequently polluting the gratings lobes. This is not desirable. This scenario can be thwarted by concentrating on a range of temporal frequencies, this achieved with a temporal filter designed to pass frequencies of interest and filter those out which are undesirable.

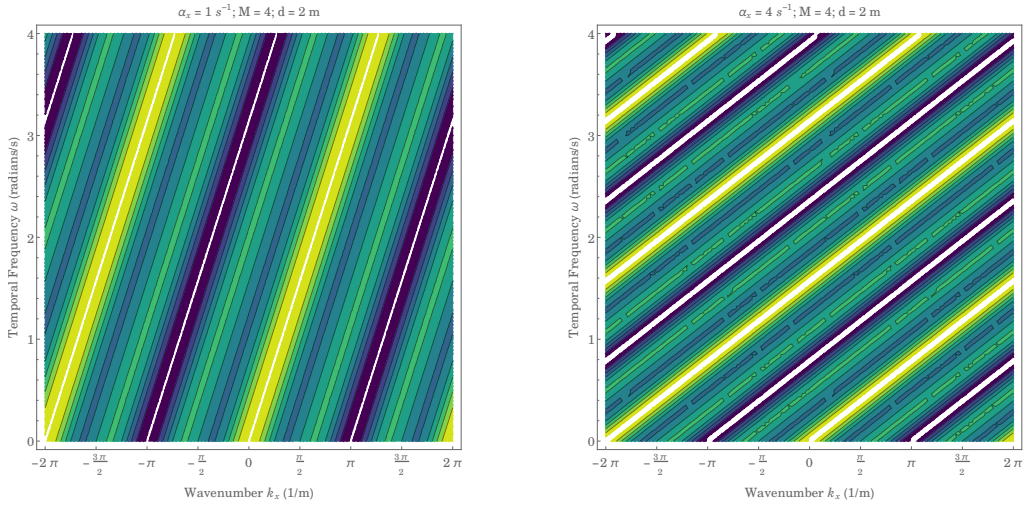


Figure 19: Two contour plots the first of which has been steered in direction $\alpha_x = 1$ and the second in $\alpha_x = 4$.

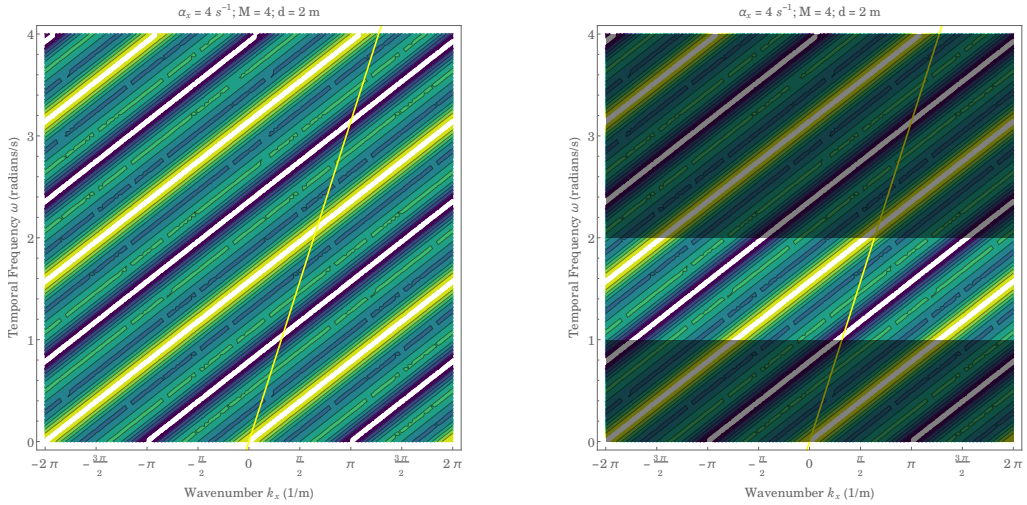


Figure 20: On the left, the main lobe for slowness $\alpha_x = 1$ has been superimposed over the wavenumber-frequency response of the array steered to look in direction α_x =4. The figure on the right shows the effect of temporal filtering. On the response of the array.

5 The array

The hydrophones used in this study were manufactured by High Tech, inc. (Long Beach, MS), and the array was constructed by William Holland and Chris Bonnerup at East Carolina University Department of Physics and consisted of seven omnidirectional hydrophones (model HTI-96-MIN) for signal acquisition. Macartney, inc. built the connectors and cables. Detailed specifications concerning the HTI-96-MIN hydrophone have been organized into Table 1. The manufacturer reports average sensitivity values of -164 dB, with a usual reference factor 1 V/ μ Pa. Hydrophone calibration occurred on site at High Tech, inc. Individual characteristics of each hydrophone channel, including specific signal-to-noise and sensitivity data, are itemized in Table 2. All hydrophones have sensitivities within 0.4 dB of the average value.

Specifications [26] :	
Sensitivity w/ preamp:	-164 dB ref: 1 V/ μ Pa
Frequency response:	2 Hz to 30 kHz
Preamplifier type:	Voltage mode (3 wire: Power, Ground, Signal Out)
Operating Voltage:	5 VDC to 15 VDC
Connector:	XLR
Maximum Operating Depth:	3,048 meters
Size:	2.50" length x 0.75" dia.

Table 1: Specifications of the HTI-96-Min hydrophone manufactured by High Tech, inc.

Seven Channel Hydrophone Array:

Array Channel	Hydrophone S/N	Sensitivity (dB re 1 V/ μ Pa)
1	408001	-164.0
2	408009	-163.9
3	408006	-164.1
4	408007	-163.9
5	408002	-164.4
6	408004	-164.1
7	408008	-163.7

Table 2: Sensitivites and signal-to-noise characteristics of each hydrophone channel [26].

The frequency response of each hydrophone is 2 Hz to 30 kHz; the segment of the audio spectrum of interest for those surveying soniferous fish is typically 50 Hz to 5 kHz. With an operating voltage range of 5 V to 15 V, the power was supplied by a convenient 12-V marine battery. An additional cable was custom ordered to a length of $\approx$ 200 m to ensure an ample area of surveillance to achieve a broad aperture window. Each hydrophone is rated for a maximum depth of 3 048 m, well in excess to the conditions an estaurine environment of very-shallow water.

Audio was recorded with a ZOOM F8 field recorder, weatherproofed in a customized Pelican$^{\text{TM}}$ case housing. An adapter prototyped by the East Carolina University Department of Physics's electronics shop was fixed into the side of the housing and converted output from the array to be compatible with the 8-channel input of the ZOOM field recorder. Audio data was recorded at a resolution of 24-bits with a sampling frequency of 44.1 kHz in 3600 s (1 h) long segments written to a removable SD card.

Figure 21: Slant forward image of the ZOOM F8 field recorder [27].

5.1 The array geometry and setting for deployment

The ZOOM F8 field recorder was stowed in the weatherproof Pelican^{TM} case stationed on land away from the banks of the marsh creek partially concealed from open view. The seven hydrophones were connected to seven of the eight tracks on the field recorder and powered by a 12 V marine battery. With the aid of a small boat, the loops of cable were unwound completely to their full length, and each of the HTI-96-MIN hydrophones were zip-tied to a concrete cinder block and positioned at random locations in an extensive line array that stretched from the bank into the middle of the channel. Lengthy pieces of PVC pipe, which had been ruled, and wide flat surfaces of material, which were fastened atop, were used to construct a tool that served to measure the water depth and as targets for the laser range finding device. The entire system was deployed during low tide and stationed for 24 h at the Clam Bank Monitoring site (latitude 33.335 and longitude -79.194) before recovery. To establish a coordinate system, extensive field measurements were obtained by a portable GPS de-

vice and a laser-range finding device, TruPulse 360 R, Laser technology, inc.,accurate to ± 0.1 m in tandem. The measured positions of important markers and landmarks are displayed in the data plot in Figure 3. The site contained a GPS reference station which was used as first reference and origin of the coordinate system, labeled R1, and GPS and distance relative to this origin were measured for other landmarks in the neighboring area (R2, R3, R4). A tall target sign was planted at the site where each hydrophone was deployed and both GPS position and distance relative to each of the reference land marks recorded. Initial depth was recorded for each hydrophone (the depth value changed based on tide conditions), and declination was measured from the creek to the height of the target, so geometric calculation could obtain two dimensional projections of the measured distances. Azimuth readings were corrected for the magnetic declination by subtracting the declination value from the rangefinder reading. Relevant data has been collected and presented in Tables 3 and 4.

Reference landmark coordinates:

Reference	Lat. [°]	Long. [°]	x [m]	y [m]
R1	33.33399257	-79.19289139	0.0	0.0
R2	33.33427748	-79.19282694	6.0	31.6
R3	33.33459035	-79.19201705	81.4	66.3
R4	33.33419092	-79.19191931	90.5	22.0

Table 3: Coordinates used for reference to measure distances to hydrophone position.

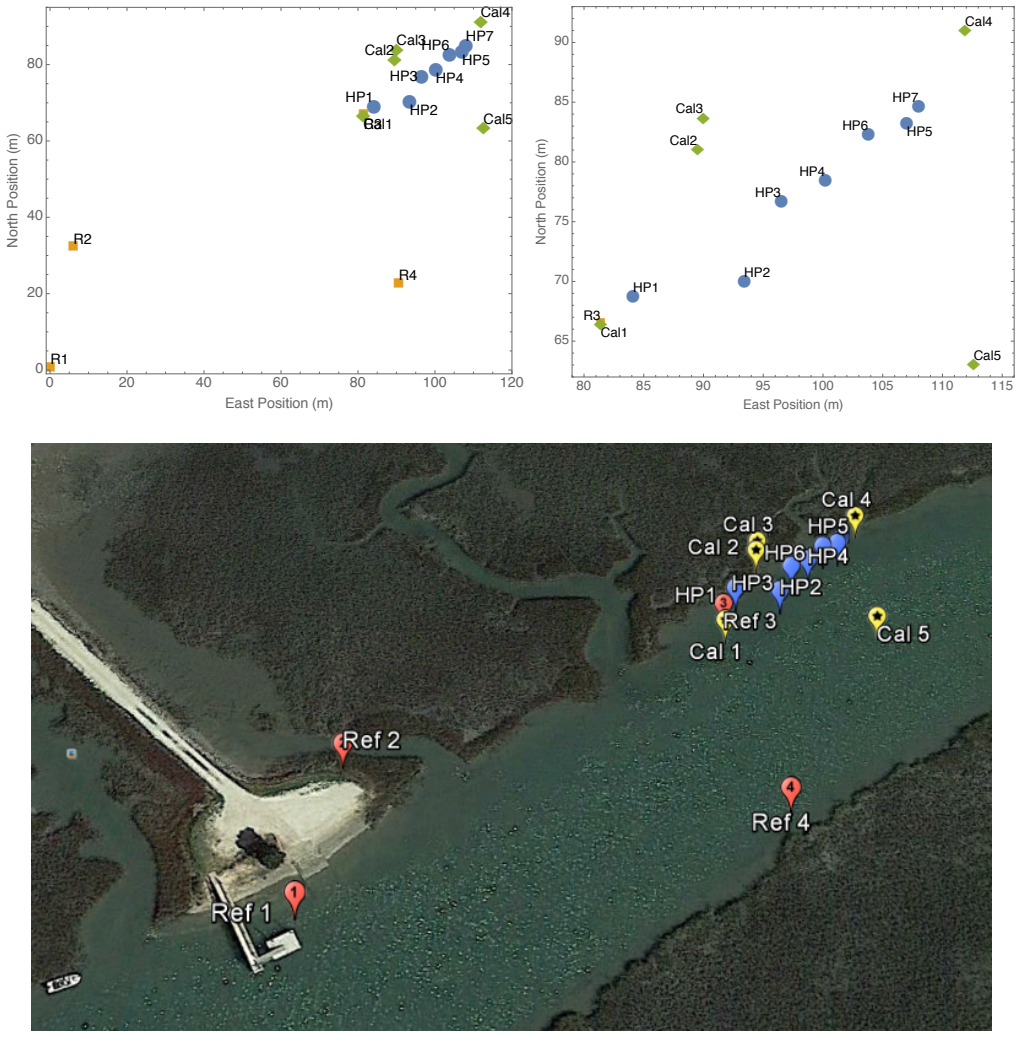


Figure 22: Array plots of the hydrophone geometry at Clam Bank in the North Inlet-Winyah Bay National Estaurine Research Reserve near Georgetown, SC. The reference points are displayed as orange squares. The blue disks show the location of each hydrophone in the array. The green diamonds are the location of the calibration clanks. The reference points were very far away (*left*) and a plot excluding these points has been supplied (*right*). GPS Locations were recorded and submitted into Google Earth (*bottom*) *The plots and map are credited to M.W. Sprague.*

Hydrophone Points:

	Lat. [°]	Long. [°]	x [m]	y [m]	z [m]
HP1	33.3346119855189	-79.1919880492172	84.1	68.7	-0.83
HP2	33.3346228042941	-79.1918881552849	93.4	69.9	-1.35
HP3	33.3346832131291	-79.1918548566292	96.5	76.6	-0.89
HP4	33.3346994421037	-79.1918151136994	100.2	78.4	-1.32
HP5	33.3347418179877	-79.1917420724359	107.0	83.1	-1.37
HP6	33.3347337036447	-79.1917764446533	103.8	82.2	-1.07
HP7	33.3347553423149	-79.1917313309710	108.0	84.6	-0.86

Table 4: Coordinates for hydrophone positions based on field measurements.

5.2 Calibrating the array

Field measurements were admittedly difficult. The conditions under which they were made were arduous; while the array was deployed by wading into the water during low-tide, the measurements were made during a low-high tide transition, and areas in the middle of the channel were increasing in depth – much of the research team had difficulty maneuvering in the water as the water deepened. This limited the time of deployment and hastened the efforts of researchers. Deployment form was skillful and the research team performed with purpose so that each move was carried out deliberately as planned in the briefing. In addition, the hydrophones were weighted to the bottom of the channel, and were not visible to measurement. Targets were lazed with a range-finding device in a competent manner, so a certain degree of accuracy is safely assumed, and geometric calculations were double checked so that there was good faith in the azimuthal components of the measurement. But, GPS measurements with hand-held field devices have reasonable large tolerances for position and the reported precision of the range-finding device was questionable,

not spurious, but questionable. The rangefinder reported a resolution of ± 0.1 m; however, for measurement greater than 100 meters it often reported integers – not always, strangely enough, but most of the time. In the event that there is a low quality target, the precision defaults to ± 1 m. With all of these factors combined, and, with full premonitory awareness of these challenges, ahead of time it was decided a method would be used to calculate sensor positions based on sound speed in the medium; this would yield the *acoustic positions* of the sensors, and all localization efforts would be carried out in this *configuration space*.

To this end, five calibration sounds were produced within the array, just beneath the surface of the water. This was completed in a small boat, at high-tide, and all measurements were made without any altitude angle to reference points in the landscape, using the laser range finder. One member of a two man team leaned over the side of the small boat, placing their hands beneath the surface of the water and striking two pipes together several times. The position and time for each calibration tone was recorded so that the array could be tested, and tweaked – that, these measurements could be used to calculate the true positions of every sensor, so that whichever algorithm was settled on would give accurate results calibrated to these measurements.

The distance between an acoustic source and a hydrophone is given by the distance formula:

$$r_{i,j} = \sqrt{(s_{i,x} - x_{j,x})^2 + (s_{i,y} - x_{j,y})^2}, \tag{5.2.1}$$

where i is an index denoting the calibration, $i = \{1, 2, 3, 4, 5\}$, and j is an index denoting the sensor from which the distance is measured. The position for $\{x_{1,x}, x_{1,y}\}$

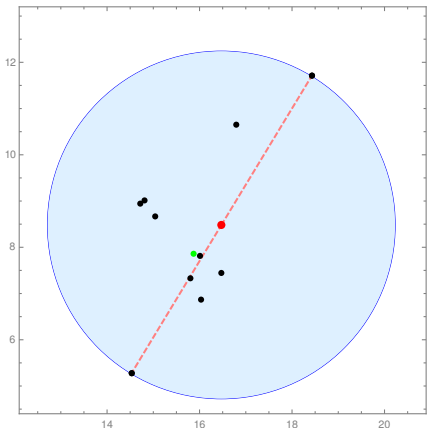


Figure 23: The red point is the center of a circle of minimal area that encloses all the intersection points for the system, and the green point is the spatial median.

is taken as $\{0, 0\}$ for the calibration, and the results are translated back to the coordinates of the measured space.

$$\frac{f_s}{c}\ (r_{i,j'} - r_{i,j}) = \tau_{j'-j} \tag{5.2.2}$$

These delays, $\tau_{j'-j}$, are calculated using the generalized cross-correlation method, and are an integer number of samples. The resulting system of equations was then solved numerically for the hydrophone positions, $\{x_{j,x}, x_{j,y}\}$. Total consistency is not guaranteed, and measurement error and this approach using a whole number value of samples often gives a collection of intersections which are solutions for $\{x_{j,x}, x_{j,y}\}$. A multivariate spatial median method was used to determine a location that minimized the distance to each intersection. For a set of points (x_i, y_i), where $i = 1, 2, ..., n$, the spatial median be a choice $\hat{\mu}$ of $\mu = (\mu_x, \mu_y)$ that minimizes the function:

$$T(\mu) = \sum_{i=1}^{n} \left((x_i - \mu_x)^2 + (y_i - \mu_y)^2\right)^{1/2} \tag{5.2.3}$$

The specific choice of $\mu = (\mu_x, \mu_y)$ that satisfies

$$\frac{\partial T}{\partial \mu_x} = \frac{\partial T}{\partial \mu_y} = 0 \tag{5.2.4}$$

is $\hat{\mu}$, the spatial median, see Figure 23. It is useful in circumstances when an average is taken so to be unaffected by magnitudes instead of residuals.

6 Steered power response method

If a source's location is unknown, an array set up as a beamformer may be steered over a predefined region of space by adjusting delays, the output of a beamformer when utilizing this method is known as the *steered response* [28]. The power of the steered response is maximized when the delays implemented on the beamformer match the propagation delays for a sound wave traveling to each sensor in the array. Discovering which delays satisfy this condition make it possible to map the result from delay space to location space, and ultimately provide the original location for the sound source.

Consider a system of M hydrophones, following the convention that each sensor is denoted by an index $m = \{0, 1, ..., M\}$. As previously discussed the discrete time signal recorded by each hydrophone is $s_m(n)$. Steered power method seeks to maximize the power of a steered beam across the location space. In Cartesian coordinates, the delay between a pair of hydrophones is given by:

$$f_{i,j}(\vec{x}) = \frac{1}{c}\left[d_{0,i}(\vec{x}) - d_{0,j}(\vec{x})\right], \tag{6.0.1}$$

where $f_{i,j}$ represents the delay between microphones m_i and m_j and $d_{0,i}$ is the distance between the sound source and microphone m_i.
The search grid will be defined by [29]:

$$\mathcal{G} = G_x \times G_y \times G_z = \{x_0, x_1, ..., x_{N_x}\} \times \{y_0, y_1, ..., y_{N_y}\} \times \{z_0, z_1, ..., z_{N_z}\}, \tag{6.0.2}$$

where $G_x = \{x_0, x_1, ..., x_{n_x}, x_{N_x}\}$ is a set of x-dimension candidate locations with cardinality $|G_x| = N_x$, $G_y = \{y_0, y_1, ..., y_{n_y}, y_{N_y}\}$ is a set of y-dimension candidate locations with cardinality $|G_y| = N_y$, and $G_z = \{z_0, z_1, ..., z_{n_z}, z_{N_z}\}$ is a set of z-dimension candidate locations with cardinality $|G_z| = N_z$. The number of candidate positions is given by the cardinality of $\mathcal{G}$, is given by $|\mathcal{G}| = N_x N_y N_z$. The indices n_x, n_y, n_z, refer to the position of each element inside of the corresponding set of the given subscript. A specific coordinate that is an element of $\mathcal{G}$ will be denoted g_{n_x,n_y,n_z}.

The SRP is carried out not in location space, but in the delay space, defined at each search location by $f_{i,j}$, the delays calculated, in general, for all unique, order-independent microphone pairs, given by set P. Whose cardinality is given by the binomial,

$$|P| = \binom{M}{2} = \frac{M(M-1)}{2}, \tag{6.0.3}$$

For larger arrays, the cardinality of this set can be quite large, considering that an SRP algorithm takes place for delays for each microphone pair for each candidate position in $\mathcal{G}$. This places some constraints on the other parameters of the computation. A fine grid resolution is important for more precise localization; moreover, the SRP is carried out using audio excerpts from each channel, and it is desirable to use longer audio tracks for more accurate results, which is computationally costly even more so. It can be possible to locate an acoustic source using less pairwise delays, sacrificing a cleaner output. It is not that there is redundancy found in using greater amounts of pairs, but that eventually after combing several pairwise calculations a location

begins to emerge and involving higher pairs give the same result. In this application, HP1 is used as a reference microphone for all pairwise calculations, and P will be represented by:

$$P = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \\ 1 & 7 \end{pmatrix} \tag{6.0.4}$$

The set that denotes the delay space of pairwise delays for microphones m_i and m_j is denoted $D_{i,j}$. There is a set like this for each microphone pairing, and it takes the form:

$$D_{i,j} = \{-\tau_{\max_{i,j}}, -\tau_{\max_{i,j}} + 1, ..., -1, 0, 1, ..., \tau_{\max_{i,j}} - 1, \tau_{\max_{i,j}})\} \tag{6.0.5}$$

The set $D_{i,j}$ in (5) represents the best attainable resolution in delay space. The elements inside of this set are pairwise delays with the units of samples – the audio being analyzed are discrete-time signals. A search grid $\mathcal{G}$ can only be defined with a resolution up to a corresponding resolution of one sample in delay space. This is because the time information embedded in each signal is only accurate to intervals of the inverse of the sampling frequency, linear connections are filled in by the audio recording device whilst writing the digital recording file. In general, the resolution in

delay space will be determined by the spatial grid resolution, which is chosen at the beginning of the algorithm. Finally the SRP seeks to maximize the power output of a conventional beamformer. For a typical array the output signal is given by,

$$z[n] = \sum_{m=1}^{M} w_m y_m[n - \tau_m], \tag{6.0.6}$$

where, $z[n]$, is the array's output, w_m are amplitude weights applied to the output of each sensor to taper the beampattern and reduce sidelobe levels, y_m is the signal recorded at each hydrophone m, and Δ_m is some delayed applied to each signal to steer the focus of the array on signals propagating from a particular direction $\vec{\zeta^0}$ or from a particular point in space, $\vec{x^0}$. The steered power response at an arbitrary position $\vec{x} = \{x, y, z\}$ is expressed as [8]

$$P(\vec{x}) = \sum_{n} \left| \sum_{m=1}^{M} y_m[n - \tau_m[\vec{x}]] \right|^2, \tag{6.0.7}$$

Implementing this within the realm of the proposed search algorithm using grid points in grid $\mathcal{G}$,

$$\mathcal{P}(\vec{x}) \equiv \sum_{n \in \mathcal{Z}} \left| \sum_{m=1}^{M} y_m[n - \tau_m[\vec{g}_{n_x,n_y,n_z}]] \right|^2 \tag{6.0.8}$$

where, this represents the power of all channels summed after being delay by an amount $\tau_m(\vec{x})$, given by:

$$\tau_m(\vec{x}) = \lfloor f_s \frac{|\vec{x} - \vec{x}_1| - |\vec{x} - \vec{x}_m|}{c} \rceil \tag{6.0.9}$$

The sampling frequency f_s is included, with the propagation speed c because time

in this case has units of samples. The algorithm is limited to time units of sampling period, and since they are whole numbers, the round operation, $\lfloor \cdot \rceil$ is used. Incidentally, this equation can be related to the cross-correlations between the signal at each hydrophone [30].

$$\mathcal{P}(\vec{g}_{n_x,n_y,n_z}) = \frac{1}{2\pi} \sum_{m_i=1}^{M} \sum_{m_j=1}^{M} \int_{-\pi}^{\pi} \Phi_{m_i,m_j}(e^{i\omega}) Y_{m_i}(e^{i\omega}) Y^*_{m_j}(e^{i\omega}) e^{i\omega\tau_{m_i,m_j}(\vec{g}_{n_x,n_y,n_z})} \mathrm{d}\omega \tag{6.0.10}$$

where,

$$R_{m_i,m_j}(\tau_{m_i,m_j}) \equiv \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_{m_i,m_j}(e^{i\omega}) Y_{m_i}(e^{i\omega}) Y_{m_j}(e^{i\omega}) e^{i\omega\tau_{m_i,m_j}} \mathrm{d}\omega \tag{6.0.11}$$

so,

$$\mathcal{P}(\vec{g}_{n_x,n_y,n_z}) = \sum_{m_i=1}^{M} \sum_{m_j=1}^{M} R_{m_i,m_j}(\tau_{m_i,m_j}(\vec{g}_{n_x,n_y,n_z})) \tag{6.0.12}$$

To estimate source location, the maximum of this set is taken:

$$\vec{x^0} \approx \vec{g}_{n_x,n_y,n_z} = \underset{x^0 \in\ \mathcal{G}}{\arg\max}\ \mathcal{P}(\vec{g}_{n_x,n_y,n_z}) \tag{6.0.13}$$

where the function argmax identifies the maximum of set $\mathcal{P}$. The position of the maximum power in set $\mathcal{P}$ yields the grid point $\vec{g}_{n_x,n_y,n_z} \in \mathcal{G}$ that corresponds to the most likely location of the source $\vec{x}^0$. Note, this is only an approximate source location; presumably the grid point that produces the maximum power is the closest

location to the actual spot the source made the sound.

6.1 The steered-beamformer algorithm

1. Import the audio excerpt

The audio was recorded in 3600 s (1 h) segments. These are trimmed using a Python program prior to analysis in Mathematica. There is much when it comes to audio processing that Python surpasses Mathematica in ease and computation timing. The length of time isn't specific, but it has to include the sound one wishes to locate.

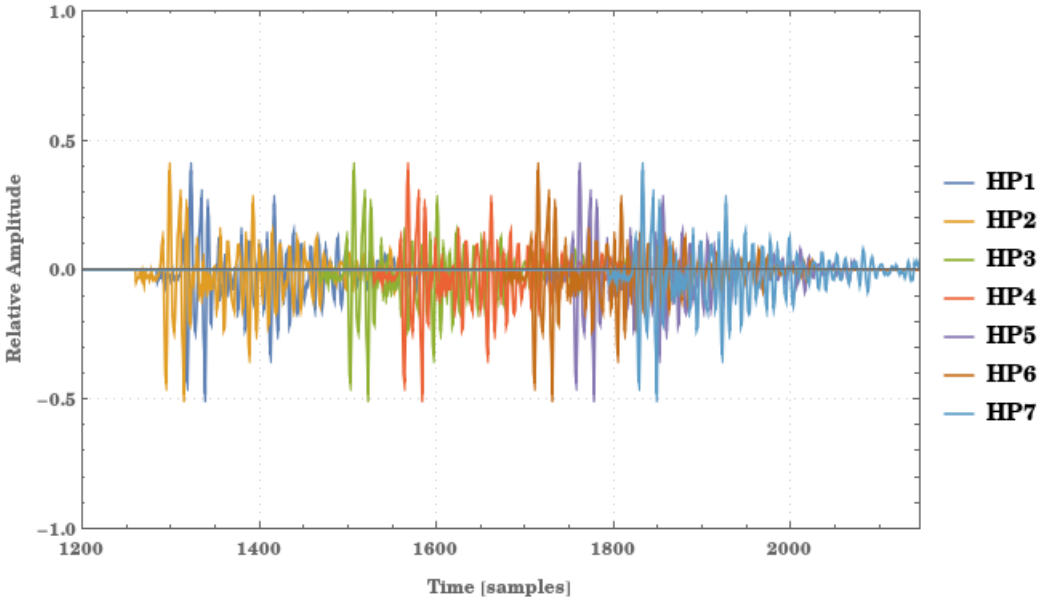


Figure 24: Typical waveform plot of a source signal recorded on the seven hydrophone channels.

2. Trim the audio even further

Only the arrival of the first maximum is of interest, the final clip should contain

no echo and should be truncated before another sound is issued. The clip should be of a reasonable length, so that it is more than just a few amplitudes. This is more of an art than a science and is often very difficult to achieve a perfect length in this noisy environment. Ideally, the target is to have as long an excerpt as the program allows; computation timing increases dramatically for longer tracks, and if the datasets are too long it becomes impractical – memory is filled up quickly, and processes can run indefinitely until a crash occurs. Isolating a specific sound is very difficult in such circumstances. The sounds of interest have amplitudes much less than the percussive shrimp clicks, and the clip will be far less than can be perceived at a normal playback rate.

3. Filter the audio

In an environment this noisy, filtering becomes extremely important. The shrimps produce percussive sounds that dominate the full range of the spectrum. There are countless of these creatures in the waters of the marsh, and several are clicking at a given instant. Those near the microphone are recorded very loudly. Many times, though, it is known at which frequency a fish emits its call and power spectra can reveal the peak frequency of the call the fish makes.

The filter used is very rudimentary. It is a brick wall filter [31] with a very small bandwidth that can be adjusted in width and central frequency, see Figure 26. If there exists in the call higher harmonics, the filter can be altered to include this part of the spectrum as well. In general, at higher harmonics, the sound from other sources and noise dominates, so this was considered a reasonable approach.

4. Generate a grid of desired resolution over the physical domain

This is simply a set of ordered pairs that serve as candidate source locations [29].

The level of resolution specified is a nuanced decision as well. A finer grid means longer computation times. To overcome this restriction, the process can be cascaded so that the localization occurs in stages with increasing grid resolution [32]. The resolution for the first stage must be chosen carefully, because too broad of a spacing will distort the location the algorithm reports. For longer audio clips, a broader resolution must be used. Stage 1 was always implemented with a 1 m grid spacing and resolution was increased from there. The grid is generated using a simple iterating method within Mathematica.

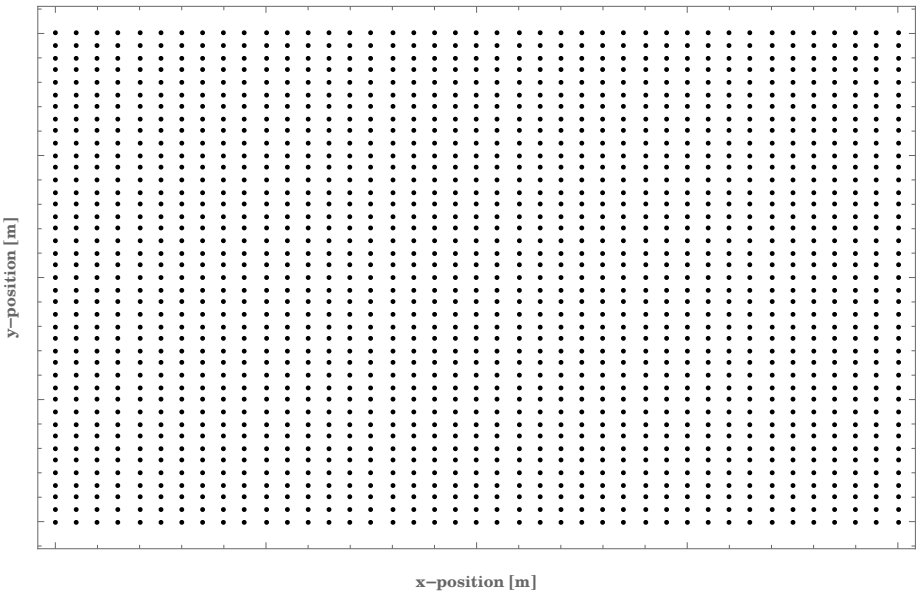


Figure 25: A typical grid spacing for stage 1 of the localization algorithm.The filtered spectrum is shown in red.

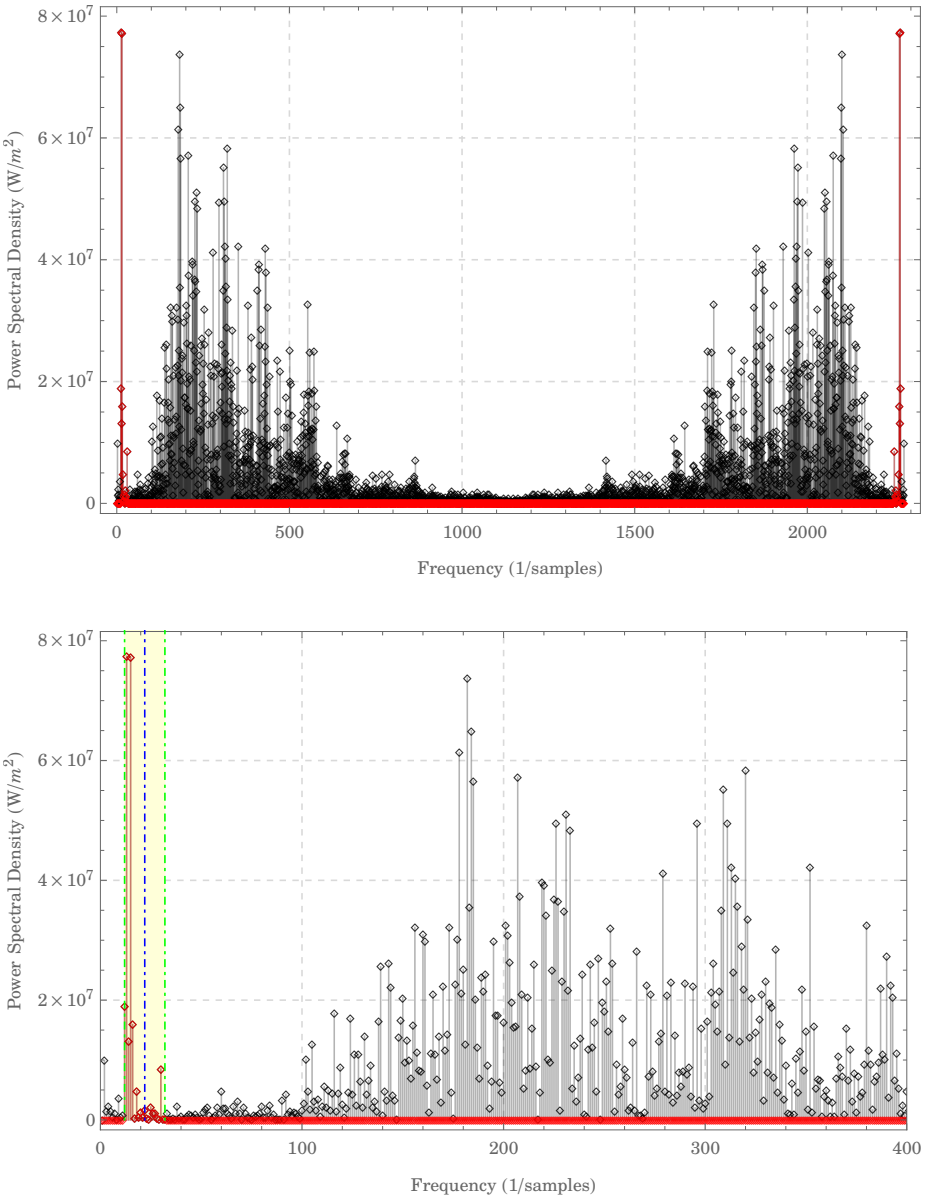


Figure 26: The brick-wall filter's response.The yellow region is the passband of the filter. The filtered spectrum is shown in red.

5. Calculate distance from each gridpoint and time delays

To implement the pair-wise time delay method outlined in the background section requires the difference between the distance to the source and the distance to each microphone. The sound speed of water is calculated from the water quality data (salinity, depth, temp.) for the time of day it was when the recording was made. The sample rate is used to get a time-delay in samples. This calculation must be carried out for each gridpoint and each hydrophone pair being considered. Here, hydrophone one (H1) is treated as a reference hydrophone and pairs $\{\{1,2\},\{1,3\},\{1,4\},\{1,5\},\{1,6\}\}$ are considered only for simplicity and ease of computation (it can be possible to locate with this few number of pairs). DO loops are ideal for this purpose.

6. Implement the delays

Many options exist for the signal processor to impose the delays appropriately on the audio clips, but only one was considered a feasible choice. Ultimately, a sum between all channels is the destination of the delayed audio clips; therefore, it is critical the amplitude datasets are of equal length. Through zero-padding, or the addition an amount of zeros to the front and back of a dataset, the delay can be achieved and the common length requirement met. It is important to pad each clip with the same amount of zeros also to not artificially lessen the impact of any track to the spectrum of the sum; this is important for the average power calculation. To automate this process is important so that the same script can be carried out again for each point in the grid. Depending upon where a soniferous fish is located within the physical domain, there are three possibilities for how to pad the channels, and an IF

statement is used to distinguish between each case. What results from this process is 7 lists of length congruent with the number of grid points, so choosing an appropriate grid resolution and clip length for stage 1 is very important. The elements of this list represent how each clip would be recorded at each grid point in location space, and an investigator can "listen" at each grid point if so desired.

7. Sum channels to find maximum average power

The process revealing the spatial maximum likelihood is one of the more straightforward sections of the algorithm. At each grid point, the channels are summed, and the average power is taken for each grid point in location space. The maximum of this list gives the point in location spatial grid where the power is highest, corresponding to a zero-delay between all channels. The grid point that satisfies this condition is retrieved and a density plot can show the effective average power for each grid-point along the grid.

8. Implement a second stage

The resulting location from stage 1 can be used to generate a grid of smaller domain and range around the grid-point with a finer resolution [33, 34]. In this way, many successive approximations of location can be achieved to the desired resolution. The limitation here, would be the resolution with which the landscape was surveyed and measured, and reporting a location estimate of finer precision than is possible with the surveying instruments.

The figures displaying sample outputs of the 2-stage localization algorithm resulted from tests with an unfiltered exponentially decaying sinusoid – without any Gaussian noise. The delays were implemented to suggest the source emanated from

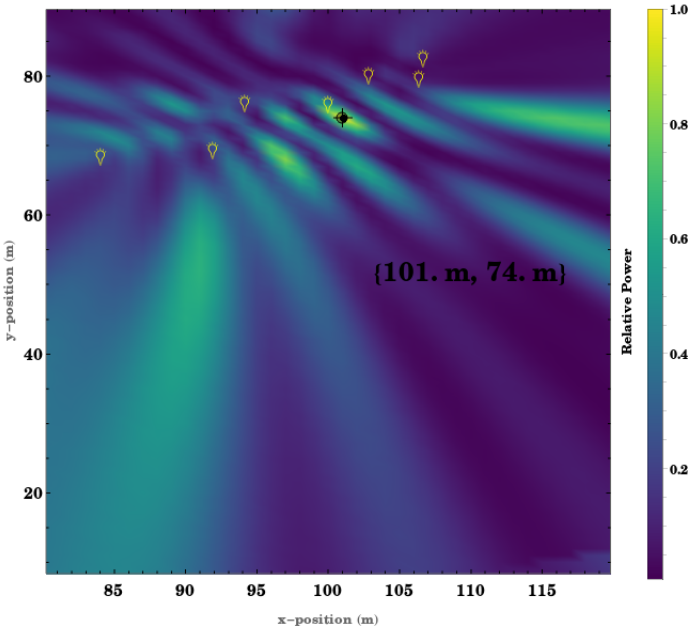


Figure 27: A sample output of the steered-beamformer algorithm after stage 1.

a point (101. m, 74. m) in location space, Figure 27. Under these perfect conditions, the density plot is ideal in the second stage. While there is a very nice gradation in color, the area of highest intensity is about 2 meters in breadth and 2 in height, and a tighter focus can be achieved through bandwidth filtering.

6.2 Testing the algorithm

To make sure the algorithm itself operated as desired source signals were simulated, delayed by amounts appropriate to seem as if they occurred from a pre-known location, and then buried in different degrees of Gaussian noise. The test process was carried out with exponentially decaying sinousoids and example real-world audio data included with the Wolfram library.

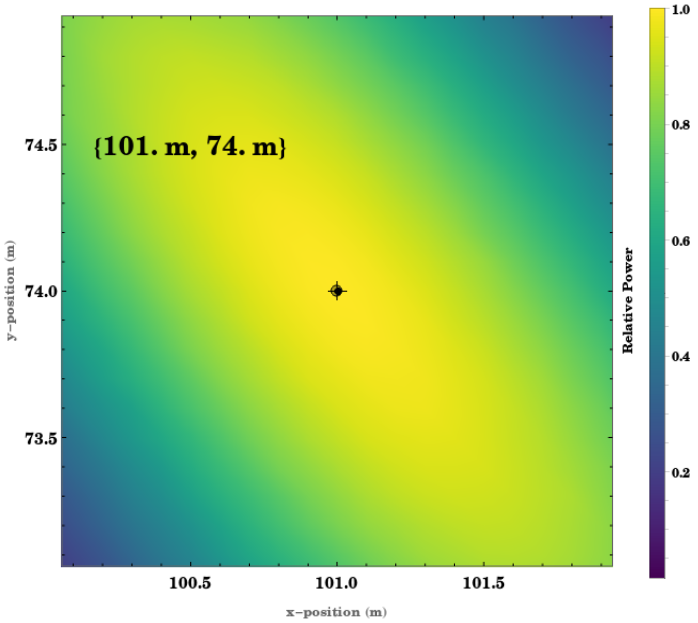


Figure 28: A sample output of the steered-beamformer algorithm after stage 2.

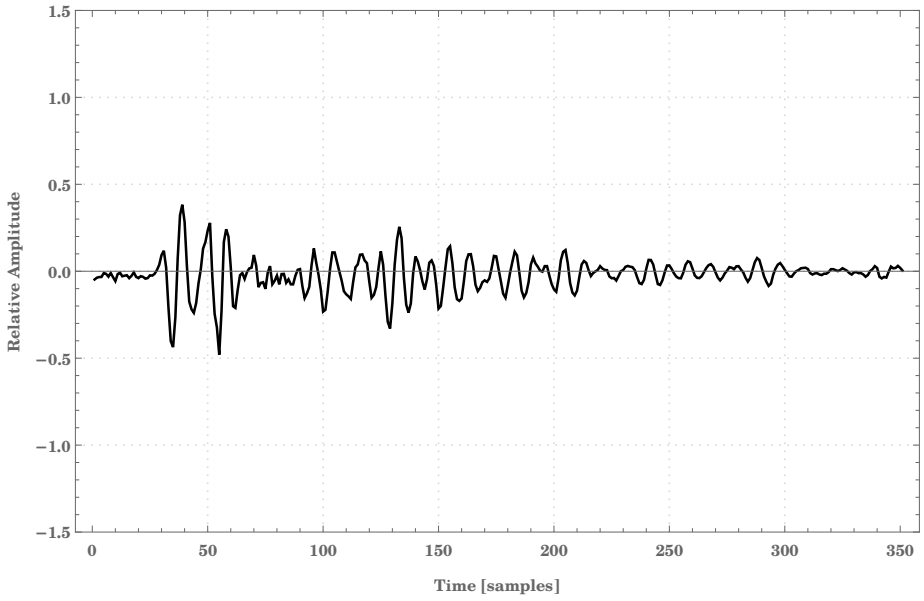
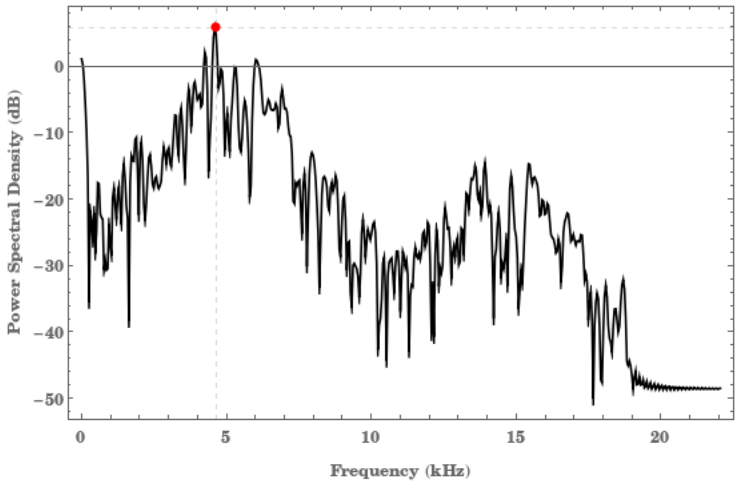


Figure 29: The sample sound used to text the extent of the beamformer algorithm.

The delays for a given source location can be calculated regardless of the audio clips; they depend only on the geometry of the array, the source position, and speed of sound in the medium. The real-world audio clip used most extensively for testing was a tone which indicates a service truck is driving in reverse accompanied by engine and driving sounds.



	Frequency (kHz)	PSD (dB)
1	4.63	6.

Figure 30: Welch plot of the sample sound.

The frequency content of the clip is shown in a Welch plot, taken using a Hamming window of 4096 bit with standard half-window width overlap. The peak frequency is around 4,630 Hz, and the clip will be filtered with a bandwidth that includes all spectra above 0 dB. With the added Gaussian noise, the spectrum itself is unrecognizable.

The testing process was actually fairly extensive. Several positions were chosen to investigate how the delays were related to positions in location space and to ensure the algorithm was implementing the delays correcting in lag-space. Furthermore, a single position was used while decreasing the signal-to-noise ratio of the simulated

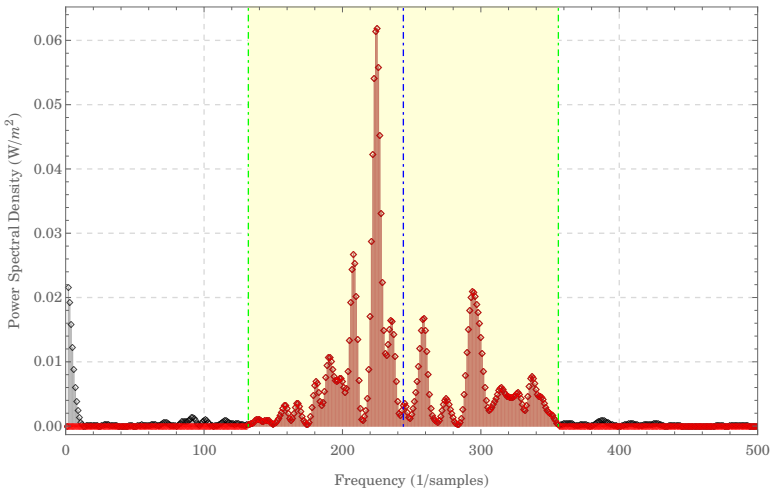


Figure 31: EXAMPLE: The filtered test sound without noise.

source signal to get an impression of how noise affected the algorithm.

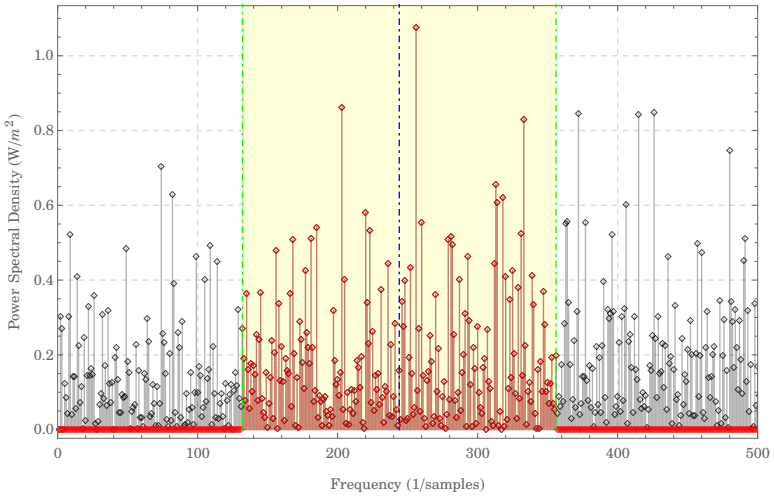


Figure 32: EXAMPLE: The filtered test sound with Gaussian noise.

The aforementioned real-world audio sample was delayed to be perceived as sourced at position (90. m, 60. m) within the grid, woven with varying amounts of noise, and filtered so that the algorithm only maximized power using frequency components

within the appropriate bandwidth.

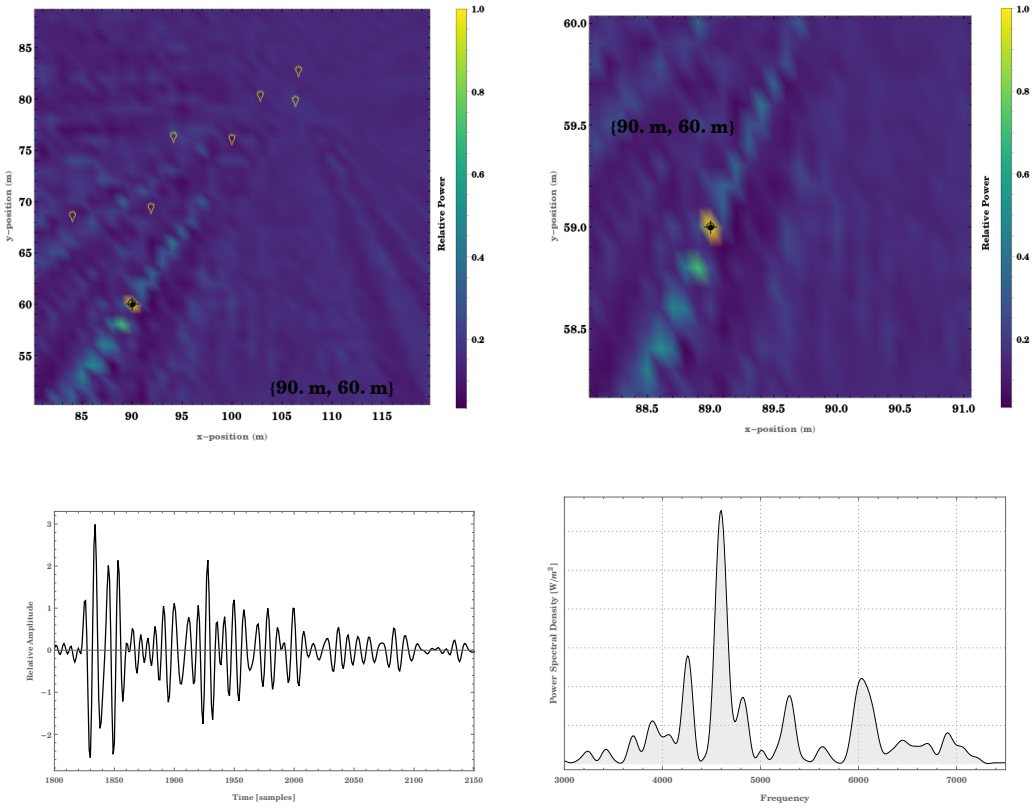


Figure 33: EXAMPLE: The filtered test sound without noise.

To be sure the sound has been recovered, a revealing source of information is the waveform and spectrum plots of the combined signal, which are both displayed in the figure. Comparing both the waveform and spectrum that result to those of the original reveals a couple of things. First, at least with a signal this ideally simple, the spectrum that results is indicative of information in the spectrum of the unadulterated signal; in fact, it is identical. For unknown signals, this fact is not guaranteed, although it has been established previously that frequency content within the resulting signal will be related. Notice in the waveform, which is virtually identical to the test signal in

behavior, an increased amplitude, scaled up by a factor of 7. This is of course expected in an example where seven channels are being summed, and it is conventional to shade the outputs in such an algorithm by $1/M$.

No algorithm is impervious to noise, and it becomes increasingly difficult when multiple sources are present in the physical domain being surveyed. As in our survey, the space was heavily populated with soniferous creatures, to such an extent that at any given instant there were several vocalizations in the field emanating from several locations. The algorithm itself seems very robust to noise; however, there is definitely an extent. To what extent of noise the algorithm itself is sufficiently robust has not been quantifiably determined. Several processes were run with the test signal being buried in varying strengths of Gaussian noise, and localization was consistently accurate up to a SNR of -13 dB. All things considered this is pretty noisy, but the noise itself is spread flatly across the spectrum. In the case of multiple sources; however, if all sources other than the one of interest are considered to be noise, there would be high magnitude intensities at other frequencies as well, possibly in the bandwidth of interest as well, which would have an affect on the algorithm and the steered-response of the array designed to operate as a beamformer. What would determine that case is how many of a certain species of fish are located in the physical domain, as well as other species who make vocalizations in the same range of the spectrum.

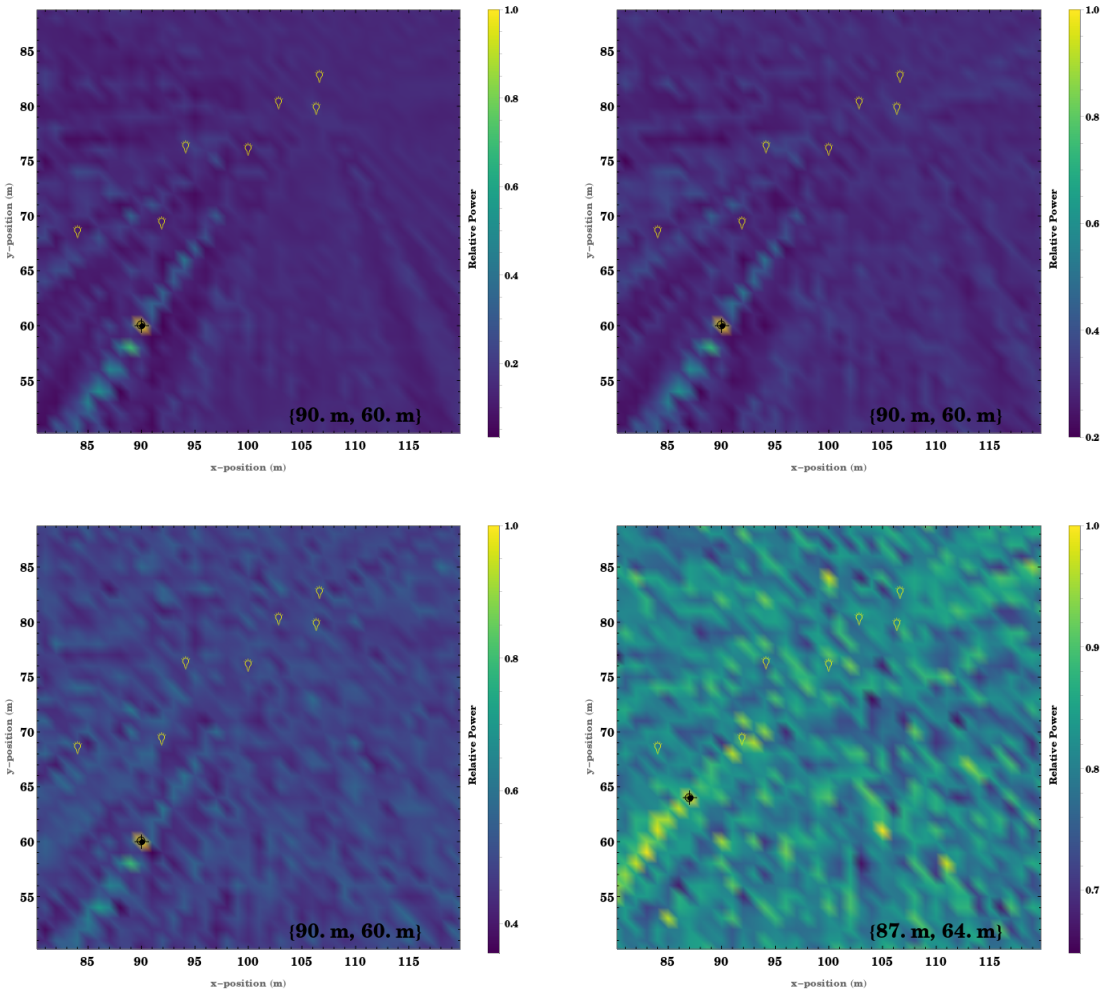


Figure 34: *Top left.* Localization result of pure signal. *Top right.* Localization result with $10 \log \mathrm{SNR} = 0$ dB. *Bottom left* Localization result with $10 \log \mathrm{SNR} = -5$ dB. *Bottom right.* Localization result with $10 \log \mathrm{SNR} = -13$ dB.

7 Results

Array analysis and SRP was done with Wolfram Mathematica after processing in Python was used to edit the audio tracks to a manageable length. Power spectra of the array's output were calculated by Mathematica and Audacity, the latter of which proved very useful. Audacity is an open source audio analysis/editing program that has a plethora of useful features including power spectra and spectograms. Although it offered less control, the default options worked immensely well for producing power spectrum, and Audacity allows the data to be exported and analyzed through other avenues. Spectrograms were taken with Mathematica, Audacity, another open source audio analysis option called Sonic Visualizer `https://www.sonicvisualiser.org/`, and in Python. This program offered a variety of spectrogram options that were extraordinarily convenient to implement; it didn't suffer compared to other options, and its efficacy was largely apparent while searching for vocalization features buried in a noisy background. Due to the large number of sound producers in these waters, this faculty of Sonic Visualizer was absolutely critical.

7.1 Array analysis

Analysis of the array is shown in the figures 35 and 36. The beampattern was plotted for frequencies up to 400 Hz, and the directivity pattern was plotted for frequencies up to 1000 Hz. The main lobe is distinct to about 500 Hz. After that, grating lobes dominate. This means that the array geometry used can beamform reliably up to 500 Hz, see Figure 37. A ratio comparison of the heights of the main lobes to the tallest sidelobe in the beampattern is given in Table 5. This is adequate

Ratio of mainlobe height to height of the tallest sidelobe:	
50 Hz	13.3
100 Hz	8.45
200 Hz	1.83
300 Hz	3.18
400 Hz	1.53
500 Hz	0.61

Table 5: This parameter shows the arrays ability to distinguish signals propagating in at the angle of acceptance and reject waves coming in from other directions.

for a few sound producing fish species such as toadfish and certain species of drums providing that the most of the energy in its vocalization is located below this frequency range.

7.2 Calibration results

The results from the calibration were used to determine new hydrophone positions in efforts to increase accuracy by using the acoustic position of the hydrophones that may have shifted during deployment. To this end, analysis was carried out as described in subsection 5.2. Data from field deployment and initial analysis is organized in Table 6.

Calibration Sound	Time	Position
1	3:07 PM	(81.4 m, 66.3 m)
2	3:12 PM	(89.0 m, 81.0 m)
3	3:33 PM	(90 m, 83.6 m)
4	3:41 PM	(111.9 m, 90.9 m)
5	3:49 PM	(112.6 m, 63.0 m)

Table 6: Field calibration data.

In Figure 38 are the results of the algorithm on the array output for the calibration

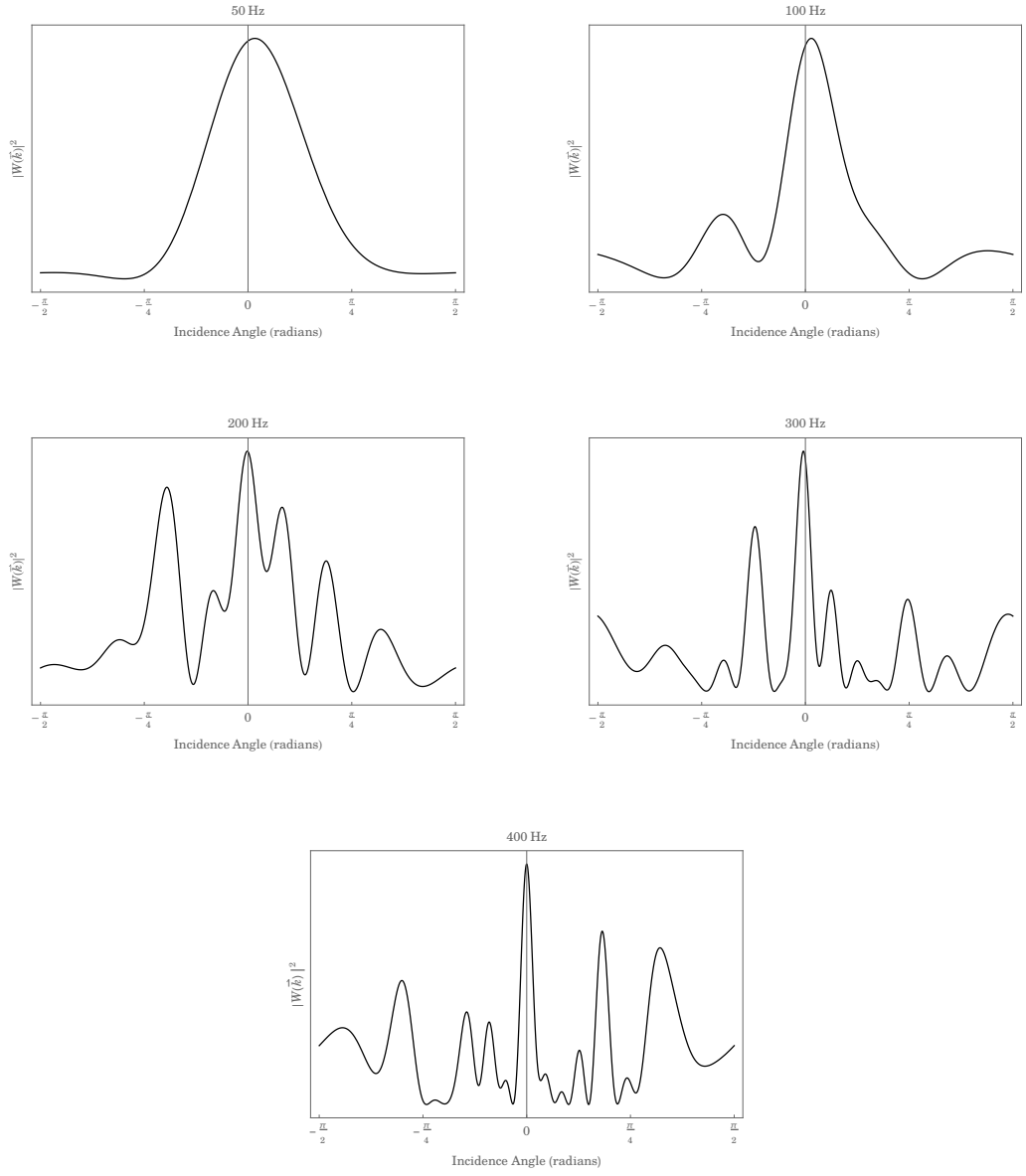


Figure 35: The array pattern $|W(\vec{k})|^2$ is plotted as a function of incidence angle for several given frequencies. The array pattern is aperiodic, and increasing chaotic for higher frequencies.

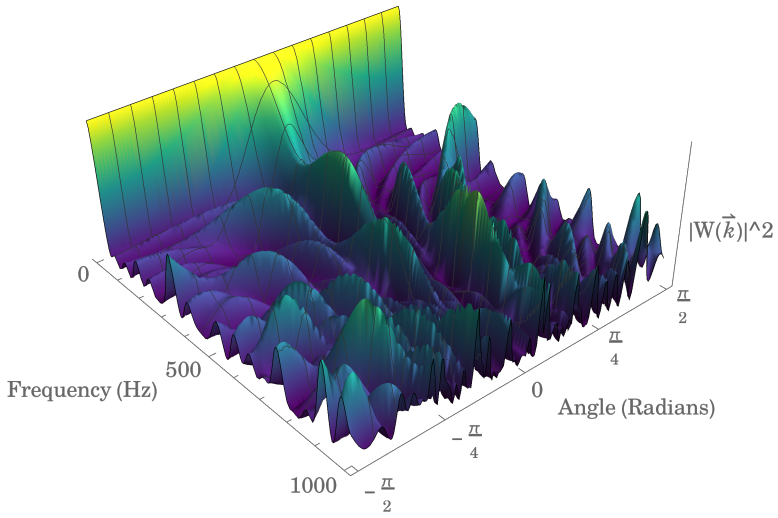


Figure 36: The Clam Bank array's directivity pattern.

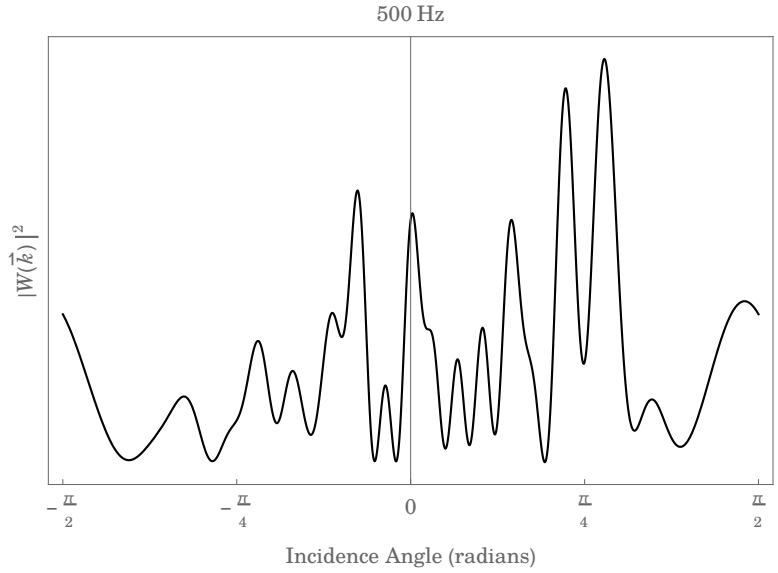


Figure 37: The beampattern for waves with frequency 500 Hz. The sidelobes have greater maximum here and suggest a diminishing ability to beamform accurately above these frequencies.

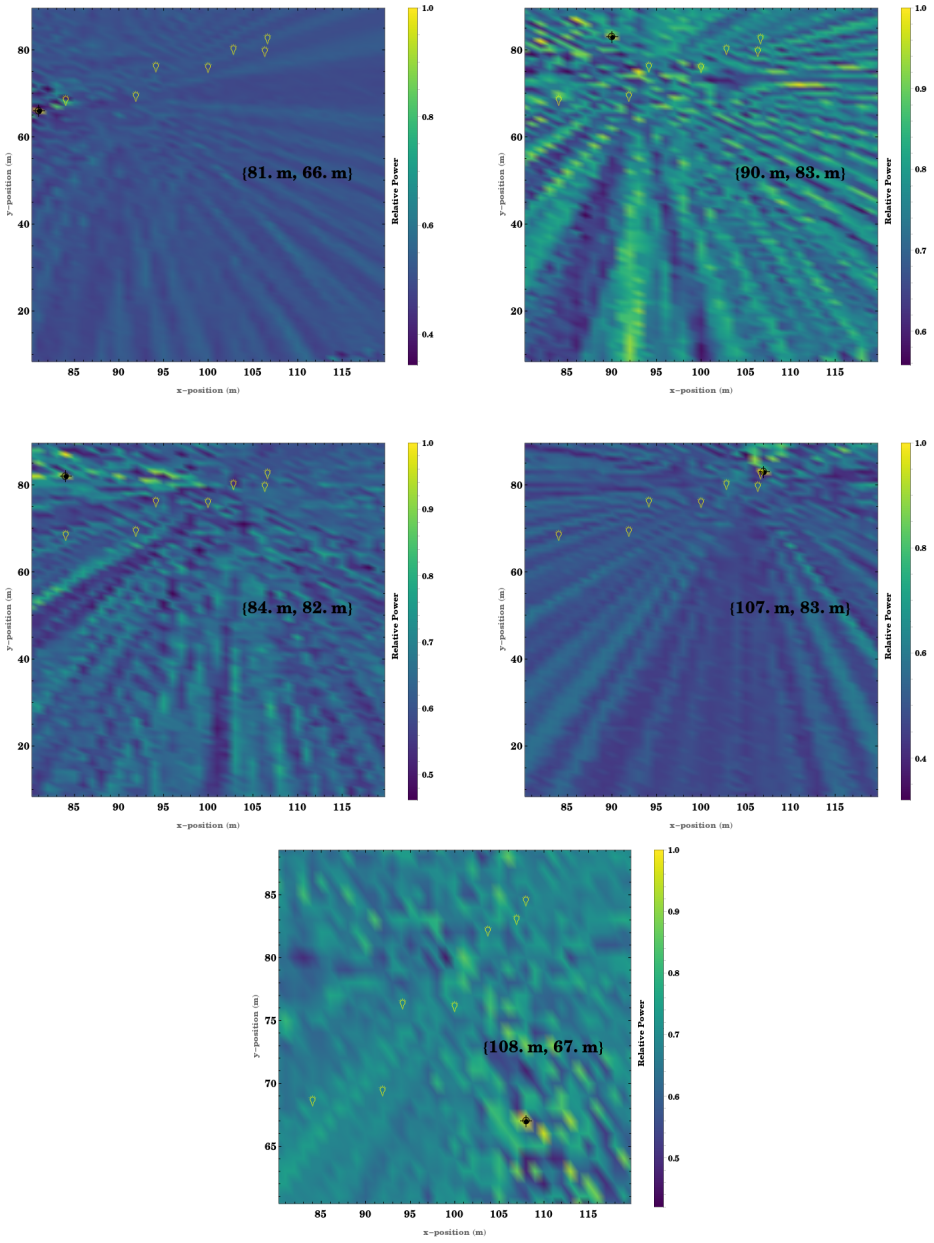


Figure 38: The results of the localization algorithm on the measured calibration tones. Hydrophone positions were adjusted until distance from true position was as minimal as possible using the spatial median.

sounds. For this analysis the cross correlation provided the delay between microphone pair which were used in a geometrical approach to triangulate the hydrophone positions. Overall, the solutions were inconsistent with each other, likely due to position error, and the spatial median was used to approximate a position for where the solutions would overlap. Incidentally, a least-squares method was implemented, among others, but the spatial median produced the hydrophone positions that yielded results closest to known measurements of the calibration sound positions. Limitations and possibles sources of error are discussed in a future section.

7.3 Steered response power results

Analysis was carried out on recordings made at low tide 8:00 P.M. to 9:00 P.M. on 06/01/17 by the hydrophone array at Clam Bank. The SRP algorithm will locate a source in two dimensions, so an environment that provided a low z-component to position fit best with the assumption of a two dimensional space. The recording at 8:00 PM was by far the most eventful compared to other low-tide recordings throughout the twenty four hours of survey. Two oyster toadfish (*Opsanus tau*) were issuing the unmistakable boatwhistle vocalization, along with snapping shrimp, and there is a silver perch (*Bairdiella chrysoura*) producing the high pitched knocking sound used to attract mates.

7.3.1 Oyster Toadfish

Oyster Toadfish are bottom dwelling species that reside amongst rocky substrate or debris (see Figure 39.) which serves as shelter and nests for spawning, and which usually happens during June and July [35]. These nests are discovered and made

ready by the male of the species, and upon completion they will begin to call out with its boatwhistle by contracting rapidly the muscles that surround the swim bladder [36, 35]. Previous research has determined the dominant frequency of this call to be around 240 Hz. This behavior is meant to encourage the female of the species to come lay eggs. The male will guard the eggs until they hatch.

Figure 39: Photo from NOAA, Credit: Andrew David, NOAA/NMFS/SEFSC Panama City; Lance Horn, UNCW/NURC – Phantom II ROV operator

During this recording there are two distinct toadfish boatwhistles occurring periodically. The first toadfish is interesting; its call is unmistakable because an overtone is produced around 500 Hz in addition to the first harmonic at 240 Hz. The second specimen is located more central to the array, and so its boatwhistle arrives more clearly, at higher sound level. The results of the SRP algorithm are plotted in Figure 40, the most likely source location is $\{107\text{ m}, 65\text{ m}\}$ from the first reference landmark that served as the origin for the physical domain. This is a position central to the channel. It is highly likely this fish has made a habitat of the derelict oyster reef that

comprises most of the rocky bottom in the creek.

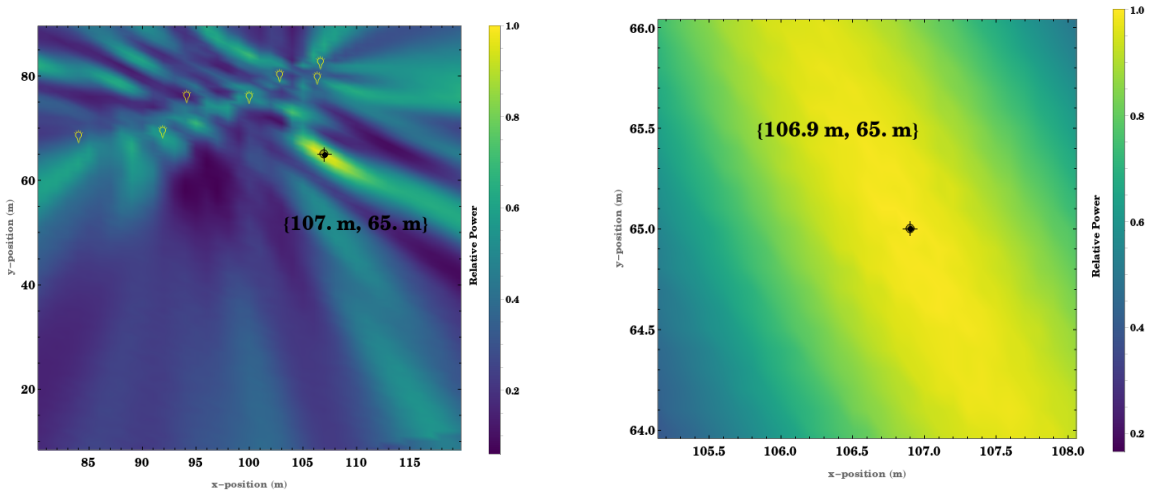


Figure 40: SRP analysis of the first Toadfish.

The waveform is still dominated by spikes due to randomly summed shrimp clicks. Low frequency oscillations are apparent. There are many sound producers, mainly snapping shrimp, and these concerns will be addressed in Section 8.

A power spectrum, displayed in Figure 41, was taken with a 1024-sample wide Hanning window reveals the dominant frequency around 240 Hz and the presence of its overtone at 500 Hz. A logarithmic scale has been used to emphasize the width of the lower frequencies as opposed to the noise and undesired signal energy at higher frequencies, as this broadens the axes.

The sonogram in Figure 42 shows the spectrum of the array's output in time. The two lighter shade creases of -15 dB are the fundamental and harmonic of the first toadfishes boatwhistle.

The second toadfish is located central to the array, (92 m,74 m), see Figure 43. It is located in a region of some ambiguity; in addition, several maxima occur in the

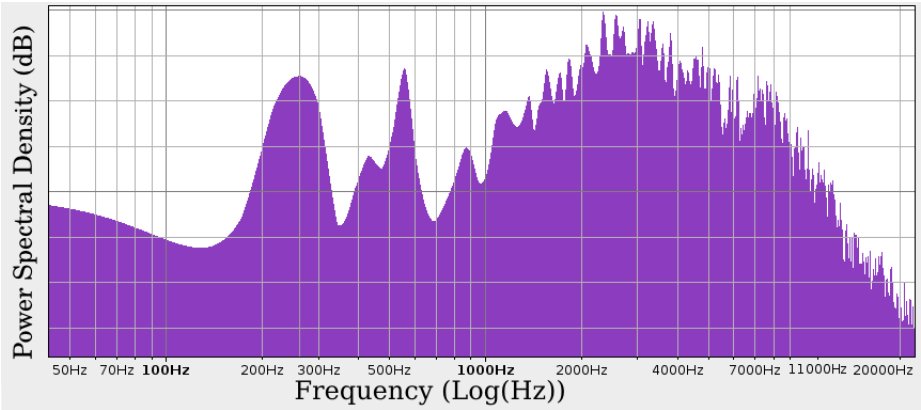


Figure 41: Power Spectrum of the array output of the first toadfish taken with a 1024 sample width window.

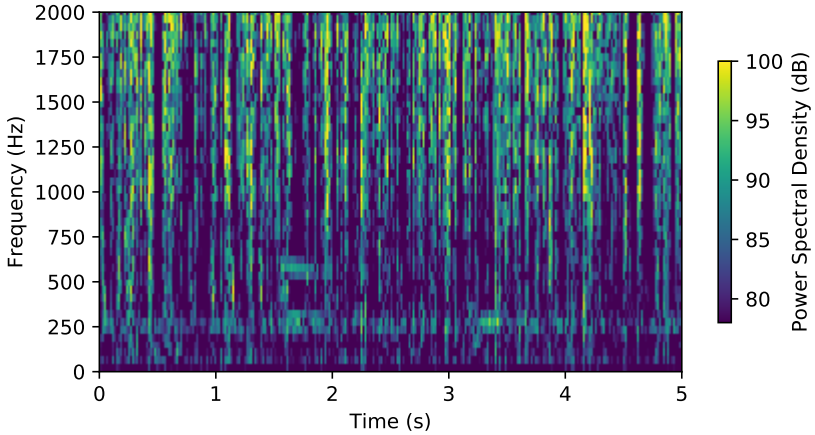


Figure 42: Sonogram plot of the array output of the first toadfish. Taken with a 1024 sample window that overlaps by 512 samples.

steered-response because the dominant frequency, being so prevalent in this recording, has been isolated so well. A sonogram taken of the array's output reveals this is a good choice for a likely position.

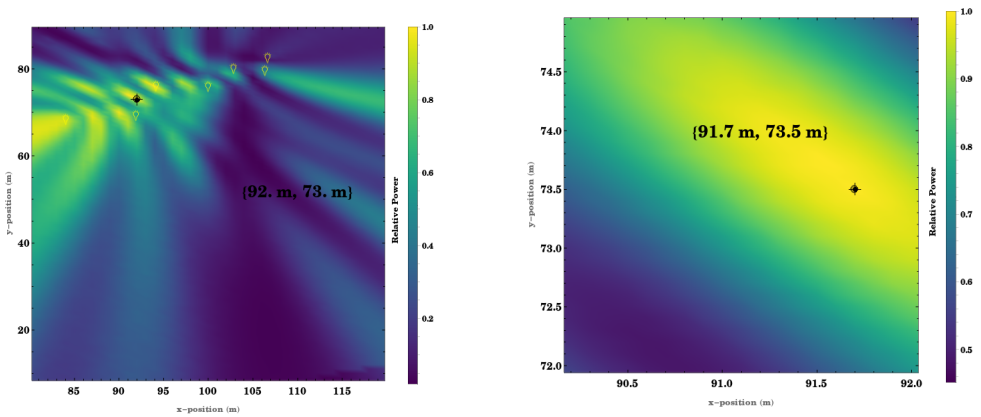


Figure 43: SRP analysis of the second Toadfish.

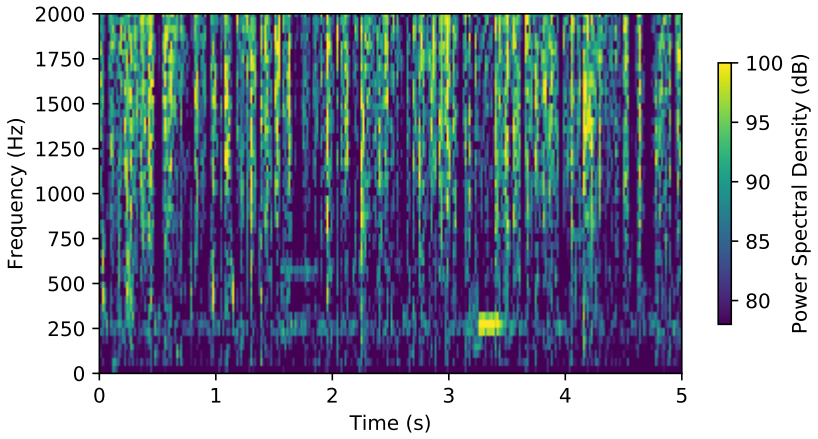


Figure 44: A sonogram of the array's output during the time the second toadfish makes its boatwhistle taken with a 1024 sample wide window that overlaps by 512 samples.

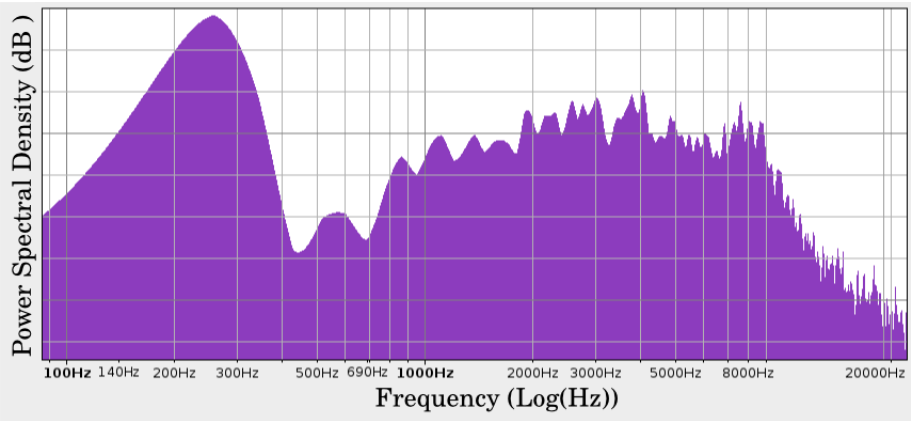


Figure 45: A power spectrum of the array's output of the toadfish with the delays implemented appropriately from the SRP algorithm. The power spectrum is taken with a 1024 samples wide Hamming window.

Further analysis reveals a very prominent peak around the dominant frequency of the boatwhistle approximately 240 Hz to 260 Hz, see Figure 45. The waveform plot is dominated by shrimp clicks, for sustained calls this is almost always an issue. Figure 44 contains the spectrogram of the array's output.

7.3.2 Silver Perch

The silver perch is a drum and makes the stacatto sounds associated with many fish of these type [37, 38]. It is commonly found on the Atlantic coast from New York all the way to Florida and in the Gulf of Mexico. It is a coastal fish, typically found over soft bottom areas. Silver perch are common to intertidal creeks, but can be found in deeper channels, tolerating a broad range of salinities. Spawning occurs in shallow coastal waters, where the juvenile migrates upstream to fresher water to grow and development [39]; adult silver perch will return to the coast after growth [35]. Figure 47 contains the localization results. This localization was carried out on

Figure 46: A silver perch. Brandi Noble/NOAA NMFS SEFSC Pascagoula Laboratory.

a silver perch drum 5 seconds into the recording. The silver perch is located in the very middle of the channel approaching the array to the northeast. Many silver perch continued to called continuously in short bursts throughout the recording.

7.4 Tracking a Silver Perch

A silver perched was observed drumming 0.5 seconds into the record at low-tide around 9:00 PM, and a silver perch was observed drumming later at around 30 seconds into the recording. The result of the localization placed the two results within 11.7 m of each other. With recording time lapsing approximately 40 s; this gives an estimated .4 m/s travel speed for the silver perch, which is within reasonable limits. Nominal speeds for silver have been reported up to 2.0 m/s [40]. Tracking is possible; however, it is not built directly into the algorithm. In order to achieve this, the vocalizations of many fish must be analyzed, and positions considered carefully. The results of the

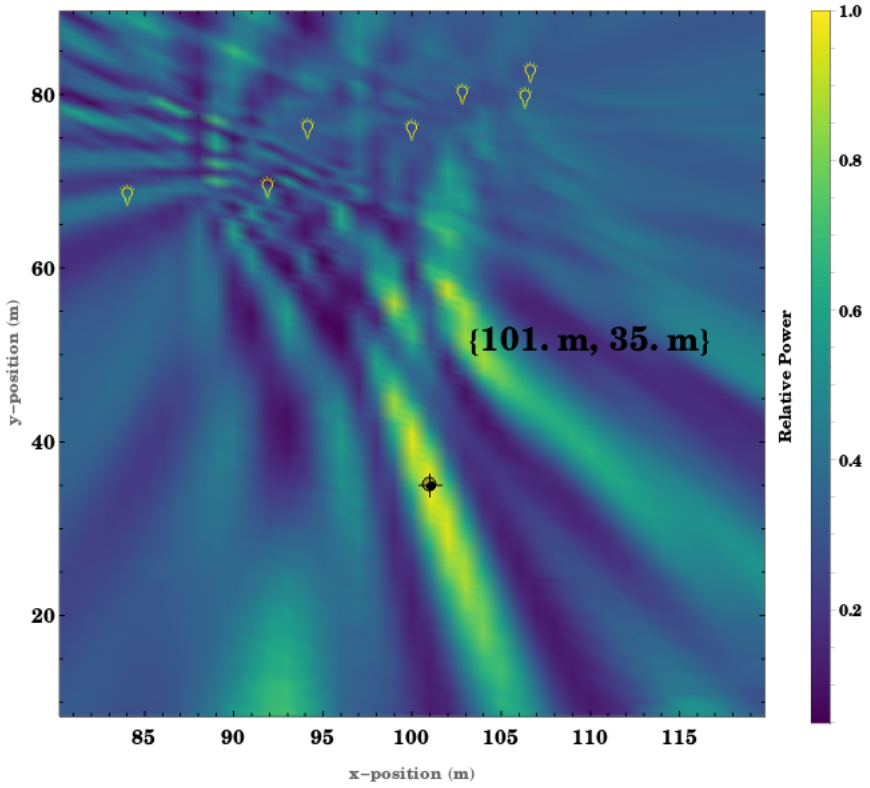


Figure 47: SRP analysis of the silver perch recorded at 5 s.

algorithm are displayed in Figure 50, and the distances from each hydrophone to the source are given in Table 7.

	H1	H2	H3	H4	H5	H6	H7
Toadfish 1	23.2 m	15.8 m	17.1 m	13.2 m	14.91 m	15.9 m	17.8 m
Toadfish 2	9.5 m	4.5 m	3.2 m	8.3 m	15.6 m	12.6 m	17.1 m
Silver Perch 1	37.7 m	35.68 m	41.95 m	41.21 m	45.44 m	48.14 m	48.1 m
Silver Perch 2a	32.2 m	25.05 m	19.9 m	14.6 m	7.29 m	10.43 m	6.3 m
Silver Perch 2b	33.07 m	25.22 m	23.22 m	17.51 m	13.22 m	16.41 m	14.9 m

Table 7: The calculated distances of each fish from the hydrophone array.

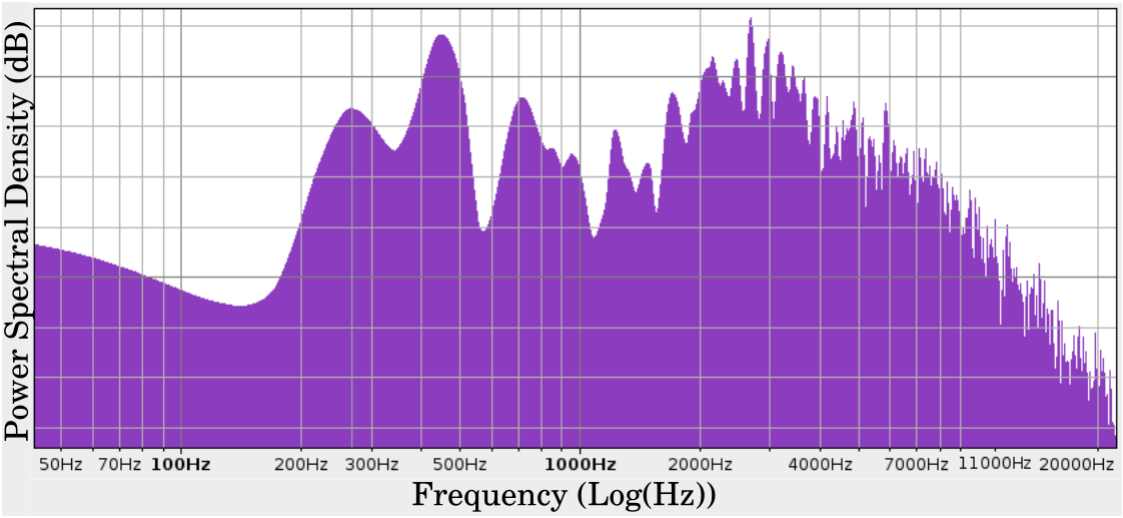


Figure 48: A power spectrum of the array's output of the silver perch with the delays implemented appropriately from the SRP algorithm. The power spectrum is taken with a 1024 samples wide Hamming window.

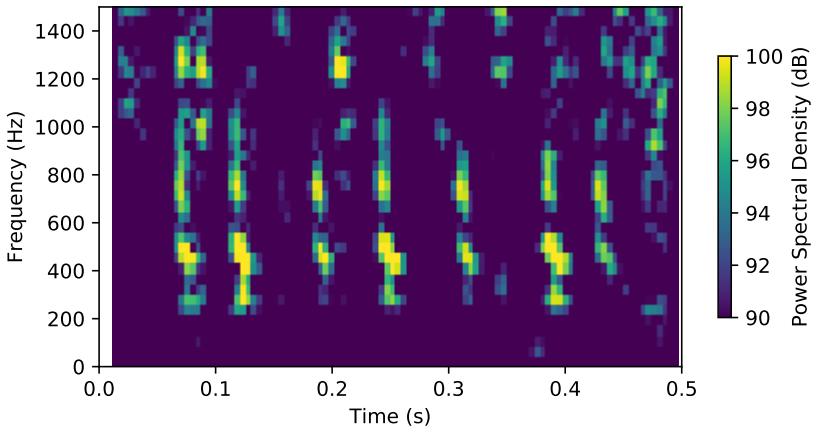


Figure 49: A sonogram during the time the silver perch's drumming taken with a 1024 sample wide window that overlaps by 512 samples.

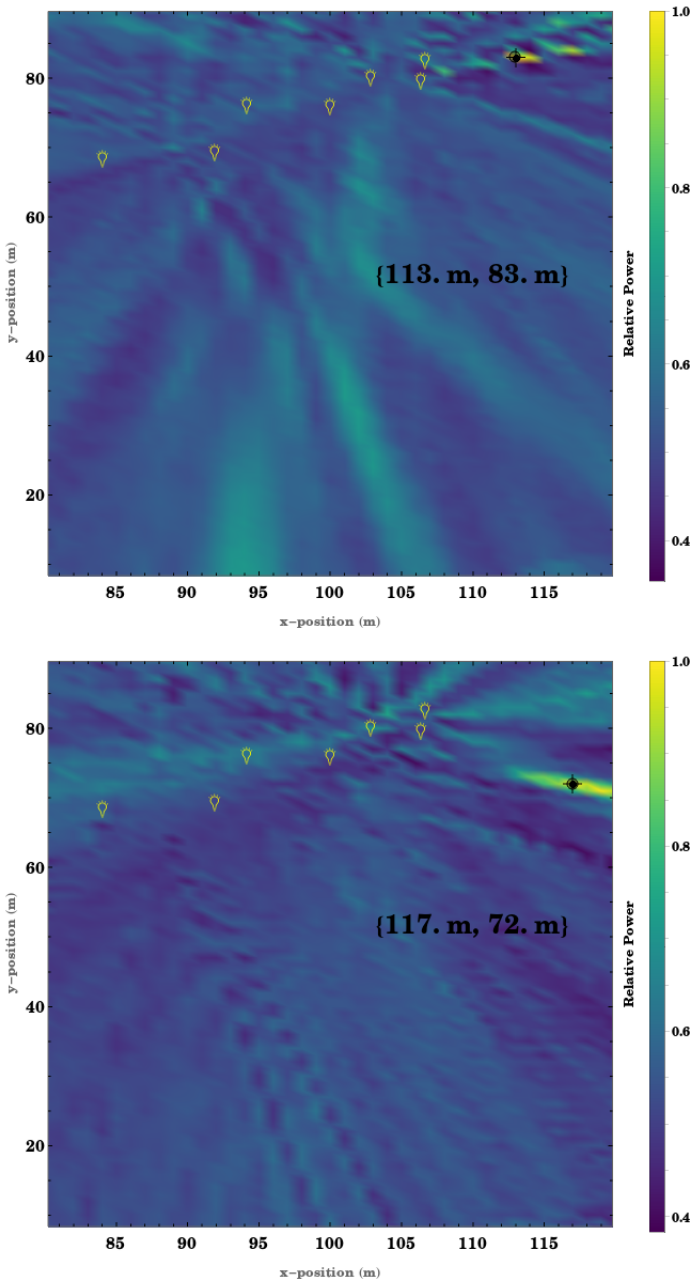


Figure 50: Another silver perch was localized (*top*) a 0.5 s into the recording, and then located 32 seconds (*bottom* into the recording having moved 11.7 m away.

8 Discussion

In conclusion, a seven channel hydrophone array was deployed at Clambank in Hobcaw Barony, SC in an intertidal creek located at the North Inlet-Winyah Bay Nature Reserve. Our objective was to locate sound producing species of fish. The biggest challenges were the field measurements and the amount of noise in the waters. The array performed adequately, but more work could be done to fully unlock its potential. The promise of this technique to future research and the capabilities the array has to study fish is important to us and should be researched more extensively.

Reported instrument precision was concerning, and the true accuracy with which the surrounding area in clam bank is questionable. Most likely, due to the breadth of the space that was being surveyed, the laser range-finding device rounded to the nearest meter. Regardless of what the measurement device claimed as its precision, it did not perform over this mark consistently enough to be trusted to the 0.1 m reported precision value. Ultimately, the hydrophone was unobserved beneath the water where we deployed it, and cautious attempts were made to measure carefully, and calculate a true position of the hydrophone. This was done through correlation analysis of audio tracks which contained a calibration sound played after deployment from known positions. A geometric approach was used considering the distance formula, and the spatial median was used to estimate a location of hydrophones 1 through 7 for SRP analysis. The results of the calibration were a bit tenuous. The localization algorithm did not return the exact location we had measured each calibration to occur. Being off as much as several meters on a single coordinate as is the case for calibrations 2, 4, and 5. There are a few possible reasons for this. The two person team in the small boat consisted of one researcher to make the calibration sound and one researcher

to measure the distance from the reference landmarks to the small boat with the laser range finder. This means that there was a little over a meter distance between the measurement recorded and the researcher producing the calibration sound. This distance was considered negligible; however, the angle the small boat was making with respect to the array and which side of the small boat the sound was produced would have introduced variations from what was actually measured. Also, analysis was implemented in two dimensions and the hydrophone depth was not considered. The calibration was done at high tide with depths of at most two meters, that distance may have impacted the results of the beamformer algorithm more than expected. Also, calibration occurred over an hr, where moving about in the creek and putting sounds into the environment may have made the waters noisier. This is supported by the steered response power calibration results. The first calibration sound came through very cleanly, and in the recording there is noticeably more shrimp during later calibrations.

Apart from the wealth of soniferous fish in the waters at Clam Bank to survey, which would have been no trouble to separate and identify in this analysis, there were thousands upon thousands of snapping shrimp. The animals dominate the recording in the time domain, where close to the hydrophone and far away there is irregular but constantly occurring shrimp pops. These animals dominate the spectrum of the signal as well producing a sound that has intensity at all audible wavelengths (20Hz-20kHz). The pops from shrimp to shrimp are essentially identical, and this masked many fish calls, which were of lower amplitude and buried in a frequency range where shrimp sound still was the dominate form of acoustic energy. Cross correlations and steered-response were affected by them and they never ceased their snapping. This

made localization of calls with any sustain very difficult. Many times a portion of the waveform had to be used that was in between pops from shrimp that were located in the array. This was arduous to do. The attack of a sound is often the most interesting part with the unique structure and features that would be ideal to correlate. As such, staccato calls of fish are likely to perform much better under this type of analysis. Those calls that do coincide with silence from the shrimp can be isolated and used for localization effectively.

The array was a random sparse array. Our goal was to aim for placement of the hydrophones in a line and attempt to replicate array geometries that were known sparse arrays with coarrays that would allow for us to push the array's capabilities past those predicted by the spatial sampling theorem. The space was vastly undersampled, but we were successful in part because the beam pattern produces a prominent mainlobe located at an angle of arrival of zero degrees up to 500 Hz. Toadfish make sounds in this region and there are components of silver perch sounds in this region too, so that through temporal filtering there was assurance of some ability to beamform at these frequency ranges. A sparse array to achieve this was not found. Random arrays and sparse arrays are designed based of a hypothetical regular where the sensors are placed randomly on a grid that represents the idealized array or removed randomly from a grid of regularly placed sensors. The requirement for proper sampling of location space $d/2 \leq \lambda$ does not relate any information about how many sensors a surveyor is using. With enough hydrophones, there begins to be redundancy [41]. Sparse arrays are based on replicating beam patterns of arrays with as few microphones as possible. There are a few known sparse arrays one sees, and that are used frequently in analysis. We were not able to achieve a sparse array

that would allow us to do this. That does not mean it is impossible. There are other techniques that push the array's capabilities further, like the use of co-prime arrays [42], which is basically a composite array with two arrays designed to act in tandem, whose combined output is a desired beampattern. All of this basically happens in the design phase. Future research would involve the exploration of such analyses.

9 Conclusion

Microphone arrays allow for a more extensive survey of an environment, providing the ability to locate sound producing species of animals, and they are a viable option to bioacousticians that wish to know more about the population density in a given habitat. It is relatively noninvasive and convenient to deploy an array that will remained stationed there for an amount of time. Species can be identified based off vocalization, and analysis can be done some time after deployment as a post-process. Any drawbacks in portability of the heavy array equipment are outweighed by the arrays ability to study species that would be very difficult to observe otherwise like nocturnal species or those who stayed hidden within the environment. Another benefit is the arrays ability to study more than one animal at once. The array samples the entire wavefield in the region near the array. The call and response of many animals can be studied. Once position data is known, sound analysis can be done, as well as further analysis with the beamformer. Counting the number of organisms of sound-producing species and tracking species movement is also possible [4] and can reveal even more about the relationships between animals in the ecosystem. Arrays are highly useful and much can be learned about fish by using them.

References

[1] J. F. Webb, R. R. Fay, and A. N. Popper, *Fish Bioacoustics*, vol. 32. Springer Science & Business Media, 2008.

[2] N. H. Fletcher, "Animal bioacoustics," in *Springer handbook of acoustics*, pp. 821–841, Springer, 2014.

[3] R. Rountree, R. Gilmore, C. Goudey, A. Hawkins, J. Luczkovich, and D. Mann, "Listening to fish," *Fisheries*, vol. 31, pp. 433–446, 09 2006.

[4] I. R. Urazghildiiev and D. E. Hannay, "Passive acoustic detection and estimation of the number of sources using compact arrays," *The Journal of the Acoustical Society of America*, vol. 143, no. 5, pp. 2825–2833, 2018.

[5] A. Hughes, *Acoustic source localisation and tracking using microphone arrays*. PhD thesis, 2015.

[6] J. Neering, *Optimization and estimation techniques for passive acoustic source localization*. PhD thesis, École Nationale Supérieure des Mines de Paris, 2009.

[7] N. Madhu and R. Martin, *Acoustic Source Localization with Microphone Arrays*, ch. 6, pp. 135–170. John Wiley & Sons, Ltd, 2008.

[8] J. H. DiBiase, *A high-accuracy, low-latency technique for talker localization in reverberant environments*. PhD thesis, 2000.

[9] D. R. Dowling, "Revealing hidden information with quadratic products of acoustic field amplitudes," *Phys. Rev. Fluids*, vol. 3, p. 110506, Nov 2018.

[10] J. Locascio and D. Mann, "Localization and source level estimates of black drum (pogonias cromis) calls," *The Journal of the Acoustical Society of America*, vol. 130, pp. 1868–79, 10 2011.

[11] D. G. Zeddies, R. R. Fay, M. D. Gray, P. W. Alderks, A. Acob, and J. A. Sisneros, "Local acoustic particle motion guides sound-source localization behavior in the plainfin midshipman fish, porichthys notatus," *Journal of Experimental Biology*, vol. 215, no. 1, pp. 152–160, 2012.

[12] R. Putland, A. Mackiewicz, and A. Mensinger, "Localizing individual soniferous fish using passive acoustic monitoring," *Ecological Informatics*, vol. 48, pp. 60–68, 2018.

[13] A. D. Pierce, *Acoustics an introduction to its physical principles and applications*. Acoustical Society of America, 3rd ed., 1989.

[14] K. V. MacKenzie, "Discussion of sea water sound-speed determinations," *Acoustical Society of America Journal*, vol. 70, pp. 801–806, 1981.

[15] P. C. Etter, *Underwater acoustic modeling and simulation*. CRC Press, 2013.

[16] F. A. Everest and K. C. Pohlmann, *Master handbook of acoustics / F. Alton Everest, Ken C. Pohlmann*. McGraw-Hill Education, 2015.

[17] B. D. Steinberg, *Principles of aperture and array system design: including random and adaptive arrays*. John Wiley and Sons, Inc., 1976.

[18] J. G. Proakis, *Digital signal processing: principles algorithms and applications*. Pearson Education India, 2001.

[19] E. A. P. Habets and S. Gannot, "Generating sensor signals in isotropic noise fields.," *The Journal of the Acoustical Society of America*, vol. 122 6, pp. 3464–70, 2007.

[20] D. H. Johnson and D. E. Dudgeon, *Array signal processing concepts and techniques*. PTR Prentice Hall, Inc., 1993.

[21] L. L. Beranek, *ACOUSTICS: sound fields, transducers and vibration*. Elsevier Academic Press, 2019.

[22] R. L. Culver, B. E. Fowler, and D. C. Barber, "Near- and far-field beam forming using a linear array in deep and shallow water," *The Journal of the Acoustical Society of America*, vol. 135, no. 4, pp. 2393–2393, 2014.

[23] I. McCowan, "Microphone arrays: A tutorial," *Queensland University, Australia*, pp. 1–38, 2001.

[24] J. W. Strut, "Xxxi. investigations in optics, with special reference to the spectroscope," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 8, no. 49, pp. 261–274, 1879.

[25] A. Manikas, C. Proukakis, and V. Lefkaditis, "Investigative study of planar array ambiguities based on "hyperhelical" parameterization," *Signal Processing, IEEE Transactions on*, vol. 47, pp. 1532 – 1541, 07 1999.

[26] High Tech , inc., "HTI-96-min hydrophone series," 2019.

[27] ZOOM North America, "Zoom f8 field recorder," 2019.

[28] J. Dmochowski, J. Benesty, and S. Affes, "A generalized steered response power method for computationally viable source localization," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, pp. 2510–2526, 2007.

[29] J. Lee, S.-W. Chung, H.-G. Kang, and M.-S. Choi, "Generic uniform search grid generation algorithm for far-field source localization," *The Journal of the Acoustical Society of America*, vol. 143, no. 1, pp. EL37–EL42, 2018.

[30] H. Pan, R. Scheibler, E. Bezzam, I. Dokmanic, and M. Vetterli, "Frida: Fri-based doa estimation for arbitrary array layouts," *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar 2017.

[31] M. Owen, *Practical signal processing*. Cambridge University Press, 2012.

[32] H. Do, H. F. Silverman, and Y. Yu, "A real-time srp-phat source location implementation using stochastic region contraction(src) on a large-aperture microphone array," *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, vol. 1, pp. I–121–I–124, 2007.

[33] M. Cobos, A. Marti, and J. J. Lopez, "A modified srp-phat functional for robust real-time sound source localization with scalable spatial sampling," *IEEE Signal Processing Letters*, vol. 18, pp. 71–74, Jan 2011.

[34] M. V. Lima, W. A. Martins, L. O. Nunes, L. W. Biscainho, T. N. Ferreira, M. V. Costa, and B. Lee, "Efficient steered-response power methods for sound source localization using microphone arrays," *IEEE Signal Processing Letters*, vol. 22, no. 8, 2014.

[35] R. Simpson, D. Allen, S. Sherman, and K. Edwards, *Fishes of the North Inlet estuary: a guide to their identification and ecology.* Belle W. Baruch Institute, Special Publication., 2015.

[36] C. P. H. Elemans, A. F. Mensinger, and L. C. Rome, "Vocal production complexity correlates with neural instructions in the oyster toadfish (opsanus tau)," *Journal of Experimental Biology*, vol. 217, no. 11, pp. 1887–1893, 2014.

[37] M. W. Sprague, J. J. Luczkovich, R. C. Pullinger, S. E. Johnson, T. Jenkins, and H. J. Daniel, "Using spectral analysis to identify drumming sounds of some north carolina fishes in the family sciaenidae," *Journal of the Elisha Mitchell Scientific Society*, vol. 116, no. 2, pp. 124–145, 2000.

[38] M. W. Sprague and J. J. Luczkovich, "Measurement of an individual silver perch bairdiella chrysoura sound pressure level in a field recording," *The Journal of the Acoustical Society of America*, vol. 116, no. 5, pp. 3186–3191, 2004.

[39] J. J. Luczkovich, H. J. D. III, M. Hutchinson, T. Jenkins, S. E. Johnson, R. C. Pullinger, and M. W. Sprague, "Sounds of sex and death in the sea: Bottlenose dolphin whistles suppress mating choruses of silver perch," *Bioacoustics*, vol. 10, no. 4, pp. 323–334, 2000.

[40] M. Mallen-Cooper, "Swimming ability of adult golden perch, macquaria ambigua (percichthyidae), and adult silver perch, bidyanus bidyanus (teraponidae), in an experimental vertical-slot fishway," *Marine and Freshwater Research*, vol. 45, no. 2, p. 191, 1994.

[41] T. Padois, "Acoustic source localization based on the generalized cross-

correlation and the generalized mean with few microphones," *The Journal of the Acoustical Society of America*, vol. 143, no. 5, pp. EL393–EL398, 2018.

[42] K. Adhikari, "Beamforming with semi-coprime arrays," *The Journal of the Acoustical Society of America*, vol. 145, no. 5, pp. 2841–2850, 2019.

[43] M. R. Bai, J.-G. Ih, and J. Benesty, *Acoustic array systems: theory, implementation, and application*. John Wiley & Sons, 2013.

A Physical water quality data from Clam Bank monitoring site

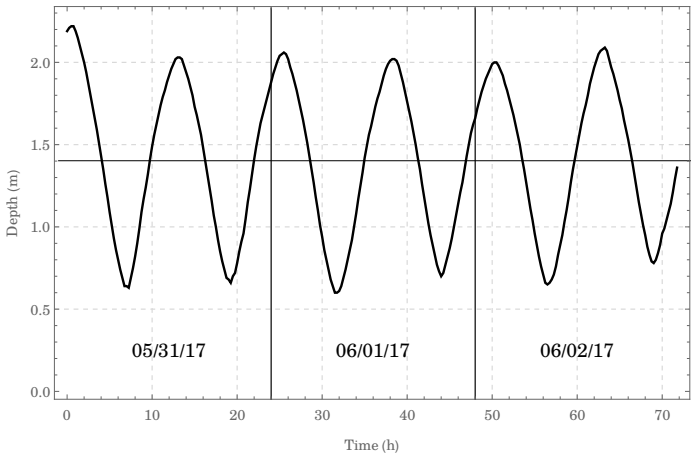


Figure 51: Depth measurements [m] for three days of monitoring.

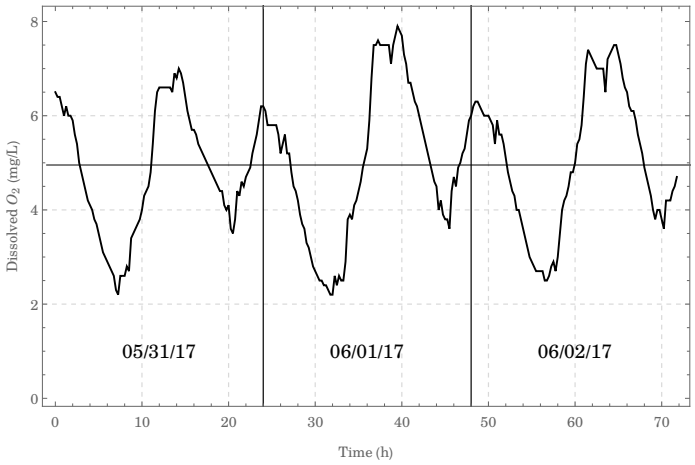


Figure 52: Dissolved oxygen measurements [mg/L] for three days of monitoring.

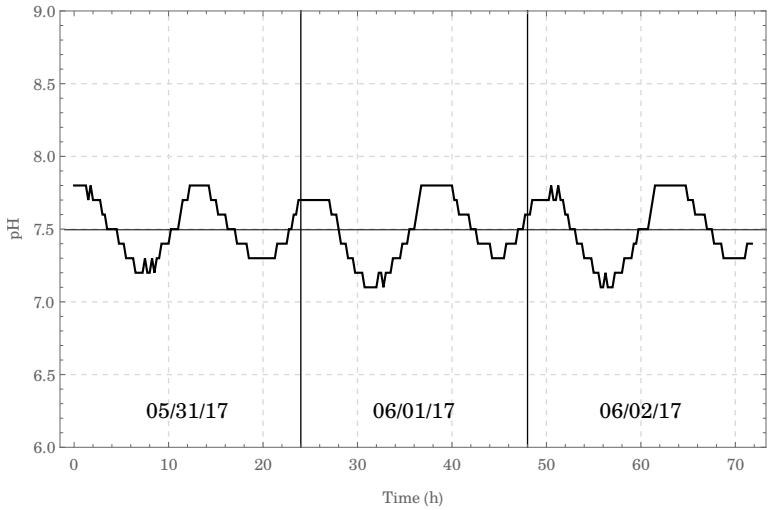


Figure 53: pH measurements for three days of monitoring.

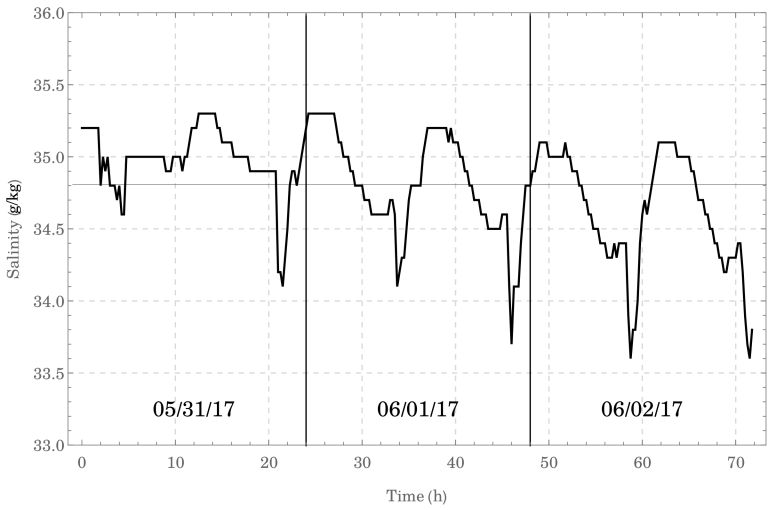


Figure 54: Salinity measurements [g/kg] for three days of monitoring.

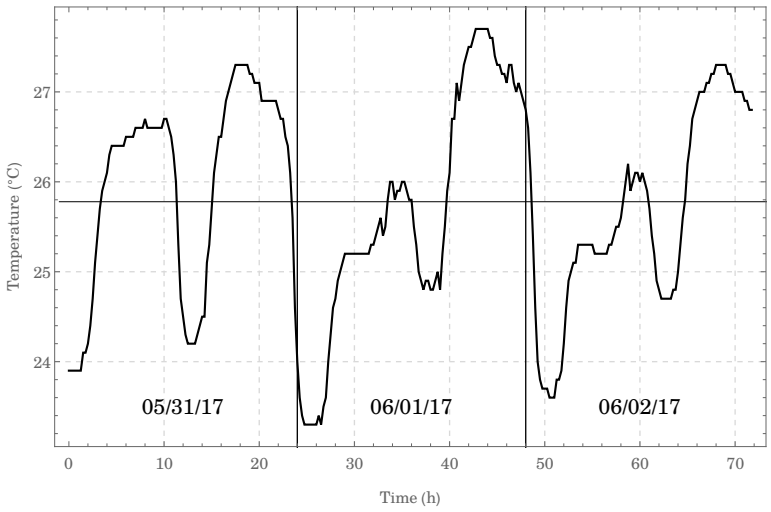


Figure 55: Temperature measurements [°C]for three days of monitoring.

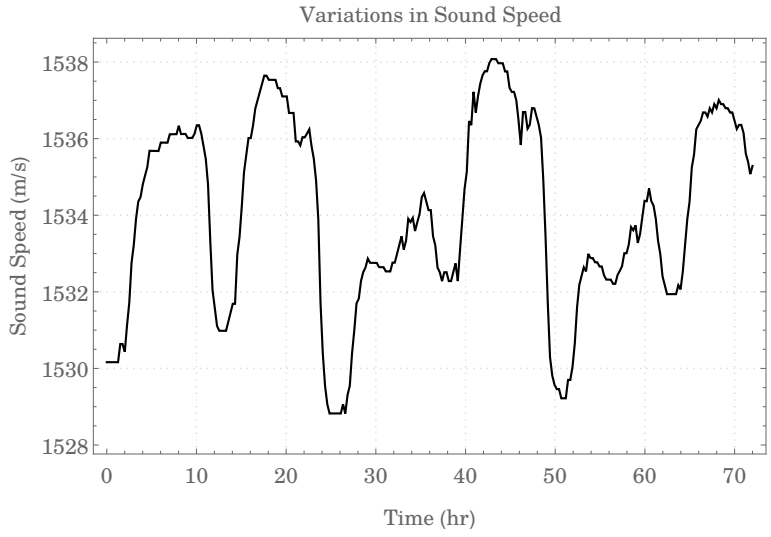


Figure 56: The sound speed (m/s) as it varied over the 72 hours of observation.

B HTI-96-MIN

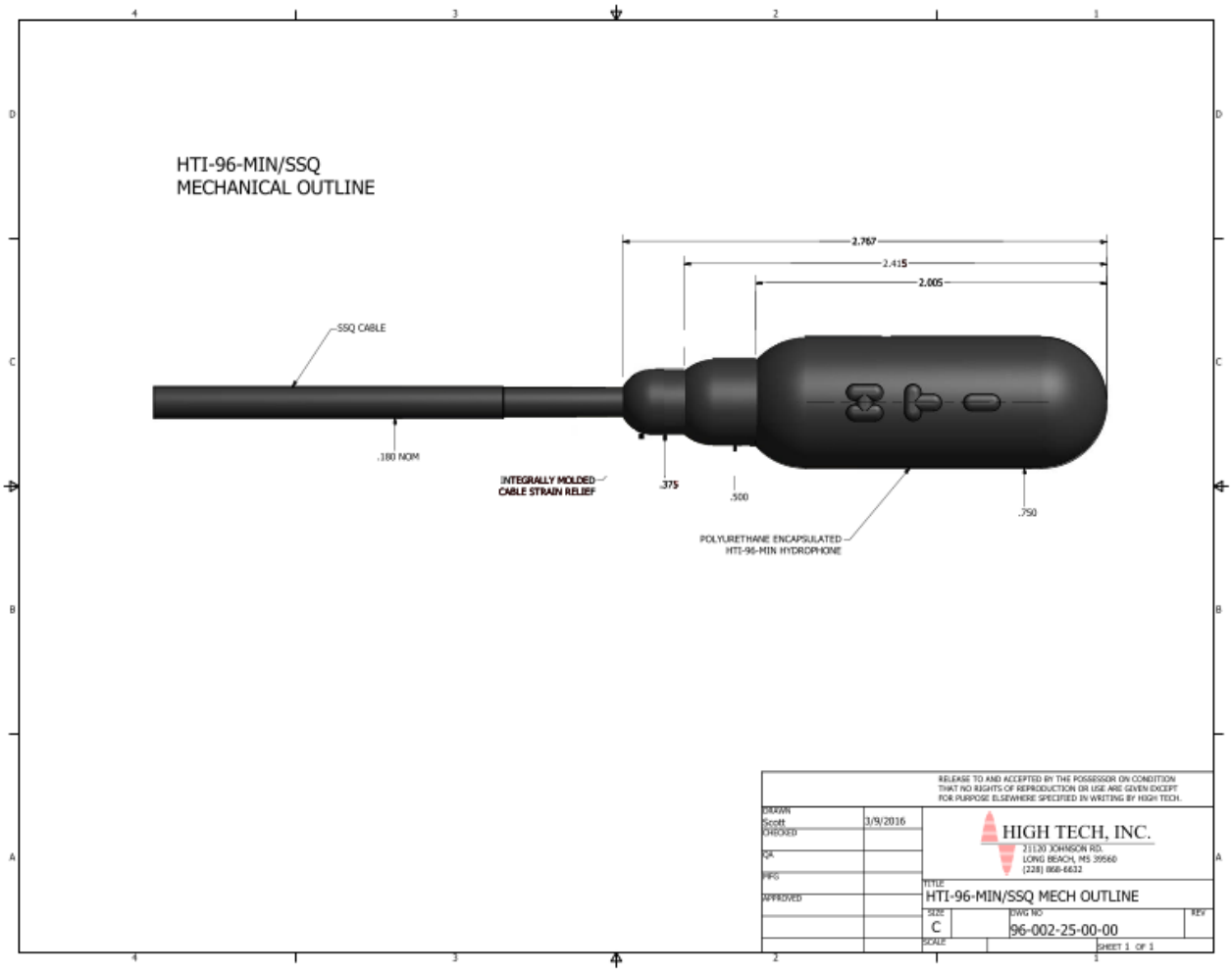


Figure 57: Mechanical outline of HTI-96-MIN hydrophones used for the seven sensor array deployed at Clam Bank [26]