Graduate thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Engineering Cybernetics

Kristian Sørensen Stene

Underwater recording and beamforming methods applicable for suppressing surface sounds from aquaculture facilities

Graduate thesis in MTTK

Supervisor: Martin Føre, John Reidar Mathiassen

June 2020

NTNU
Norwegian University of Science and Technology

Kristian Sørensen Stene

Underwater recording and beamforming methods applicable for suppressing surface sounds from aquaculture facilities

Graduate thesis in MTTK
Supervisor: Martin Føre, John Reidar Mathiassen
June 2020

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Engineering Cybernetics

NTNU
Norwegian University of
Science and Technology

NTNU
Norges teknisk-naturvitenskapelige universitet

Fakultet for informasjonsteknologi, matematikk og elektroteknikk
Institutt for teknisk kybernetikk

MASTER'S THESIS

Name of the candidate: Kristian Sørensen Stene

Fag: Engineering Cybernetics

Oppgavens tittel (norsk): Metoder for undervannsopptak og stråleforming egnet til å dempe overflatestøy fra oppdrettsanlegg.

Oppgavens tittel (engelsk): Underwater recording and beamforming methods applicable for suppressing surface sounds from aquaculture facilities

Background:
Acoustic signals are known to propagate well in water, exhibiting both higher speeds (about 1500 m s^{-1}) and ranges (kilometres) than in air. Since electromagnetic signals such as radio and visual light are heavily attenuated by passing through water, this implies that acoustic signals might be a more suitable method for monitoring larger volumes underwater than e.g. camera solutions. Although this is already being exploited in applications such as echo sounder and sonar technology and acoustic telemetry methods, there might also be a potential in analysing passive recordings of acoustic signals as a method for studying underwater dynamics.

The salmon aquaculture industry is a significant national and global provider of seafood, and as the industry has grown, so has the need for technological methods for monitoring the fish in the cages as it is now impossible to properly observe the fish using direct observation. This need has been underlined by an increasing focus on animal welfare from the customer segment and authorities, requiring that farmers actively seek to acquire a certain level of control over the welfare status in cages.
While camera-based methods, active acoustic methods (e.g. sonars) and telemetry/biosensors are today becoming increasingly industrialised and being further developed in research and innovation processes, passive acoustics has not yet been explored as a tool for assessing fish states in sea-cages. Earlier studies suggest that several species of fish can emit characteristic sound signals that may be linked with their behaviour and/or status. These signals could either arise as a voluntary sound generated using specific organs, or due to the mechanical effects of body movement and interaction with the surrounding water. Although farmed salmon have not been explicitly studied in this regard, it is likely that a similar relationship between activity/status and emitted sounds can be found for this species.

SoundWell is a research project where NINA, SINTEF, NTNU and international partners collaborate in investigating the potential of using automated analyses of soundscapes in aquaculture production facilities as a tool for monitoring the fish. The SoundWell project will explore if population level (i.e. "roar of the crowd") and individual (i.e. "cocktail party problem") sound profiles contains information on the states and welfare of the fish during production.

This master's thesis will focus on developing fundamental methods and knowledge required to develop methods for using arrays of multiple hydrophones to monitor the soundscape at specific

regions in the cage volume. Specifically, the project will target the challenges of isolating the sounds emitted in the mid-range region of the cage (where the fish are more likely to stay) from the noise generated at the surface. Since this approach to fish monitoring has never been tried before, the study will feature aspects from basic literature studies on acoustic propagation in water, through mathematical modelling of underwater soundscapes, to developing algorithms for removing background noise from hydrophone array measurements. The work will therefore comprise both purely theoretical studies and the use of actual soundscape data collected in the sea.

The project will contain the following elements:

- Literature study covering:
    - Acoustic propagation in water and marine hydroacoustics
    - Beamforming methods for spatial filtering of acoustic signals
- Implement candidate beamforming methods for spatial filtering of acoustic underwater signals
- Virtual experiments
    - Develop theoretical mathematical model of underwater soundscapes, describing sound propagation from point source subjected to background noise
    - Explore the theoretical efficacy of beamforming methods and hydrophone array configurations
    - Based on this, select beamforming methods and array configurations suitable for isolating the sound of the fish population in a sea-cage by dampening noise from the surface/bottom
- Physical experiment to collect validation data for selected methods
    - Deploy rig with the selected hydrophone array configuration in a marine environment
    - Deploy point source emitting a known sequence of signals with similar bandwidth as those expected from a real fish population
- Apply the implemented beamforming approaches to the validation data to explore the actual efficacy of these in isolating the desired signal from background noise

Thesis start: 06. January 2020

Delivery deadline: 08. June 2020

Delivered at the Department of Engineering Cybernetics

Main supervisor: Martin Føre
Co-supervisor: Dr. John Reidar Mathiassen, SINTEF Ocean

Trondheim, 17. January 2020

Faglærer

Abstract

Norwegian version
Nylig økt fokus på produksjonseffektivitet og fiskevelferd har skapt behovet for ny teknologi for å overvåke statusen til fisk under forskjellige driftsforhold innen fiskeoppdett. Innspillinger av kjente retningsbestemte undervannssignaler ved bruk av en vertikal uniform lineær mottager (ULM) med tre hydrofoner behandles ved hjelp av stråleforming for å sjekke metodens evne til å passivt lytte horisontalt samtidig som den undertrykker overflatestøy / interferens. Resultater fra dette fysiske eksperimentet i tillegg til to virtuelle eksperimenter, der alle eksperimenter bruker tre forskjellige stråledannende teknikker, presenteres i denne oppgaven. De to virtuelle eksperimentene er forskjellige med tanke på antall mottagerelementer som er brukt, dvs. tre og fem. Stråleformingsteknikkene som brukes er filtrer-og-summer (FOS) stråleforming, adaptiv lineært begrenset minimum varians (LBMV) stråleforming og adaptiv generalisert sidelobe kansellerende (GSK) stråledannelse.

Fysisk testing med signal av interesse (SAI) som ankommer mottageren med 0 høyde, og overflateforstyrrelser ankommer fra en vinkel på 30 høyde, viser at adaptiv stråleforming, og mest fremtredende LBMV-stråledannelse, er i stand til å øke signalstyrken mens den undertrykker interferensen samt senke det generelle støynivået. Virtuell testing antyder ytterligere forbedringer av signal til støyforhold (STS) ved bruk av fem hydrofoner i stedet for tre. Denne teknologien antas dermed å være en verdifull ressurs innen presisjonsfiskeoppdrett for akustisk fiskestimovervåking.

English version
Recent increased focus on production efficiency and fish welfare urge the need for new technology for monitoring the status of fish under different operational conditions within sea cage aquaculture. Recordings of known directional underwater signals using a vertical uniform linear array (ULA) with three hydrophones is processed using beamforming to check the method's ability to passively listen horizontally while suppressing surface noise/interference. Results from the physical experiment in addition to two virtual experiments, where all experiments applies three different beamforming techniques, are presented in this thesis. The two virtual experiments differs by the number of array elements used, i.e. three and five. The beamforming techniques used are fixed filter and sum (FAS) beamforming, adaptive *linearly constrained minimum variance* (LCMV) beamforming and adaptive generalized sidelobe canceller (GSC) beamforming.

Physical testing with *signal of interest* (SOI) arriving at the array at 0° elevation, and surface interference coming in from an angle $\sim 30^\circ$ elevation proves that adaptive beamforming, and most prominently LCMV beamforming, is able to increase the signal gain while suppressing the interference as well as lowering the overall noise level. Virtual testing suggests further *signal-to-noise ratio* (SNR) improvements using five hydrophones instead of three. This technology is consequently believed to be a valuable asset in a precision fish farm setting for acoustic fish school monitoring.

Preface

This thesis is my final work in connection with my Master of Science degree at the Department of Engineering Cybernetics at the Norwegian University of Science and Technology. The work is done as part of a project funded by the Research Council of Norway (RCN grant number 280512). Despite the unfortunate circumstances that the Covid-19 outbreak brought upon us during the time frame of this thesis, I managed to finalize the thesis in time.

The completion of this thesis was done under the supervision of Associate Professor Martin Føre and Senior Research Scientist John Reidar Mathiassen. I am most grateful for their professionality, invaluable opinions and help leading my research in the right direction. I would not have ended up with the final results if they had not helped me a long the way. I would also like to show my appreciation to friends and family for their support, especially Arild and Nicholas for helping me and lending me equipment in regards to the physical experiment conducted in this thesis. I would also like to thank and girlfriend, Silje, who has been a victim of hours of topic related gibberish, for her motivation and patience.

Table of Contents

List of Tables

List of Figures

Abbreviations

AG	=	Array gain
DAS	=	Delay and sum
DI	=	Directivity index
DOA	=	Direction of arrival
FAS	=	Filter and sum
FBW	=	Fractional beam-width
FIR	=	Signal of interest
HPBW	=	Half power Beam-width
LCMV	=	Linearly constrained minimum variance
LMS	=	Least mean square
NLMS	=	Normalized least mean square
PCC	=	Person's correlation coefficient
PR	=	(signal-to-interference) peak ratio
PSD	=	Power spectral density
RMS	=	Root mean square
SOI	=	Signal of interest
SONI	=	Signal of no interest
ULA	=	Uniform linear array
VAD	=	Voice activity detection

Chapter 1

Introduction

1.1 Norwegian salmon farming

Norwegian salmon production has been facing strong production and economic growth over the last few decades [33; 34]. To this date, a lot of the fish farm processes are manually controlled and the operations are mostly experience based. Technological solutions for fish monitoring in Norwegian fish farms are often quite limited. Cameras are commonly used for fish monitoring within the industry, but this technology is limited to the visual aspect of monitoring. Researchers are constantly working to find new ways to collect data that can be utilized within aquaculture, e.g. active sonar/echo sounding, hydroacoustic telemetry, hyper- and multispectral imaging [7]. This development is needed to acquire better control over the different operational processes and phases within aquaculture. Complete knowledge about the variables within the fish cage system could allow for complete automatic control which could cut costs as well as improve the fish welfare in the cage. Therefore, the development of new sensors are needed for better observation of the system to enhance control following the precision fish farming philosophy.

1.2 Fish bioacoustics

Fish bioacoustics is a ever developing field of research dwelling upon biological vocalization and auditory mechanisms with regards to fish. How fish respond to sounds has probably been a question to man since they started fishing, as it is embedded in our language that one should be quiet while fishing to not scare the fish. However, one had to wait until the 1960s for the field of marine bioacoustics to really become established [46]. Focusing on fish vocalization, some research has been conducted to map fish sounds and how they are produced. Most researched fish sounds are linked to the reproductive stages of the fish such as courtship and spawning as well as agonistic behaviour commonly tied to male territorial disputes [11; 14; 17; 46]. Apart from the ever increasing interest of the marine soundscape, there is still (to the best of the author's knowledge) a lack of informa-

tion about the vocal mechanisms of Atlantic salmon (*Salmo salar*), and especially tied to different operational conditions in sea cage aquaculture.

1.3 Motivation and challenges

Combining interests from precision fish farming, fish welfare and bioacoustic opens up a new landscape of opportunities, research and, hopefully, solutions that could give a better understanding of how the fish interact with each other within a sea cage, as well as increase the level of control and hence automation. Still, acquiring profound and useful fish audio data from a possibly noisy and harsh environment such as a fish cage containing hundred of thousands of fish can be cumbersome. Mechanical/industrial noise from boats and the feeding barge, wind, rain, biological noise from outside the sea cage and structural noise from the cage itself are some of the noise/interference sources in near vicinity of the fish farm, leading the question of how to isolate the SOI from the noise in such an environment.

1.4 Array processing and beamforming

Beamforming is a space-time filtering technique that utilizes multiple spatially distributed transmitting or receiving elements, combined known as an *array*, for directional transmission or reception of signals. Different formations of the array elements creates different transmitting or receiving capabilities. Highlighting signal reception, beamforming enhances the *signal of interest* (SOI) while suppressing interference, background noise, and reverberation, by steering its listening direction towards desired directions. For this reason, the technique is used in a wide variety of disciplines such as telecommunication, astronomy, medical imaging, acoustics, radar systems and sonar systems [15; 19; 36; 39; 42; 49; 51].

In *passive* acoustics (i.e. the act of listening to sounds sources opposed to *active* acoustics where sound is generated and received as echos), beamforming has been applied in applications such as directional noise cancelling [16], source separation and signal enhancement [30; 35] and source detection, localization and tracking [23; 41]. An great example is Norsonic's Acoustic Camera Nor848A equipped with hundreds of microphones and a single camera mounted on a circular disk with 0.4m in diameter [27], making it useful in sound dispalying applications such as room acoustics, locating sound leakage from e.g. offices or buildings, pinpointing tonal noise sources at processing facilities. The most profound use of passive arrays and beamforming in underwater acoustics has been within military application for detecting, locating and/or tracking enemy ships and submarines [39], but it has also been used in passive acoustic monitoring of cetaceans [2; 3], survey the distribution of spawning red drum fish (*Sciaenops ocellatus*) using a towed *hydrophone* (or underwater microphone) array [12] and to observe vertical migration of possibly marine mammals, fish, and invertebrates using a large vertical array of 131 hydrophones [5].

1.5 Problem formulation and objectives

As the fish is believed to most likely stay in the (vertically) mid range region of the sea cage, this thesis wish to determine if a vertically placed *uniform linear array* (ULA) with an inherent axisymmetrical response pattern in the horizontal plane is able to isolate the SOI from directional noise and interference (e.g. surface and seabed noise arriving from angles other than the listening direction) using beamforming. Figure 1.1 visualizes the imagined application scenario within a sea cage.

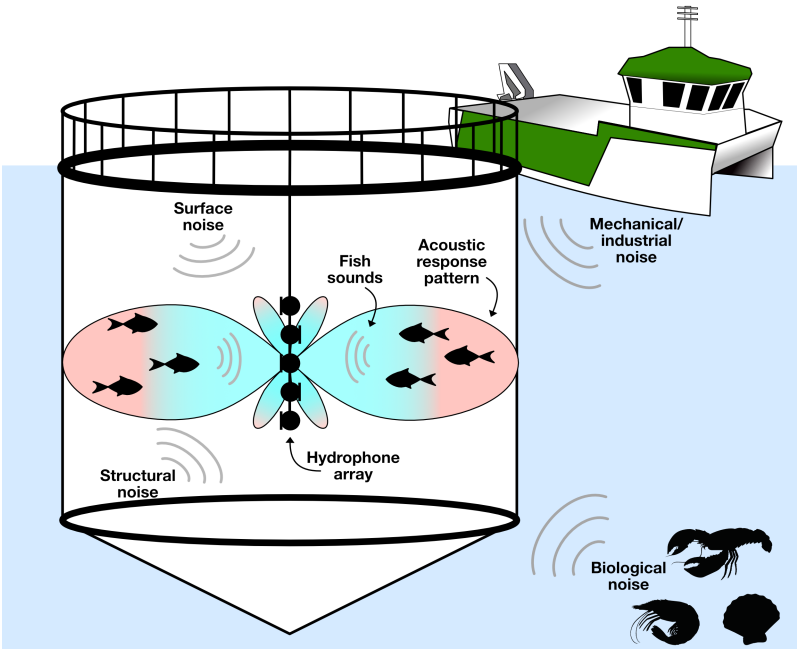


Figure 1.1: Illustration of a potential use case of the proposed technology within a fish farm sea cage.

The main objective with this thesis is to conduct an physical beamforming experiment with three hydrophones in an marine environment imitating a oversimplified sea cage setting. The execution of this experiment, should give insight into the questions:

1. Is there an advantage of using several hydrophones over a single hydrophone?
2. Does hydrophone array beamforming increase sensitivity in the listening direction while suppressing noise/interference from other directions.

Secondary objectives are

1. Compare the results from the physical experiment with results form two virtual experiments with three and five virtual hydrophones.
2. Implement the beamforming techniques and necessary evaluation code independently from major code libraries and toolboxes.

1.6 Historical perspective of underwater sound

Humans listening to sounds beneath the surface of the sea can be traced back to 1490 when one of humanities finest polyhistors, Leonardo da Vinci, wrote the first known report on the subject [39].

> *"If you cause your ship to stop and place the head of a long tube in the water and place the outer extremity to your ear, you will hear ships at a great distance from you."*
>
> — Leonardo da Vinci

Further on, Newton is credited as the first person that tried to describe the properties of sound waves in 1686 in proposition 49 of book II of the *Principia Mathematica* [25], before Swiss physicist, Jean-Daniel Colladon, and French mathematician, Jacques Charles François Sturm, collaborated and was the first to measure the propagation velocity of sound underwater to a surprising degree of accuracy in 1827. Moving along to the later part of the nineteenth century, many scientist took interest in subjects that indirectly led up to the development of the *transducer*, i.e. a material that convert vibrations into electricity and vice versa. *Piezoelectricity* was first discovered in 1880 by Jacques and Pierre Curie as they demonstrated that stressing certain crystals produces electric charge between the crystal faces [39].

Regardless of the development of the nineteenth century, it was not until the World-War-I era with the rise of submarine technology, such as early echo ranging schemes and underwater signaling, that the interest in underwater sound really escalated. As the war ended in 1919, the first scientific paper on underwater sound was published by the German scientist Hugo Lichte giving the, ahead-of-its-time, theoretical description of refracting sound waves in deep sea due to temperature and salinity differences [18]. Urick [39] claimed that the work of Lichte remained unrecognised for 60 years.

In between World-War-I and World-War-II better knowledge about the fundamentals of sound propagation in the sea was gained. The sonar hardware advanced, e.g. natural quartz was replaced with synthetic Rochelle salt as the piezoelectric material of choice in transducers. Echosounders found its place, both military and commercially. Non-military applications was sea floor mapping in shallow waters, seismic mapping using low frequent sound and fishery echosounding for school detection [37].

Advances within the field continued under World-War II, but behind closed curtains. But, the research done at that time became the cornerstone for what we call *the sonar equations* today, including topics like source strength, noise levels, receiver characteristics, reverberations and sound absorption. Military equipment such as homing torpedoes, acoustic mines, active sonar and stealth coating for submarines was also deployed under the second world war [38; 39].

More advancements and refinement of the the research and technology developed since World-War-I was continued after the second world war. During the cold war, large hydrophone arrays was placed along the continental shelves (first) on the east coast and (then) the west coast of North America by the United States to listen for Soviet submarines. The cultivation and maturing of complex signal processing during the 60s and 70s resulted in advanced digital computing and the introduction of adaptive optimal beamformers among

other acoustic processing successes. The blossom of the digital revolution in the 80s led to great advances in hardware, and hence new areas of research and applications [24].

As explained above, underwater acoustics has its roots buried deep into history with a lot of its technology tied to military interest. Commercially today, underwater acoustics are used for seabed scanning, seismic mapping, acoustic speedometers and flow-meters, aquatic animal finding, counting, luring and tagging, divers' aids, underwater communication and telemetry [39]. These applications are in many ways different to its "dry" counter part because of the differences between the properties of air and water, and because of the fact that humans prefer to stay out of the water. Although humans are mammals, the interests in exploring and utilizing the oceans is trending, and humans tend to find new uses for old technology.

Chapter 2

Theory

Sound is simply mechanical vibrations that propagate through a transmission medium by the compression and decompression of molecules. The transmission medium may be of gas, liquid or solid state. Just like radio waves are good for transmitting information in air, sound is good for transmitting information in water. In fact, sound propagates best in water compared to all other forms of radiation that tend to attenuate more quickly in this medium [39]. When sound is transmitted at point A and travels to point B it may travel directly taking the direct path, or by reflecting of the sea surface, seabed or other objects in the water. This combination of paths is called *multipath* transmission and may be perceived as echoes at point B. Environments that suffers from this, e.g. confined spaces, are called *echoic* while the absence of echoes gives *anechoic* environments, e.g. deep sea. If the reflection surface is rough the sound may reflect in multiple directions, or simply put, acoustic *scattering*. As the signal propagate through the water it may experience attenuation due to sound energy being converted into heat as a result of the viscosity and thermal conduction of water. This is called *acoustic absorption*. Another effect that the sound may experience is *refraction*, or bending of propagation direction caused by differences in density and hence propagation speed in the water column. All these effects add up to a well composed and complex medium that is important in the work of underwater acoustics.

This chapter will provide the reader with all the basic and maybe not so basic theory needed to understand how to listen to sounds underwater with array technology. Starting off by giving a quick introduction to underwater noise will provide the reader with underlying noise characteristics found in the ocean. Next, explanations of the mechanisms that fish use to produce sounds will be given in addition to briefly portraying the character of fish sounds and what to expect in a potential fish farm setting. Then, the next two sections will cover how to record sounds with hydrophones, common array geometries and the concept of temporal and spatial aliasing. A mathematical and visual explanation of wave propagation will henceforth be given before explaining the underwater signal model used to simulate propagation, and for making important assumptions in the development of beamforming techniques introduced after the concept of time shifting is delivered. Lastly, important array and sound performance measurements are introduced.

2.1 Underwater noise

2.1.1 Ambient noise

Just like any other natural environment the ocean contains ambient background noise. This ambient noise is the part of the observed total noise that is not due to the hydrophone itself, i.e. *self-noise*, or identifiable sources of noise/interference. What causes this noise is everything from tides, waves, rain, tectonic movement and seismic activity, distant shipping and human activity, biological activity, currents and molecular thermal motion [39]. The different factors contribute to ambient noise in different frequency spectra, and varies with oceanic location, depths and weather. Giving a good description of ambient noise in the ocean is quite complex but has been one of the most important fields of study within underwater acoustics because of the importance for underwater warfare [39]. Figure 2.1 is an adapted illustration of studied noise sources summarised by Wenz in [47].

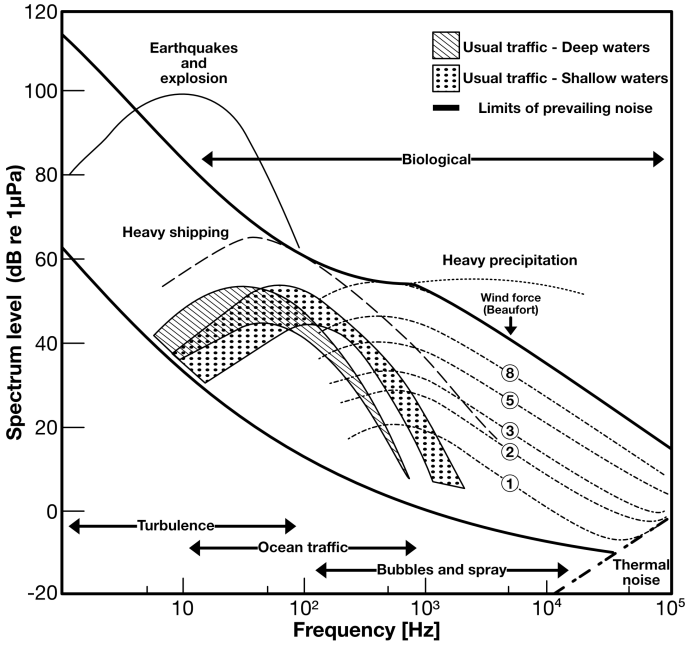


Figure 2.1: Composite of ambient noise spectra (Adapted from Wenz, 1962 [47]).

2.1.2 Noise field

Noise field mapping is an important design consideration when developing the array geometry and choosing array processing method. This mapping includes characteristics such as number of sources, movement, propagation paths, reverberation, bandwidth and type of noise field [36]. There are four types of noise fields that is differentiated based on their statistical character. The first type is called *coherent noise field* as spatially separated recordings are strongly correlated. Typically, this is the case when the radiated sound do

not experience reflections, attenuation or scattering, e.g. an open environment or anechoic chamber. The opposite situation in the case of strongly uncorrelated sensor signals is called *incoherent noise field*. An example here is spatially white sources of noise such as independent electrical sensory noise. Both coherent and incoherent noise fields are rare in nature. Environments where reverberations occur, e.g. confined spaces with walls and obstacles, will have a inherent background noise due to constructive and destructive interference of reverberated sound over longer periods of time. This is called a *diffuse noise field* because sound has been "scattered" into noise within the limits of this environment, giving it characteristics like (i) Weakly correlated sensor signals and (ii) spherical or cylindrical symmetric in energy. The last type of noise field is a combination of all of the above mentioned noise fields. Coherence from direct path propagation from a noise source to the receiving array overlapped with incoherent noise and diffuse noise from reverberation.

2.2 Fish sounds

There exists 30 thousand known species of fish to this date [44], and they hold the largest diversity of sound generation mechanisms among vertebrates [46]. Due to the wast diversity of fish species and the progressiveness of the field, new mechanisms are still being reported. In research papers, tonal fish sounds are sometimes described with words like *croaks*, *clicks*, *chirps*, *drumming*, *hissing*, *humming*, *moans*, *rasps*, *thumps*, etc. e.g. in [6]. The sounds are related to species, and subsequently, their corresponding sound generating mechanisms. Dividing the mechanisms into groups, the main vocalization method is linked with the vibration of the swimbladder. Within this group there are both direct and indirect mechanisms that produce sounds. These two variations are linked with intrinsic and extrinsic muscles respectively. Intrinsic muscles interact with the swimbladder directly, and extrinsic muscles attach to other structures such as the scull, vertebrae and ribs [13; 46].

Another group of sound generation is connected to pectoral griddle, and exterior structures as fins and tendon. For instance, rapid muscle contraction of muscles attached to the scull and cleithrum of the fish can generate sound through stridulation.

Newer suggested sound generating mechanism include pharyngeal teeth grinding [1], gill chamber water ejection [32] and anal swimbladder gas release [28; 45; 50].

A lot of the sounds mentioned have been observed as broadband signals ranging from 20 Hz-5 kHz, where most of the recordings are tied to low frequent (less than 3 kHz) sounds [6; 11; 14]. However, the scenery of the majority of those recordings has been associated with reproductive and agonistic behavior of only a handful of the species that exists. Newer preliminary studies of herring (*Clupea harengus* and *Clupea pallasii*) show that the fish is able to produce broadband signals of higher frequencies due to the mechanism of swimbladder air release in relation to pressure differences. For instance, Wahlberg and Westerberg tested how herring reacted to differential pressures in a pressure chamber, producing chirp-like sounds with center frequencies ranging from 3 kHz to 4.1 kHz [28; 45]. Wilson, Batty and Dill observed herring in captivity producing pulse trains with pulses ranging from 1.7 kHz-22 kHz [50], which is much higher than what is observed in [6]. In another study including cod (*Gadus morhua*) and Atlantic salmon (*Salmo salar*) it was concluded that the observed broadband clicks, with average center frequencies of 7

kHz, was produced by cod as the cod was approached by seals (*Pagophilus groenlandicus* and *Cystophora cristata*) or human divers [43]. It was not observed any vocal fish sounds in the test conducted with salmon. These are still single studies and more research is needed. If the sounds are communicative or not is not conclusive, at least for the herrings gas release mechanism, but it is an indication that fish are able to produce sounds in large parts of the auditory spectrum.

Little information has been found about the vocalization and sounds of Atlantic salmon. There is however reason to believe that salmon is able to produce sounds as one recording from a indoor fish tank at NINA research station Ims (Sandnes, Norway) measured chirp and click-like sounds active in the 4 kHz-10 kHz region.

2.3 Hydrophones

A hydrophone is a underwater microphone. It is called a transducer as it consequently converts one form of energy into another, specifically sound energy into electric energy. The hydrophone inhabits this property because it consists of materials that are piezoelectric or *magnetostrictive*. Piezoelectric materials are usually crystalline solids or ceramics that produces electric charges when it is exposed to mechanical stress. Magnetostrictive materials consists of small ferromagnets that induces a magnetic field when exposed to a mechanical force, called the Villari effect, that can be used to create an electric current [39]. An advantage of using piezoelectric or magnetostrictive materials in hydrophones instead of other kids of transducers common for microphones operating in air, e.g. moving-coil transducers, is that it has better acoustic impedance match in water.

The design of the hydrophone is important for its directional sensitivity. By varying the casing and placement of the piezoelectric/magnetostrictive elements within the casing, different directional sensitivity patterns emerges. Common directional sensitivity pattern found for microphones are presented in fig. 2.2, which show their cross cut directionality. A hydrophone having flat frequency response in all directions is called isotropic [36]. Isotropic sensors have omnidirectionality, as seen in the figure, in three dimensions for all frequencies making the directional sensitivity formed like a sphere. Only isotropic hydrophones will be considered in this thesis.

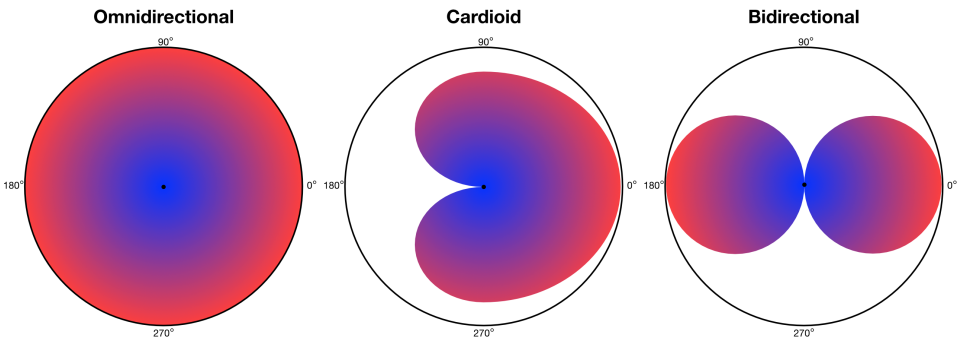


Figure 2.2: Directional sensitivity pattern for omnidirectional, cardioid and bidirectional transducer elements.

2.3.1 Array geometry

Several hydrophones may be used together to from a hydrophone array. How the hydrophones are placed relative to each other defines the array geometry. This geometry is very important since it affects the arrays listening sensitivity in different directions dependent on the array processing methodology. Several common array geometries are presented in fig. 2.3. Only the uniform linear array (ULA) will be of interest in this thesis, which gives an axisymmetrical sensitivity along the axis of hydrophone placement, i.e doughnut shaped response. However, using other types of geometries can give other types of response patterns. A back baffled planar array may give more of a spotlight formed response.

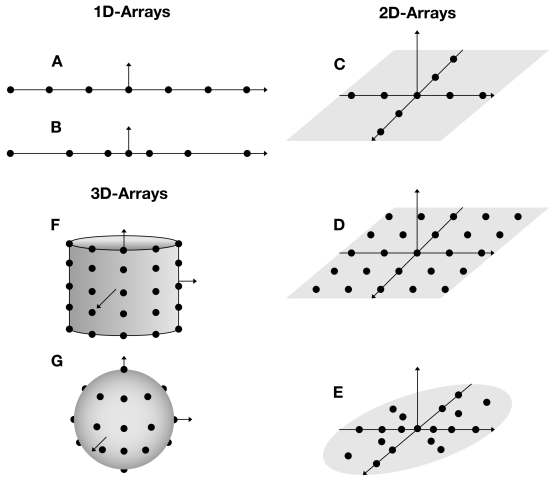


Figure 2.3: Illustration of different types of array geometries. The black dots represent the array sensors. A: Uniform linear array, B: Linear nested array, C: Planar cross array, D: Planar square array, E: Planar circular array, F: Cylindrical array, G: Sphere array.

2.4 Aliasing

Array processing involves both spatial sampling and temporal sampling of the impinging signal and is subsequently subject to both spatial and temporal aliasing. These two phenomenons will be explained and discussed in the two following subsections.

2.4.1 Temporal aliasing

Temporal aliasing occurs when a continuous time signal is sampled too rarely making the reconstruction process of the continuous time signal unambiguous. In other words, if the continuous signal is sampled at a rate that is lower than the frequency specified by the Nyquist-shannon sampling criterion it is not possible to mathematically reconstruct the original signal from the sampled signal without ambiguity [31]. The Nyquist-shannon sampling criterion is

$$\frac{f_s}{2} > f_{max} \tag{2.1}$$

where f_s is the sampling frequency and f_{max} is the highest frequency component of interest present in the continuous signal that is being discretized.

In fig. 2.4 the sampling frequency f_s is lower than the frequency of the signal being sampled resulting in aliasing. The figure also include an aliased signal reconstructed from the samples, which indicates that there is no unique way to reconstruct the signal.

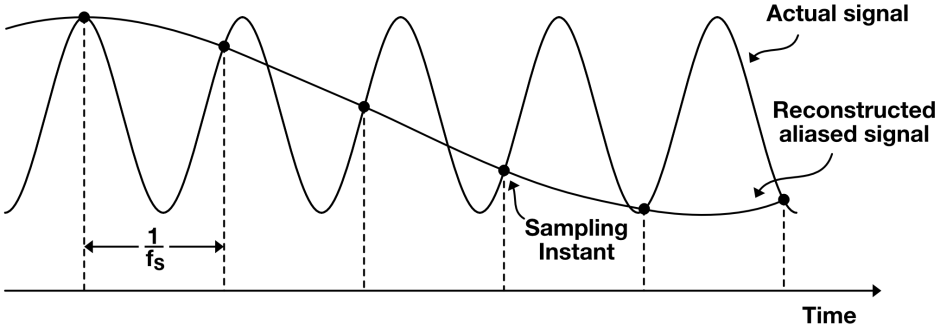


Figure 2.4: Temporal aliasing

2.4.2 Spatial aliasing

Positioning of the elements in an array is the deciding element linked to spatial aliasing. The elements should be spaced dependent on the smallest wavelength component of the signal incident on the array. For simplicity, consider an ULA where the array elements are distributed along a single direction. Figure 2.5 illustrates the case where a wavefront hits the ULA parallel to the array element axis such that the propagation delay between the elements is at maximum. The important remark in this case is that the maximum frequency component of the impinging plane wave signal will have the smallest wavelength λ_{min} of all the components in the signal and subsequently needs to be spatially properly

sampled. The spacing between the array elements, d, needs to be less than $\lambda_{min}/2$ to avoid directional ambiguity.

To draw parallels to temporal sampling imaging that the M receiving elements takes a single snapshot of the impinging wave simultaneously. The snapshot contains M samples that is sampled at different parts of the impinging wave due to the dispersion of the receivers. Analogously, the snapshot of M samples is similar to the set of temporal samples of a continuous time signal. Followingly, To be able to reconstruct the impinging wave or signal the spacing should follow the Nyquist-Shannon sampling theorem for the spatial domain which states

$$d < \frac{\lambda_{min}}{2} \tag{2.2}$$

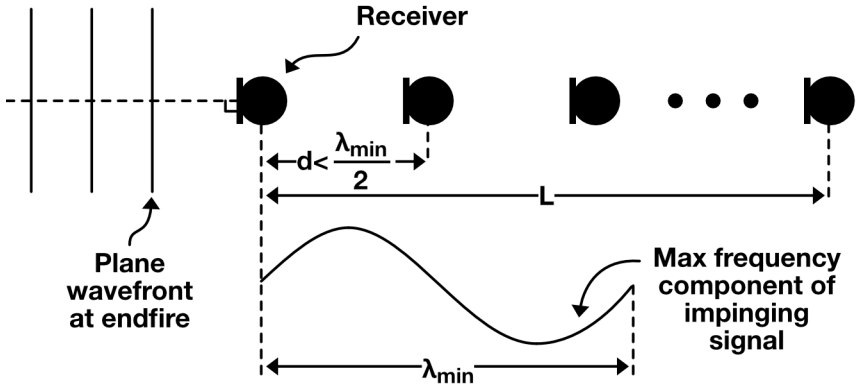


Figure 2.5: Spatial aliasing

2.5 Wave propagation model

The pressure waves in the medium can be expressed in terms of a pressure field $p(t, \mathbf{r})$ which describes the pressure in the medium at any time t and point $\mathbf{r}$ in space, and is found by solving the wave equation represented by eq. (2.3). In fact, one could use the the wave equation to find a solution for any kind of mechanical or electromagnetic wave propagationg through a medium [36].

$$\nabla^2 p(t, \mathbf{r}) = \frac{1}{c^2} \frac{\partial^2}{\partial t^2} p(t, \mathbf{r}) \tag{2.3}$$

∇^2 is the *Laplace-operator*, c is the propagation speed in the medium and $\mathbf{r}$ is the *position vector*

$$\mathbf{r} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{2.4}$$

in the Cartesian coordinate system.

Acoustic waves are spherical in nature. When sound is created at a point in space, it will normally propagate in a spherical manner away from the point source. The spherical *monochromatic* (single frequency) solution to eq. (2.3) is

$$p(t, \mathbf{r}) = \frac{A}{4\pi|\mathbf{r}|} e^{j(2\pi f t - k|\mathbf{r}|)} \tag{2.5}$$

where j denotes the imaginary unit, f is the frequency of the wave, k is called the *wavenumber* equal to $2\pi/\lambda$ where λ is the wavelength of the wave [36]. It is easy to see that the amplitude A of the spherical solution is highly dependent on the radial distance $|\mathbf{r}|$ from the sound source.

In many cases it is reasonable to assume plane wave propagation. The wave will reassemble a plane wave on the receiving array as long as the distance between the source and receiver is much larger than the physical size of the receiver. Figure 2.6 illustrates the concept as four omnidirectional point sources radiate waves forming an irregular wavefront in the near-field region (Fresnel field) before transitioning to become smooth plane waves in the far-field region (Fraunhofer field) [39]. The far-field region is assumed to start at certain distance from the source called the Fraunhofer distance, defined as eq. (2.6), which is dependent on the size/diameter L of the source and the wavelength λ of the wave at a certain frequency. E.g. a sound source with diameter 0.5 m radiating a sound at 6 kHz underwater will give a Fraunhofer distance of 2 m assuming propagation speed in water set to 1500 m/s.

$$D_{Fraunhofer} = \frac{2L^2}{\lambda} \tag{2.6}$$

Plane wave propagation is desirable when operating with hydrophone arrays since calculating with plane waves is much easier than handling near-field effects.

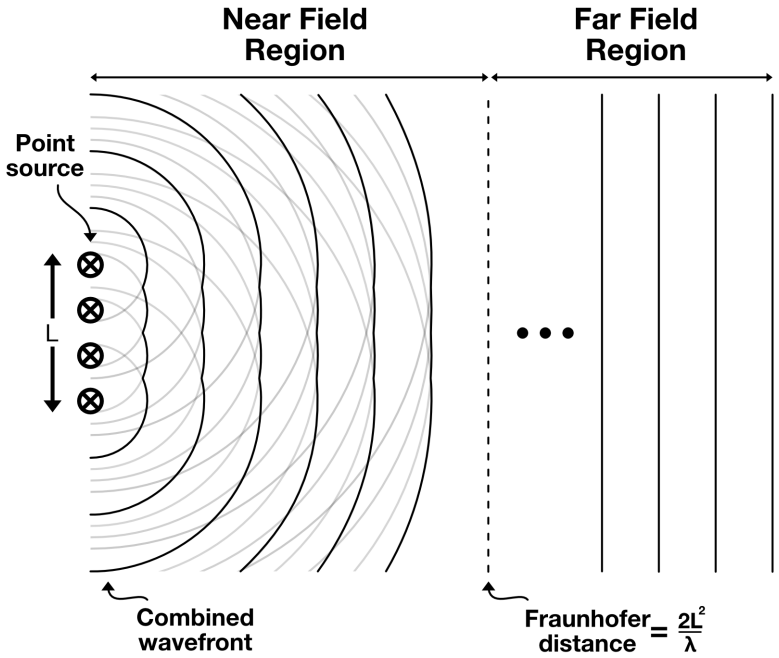


Figure 2.6: Near-field vs far-field illustration. L is the diameter of the source, and λ is the wavelength. The combined wavefronts are simplified for illustration purposes. They will experience destructive and constructive interference as well.

As for spherical propagation, one can solve the wave equation, eq. (2.3), to find the space-time representation of the pressure field due to plane wave propagation. Since the wave fronts in this case is plane and perpendicular to the direction of propagation, the direction of propagation must be specified explicitly. In the monochromatic plane wave case the Cartesian coordinate solution becomes

$$p(t, \mathbf{r}) = Ae^{j(2\pi ft - \mathbf{k}\cdot\mathbf{r})} \tag{2.7}$$

where $\mathbf{r}$ is the position in space defined as eq. (2.4) and $\mathbf{k}$ is the *wavenumber vector* defined as

$$\mathbf{k} = \frac{2\pi}{\lambda}\begin{bmatrix}\cos(\theta)\cos(\phi)\\ \cos(\theta)\sin(\phi)\\ \sin(\theta)\end{bmatrix} = k\mathbf{d}(\Theta) \tag{2.8}$$

which gives speed and direction to the wave propagation. Here, $\lambda = c/f$ is the wavelength, the unit vector $\mathbf{d}(\Theta)$ is the *direction of arrival* (DOA) vector (assuming signal reception) and $\Theta = (\phi, \theta)$ is the incident angle represented by azimuth and elevation angles following the definition illustrated by fig. 2.7.

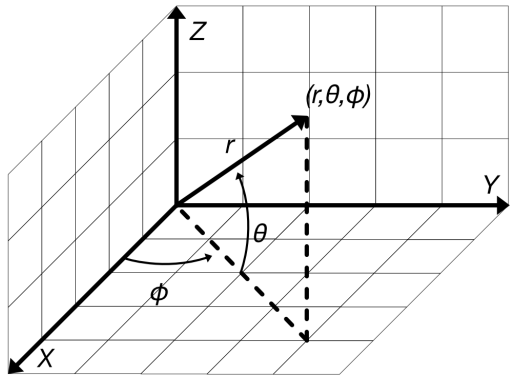


Figure 2.7: Spherical coordinates using azimuth (ϕ) and elevation (θ) angles. Notice that when $\theta = 0$ implies that the position vector is perpendicular to the z-axis rather than parallel to the z-axis for regular spherical coordinates.

Consider now an array that receives a plane wave in the far-field omitting noise. The m'th array element will experience the signal

$$p(t, \mathbf{r}) \mid_{\mathbf{r}=\mathbf{r}_m} = Ae^{j(2\pi f t - \mathbf{k}\cdot\mathbf{r}_m)} \tag{2.9}$$

calculated from eq. (2.7) and eq. (2.3) and the fixed spatial position of the receiver element $\mathbf{r}_m = \begin{bmatrix} x_m & y_m & z_m \end{bmatrix}^T$. It is common to use one of the array elements as a reference point indicating the origin of the reference frame, e.g. $r_1 = [0, 0, 0]^T$. Further on, as a wave is impinging on the array from a fixed direction $\Theta_0 = (\phi_0, \theta_0)$ propagation delays will be present due to spatial differences between the array elements. The m'th propagation delay relative to the reference point (origin) becomes

$$\tau_m = \frac{1}{c}\mathbf{d}(\Theta_0)\mathbf{r}_m \tag{2.10}$$

Assume that a source transmits a unmodulated signal of a single frequency f_s with unit amplitude, i.e. $p(t) = e^{j2\pi f_s t}$. Inserting this as well as eq. (2.10) into eq. (2.9) implies that the signal impinging on the m'th element becomes

$$p(t - \tau_m) = p(t)e^{-j2\pi f_s \tau_m} \tag{2.11}$$

2.5.1 Steering vector

Continuing the discussion of the wave equation solution at the m'th array element presented in eq. (2.9), the phasor part $e^{-j\mathbf{k}\cdot\mathbf{r}_m}$ of the solution contains the geometric information about the wave impinging on the m'th array element. Assume now that the array contains M sensors, and experiences a monochromatic wave from a single radiation source. Every array element will experience the wave differently due to their unique element positions, hence their phasors will be different. Equation (2.12) is the collection of the phasors

in a single vector called the *steering vector* or *direction vector* [36].

$$\mathbf{v}(\mathbf{k}) = \begin{bmatrix} e^{-j\mathbf{k}\cdot\mathbf{r}_1} \\ e^{-j\mathbf{k}\cdot\mathbf{r}_2} \\ \vdots \\ e^{-j\mathbf{k}\cdot\mathbf{r}_M} \end{bmatrix} \tag{2.12}$$

This vector represent the phase delays that a plane wave experiences at the receiver array with respect to an arbitrary reference point.

The reception of signals for an array of M spatially spaced receiver elements exposed to a signal $p(t)$ can be written as

$$\mathbf{x}(t) = p(t)\mathbf{v}(\mathbf{k}) \tag{2.13}$$

defined by eq. (2.9) and eq. (2.12) where

$$\mathbf{x}(t) = \begin{bmatrix} p(t,\mathbf{r}) \mid_{\mathbf{r}=\mathbf{r}_1} \\ p(t,\mathbf{r}) \mid_{\mathbf{r}=\mathbf{r}_2} \\ \vdots \\ p(t,\mathbf{r}) \mid_{\mathbf{r}=\mathbf{r}_3} \end{bmatrix} \tag{2.14}$$

Do not confuse $\mathbf{x}(t)$ with the cartesian x-axis position in eq. (2.4). It should also be noted that this representation of the steering vector is only valid for isotropic sensors. If non-isotropic sensors are of interest, then the steering vector is a product of the steering vector specified in eq. (2.12) and the sensors radiation diagram, which is a representation of the sensors inherent sensitivity in different directions (which simply is 1 for isotropic sensors) [36].

One way of thinking about the purpose of the steering vector is that it aligns the source signal at the receiver end by compensating for propagation delay. Remember that the wavenumber vector $\mathbf{k}$ depends on DOA, frequency (or conversely wavelength) and propagation speed. If these values are known, they can simply be inserted into eq. (2.13) to produce M in-phase receiver signals. The concept of time adjusting signals using steering vectors is simply called *steering*.

2.6 Signal model

The receiving array converts continuous space-time information that is carried by propagating waves into discretized time signals coupled with spatial positions. To describe the signal model some assumptions are made:

1. Plane wave propagation.
2. Isotropic receivers.
3. Propagation channel/environment is imagined to be anechoic or free-field with no reflections.

4. Independent and different measured noise at each array element.

5. Linear propagation medium.

In an anechoic environment, only the direct path between the source and receiver is considered. In such environment the signal model for the m'th array element is

$$x_m(t) = p(t, \mathbf{r}) \mid_{\mathbf{r}=\mathbf{r}_m} + n_m(t) \tag{2.15}$$
$$= v_m(\mathbf{k})p(t) + n_m(t) \tag{2.16}$$

where v_m is the steering of the m'th array element, $p(t)$ is the signal radiated from the source, and $n_m(t)$ is the measured noise following the stated assumptions above. In generalized vector format the received array signals are

$$\mathbf{x}(t) = \mathbf{v}(\mathbf{k})p(t) + \mathbf{n}(t) \tag{2.17}$$

where $\mathbf{x}(t) = [x_1(t), \ldots, x_M(t)]^T$ and $\mathbf{n}(t) = [n_1(t), \ldots, n_M(t)]^T$.

As the propagation medium is water in this thesis, the propagation is considered linear if it has steady pressure, temperature and salinity. Thus, several signals from N different sources can be added together using the *superposition principle* [36].

$$\mathbf{x}(t) = \sum_{i=1}^{N} \mathbf{v}_i(\mathbf{k})p_i(t) + \mathbf{n}(t) \tag{2.18}$$

or in vector format

$$\mathbf{x}(t) = \mathbf{V}\mathbf{p}(t) + \mathbf{n}(t) \tag{2.19}$$

where $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_N] \in \mathbb{C}^{M \times N}$ is the *steering matrix*. As N spatially separated signal sources radiate signals towards the M array elements following the model hypothesis stated above it is evident from eq. (2.9), eq. (2.10) and eq. (2.11) that the propagation delay present at the m'th sensor for the k'th signal source is

$$\tau_{m,k} = \frac{1}{c}\mathbf{d}(\Theta_k)\mathbf{r}_m \tag{2.20}$$

and hence

$$x_m(t) = \sum_{i=1}^{N} p_i(t - \tau_{m,i}) + n_m(t), \quad m = 1, \ldots, M \tag{2.21}$$

Thus, the impulse response between the source and receiver is simply

$$v_{m,k} = \delta(t - \tau_{m,k}) \tag{2.22}$$

given the anechoic model for plane wave propagation. In other words, the steering vector models the propagation delays at the array in this case.

2.6.1 Spatial covariance

Signals received at each array element are not equal. As seen in eq. (2.17) and eq. (2.19) the receiver signals $\mathbf{x}(t)$ consists of *signal of interest* (SOI) $p(t)$ and noise $\mathbf{n}(t)$. The sensors will experience different noise and interference due to the spatial and temporal characteristics of the surrounding noise field as well as experienced self noise. The SOI will also differ in time and power because of the spatial differences and propagation degradation. Defining statistical space-time relations is consequently useful for theoretical analysis.

Spatial covariance is a measure of joint variability of spatially separated signals [15]. Variability for two signals can be represented in a square matrix $\mathbf{R}_{xx} \triangleq \mathbb{E}\{\mathbf{x}\mathbf{x}^H\}$ called the *spatial covariance matrix*. The superscript H denotes the Hermitian transpose for complex conjugate numbers. Taking the multi-signal generic signal model, eq. (2.19), into consideration, the spatial covariance matrix becomes

$$\mathbf{R}_{xx} = \mathbb{E}\{\mathbf{x}\mathbf{x}^H\} \tag{2.23}$$

$$= \mathbf{V}\mathbf{R}_{pp}\mathbf{V}^H + \mathbf{R}_{nn} \tag{2.24}$$

where $\mathbf{R}_{pp} = \mathbb{E}\{\mathbf{p}\mathbf{p}^H\} \in (\mathbb{R}, \mathbb{C})^{N\times N}$ is the source signal covariance matrix and $\mathbf{R}_{nn} = \mathbb{E}\{\mathbf{n}\mathbf{n}^H\} \in (\mathbb{R}, \mathbb{C})^{N\times N}$ is the noise covariance matrix. A single signal source implies that $N = 1$ assuming anechoic modeling. It follows that $\mathbf{V} = \mathbf{v}(\mathbf{k})$ in eq. (2.23).

In many cases it is convenient to analyse signals in the frequency domain. Followingly, the frequency domain analogue to the spatial covariance matrix is the *power spectral density* (PSD) matrix

$$\mathbf{R}_{xx}(j\Omega) = \begin{bmatrix} R_{x_1x_1}(j\Omega) & \dots & R_{x_1x_M}(j\Omega) \\ \vdots & \ddots & \vdots \\ R_{x_Mx_1}(j\Omega) & \dots & R_{x_Mx_M}(j\Omega) \end{bmatrix} \tag{2.25}$$

2.7 Time shifting

High quality space-time array processing requires the ability of shifting signals in time. It is convenient to use *time-shift operators*, Δ, to shift a signal by a certain amount of time, τ. The continuous time signal expression is given as

$$s(t+\tau) = \Delta(\tau)s(t) \tag{2.26}$$

2.7.1 Unit delay

The simplest form of time shifting a discrete signal is by shifting it by a defined amount of units or samples. In this case, the time-shift operator will be an column shifted identity matrix on Toeplitz form [29] where the column shift is specified by the number of samples to delay the signal with. E.g. in the simplest case where the time shift $\tau = 0$ implies that the time-shift operator Δ is the identity matrix. Followingly a time shift of one sample

gives the time-shift operator

$$\Delta(1) = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & \dots \\ 0 & 0 & 0 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{2.27}$$

Unit delaying requires that the τ is an integer since the signal is discretized with finite amount of units/samples. The continuous expression in eq. (2.26) in discrete form is

$$s[n + \tau_n] = \Delta(\tau_n)s[n] \tag{2.28}$$

where n subscript of τ_n denotes discrete time shift. For mathematical convenience $\Delta(\tau_n)$ should be an unitary matrix such that the inverse is simply the complex conjugate, thus $\Delta(\tau_n) \in \mathbb{R}^{N \times N}$ where $s[n]$ has N samples.

When implementing this in code, it is better to just prepend or append zeros to the start or end of the digital signal sequence to avoid memory issues and reduce computational time. Large signal sequences requires a unit time-shift operator that is of squared size as the signal sequence it self.

2.7.2 Fractional delay

Assume that the time difference between reception at two array elements is 0.34 milliseconds and that the signals are sampled at frequency 42 000 Hz. This corresponds to 0.34 ms$\times$42000 Hz $=$16.32 samples which is a fractional number of samples. In other words, by performing unit time-shift on this signal, there will still be 0.32 seconds time difference between the signals. Fractional time shifting accounts for this fractional time difference.

Based on a topic touched upon in section 2.4.1, the Nyquist-Shannon sampling theorem states that it is possible to recreate the original continuous signal from a sampled version of the signal by multiplying each sample by a scaled sinc-function [31]. For this to be true, the original signal needs to have an upper bandlimit that is less than half of the sampling frequency. The fractional delaying of the signal can then be done by ”reconstructing” the signal by *finite inpulse response* (FIR) filtering it using a shifted and windowed sinc function to evaluate the filter coefficient. Figure 2.8 shows a 0.25 fractionally shifted sinc function. The FIR filter coefficient that is used to fractionally shift a desired signal (through the process of FIR filtering) is marked with black dots at every whole integer along the horizontal axis. However, there is a problem that the sinc is a continuous function of infinite length because the FIR filter is finite in length. The solution to this is to use a windowing function e.g. the Hamming window that such that only a portion of the impulse response is used [40].

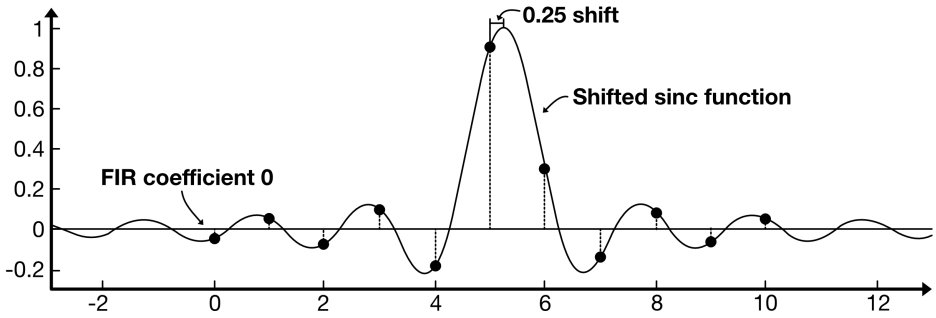


Figure 2.8: Sinc function that is fractionally shifted by 0.25 samples.

2.8 Beamforming

At this point the propagating space-time signals have reached the array and have resulted in multiple array signals that have differences both in time and content, but still contain SOI correlating features. Representing the received signals as time-shifted versions of the original radiated signal in addition to noise and interference, as in eq. (2.21), the underlying problem of *beamforming* emerges.

Beamforming is the name given to a set of array processing algorithms that steers the listening direction (time shift signals) of the receiving array and tries to increase the *signal-to-noise ratio* (SNR) using a variety of processing architectures. Common for all beamformers is that they (hopefully) produce a single enhanced signal from several sensor inputs.

Since the main beamforming problem is highly associated with propagation delay at the receivers it is natural to process the data in time domain. Time delaying can be handled using tools like time shift operators and FIR filtering, see section 2.7. Time delaying in time domain is analogous to a single phase shift in frequency domain assuming signals of single frequency. Hence, most time domain beamforming techniques has a frequency domain counterpart. Phase shift is however frequency dependent, and since acoustic signals tend to consist of multiple frequencies subband processing is needed to compensate for the phase shift difference at different frequencies.

Only time domain beamformers will be discussed in the remainder of this thesis.

2.8.1 Narrowband vs wideband beamforming

The choice of beamformer technique highly depends on the nature of the SOI. In some cases the signal of interest only contain a single frequency or a narrow range of frequencies, i.e. the signal has narrow *bandwidth* and are called *narrowband* signals. In the opposite situation where the signal bandwidth includes a wide range of frequencies, it is called *broadband* or *wideband*. The two different signal types determines what kind of processing structure is needed. Narrowband beamforming tends to have a simpler structure than wideband beamforming because narrowband signals do not require the same temporal (or frequency) discrimination of the signal like wideband beamforming do. The effect this

has on the processing structure is illustrated in fig. 2.9. Here, narrowband beamforming is illustrated as a complex linear combination of the aligned array receiver signals, while wideband beamforming has a FIR filter, sometimes called *Tapped delay line*, operating in the time domain. The two different structures are sometimes referred to as *delay-and-sum beamforming* (DAS) and *filter-and-sum beamforming* (FAS) respectively. The weighting for both of the structures follows the chosen design ideology which is discussed when the different beamformers are introduced.

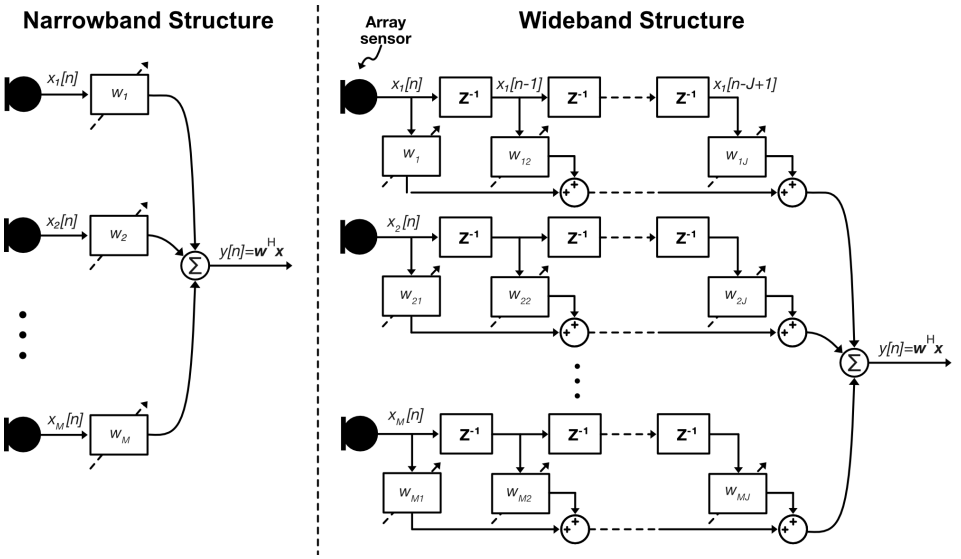


Figure 2.9: Beamforming structure of narrowband and wideband signals. (Left) Narrowband delay-and-sum beamforming structure showing that the output is a complex linear combination of M aligned receiver signals. (Right) Wideband filter-and-sum beamforming structure showing that the output signal is a sum of M FIR-filtered aligned signals. The dashed arrows over the weights notify that the weights may be adaptive.

2.8.2 Delay-and-sum beamformer

The simplest beamformer is called delay-and-sum beamformer because that is exactly what it does with the incoming signals. It is a simple narrowband beamformer that goes under different names such as *delay-and-weighted-sum beamformer*, *narrowband conventional beamformer*, the *Barlett beamformer* or commonly just the abbreviation DAS. A simple illustration of the working principle of this beamformer is given in fig. 2.10. The incoming signal $p(t, \mathbf{r})$ hits the array elements at different times which means that the information in the signal will be delayed between the array elements. To compensate for this delay the beamformer time shifts each signal $p_i(t)$ (where i denotes the i'th receiver signal) by the amount of time, τ_i, the plane wave uses to travel to the different array elements relative to a reference point, which usually is one of the elements, before summing the weighted versions of the delayed signals creating the output signal $y(t)$.

The DAS beamformer can be summarised by eq. (2.29). The weights w_i are fixed and

are called *shading weights* as they adjust the impact of the signals from different array elements and hence modifies the spatial sensitivity of the beamformer.

$$y(t) = \sum_{i=1}^{M} w_i x_i \tag{2.29}$$

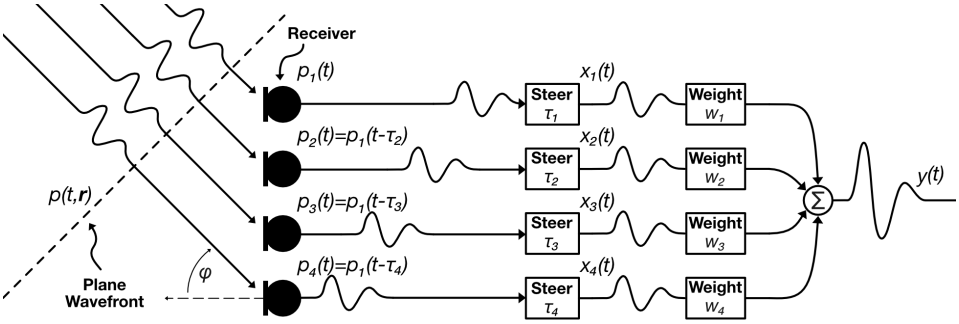


Figure 2.10: Block diagram of the delay-and-sum beamformer. The reference frame is placed at the top receiver. Only the SOI, $p(t)$, component taken into account here to emphasize the working method.

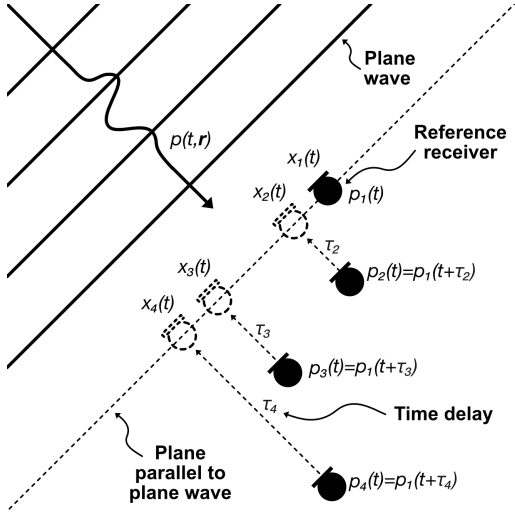


Figure 2.11: Alternative visualization of the Delay-and-sum beamformers working method. If propagation attenuation and noise is neglected and the plane wave assumption holds then steering of the array is analogous to artificially moving the array elements from their respective positions to a plane parallel to the incoming plane wave.

As described in section 2.5.1, the concept of steering involves time shifting of the receiver signals. A way to spatially visualize how array steering works is to think of time shifting as a way to map the corresponding array element position onto a plane parallel to the incoming plane wave (neglecting noise and attenuation from propagation), see fig. 2.11. Intuitively, summation of signals collected by array elements that are positioned in a plane parallel to the incoming plane wave will result in a signal with increased SNR since the aligned information in the signals add constructively while the noise add destructively (if the noise is weakly correlated).

It is important to note that the way the signal is received by the array elements differs from each other due to propagation attenuation differences and noise. Put differently, the i'th array element signal , $x_i(t)$, is not equal to the j'th array element time adjusted signal, $x_j(t + \Delta_{ij})$, where Δ_{ij} is the reception time difference between the array elements. But, the time shift is important for aligning the signals before adding them together, as shown in fig fig. 2.12.

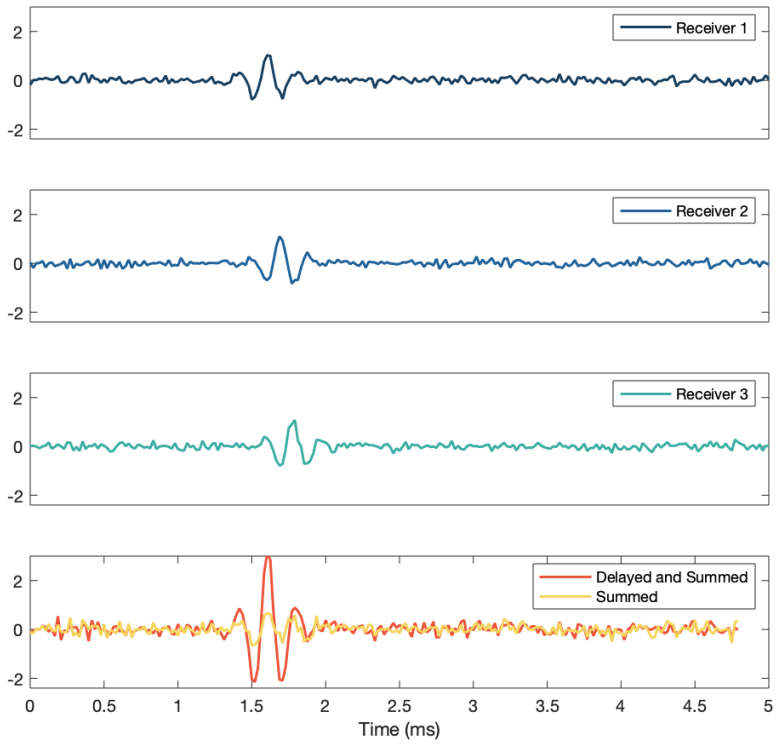


Figure 2.12: This figure shows how three receiver signals contain the same shifted information due to spatial differences between the receivers. Time adjusted signals will add constructively, while no time adjustments will add destructively.

To elaborate on the DAS beamformer assume that the SOI impinging on the m'th array element is the single frequency monochromatic solution of the wave equation, as in eq. (2.7). In other word, the input signal is a pure frequency signal

$$p_m(t) = A_m e^{j(2\pi f_s t - \mathbf{k}\cdot\mathbf{r}_m)} \tag{2.30}$$

Where f_s is the signal frequency, A_m and $\mathbf{r}_m$ are the amplitude and position specific for the m'th element respectively. The DAS beamformer delays the signal by τ_m due to the relative position to the pre-picked reference point. Accordingly, the time shifted signal without noise can be expressed as

$$x_m(t) = p_m(t - \tau_m) = p(t)e^{-j2\pi f_s \tau_m} \tag{2.31}$$

where $p(t) = e^{j2\pi f_s t}$ is the signal impinging on the reference point. It follows from eq. (2.29) that the output becomes

$$y(t) = \sum_{i=0}^{M} w_i p(t) e^{-j2\pi f_s \tau_i} \tag{2.32}$$

arranging the components as vectors, i.e.

$$\mathbf{w} = \begin{bmatrix} w_1 & w_2 & \dots & w_M \end{bmatrix}^T \tag{2.33a}$$

$$\mathbf{x}(t) = \begin{bmatrix} p(t)e^{-j2\pi f_s \tau_1} & s(t)e^{-j2\pi f_s \tau_2} & \dots & p(t)e^{-j2\pi f_s \tau_M} \end{bmatrix}^T \tag{2.33b}$$

The vector format of eq. (2.32) becomes

$$y(t) = \mathbf{w}^H \mathbf{x}(t) \tag{2.34}$$

Since the array element position vector $\mathbf{r}$ is know a priori, and assuming the array is steered in a pre defined look direction $\Theta_0 = (\phi_0, \theta_0)$ it is possible to calculate the zero-mean power, or variance, of the output signal as

$$P(y) = \mathbb{E}[|y(t)|^2] = \sigma_y^2 = \mathbf{w}^H \mathbf{R}_{\hat{x}\hat{x}} \mathbf{w} \tag{2.35}$$

where $\mathbf{R}_{xx} = \mathbb{E}[\mathbf{x}(t)\mathbf{x}^H(t)]$ is the correlation matrix of the element signals. Sometimes it is convenient to represent the output power as a function of the steering vector, i.e the *steered array response*. By looking at fig. 2.10 it is easy to see that it is possible to just combine the steering blocks and the weights by multiplying it together. If the shading weights are 1, then the steered response is simply

$$P(y) = \mathbb{E}[|y(t)|^2] = \sigma_y^2 = \mathbf{v}^H \mathbf{R}_{\hat{x}\hat{x}} \mathbf{v} \tag{2.36}$$

where $\mathbf{v}$ is the steering vector.

2.8.3 Filter-and-sum beamformer

The Filter-and-sum (FAS) beamformer is the wideband analogue of DAS, and is sometimes called the *wideband conventional beamformer*. The only difference between the DAS and FAS beamforming is that the shading weights are traded with fixed FIR filters. When a FIR filter is associated with each receiver signal, it is possible to predefine the frequency range of interest. E.g. letting each FIR filter correspond to a bandpass filter allowing frequencies in a certain range to pass before summing all the filtered receiver signals together. The FAS structure is presented in fig. 2.13.

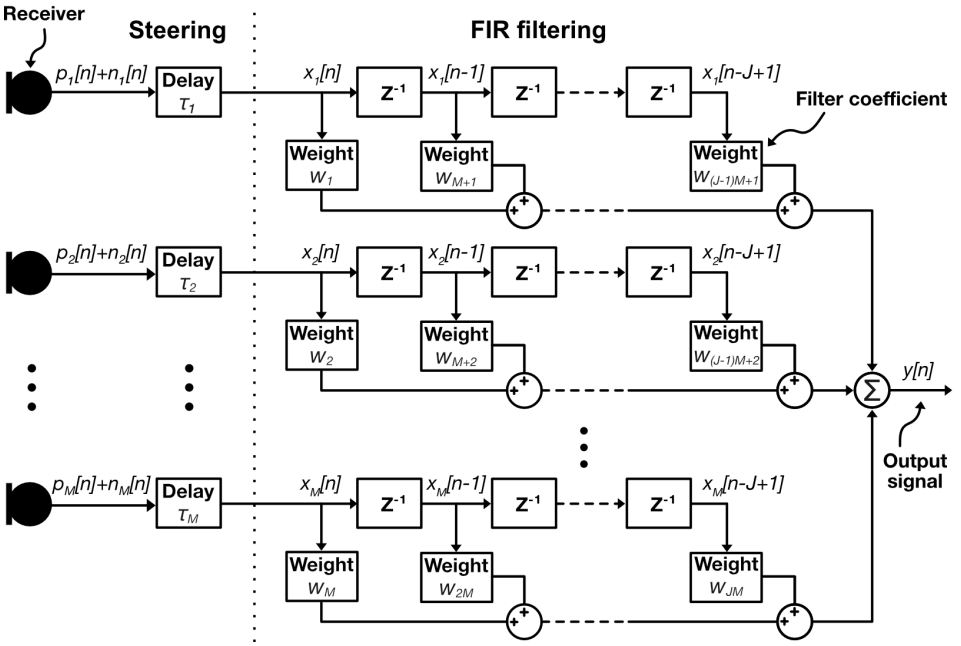


Figure 2.13: Block diagram of the filter-and-sum beamformer. Each receiver signal is first time delayed and then FIR filtered before summation.

The discretized beamformer output is simply calculated as

$$y[n] = \mathbf{w}^H \mathbf{X}[n] \tag{2.37}$$

where $\mathbf{w}$ holds the filter weights/coefficients, and $\mathbf{X}[n]$ are the filter *tap voltages* at the n'th sample of the discretized steered receiver signals. A tap voltage is the name for the

value at a certain filter tap. Explicitly noted, the weight vector and tap voltage vector are

$$\mathbf{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_M \\ w_{M+1} \\ \vdots \\ w_{2M} \\ \vdots \\ w_{(J-1)M+1} \\ \vdots \\ w_{JM} \end{bmatrix}, \mathbf{X}[n] = \begin{bmatrix} x_1[n] \\ \vdots \\ x_M[n] \\ x_1[n-1] \\ \vdots \\ x_M[n-1] \\ \vdots \\ x_1[n-J+1] \\ \vdots \\ x_M[n-J+1] \end{bmatrix} \tag{2.38}$$

Here, $x_m[n-j]$ indicates the (j+1)'th FIR filter tap voltage for the n'th sample of the m'th steered receiver signal. Assuming that there is M array elements, the weight matrix $\mathbf{w}$ is arranged such that the M first entries is the first vertical column of filter weights in fig. 2.13, the next M entries are the second column and so forth. By arranging the weight like this has some structural advantage in the following subsections section 2.8.4 and section 2.8.5. However, when implementing the FAS it is probably more useful to just do a discrete convolution of each signal since the filters have fixed weights.

The FAS structure has $M \times J$ free coefficients when M is the number of array elements and J is the filter length or number of filter coefficients. Setting these coefficient can be done in several different ways, but they all involve building a filter (lowpass, highpass or bandpass) using windowing.

The FAS algorithm is as follows

Algorithm 1: FAS Beamforming

Result: Beamformed signal y
Align signals $\mathbf{x}(t) = \mathbf{p}(t-\tau)$;
Initialize weights filters $\mathbf{w}$;
Filter signals $\mathbf{x}(t)$;
Sum signals to obtain output y;

2.8.4 Linearly constrained minimum variance beamformer

In this section an adaptive beamforming approach proposed by Frost [8] is described. The method presents beamforming as a linearly constrained optimization problem for wideband signals with unknown statistics. What this means in practise is that the method tries to optimize the SNR of the output signal while having an opinion on the desired frequency response in terms of constraints on the adaptive weights. The general processing technique introduced by Frost is suitingly called *linearly constrained minimum variance* (LCMV) wideband beamforming.

Before jumping into the mathematics, it can be handy to have the LCMV structure fresh in mind. Figure 2.14 shows that the LCMV beamformer consists of two stages. The first stage steers the array response towards the listening direction by delaying the

signals just like for the DAS and FAS beamformer. When the signals are aligned they enters the FIR filtering stage where each signal passes a adaptive FIR filter before being summed together. The structure itself reassembles the FAS structure, but includes the adaptive weights that are constrained. The clever part of this algorithm is how the adaptive weights are updated. From the same figure, fig. 2.14, it is apparent that the beamformer can be summarised in an equivalent processor with fixed weights stated as the (optimization problem) constraints. Each vertical column of LCMV weights are constrained to a corresponding fixed FIR filter coefficient of the equivalent processor which is designed a priori with desired frequency response for the specified listening direction. These constraints are highlighted in the bottom part as the FIR filter constraints.

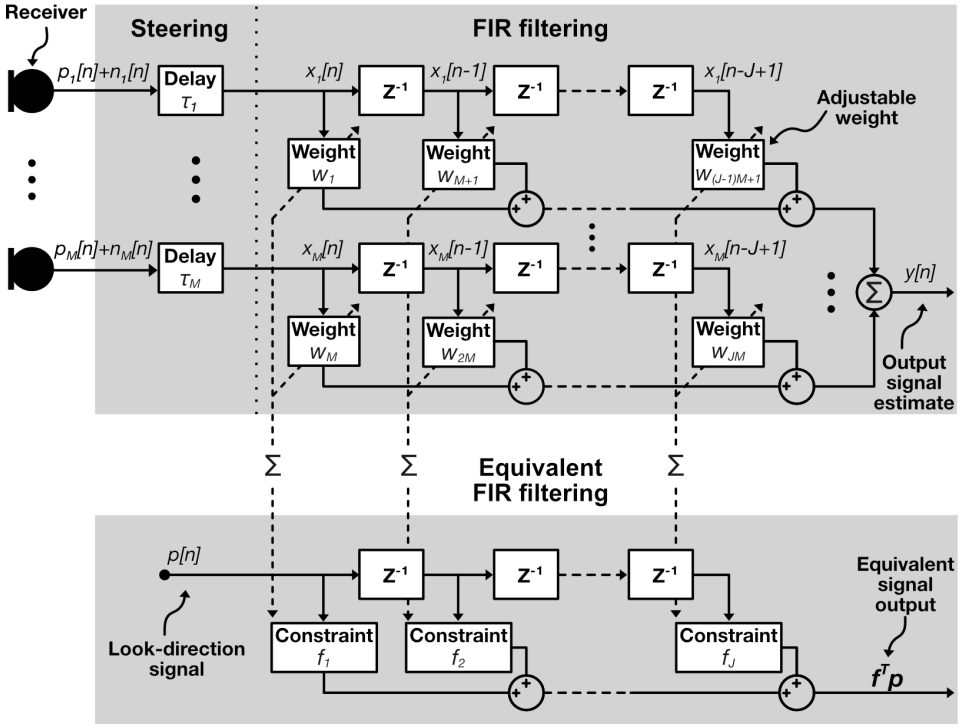


Figure 2.14: (Top) Wideband linearly constrained minimum variance beamformer processor structure and (Bottom) equivalent representation imposed as tapped delay constraints for signals coming from the look direction.

In Frost's paper [8] there is no steering stage involved in the description as the input signals were assumed in-phase. In this thesis discrete in-phase signals are denoted $x_m[n]$ where the subscript m indicate that it is from the m'th receiver and n indicate the n'th sample of the discrete signal. The output of Frost's beamformer can be identically expressed as the output of the FAS, eq. (2.37), with the same tap voltage vector ($\mathbf{X}[n]$) and weight vector ($\mathbf{w}$) as in eq. (2.38). Subsequently, these weights are updated adaptively, or in other words, the weights are adjusted based on the output value for the most recent input values

using a optimization criteria.

The adaptiveness is confined to follow the a priori constraints

$$\mathbf{f} = \begin{bmatrix} f_1 & f_2 & \dots & f_J \end{bmatrix}^T \tag{2.39}$$

The sum of each vertical column of filter weights in fig. 2.14 are constrained by f_j where j denotes the j'th column as presented by eq. (2.40) where M is the number of input signals.

$$f_j = \sum_{i=(j-1)M+1}^{jM} w_i, \quad \text{for } 1 \leq j \leq J \tag{2.40}$$

By introducing a matrix $\mathbf{C}$ on the form

$$\mathbf{C}^H = \begin{bmatrix} 1 & \dots & 1 & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 & \dots & 0 & \dots & 0 \\ & \vdots & & & \vdots & & \ddots & & \vdots & \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & 1 & \dots & 1 \end{bmatrix} \in \mathbb{R}^{J \times JM} \tag{2.41a}$$

Or expressed with the Kronecker product $\otimes$

$$\mathbf{C}^H = \mathbf{I}_{J \times J} \otimes \mathbf{1}_{1 \times M} \tag{2.41b}$$

it is possible to write the constraints as

$$\mathbf{f} = \mathbf{C}^H \mathbf{w} \tag{2.42}$$

where $\mathbf{I}_{i \times j} \in \mathbb{R}^{i \times j}$ represent the identity matrix, $\mathbf{1}_{1 \times j} \in \mathbb{R}^{1 \times j}$ is a vector of ones, and J and M is the FIR filter length and number of input signals respectively.

The constraining methodology goes under the name *constrained power minimization* as it assigns a specific filter frequency response to the listening-direction. E.g. $\mathbf{f}$ could be a bandpass filter. The adaptive algorithm will assign (complex) values to the filter weights such that the variance (or power if zero-mean signals) along the non-listening-directions is minimized. Denoting the error function $e[n]$ as

$$e[n] = d_\Theta[n] - y[n] \tag{2.43}$$

which describes the relationship between the desired output $d_\Theta[n]$ and the actual output $y[n]$. Minimizing the non-listening-direction signal power and maximizing listening-direction signal power leads to desired output

$$d_\Theta[n] = \begin{cases} 0 & \Theta \neq \Theta_{LD}, \\ max & \Theta = \Theta_{LD}. \end{cases} \tag{2.44}$$

In the case of non-listening-direction, $\Theta \neq \Theta_{LD}$, $d_\Theta = 0$ and hence minimizing the error variance is equivalent with minimizing the output variance of the beamformer in the non-look-directions. Using the definitions of zero mean variance on the error function

$$\mathbb{E}[|e[n]|^2] = \mathbb{E}[|y[n]|^2] = \mathbf{w}^H \mathbb{E}[\mathbf{X}[n]\mathbf{X}^H[n]]\mathbf{w} = \mathbf{w}^H \mathbf{R}_{XX} \mathbf{w} \tag{2.45}$$

which leads to the LCMV optimization problem

$$\begin{aligned} \underset{\mathbf{w}}{\arg\min} \quad & \mathbf{w}^H \mathbf{R}_{XX} \mathbf{w} \\ \text{subject to} \quad & \mathbf{C}^H \mathbf{w} = \mathbf{f} \end{aligned} \tag{2.46}$$

The optimal solution to eq. (2.46) can be found using Lagrange multiplier method [8; 26; 36] giving the optimal nontrivial solution

$$\mathbf{w}_{\text{opt}} = \frac{\mathbf{R}_{XX}^{-1}\mathbf{C}}{\mathbf{C}^H \mathbf{R}_{XX}^{-1}\mathbf{C}}\mathbf{f} \tag{2.47}$$

To be able to use the Frost LCMV algorithm in practice a recursive updating scheme should be at hand. The following derivation skips some steps, see [8] for thorough derivation. Stating the recursive weight updating scheme as

$$\mathbf{w}[n+1] = \mathbf{w}[n] - \mu \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \lambda) \tag{2.48}$$

$$= \mathbf{w}[n] - \mu[\mathbf{R}_{XX}\mathbf{w}[n] + \mathbf{C}\lambda] \tag{2.49}$$

where $\mathcal{L}(\mathbf{w}, \lambda)$ is the Lagrange function with Lagrange multipliers λ, and μ is the tunable *learning rate*, or adaptation step, that determines the step size towards the local minima of the loss function in the optimization problem. Inserting eq. (2.48) into the constraint definition eq. (2.42) and rearranging gives

$$\mathbf{w}[n+1] = \mathbf{P}(\mathbf{w}[n] - \mu \mathbf{R}_{XX}\mathbf{w}[n]) + \mathbf{g} \tag{2.50}$$

where $\mathbf{P}$ and $\mathbf{g}$ are the *projection operator* and *quiescent vector* respectively and defined as

$$\mathbf{P} = \mathbf{I} - \mathbf{C}(\mathbf{C}^H\mathbf{C})^{-1}\mathbf{C} \in (\mathbb{R}, \mathbb{C})^{MJ \times MJ} \tag{2.51}$$

$$\mathbf{g} = \mathbf{C}(\mathbf{C}^H\mathbf{C})^{-1}\mathbf{f} \in (\mathbb{R}, \mathbb{C})^{MJ \times 1} \tag{2.52}$$

A square matrix $\mathbf{P}$ is a projection operator/matrix if, and only if, the square of the matrix results in the matrix itself, i.e $\mathbf{P}^2 = \mathbf{P}$ [36]. Projection operators are often associated with pseudoinverses of matrices. The quiescent vector relates to the projection matrix as it is the image space of $\mathbf{C}$. In other words, both the projection operator and quiescent vector are merely helping matrices, independent of $\mathbf{R}_{XX}$, that reduces computation complexity since they can be defined a priori.

In eq. (2.50) $\mathbf{R}_{XX}$ is assumed known a priori. This is usually not the case in real life situations, thus an approximation of the covariance matrix is desired. A simple approximation at the n'th recursive iteration is simply $\mathbf{R}_{XX} \approx \mathbf{X}[n]\mathbf{X}^H[n]$ i.e. the dot product of the tap voltage values at the n'th instant. Using this approximation turns the weight updating scheme eq. (2.50) into

$$\mathbf{w}[n+1] = \mathbf{P}(\mathbf{w}[n] - \mu y[n]\mathbf{X}[n]) + \mathbf{g} \tag{2.53}$$

Code implementation of this requires initialization of $\mathbf{w}[0]$, which can not be a zero-vector since this gives a trivial solution to eq. (2.46). The weight initialization proposed by [8] is $\mathbf{w}[0] = \mathbf{g}$. Thus the algorithm can be summarised as

Algorithm 2: Frost Beamforming

Result: Beamformed signal y
Align signals $\mathbf{x}(t) = \mathbf{p}(t - \tau)$;
Initialize C, P, g, w, $\mathbf{y}$, $\mathbf{X}[0]$;
for *n = 1:length of input signal* **do**
    Update $\mathbf{X}[n]$;
    Calculate step output $y[n]$;
    Update filter weights;
end

This is a (gradient projection) linearly constrained *least-mean-square* (LMS) algorithm [48].

Drawbacks with LMS algorithms is that the updating scheme is highly dependent on the scaling of the input ($\mathbf{X}[n]$), making it hard to select a learning rate that ensures stability. By normalizing the power of the input, $\mathbf{R}_{XX}$, the learning rate becomes upper bounded and stability is ensured [36]. This method is called *normalized least-mean-square* (NLMS). Mathematically, this is expressed as

$$\mu[n] = \frac{\alpha}{\mathbf{R}_{XX}} \approx \frac{\alpha}{\mathbf{X}[n]\mathbf{X}^H[n]} \tag{2.54}$$

where n denotes the learning rate at the n'th sample, and α is a tuning variable between 0 and 1.

The signal and noise is assumed uncorrelated and the noise is also assumed spatially zero mean. Signal-correlated noise may effect the beamforming performance by cancelling out parts or all of the signal components of interest. In environments prone to clutter or where signals have multiple propagation paths signal-correlated noise may occur. The LCMV beamformer is robust against correlated filter tap noise caused by noise/interference sources such as structural self-noise, vehicles, etc. as long as it is uncorrelated with the signal of interest. Uncorrelated noise at the filter taps caused by (e.g.) thermal noise is rejected by the beamformer naturally in two ways. Firstly the FIR filter structure rejects uncorrelated zero-mean noise at the taps since they are added destructively at the output. Secondly, the adaptive weight update have the chance of neglecting the effects of filter taps that produce sudden high uncorrelated noise power. This discriminating feature should in theory make the LCMV beamformer perform better than the FAS beamformer.

In theory, the filter $\mathbf{f}$ should effectively only treat the signal coming form the look direction as all other signals from different directions are assumed to be zero-mean noise.

2.8.5 Generalized sidelobe canceller

The generalized sidelobe canceller (GSC) was introduced by Griffits and Jim [9] as an alternative and more generalized implementation of the Frost's LCMV beamformer. It detaches unconstrained and constrained componets in two different paths. The schematics for the GSC is presented in fig. 2.15. As for all the othe beamformers, the GSC starts by

aligning the array signals using steering delays. Next the signals follows two paths: (i) Conventional fixed beamforming path and (ii) Sidelobe cancelling path.

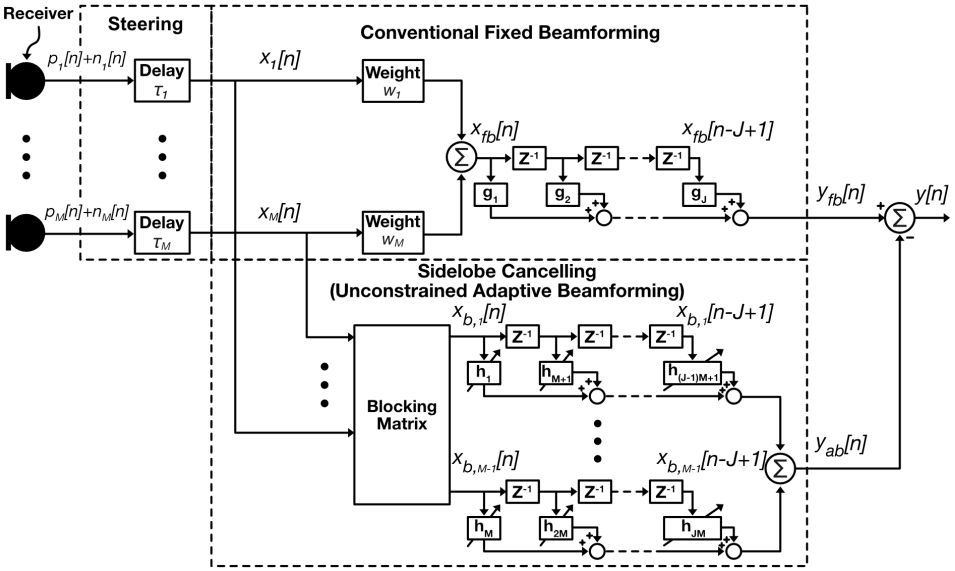


Figure 2.15: Generalized sidelobe canceller schematics.

The upper path (i) of fig. 2.15 illustrates that the steered signals $\mathbf{x}[n] = \begin{bmatrix} x_1[n] & \dots & x_m[n] \end{bmatrix}^T$ goes sums the steered signals weighted by $\mathbf{w} = \begin{bmatrix} w_1 & w_2 & \dots & w_M \end{bmatrix}^T$ just like a regular DAS, and follows up with a FIR filter with fixed filter weights $\mathbf{g} = \begin{bmatrix} g_1 & g_2 & \dots & g_J \end{bmatrix}^T$. The conventional DAS beamformer produces the FIR filter input

$$x_{fb}[n] = \mathbf{w}^H \mathbf{x}[n] \tag{2.55}$$

before it passes the filter producing the upper path output

$$y_{fb}[n] = \sum_{j=0}^{J} g_j x_{fb}[n-j] \tag{2.56}$$

The filter weights are often band pass filters that are normalized such that

$$\sum_{j=0}^{J} g_j = 1 \tag{2.57}$$

The lower path (ii) of fig. 2.15 is called the *sidelobe canceller* and/or *interference canceller* [36] since it tries to output only the noise and interference in the signal and subsequently subtracting it from the "desired" signal produced by the upper path (i) in a superposition principle manner. First the aligned signal passes the *blocking matrix* $\mathbf{B} \in (\mathbb{R}, \mathbb{C})^{(M-1)\times M}$ to remove the SOI from the steered receiver signals before the $M-1$

blocked signals
$\mathbf{x}_b = \begin{bmatrix} x_{b,1} & \dots & x_{b,M-1} \end{bmatrix}^T$ enters a set of FIR filters of length J with unconstrained adaptive filter weights $\mathbf{h} = \begin{bmatrix} h_1 & h_2 & \dots & h_{JM} \end{bmatrix}^T$. At the n'th time instant the output of the blocking matrix is

$$\mathbf{x}_b[n] = \mathbf{B}\mathbf{x}[n] \tag{2.58}$$

To determine what the $\mathbf{B}$ should be it is necessary to assume that the receiver signals follows the signal model eq. (2.17) such that the discrete-time steered signal becomes $\mathbf{x}[n] = \mathbf{v}p[n] + \mathbf{n}[n]$ where $\mathbf{v}p[n]$ is the steered SOI and $\mathbf{n}[n]$ is the noise vector. As mentioned above, the purpose of the blocking matrix is to remove the aligned SOI components from the received signals $\mathbf{x}[n]$. After steering, the SOI parts of the receiver signals are in-phase and the noise and interference are not, thus the blocking matrix needs to have row values that sum to 0 to remove the signal components while not removing the noise and interference. To illustrate that this will block the SOI component of the signals lets consider two receiver signals such that eq. (2.58) can be written out as

$$x_{b1}[n] = (b_{11} + b_{12})v_1 p_1[n] + b_{11}n_1[n] + b_{12}n_2[n] \tag{2.59}$$
$$x_{b2}[n] = (b_{21} + b_{22})v_2 p_2[n] + b_{21}n_1[n] + b_{22}n_2[n] \tag{2.60}$$

Here it is easy to see that if the blocking matrix rows sum up to zero then the SOI components $p_1[n]$ and $p_2[n]$ will be discarded. The blocking matrix rows should be linearly independent. Thus, there are many possible choises, e.g. for a 3×4 blocking matrix

$$\begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}, \begin{bmatrix} 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \tag{2.61}$$

The sidelobe canceller will output

$$y_{ab}[n] = \mathbf{h}^H \mathbf{X}_b[n] \tag{2.62}$$

where

$$\mathbf{h} = \begin{bmatrix} h_1 \\ \vdots \\ h_{M-1} \\ h_M \\ \vdots \\ h_{2(M-1)} \\ \vdots \\ h_{(J-1)M} \\ \vdots \\ h_{J(M-1)} \end{bmatrix}, \mathbf{X}_b[n] = \begin{bmatrix} x_{b,1}[n] \\ \vdots \\ x_{b,M-1}[n] \\ x_{b,1}[n-1] \\ \vdots \\ x_{b,M-1}[n-1] \\ \vdots \\ x_{b,1}[n-J+1] \\ \vdots \\ x_{b,M-1}[n-J+1] \end{bmatrix} \tag{2.63}$$

is the sidelobe canceller filter weight matrix and the tap voltage matrix containing every tap voltage value at the n'th sample of the input.

Now that the output from both paths are present, the only thing that remains is to subtract the interference signal eq. (2.62) from the desired signal eq. (2.56) to form

$$y[n] = y_{fb}[n] - y_{ab}[n] \tag{2.64}$$

To make the sidelobe cancelling adaptive the unconstrained LMS or NLMS algorithm is used [48]. The total algorithm can be expressed in pseudo code as

Algorithm 3: GSC beamforming algorithm

Result: Beamformed signal y
Align signals $\mathbf{x}(t) = \mathbf{p}(t - \tau)$;
Initialize $\mathbf{w}$ and $\mathbf{B}$;
Do weighted sum (upper path): $\mathbf{x}_{fb}$;
Do signal blocking: $\mathbf{x}_{b,1}$ ... $\mathbf{x}_{b,M-1}$;
Initialize filters: $\mathbf{g}$=desired filter response, $\mathbf{h} = \mathbf{0}$;
Initialize tap delay line: $\mathbf{X}_{ab}[0]$ and $\mathbf{X}_b[0]$;
Initialize outputs: $\mathbf{y}_{fb}$, $\mathbf{y}_{ab}$ and $\mathbf{y}$;
for *n = 1:length of input signal* **do**
    Update taps: $\mathbf{X}_{ab}[n]$ and $\mathbf{X}_b[n]$;
    Calculate step outputs: $y_{fb}[n]$, $y_{ab}[n]$ and $y[n]$;
    Update filter weights;
end

The GSC is more general than the LCMV beamformer since it allows more structural freedom such as multiple blocking matrix designs and FIR filter designs [9]. This is an advantage when it comes to practical tuning of the adaptive structure dependent on the real world statistics that may differ from theory. The LSM algorithm do not need any linear constraints, thus reducing the complexity. One should also note that it is possible to change the upper structure of the GSC to a lot of different formats, e.g. from a DAS+FIR to a FAS. A GSC approach is also considered valuable when the number of array elements become large. When this is the case, the number of input signals is also large, and henceforth makes the inversion of the spatial covariance matrix in eq. (2.47) for the LCMV computationally heavy.

2.9 Performance indices

To know how an array performs it is desirable to have methods for ”measuring” the quality of the array architecture. Continuing this discussion using the discretized time domain output of the conventional beamformer following the general input signal model eq. (2.19)

$$y[n] = \mathbf{w}^H \mathbf{x} = \mathbf{w}^H (\mathbf{V}\mathbf{p} + \mathbf{n}) \tag{2.65}$$

And in frequency domain the notation will be

$$Y = \mathbf{W}^H \mathbf{X} \tag{2.66}$$

$$= \mathbf{W}^H \mathbf{V}\mathbf{P} + \mathbf{W}^H \mathbf{N} \tag{2.67}$$

omitting the variable dependencies $j\Omega$ and Θ.

2.9.1 Beampattern

The *beampattern* or *radiation diagram* is a measure to illustrate the directional sensitivity of an array configuration. In terms of transmitting array sources (i.e. antennas) it is clear that transmission strength is directional and depends on frequency due to constructive and destructive interference creating *lobes* and *nulls* respectively at various angles. A lobe, or beam, is a formation showing the spatial radiation strength. A null is spatially where the radiation is very low, i.e. between lobes. The beampattern is a valid measure for receiving arrays as well due to Rayleigh's reciprocity theorem . The beampattern can be built using the radiation function [36; 51] defined as

$$B(j\Omega,\Theta) = \frac{Y(j\Omega,\Theta)}{P(j\Omega)} = \mathbf{W}^H(j\Omega)\mathbf{V}(j\Omega,\Theta) \tag{2.68}$$

Where $\mathbf{v}(j\Omega,\Theta)$ is the steering vector and $Y(j\Omega,\Theta)$, $P(j\Omega)$ and $\mathbf{W}(j\Omega)$ describes the frequency domain counterpart of the time domain array output $y(t)$, SOI $p(t)$ and weights $\mathbf{w}$ for isotropic sensors in an anechoic environment. The beampattern is frequently plotted in decibels, i.e.

$$B_{dB}(j\Omega,\Theta) = 20\log_{10}|B(j\Omega,\Theta)| \tag{2.69}$$

Figure 2.16 shows an illustration of beampattern plotted using polar coordinates. The figure contains markings of different beampattern formations such as lobes and nulls. The width of the lobes are often measured using half-power beamwidth (HPBW), which is the the angular separation between the two points on the lobe with a magnitude equal to half of the maximum lobe magnitude (i.e. the -3 db points). Back lobes are common for arrays that do not inhabit symmetrical properties. However, ULAs with isotropic sensors without back baffle are symmetric.

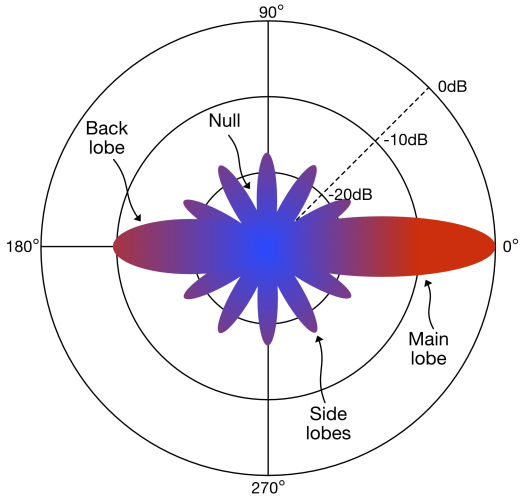


Figure 2.16: Illustration of a beampattern polar plot in decibels highlighting lobes and nulls.

Normalization with respect to the direction of maximum gain is commonly used in literature since the beampattern is more of a visual illustration of the arrays directivity.

2.9.2 Array gain

A common way to evaluate the performance of an receiver array is to measure the SNR improvement between the average array input and the array output for a given listening direction $\Theta_0 = (\phi_0, \theta_0)$. This SNR improvement is called the *array gain* $AG(j\Omega)$ and can be seen as the measure of gained signal quality using an array compared to a single sensor. That is

$$AG(j\Omega) = \left.\frac{\text{SNR}_{out}}{\text{SNR}_{in}}\right|_{\Theta=\Theta_0} \tag{2.70}$$

One approach to find the SNR ratio between the array input and output using the beampattern of the array along with the directional patterns of the signal and noise fields surrounding the array [39]. This is often cumbersome since it involves accurate knowledge about the signal and noise field which can be difficult to obtain.

An alternative approach uses measures of coherence, i.e. cross correlation in time domain or cross PSD in frequency domain, between signal and noise to calculate the array gain. Assuming that the noise spectra at each sensor is the same mutually uncorrelated with the signal. Thus, the SNR at the input can be described as

$$\text{SNR}_{in} = \frac{\mathbb{E}\{P(j\Omega)\bar{P}(j\Omega)\}}{\mathbb{E}\{N(j\Omega)\bar{N}(j\Omega)\}} = \frac{R_p(j\Omega)}{R_n(j\Omega)} \tag{2.71}$$

where $R_x(j\Omega)$ and $R_n(j\Omega)$ is the PSD of the impinging SOI and noise at a single array element respectively, and the bar-notation denotes the complex conjugate. Similarly the SNR_{out} is calculated from

$$\text{SNR}_{out} = \frac{\text{Signal part of } \mathbb{E}\{Y(j\Omega)\bar{Y}(j\Omega)\}}{\text{Noise part of } \mathbb{E}\{Y(j\Omega)\bar{Y}(j\Omega)\}} \tag{2.72}$$

where

$$\begin{aligned}
\mathbb{E}\{Y(j\Omega)\bar{Y}(j\Omega)\} &= R_y(j\Omega) && (2.73)\\
&= \mathbf{W}^H\mathbb{E}\{\mathbf{X}(j\Omega)\mathbf{X}^H(j\Omega)\}\mathbf{W} && (2.74)\\
&= \mathbf{W}^H\mathbf{V}R_p(j\Omega)\mathbf{V}^H\mathbf{W} + \mathbf{W}^H\mathbf{R}_{nn}(j\Omega)\mathbf{W} && (2.75)
\end{aligned}$$

Here, $\mathbf{R}_{nn}(j\Omega)$ is the spatial PSD matrix of the sensor noises. It is clear from eq. (2.75) that the first part represent the signal part, while the latter represent the noise. Hence,

$$\begin{aligned}
\text{SNR}_{out} &= \frac{\mathbf{W}^H\mathbf{V}R_p(j\Omega)\mathbf{V}^H\mathbf{W}}{\mathbf{W}^H\mathbf{R}_{nn}(j\Omega)\mathbf{W}} && (2.76)\\
&= \frac{R_p(j\Omega)}{R_n(j\Omega)}\frac{|\mathbf{W}^H\mathbf{V}|^2}{\mathbf{W}^H\hat{\mathbf{R}}_{nn}(j\Omega)\mathbf{W}} && (2.77)
\end{aligned}$$

$\hat{\mathbf{R}}_{nn}$ is the normalized PSD matrix for isotropic noise such that its trace is equal to the number of sensors [36].

Followingly, the array gain can be calculated by inserting eq. (2.77) into eq. (2.70) such that

$$AG(j\Omega) = \frac{|\mathbf{W}^H\mathbf{V}|^2}{\mathbf{W}^H\hat{\mathbf{R}}_{nn}(j\Omega)\mathbf{W}} \tag{2.78}$$

which clearly depends on the steering matrix $\mathbf{V}$, weight matrix $\mathbf{w}$ and the noise field characteristics.

In the best case scenario, the array is impinged with a single plane wave in a white noise field and holds the distortionless constraint $\mathbf{w}^H\mathbf{V} = 1$, which is true when the signal part of the output equals the input signal. I.e the normalized PSD noise matrix becomes the identity matrix and hence the single signal model $\mathbf{w}^H\mathbf{I}\mathbf{w} = 1/M$ for normalized weight vector $\mathbf{w}$ where M is the number of array elements. This means that

$$AG(j\Omega) = \frac{1}{\mathbf{w}^H\mathbf{w}} = M \tag{2.79}$$

Thus, the array gain will in the best case scenario be equal to the number of array elements. This is a motivation in it self to consider an array of receivers over a single element.

2.9.3 Directivity index

The *directivity Index* (DI) is one of the most important quality measurement of multiple sensor arrays. It is defined as

$$DI(j\Omega) = \frac{|B(j\Omega, \Theta_0)|^2}{\frac{1}{4\pi}\int_0^{\pi}\int_0^{2\pi}|B(j\Omega,\Theta)|^2 d\theta \cos\phi d\phi} \tag{2.80}$$

In the presence of a isotropic diffuse noise field the array gain $AG(j\Omega)$ is the same as the $DI(j\Omega)$, which means that the DI is a measure of SNR improvement as well.

$$DI(j\Omega) = \frac{|B(j\Omega, \Theta_0)|^2}{\mathbf{W}^H\rho_{\text{diffuse}}(j\Omega)\mathbf{W}} \tag{2.81}$$

Where $\rho_{\text{diffuse}}(j\Omega)$ is the normalized spatial isotropic noise matrix built up of entries $\rho(j\Omega, d_{i,j}) = \text{sinc}(j\Omega d_{i,j}/c)$ where $d_{i,j}$ denotes the distance between the i'th and the j'th array element.

Since the noise characteristics can be hard to find, it is often useful to calculate the DI instead of AG for investigating the approximate array performance.

Just like beampattern, polar plots are often used to visualize the performance in different directions.

2.9.4 Person's correlation coefficient

Pearson's correlation coefficient (PCC) is a statistical metric that measures the statistical relationship, or linear correlation, between two variables A and B with multiple observations. Assorting values ranging from -1 to 1 to describe the correlation between the variables is useful for analyzing the magnitude, association and direction of that relationship. Negative values indicate negative correlation, while positive values indicate positive correlation. No correlation is found when Person's correlation coefficient is 0.

It is defined as

$$\rho(A, B) = \frac{\text{cov}(A, B)}{\sigma_A\sigma_B} \tag{2.82}$$

where $\text{cov}(A, B)$ is the covariance of A and B, and σ_A and σ_B is the variance of A and B respectively.

2.9.5 Signal-to-interference peak ratio

In this thesis it is beneficial to quantify the peak value ratio between the SOI and interference, or *signal-of-no-interest* (SONI), coming in from an angle other than the listening direction. Defining the signal-to-interference peak ratio as

$$PR = \frac{\text{max/peak value of SOI}}{\text{max/peak value SONI}} \tag{2.83}$$

This is valuable since it will give insight into the the noise/interference suppression ability of the different processing techniques. Small values of PR under 1 means that the SONI has a higher amplitude than the SOI, and is likely to contribute more to the overall power of the signal assuming pulse like signals. If the PR is over 1 indicates the opposite situation.

Chapter 3

Material and Methods

This chapter describes code development and experiments.

3.1 Code development

One of the main challenges in the work of this thesis was to build a working code library that would grant access to key functionality used for analysis, design and visualization of sensor array systems. In the starting phase it was useful to experiment with code while reading up on the underlying theory to acquire the necessary understanding needed for further development. Therefore, the *Phased Array System Toolbox* [21] provided by MATLAB (The MathWorks, Inc., Natick, Massachusetts, USA) was helpful to get a better grasp of the theory. Since this toolbox requires a MATLAB licence that includes the *DSP System Toolbox* [20] and *Signal Processing Toolbox* [22] as well as the Phased Array System Toolbox itself, it was desirable to write code independent from these major toolboxes.

The algorithms and code used further on in this thesis is written in MATLAB programming language. The code structure is divided into different classes with one superclass and 5 subclasses that inherits functionality from the superclass. Figure 3.1 demonstrate the class structure and their dependencies. The most important functionality within the `Beamformer` class is the time shift methods `delaySignals()`, `calcTimeDelay()` and `unitTimeDelay()` that is used by the subclasses `FAS`, `Frost`, `GSC` and `PlaneWaveCollector`.

All of the beamformer subclasses, `FAS`, `Frost` and `GSC`, are implemented as described by algorithm 1,algorithm 2 and algorithm 3. Their respective beamforming implementation lies within the class method `beamform()` which essentially takes multiple array signals and look direction as input, and outputs the enhanced signals. It is however important to mention that the beamformers treat every signal as if they are collected by isotropic array elements. If the sensors radiation patters is known a priori, it could also be added to the methods as discussed in section 2.5.1.

The `PlaneWaveCollector` subclass is made for simulating spatial propagation of plane wave signals. This is a simple model for simulating propagation delay based on the

given positions for sources and the receiver array, and can be used in virtual simulations. It assumes that the transmission medium is ideal. That is, no absorption, reflections or refractions is imposed on the signal by the transmission medium. The signal model used is the anechoic signal model presented in eq. (2.19).

Lastly, the `PerformanceIndices` subclass covers metrics used to evaluate array configurations and performance introduced in section 2.9. It also includes plot functionality that allows the user to visualize beampatterns, and directivety in two or three dimensions.

The methods are well documented and it should be easy to improve, add on and reorganize the class structure of the code if desirable.

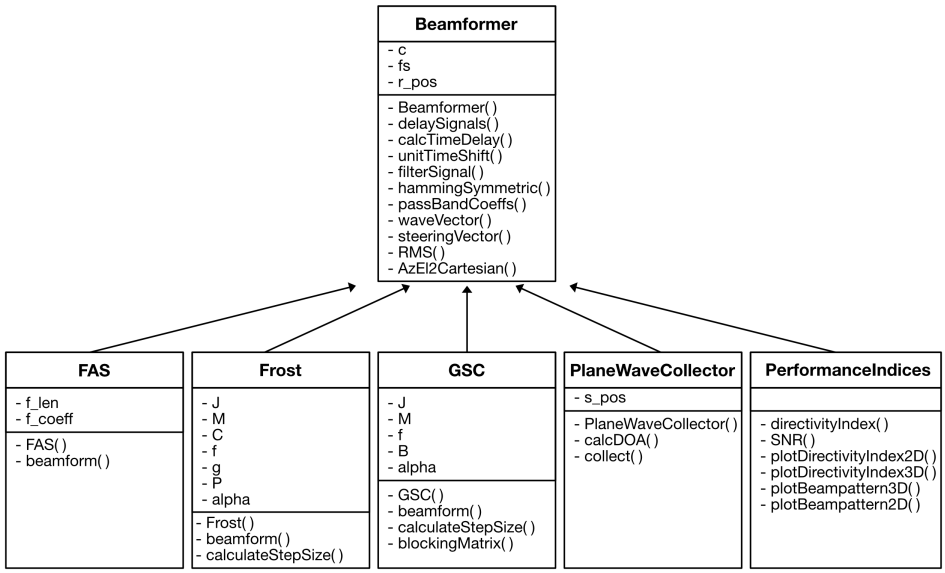


Figure 3.1: Class diagram

3.2 Experimental testing and validation

Given that the goal of this thesis was to check if a hydrophone array would be more beneficial to use than a single hydrophone in an marine sea cage environment, experimentation was crucial. At first, it was necessary to conduct virtual experiments to verify that the algorithms were working properly before physical testing could start. Underwater environments can be very hard to model since there are a lot of factors affecting the soundscape. Thus, a physical experiment taking place underwater was desirable to validate the concept of beamforming on data gathered from a marine setting with a natural noise field.

Two virtual experiments was conduced assuming the anechoic signal model eq. (2.19). One using three hydrophones, i.e. the same amount available for the physical experiment, and the second using five hydrophones. Using different sized arrays should in theory improve the array gain in the listening direction, as well as increasing the directiveness as

the main lobe gets slimmer. Moreover, the virtual experiments used the same positioning of sources and receivers as given in the design of the physical experiment. This was done to be able to give a better comparison of the performances based on the ”same” setup.

The following subsections elaborates upon the procedure of experimentation, both virtually and physically. Starting with the virtual experiments, then moving on to the physical experiment by first giving a discription of the equipment and setup before giving details about the sounds used in the experiment and data preparation.

3.2.1 Virtual experiments

It is convenient to be able to test the code within a virtual environment where it is easy to simulate the transmission and reception of signals to validate array structures and beamformer performance. It allows the user to control the spatial placement of both sound sources and array elements, source signals and adjustments of beamformer parameters in a simple manner. In this thesis two simple virtual simulation are tested using the code mentioned in section 3.1. The model parameters used in these two virtual experiments are presented in chapter 4 along side the results. The experiments models vertical ULAs of the sizes three and five isotropic hydrophones. It collects SOIs perpendicular to the array (zero elevation), and SONI (interference) at an angle of 32° elevation. Remember, the axisymmetrical property of ULAs makes the array independent of azimuth incident angles in this case. The `PlaneWaveCollector` models the collection of SOI and SONI at the array. White noise is added to the receiver signals afterwards. From here, the beamformers takes over.

There are several things that needs to be specified before the plane wave collection simulation can take place. Creating a `PlaneWaveCollector` object is done as follows

```
collector_obj = PlaneWaveCollector(c, fs, rx_pos, tx_pos);
```

where `c` is propagation speed in m/s, `fs` is the sampling frequency of singals that are to be transmitted in Hz, `rx_pos` and (optionally) `tx_pos` are the receiver and source position matrices in Cartesian coordinates with units corresponding to meters on the form

$$pos = \begin{bmatrix} x_1 & x_2 & \dots & x_M \\ y_1 & y_2 & \dots & y_M \\ z_1 & z_2 & \dots & y_M \end{bmatrix} \tag{3.1}$$

where $\begin{bmatrix} x_i & y_i & z_i \end{bmatrix}^T$ corresponds to the position vector of the i’th column of the position matrix.

Anechoic signal model collection is easily done with the code lines

```
rx_sig = collector_obj.collect(tx_sig, ang)
x = rx_sig + noise
```

where `tx_sig` is the signals transmitted from the sources at spatial positions specified by `tx_pos`. `rx_sig` is the signals received by the receivers at spatial position as in `rx_pos`. `ang` is an optional argument that specifies the incident angle of arriving signals if `tx_pos` is not specified in the `PlaneWaveCollector` object creation above. The `noise` parameter contains noise data of same size as `rx_sig`, e.g. white noise. After this, the `rx_pos` is ready to be processed by the beamformers and analysed.

3.2.2 Equipment and the physical experiment setup

Initially, a decision of having a relatively simple experimental setup with known SOI arriving from a known position relative to the receiving array was made as a first step before conducting an experiment in a more complex environment such as a fish cage. Designing the physical experiment had to be done within the limitations of equipment that was at hand at the time, and the regulations that were set due to the Covid-19 outbreak. Because of these limitations, only the simpler experiment was conducted. Equipment that were readily available and used in the experiment is presented in the list below.

- Three icListen HF hydrophones (Ocean Sonics Ltd., Nova Scotia, Canada) used for sound underwater recordings.
- SonTek's CastAway-CTD (SonTek / Xylem, Inc., San Diego, California, USA) used to obtain the sound propagation velocity profile of the underwater environment.
- Oceanears DRS-8 portable underwater speaker and EM-0611B Impedance Matching Transformer (Oceanears/O.E. Enterprises, North Canton, Ohio, USA).
- DAMGOO 100w Dual Channel Audio Amplifier (JUNTENG Co., Ltd, Dongguan, Guangdong, China) used to amplify the sound signals.
- 12V Euroglobe AGM 1,3Ah battery to drive the amplifier.
- Support frame to hold the hydrophones in place.
- Rope to hold the support frame and speaker.
- Tote bag with a brick used as a weight to hold the hydrophone support frame still.
- Apple iPhone 6 (Apple Inc., Cupertino, California, USA) used as a digital audio player.
- Portable Macbook Pro laptop computer (Apple Inc., Cupertino, California, USA) with Ethernet port.

The site of the experiment was at the dock furthest west at Skansen Marina, Trondheim. Figure 3.2 shows where the experiment was conducted with red and yellow dots indicating the placement of the speaker and hydrophones respectively.

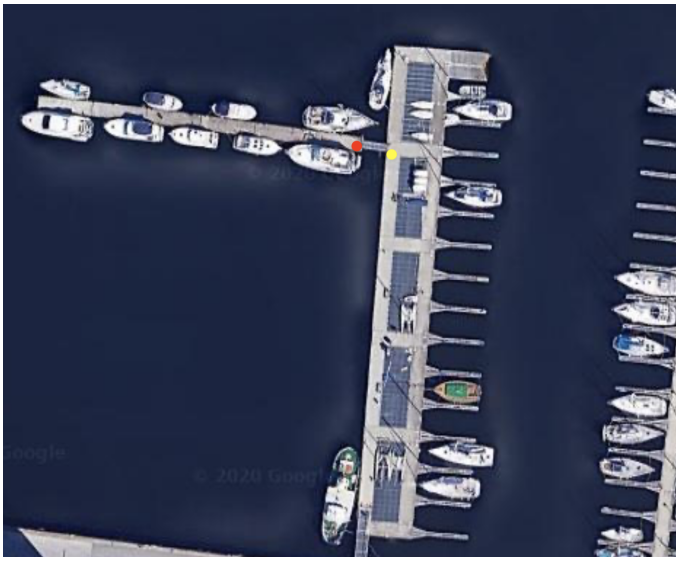


Figure 3.2: Overview of the site where the experiment was conducted at Skansen Marina, Trondheim. The red dot indicates where the underwater speaker was submerged from, and the yellow dot indicates where the hydrophone array was submerged from. (Google Maps (2020), [Skansen Marina, Trondheim, Norway]. Retrived from: `https://www.google.com/maps/search/skansen+marina/@63.4325864,10.3724952,136m/data=!3m1!1e3`)

Synchronization of the hydrophones was done at the experiment site where the internal hydrophone clocks was synchronized with the laptops internal clock before starting the recordings at a specified time.

The experimental setup is illustrated by fig. 3.3. The speaker and array were deployed from separate docks 6 m apart, and lowered to a 5 m depth. This depth was chosen because the CTD measurement showed that there were no significant rapid sound speed changes (due to depth) under 4 m. Separating the speaker and array 6 m apart guaranteed far field propagation since it is beyond the Fraunhofer distance. It was also convenient because of the placement of the docks at the site. A small rope was attached to the speaker and fastened loosely to the dock on the opposite side to ensure that the speaker was facing the hydrophone array. The same was done for the array.

All of the hydrophones was fastened to the support frame 0.1 m apart with cable ties as shown in fig. 3.4. Following the definition of Nyquist-Shannon sampling theorem for spatial aliasing, eq. (2.2), it is easy to see that a spacing of 0.1 m produces an upper frequency bound of 7.5 kHz assuming that the sound velocity was 1500 m/s.

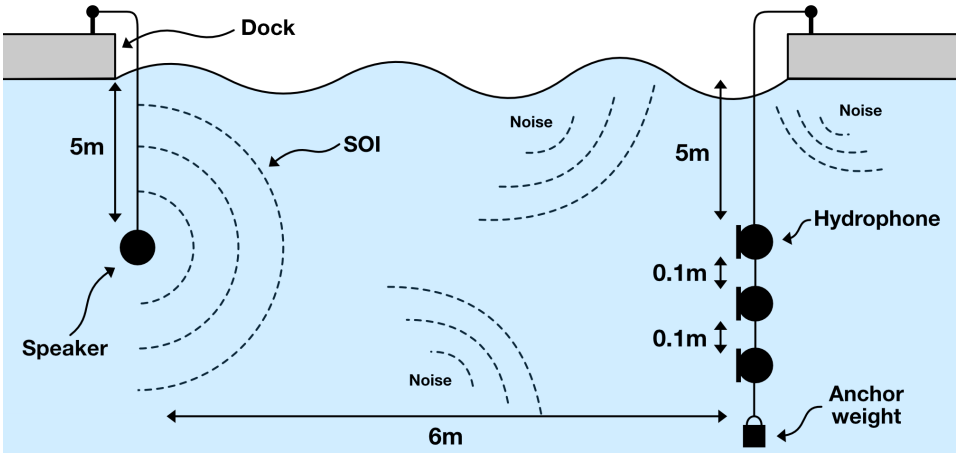


Figure 3.3: Real environment underwater experiment setup illustration.

Figure 3.4: Hydrophones mounted to support frame with 0.1 m of separation. The tote bag contains a brick and works as an anchor weight holding the array still under water. The small white rope is used to point the array in the desired direction.

3.2.3 Experiment sound signals

Since this thesis is concerned with using beamforming to improve fish SNR within sea cage environments it would be practical to experiment with sounds that imitate fish in such an environment. To the extent of the author's knowledge there is no prior information about the soundscape within a sea cage to this date. As mentioned in section 2.2, there is not much found on the sounds of Atlantic salmon either, except for the recording at the NINA research facility Ims. In this study there was an increased activity in the 4 kHz-10 kHz frequency spectrum. Listening to the recording gives an impression of small and densely packed chirps as a school class of children approach the site. In addition, research papers addressed previously showed that herring and cod was able to produce high frequent sounds in the same frequency region as well. This has been the motivation for the sound that were made for the experiments in this thesis.

Now, as the frequency spectrum of interest is considered 4kHz-10kHz, Gaussian-modulated sinusoidal pulses was chosen as the ammunition for the experiments. Gaussian pulses are pulses that consists of a variety of frequencies normally distributed around a center frequency, see fig. 3.5.

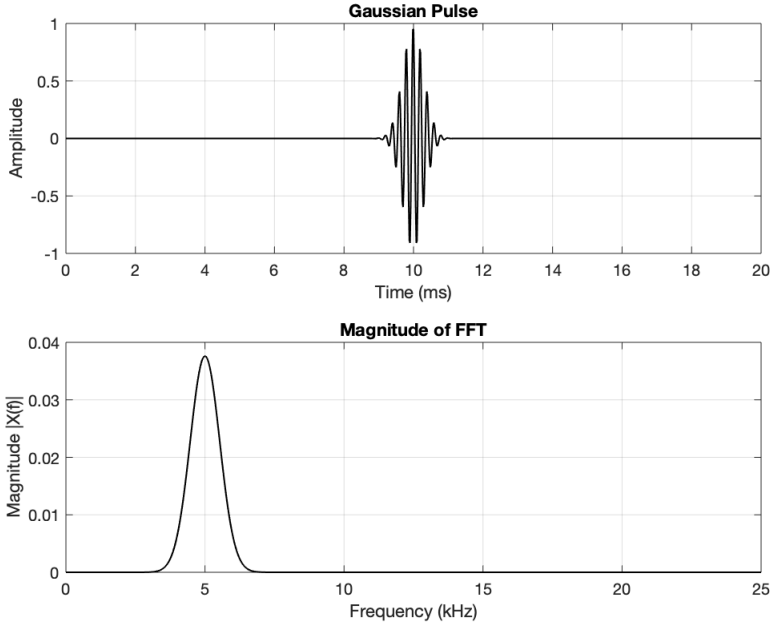


Figure 3.5: Example of an Gaussian pulse with center frequency 5 kHz and variance $\sigma = 0.0005$. The first subplot shows the time-domain plot of the pulse, and the second subplot shows the frequency domain plot.

Fractional bandwidth (FBW) is defined as the bandwidth divided by the center fre-

quency, i.e.

$$FBW = \frac{f_{max} - f_{min}}{f_c} \tag{3.2}$$

where f_{max}, f_{min} and f_c is the maximum frequency over -3 dB gain, minimum frequency over -3 dB gain and the center frequency respectively.

The sounds made for the virtual experiments were simply gaussian pulses with given center frequency and FBW. One acting as the SOI and another acting as SONI, in addition to noise as explained in section 3.2.1.

Three series of signals was made for the physical experiment: (i) 0.1 FBW Gaussian pulses (ii) 0.5 FBW Gaussian pulses, and (iii) 1 FBW Gaussian pulses. All of the series contains pulses with center frequencies at every kilohertz from 4 kHz to 10 kHz, where each kilohertz signal starts with 3.75 seconds of a 10.67 kHz sinusoidal before a pulse every 1 seconds with linearly increasing power for 60 seconds is given. The single frequency sinusoidal part of the signal is added to make time-synchronization of the hydrophones easier, see section 3.2.4. An example of one of the transmission signals is presented in fig. 3.6. The first (ca.) four seconds contains the synchronization signal while the remaining part consist of the linearly increasing Gaussian pulses every second for 60 seconds. The bottom subplot illustrate the same signal just in normalized power spectrogram format, highlighting that the synchronization part is 10.67 kHz, and that the pulses are centered at 5 kHz with (in this case) 0.5 FBW.

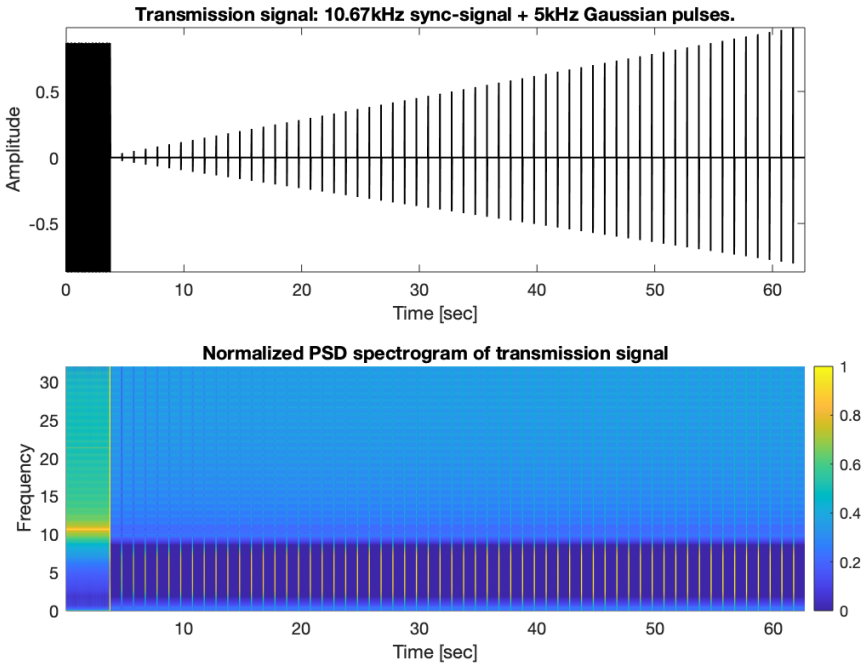


Figure 3.6: Time-domain example plot of signal used for transmission in experiment. The red section contains the 3.75 s of 10.67 kHz sinusoidal signal (used for time synchronization) while the blue section contains a Gaussian pulses every 1 s in a linearly power increasing manner.

After physical testing, it was decided to only use the 5 kHz 0.5 FBW Gaussian pulse signal for analysis. This was because this proved to be the most interesting test case for the physical experiment due to a boat located two meters behind the speaker on surface level generating motor noise/interference at the instance of recording. Thus, giving realistic surface interference, or SONI, coming in from an angle of approximately 30° elevation, and thus differs from the SOI direction of 0° elevation. The same pulse format was used for SOI and SONI in the virtual experiment for result comparison reasons.

3.2.4 Dataset preparation for physical experiment

Since the beamformers are supposed to handle raw recordings directly, there is not a lot of dataset preparation needed. But, three recording factors needs to be addressed and handled before passing the recordings to the beamformers.

The on-site synchronization would not ensure perfect synchronization since the internal hydrophone clocks might drift independently form each other. Thus, a second post-recording synchronization step was needed to ensure even better synchronization before processing and analyzing the audio signals. As mentioned in section 3.2.3, a synchronization sinusoidal signal is added to the start of each transmitter signal. Assuming that each hydrophone receives the signal at the same time implies that the recorded synchronization

signals should be in phase with each other. Shifting the signals such that the sinusoidals are aligned ensures good overall synchronization and low risk of internal clock drift between the recordings. Originally, the sync-signal was meant to be a 5 s long sinusoidal signal with center frequency of 8 kHz sampled at 48 kHz. On the other hand, the Gaussian pulse signals was sampled at 64 kHz. Thus, adding the sync-signal and Gaussian pulse signal together and treating the combined signal as as a 64 kHz sampled signal meant that the synchronization part would appear as a sinusoidal of center frequency 10.67 kHz lasting for 3.75 s instead. This did not pose any problems as the concept of aligning the sinusoidals remained the same.

Secondly, hardware differences between the hydrophones had to be accounted for. By studying the raw data it was clear that the hydrophones produced data that had different mean value. Another significant observation was that the hydrophone of the newest model type had flipped polarity of the transducer element compared to the two hydrophones of older model type. These two effects are illustrated in fig. 3.7 where raw data of one of the 6kHz Gaussian pulses with 0.5 FBW are plotted before (top subplot) and after (bottom subplot) adjusting for hardware differences. Notice, in the top subplot, that the two older model hydrophones (nr.1 and nr.2) are quite similar even though they are not zero-mean. The newer model hydrophone (nr.3) are not zero-mean either and seems to be flipped due to opposite transducer polarity.

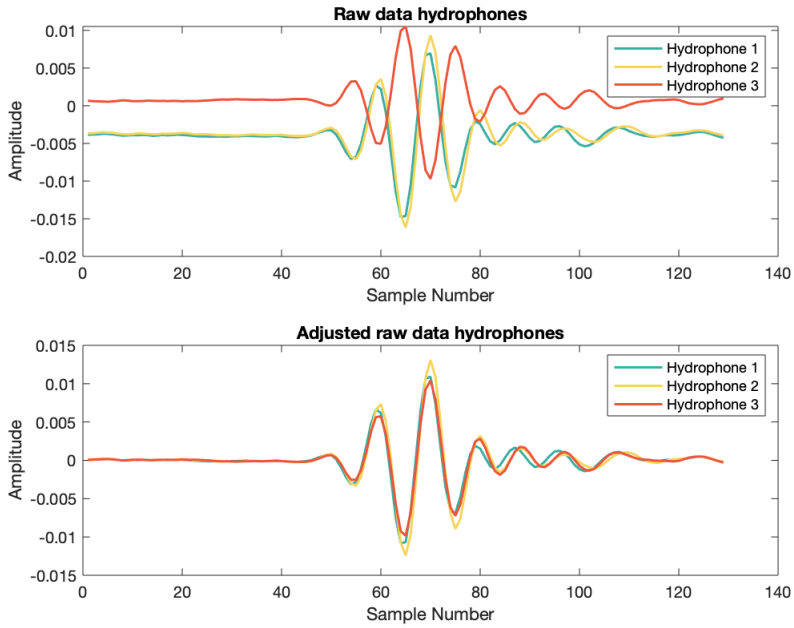


Figure 3.7: Top: Raw data signals without hardware-difference adjustments. Bottom: Hardware-difference adjusted raw data signals. Both signals are time-synchronized.

Thirdly, the raw data should be divided into smaller segments and stored in separate

audio files. Recording was done in 10 minutes intervals at a sampling frequency of 64kHz and with sample resolution of 24 bits per sample. This means that each raw data audio file consists of 38.4 million samples equating to 115.2 MB (in SI notation) of data. Since the code do not discriminate among input signals of different sizes, and not all of the recorded data was of interest it was necessary to divide the raw data in to shorter synchronized segments and save them in separate files to make it easier to analyze each signal separately and reduce processing time.

Chapter 4

Results

In this chapter the key results from processing the data with three different types of beamformers will be presented. All of the beamformers used are wideband beamformers introduced in section 2.8, i.e. FAS beamformer, Frost's LCMV beamformer and the GSC beamformer. Since SOI isolation in the listening direction is the key point of interest - AG approximations, PCC and PR (from section 2.9) will be used to quantify the performance of the beamformers in the virtual experiments. For the physical experiment only PCC and PR will be used due to the difficulty in providing good SNR, and hence, AG approximations. Further on, the performance will be compared with generic bandpass filtering of receiver signals from a single array elements to evaluate the advantage/disadvantage of using an array instead of a single element.

First, the results from a virtual experiment with three array elements will be presented. This will work as the virtual analogue to the physical experiment presented last. Following up with the five element ULA experiment before presenting ULA beampatten and DI calculation will give better basis for array design comparison later. Lastly, physical experiment results are exhibited.

4.1 Virtual experiment results: Three hydrophones

This virtual experiment was executed using the parameters specified in table 4.1.

Table 4.1: Virtual experiment parameters for experiment with three array elements.

	Environment	
Propagation speed [m/s]	1500	
	Receiver Array	
Element type	Isotropic	
Position (x,y,z) [m]	(0,0,-5) (0,0,-5.1) (0,0,-5.2)	
	SOI	SONI
Position (x,y,z) [m]	(6, 0,-5)	(8,0,0)
Signal length [s]	0.3	0.3
Pulse start time [s]	0.1	0.2
Sampling frequency [kHz]	64	64
Center frequency [kHz]	5	5
Fractional bandwidth	0.5	0.5
	Noise	
Noise type	Gaussian white noise	
Variance/power	0.15	
	Beamformers	
Listening direction: Elevation [deg]	0	
FIR filter length	51	
Lower cutoff frequency [kHz]	3.5	
Upper cutoff frequency [kHz]	7.5	
LMS step factor α_{Frost}	0.3	
LMS step factor α_{GSC}	0.2	

Conducting the experiment gave the results presented by fig. 4.1, fig. 4.2 and table 4.2. The first figure showcases the normalized shape of time-domain signals in separate plots. It includes the first unprocessed signal from receiver 1, bandpass filtered signal from receiver 1, FAS output, Frost's LCMV output and the GSC output.

The second figure display the average power normalized, or *root-mean-square* (RMS) normalized signals including the RMS normalized transmitter signal. It holds three subplots where the top plot exhibit all of the full-length signals, while the two bottom subplots display the zoomed-in wave form information of the SOI and SONI.

Table 4.1 holds the AG and PR information about all beamformer outputs as well as for the single filtered receiver signal. The PCC between the signals and the transmitted signal is also presented here.

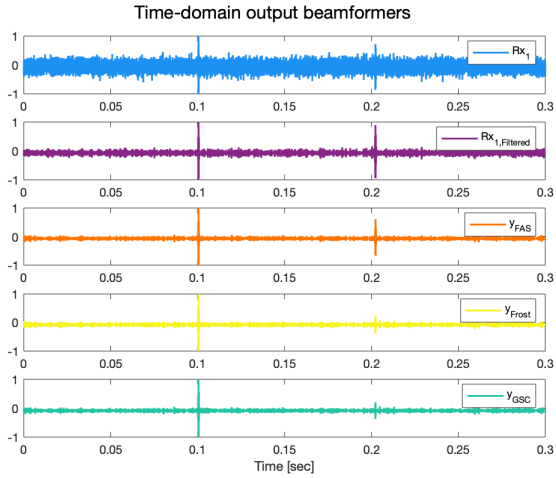


Figure 4.1: Time-domain signal shape comparison for virtual experiment with three array elements. From top to bottom: Collected signal by receiver 1, band pass filtered collected signal by receiver 1, output FAS beamformer, output Frost's LCMV beamformer, output GSC beamformer. All signals are amplitude normalized between -1 and 1.

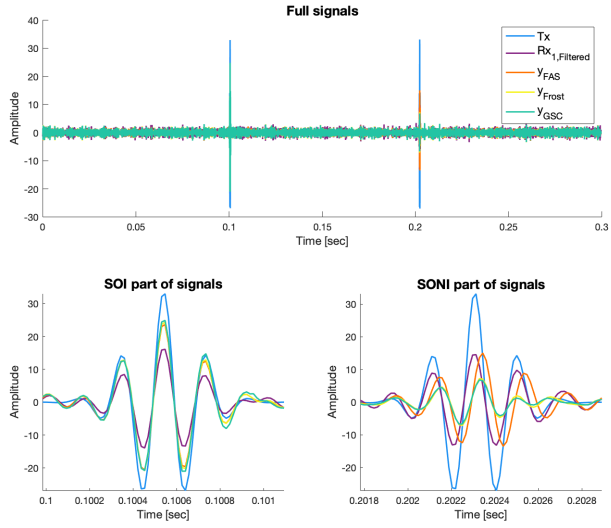


Figure 4.2: Results from virtual experiment with three array elements. Tx is the combined transmitted signals (SOI+SONI without noise) at receiver 1. $Rx_{1,Filtered}$ is the band pass filtered receiver signal collected by receiver 1. y_{FAS}, y_{Frost} and y_{GSC} is the output of FAS-, Frost's LCMV- and GSC beamformers respectively. All signals RMS normalized. Top: Full signals. Bottom left: SOI part signals. Bottom right: SONI part of signals.

Table 4.2: AG, PR and PCC between the Tx-signal and output signals for virtual experiment with three array elements.

Signal	AG	PCC	PR
Filtered Rx_1-signal	0.512 dB	0.359	1.105
Output FAS	1.397 dB	0.533	1.566
Output Frost	1.486 dB	0.546	3.355
Output GSC	1.647 dB	0.566	3.642

4.2 Virtual experiment results: Five hydrophones

This virtual experiment is executed using the parameters specified in table 4.3.

Table 4.3: Virtual experiment parameters for experiment with five array elements

	Environment	
Propagation speed [m/s]	1500	
	Receiver Array	
Element type	Isotropic	
Position (x,y,z) [m]	(0,0,-5) (0,0,-5.1) (0,0,-5.2) (0,0,-5.3) (0,0,-5.4)	
	SOI	SONI
Position (x,y,z) [m]	(6, 0,-5)	(8,0,0)
Signal length [s]	0.3	0.3
Pulse start time [s]	0.1	0.2
Sampling frequency [kHz]	64	64
Center frequency [kHz]	5	5
Fractional bandwidth	0.5	0.5
	Noise	
Noise type	Gaussian white noise	
Variance/power	0.15	
	Beamformers	
Listening direction: Elevation [deg]	0	
FIR filter length	51	
Lower cutoff frequency [kHz]	3.5	
Upper cutoff frequency [kHz]	7.5	
LMS step factor α_{Frost}	0.3	
LMS step factor α_{GSC}	0.2	

Conducting the experiment gave the results presented by fig. 4.3, fig. 4.4 and table 4.4. The first figure showcases the normalized shape of time-domain signals in separate plots. It

includes the first unprocessed signal from receiver 1, bandpass filtered signal from receiver 1, FAS output, Frost's LCMV output and the GSC output.

The second figure display the average power normalized, or *root-mean-square* (RMS) normalized signals including the RMS normalized transmitter signal. It holds three subplots where the top plot exhibit the all full-length signals, while the two bottom subplots display the zoomed-in wave form information of the SOI and SONI.

Table 4.3 holds the AG and PR information about all of the beamformer outputs as well as the single filtered receiver signal. The PCC between the signals and the transmitted signal is also presented here.

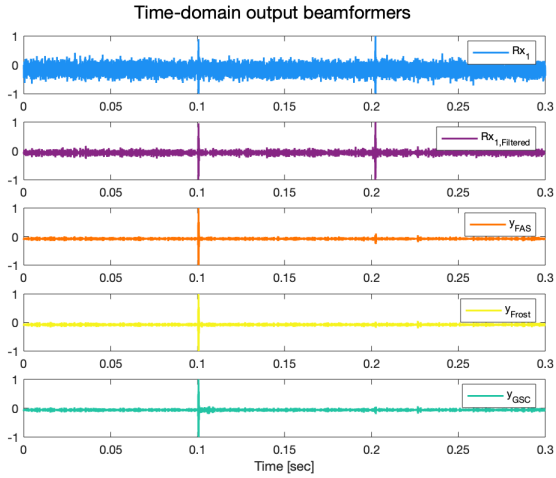


Figure 4.3: Time-domain signal shape comparison for virtual experiment with five array elements. From top to bottom: Collected signal by receiver 1, band pass filtered collected signal by receiver 1, output FAS beamformer, output Frost's LCMV beamformer, output GSC beamformer. All signals are amplitude normalized between -1 and 1.

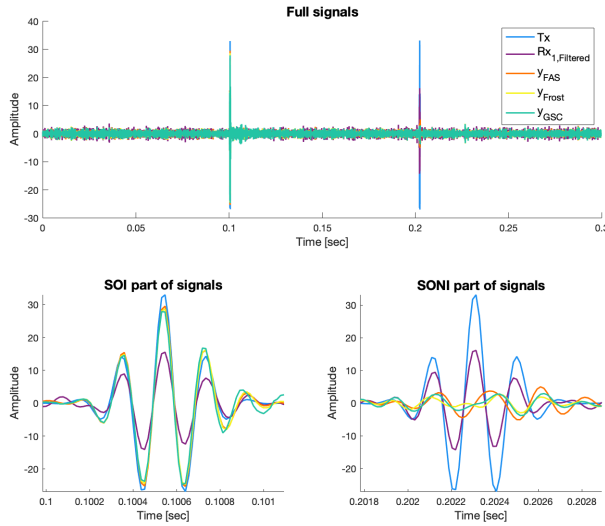


Figure 4.4: Results from virtual experiment with five array elements. Tx is the combined transmitted signals (SOI+SONI without noise) at receiver 1. $Rx_{1,Filtered}$ is the band pass filtered receiver signal collected by receiver 1. y_{FAS}, y_{Frost} and y_{GSC} is the output of FAS-, Frost's LCMV- and GSC beamformers respectively. All signals are RMS normalized. Top: Full signals. Bottom left: SOI part signals. Bottom right: SONI part of signals.

Table 4.4: AG, PR and PCC between the Tx-signal and output signals for virtual experiment with five array elements.

Signal	AG	PCC	PR
Filtered Rx_1-signal	0.491 dB	0.352	0.958
Output FAS	2.520 dB	0.666	5.877
Output Frost	2.376 dB	0.651	10.266
Output GSC	2.304 dB	0.638	9.411

4.3 ULA beampattern and directivity gain

Since the design of the array is of great importance for the performance analysis a detailed presentation of beampattern calculations and directivety is given in the following two subsections. The focus is on ULA designs with three and subsequently five uniformly distributed isotropic array elements.

The upcoming plots are calculated from the following design parameters:

Propagation speed [m/s]	1471
Element spacing along z-axis [m]	0.1
Listening direction (elevation) [deg]	0
Shading weights	Equal

Table 4.5: Design parameters for plotting three elemental ULA.

Propagation speed was set equal to the CTD propagation velocity measurement at 5 m water depth. ULA arrays are axisymmetrical about the axis of elemental distribution, z-axis in this case, making azimuth listening directions insignificant. In addition, the array performance is evaluated for Listening direction of $0°$ elevation since this is where the SOI is coming from in the experiments. An elevation angle of $30°$ is also considered for the directivity index plots since the real experiment had incoming interference from this direction.

4.3.1 ULA: Three array elements

In fig. 4.5 the beampatterns for a ULA with three elements at different frequencies are presented. Take note of the HPBW of the mainlobes for the different frequencies. For 3 kHz, the HPBW is about $100°$, and it slowly decreases with increasing frequency. For 9 kHz the HPBW is about $30°$. The second thing to take note of is the sidelobe levels which are about -10 dB for 7 kHz and 9 kHz.

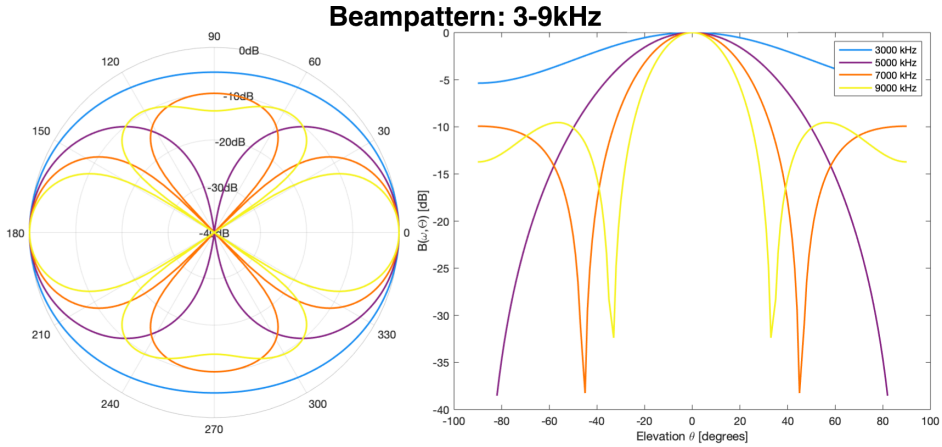


Figure 4.5: Beampattern plots for array with three elements. Frequencies ranges from 3 kHz to 9 kHz. Left: Polar representation. Right: Rectangular representation.

The next two plots, fig. 4.6 and fig. 4.7, illustrates the DI of the array. Figure 4.6 presents the DI for frequencies between 3 kHz and 10 kHz and Listening directions $-90°$to $90°$ elevation while fig. 4.7 presents only two frequency-DI lines for elevations of $0°$ and $30°$. The important remark here is the difference between the $0°$ and $30°$ elevation angle directivity.

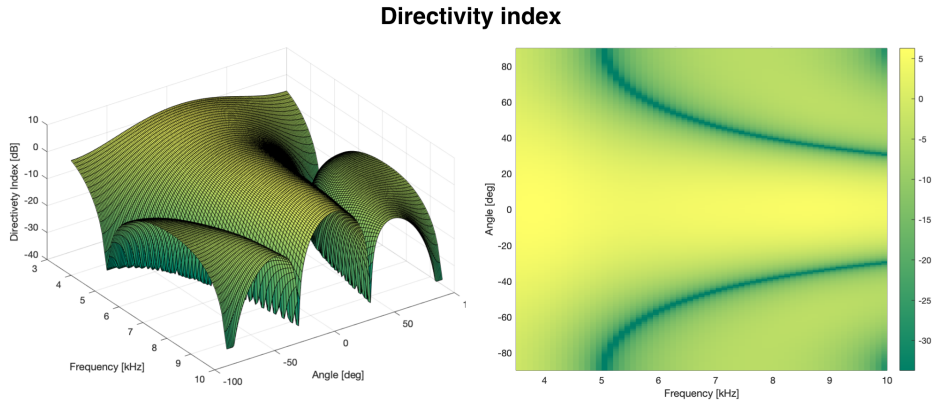


Figure 4.6: Directivity index 3D and contour plot for array with three array elements. Frequencies are in the range 3.5 kHz-10 kHz and (elevation) angles are in range -90° to 90°. Colorbarvalues are in decibels.

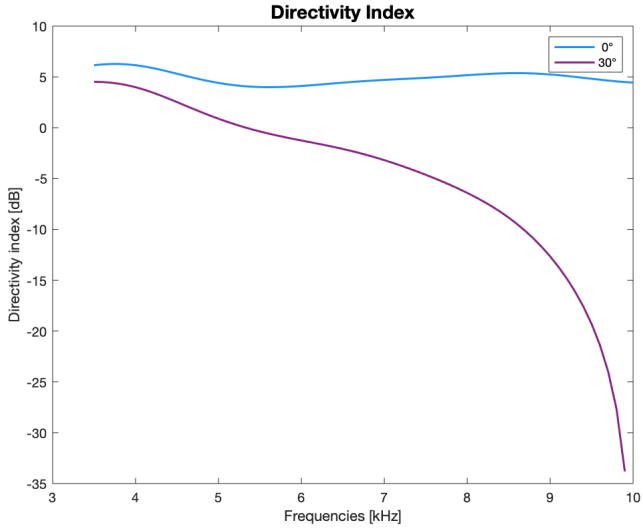


Figure 4.7: Directivity index plot for array with three elements. Frequencies are in the range 3.5 kHz-10 kHz for two elevation angles 0° and 30°.

4.3.2 ULA: Five array elements

Just like the plots pesented in section 4.3.1, the plots in this presents the same information but for an ULA with five array elements.

First, the beampattern is illustrated by fig. 4.8. Notice that the HPBW are slimmer than for the case of three array elements in fig. 4.5. The sidelobe levels are also lowered by a

few decibels.

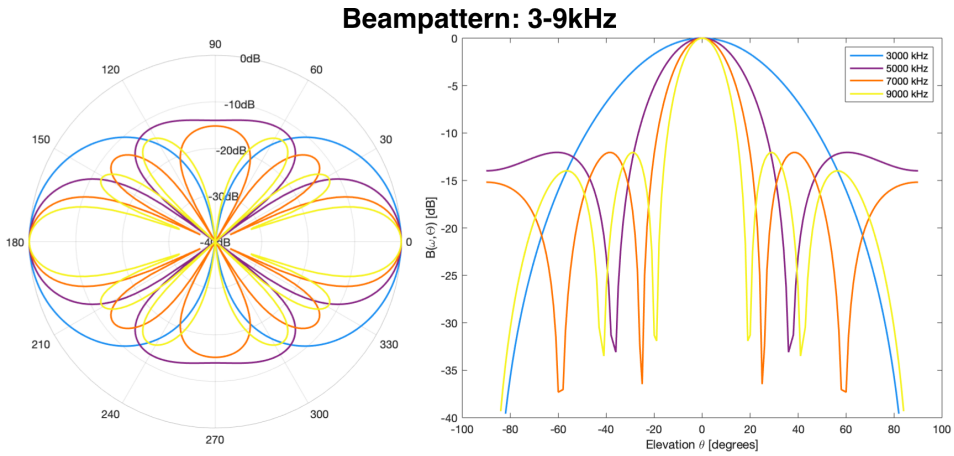


Figure 4.8: Beampattern plots for array with five elements. Frequencies ranges from 3 kHz to 9 kHz. Left: Polar representation. Right: Rectangular representation.

Next, the DI is plotted in fig. 4.9 and fig. 4.10. The plot has noticeable more sidelobe structures with lower level compared to the case with three array elements in fig. 4.6. Comparing fig. 4.10 with fig. 4.7, the 5 element array have increased directivity in the listening direction and very low directivity from 30° as it is affected by the null structure in that direction.

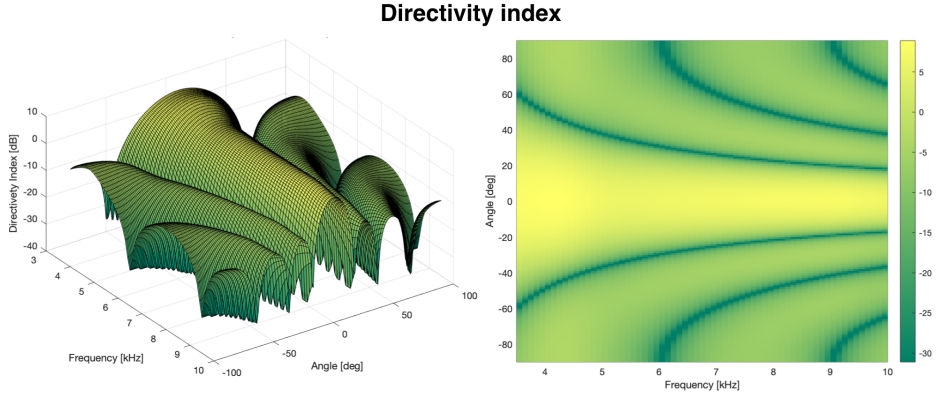


Figure 4.9: Directivity index 3D and contour plot for array with five elements. Frequencies are in the range 3.5 kHz-10 kHz and (elevation) angles in range -90° to 90°. Colorbar values are in decibels.

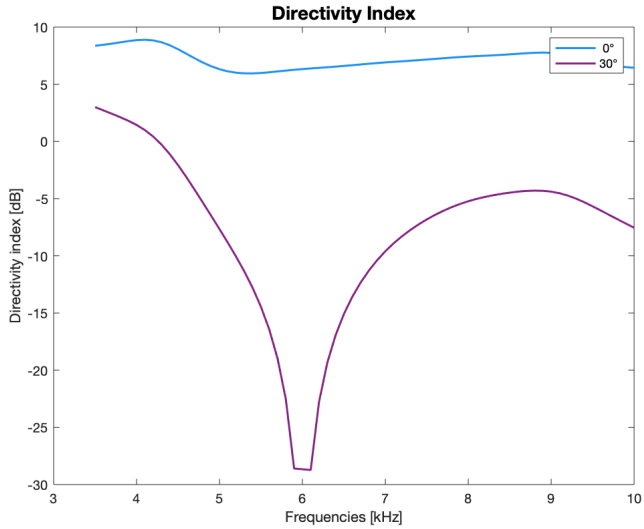


Figure 4.10: Directivity index plot for array with five elements. Frequencies are in the range 3.5 kHz-10 kHz for two elevation angles 0° and 30°.

4.4 Physical results

A summary of the physical experiment parameters are given in table 4.6. Unlike the virtual experiments, natural noise in the harbour will be present instead of white gaussian noise. In addition, the a SONI comming in from an angle is generated by a motorboat on surface level.

Table 4.6: Real experiment parameters.

	Environment	
Propagation speed [m/s]	1471	
	Receiver Array	
Element type	IcListen HF	
Position (x,y,z) [m]	(0,0,-5) (0,0,-5.1) (0,0,-5.2)	
	SOI	**SONI**
Type	Gaussian pulses	Boat motor noise
Position (x,y,z) [m]	(6, 0,-5)	ca. (8,0,0)
Center frequency [kHz]	5	-
Fractional bandwidth	0.5	-
	Beamformers	
Listening direction: Elevation [deg]	0	
FIR filter length	51	
Lower cutoff frequency [kHz]	3.5	
Upper cutoff frequency [kHz]	7.5	
LMS step factor α_{Frost}	0.3	
LMS step factor α_{GSC}	0.01	

For further elaboration on why the motorboat is a good interference source for the sake of the experiment, fig. 4.11 is included. The SOI is marked in red in the top plot, while all of the other peaks are SONI pulses made by the engine of the motorboat except for an unknown peak at the 0.86 second mark. As seen in the bottom spectrogram plot, the SONI pulses share common frequencies with the SOI pulse, mostly in the range 3 kHz-4 kHz and 6 kHz-7 kHz.

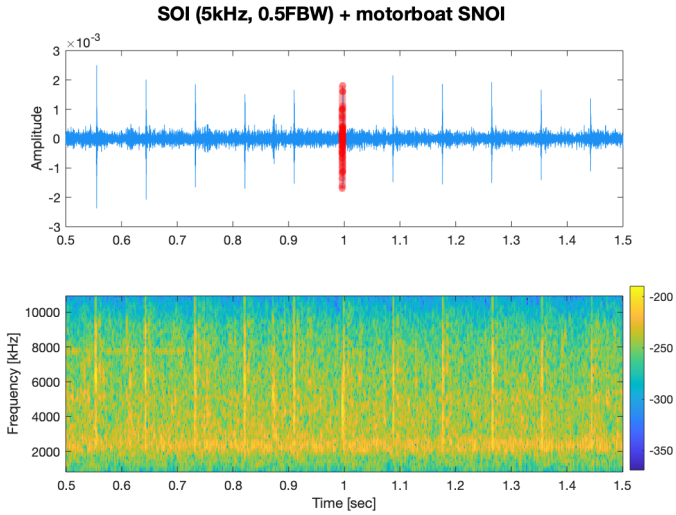


Figure 4.11: Plot of the first pulse (at time 1 second) of real experiment data with 5 kHz 0.5 FBW SOI and motorboat SONI. Top: Time-domain plot of received signal at receiver 1 bandpass-filtered between 2 kHz and 10 kHz. The SOI is illustrated in red. Bottom: Spectrogram representation of the top plot.

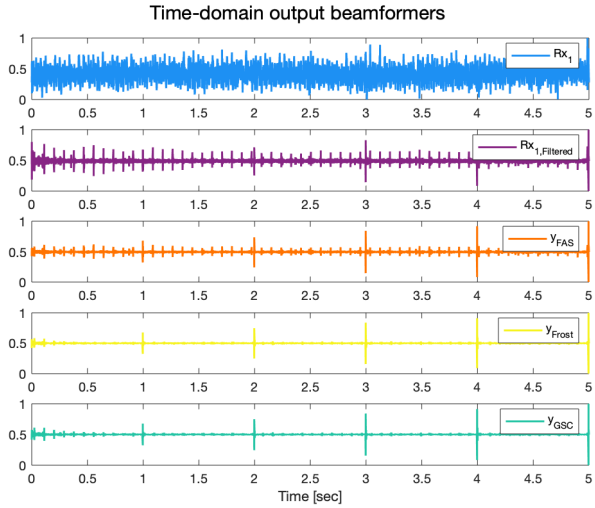


Figure 4.12: Time-domain signal shape comparison for real experiment. From top to bottom: Collected signal by receiver 1, band pass filtered collected signal by receiver 2, output FAS beamformer, output Frost's LCMV beamformer, output GSC beamformer. All signals are amplitude normalized between -1 and 1.

Figure 4.12 compares the shape of the time domain signals of the first five SOI pulses that arises every whole second. Observe the differences in SONI amplitude level between the different signals. It is clear that the two adaptive algorithms are able to suppress the motorboat noise to a greater degree than the FAS beamformer and single element bandpass filtering.

Zooming in on the first SOI pulse and the second SONI pulses after this SOI pulse in fig. 4.13, where the RMS normalized outputs are plotted alongside the equally normalized transmitter signal, a visual comparison of the different processing methods can be conducted. PR calulations is done between the same pulses stated above and presented in table 4.7. Followingly, PCC is calculated between the transmitter signal and the output of the different methods and presented in the same table. The calculation is done for the same time section as plotted in fig. 4.13.

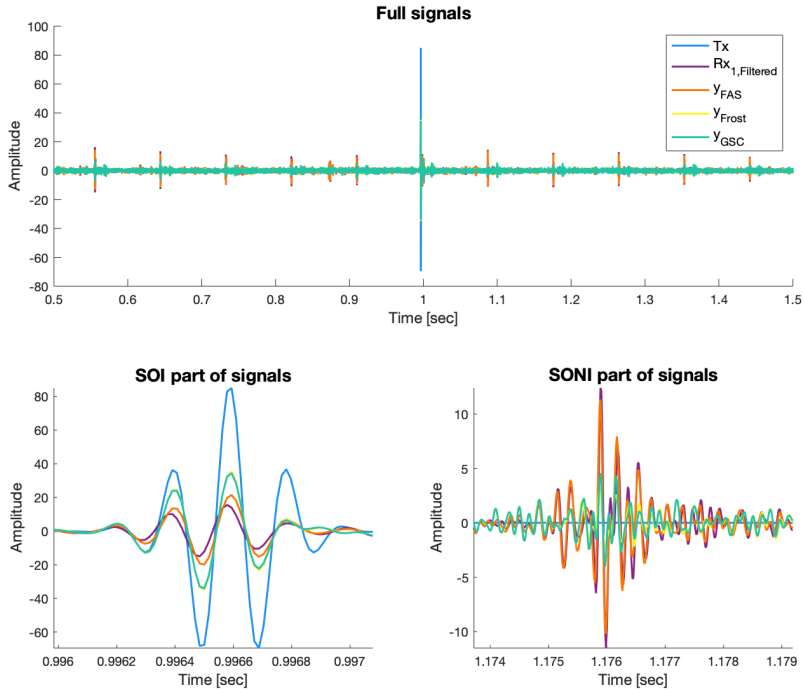


Figure 4.13: Results from real experiment. Tx is the signal transmitted from the source. $Rx_{1,Filtered}$ is the bandpass filtered receiver signal collected by receiver 1. y_{FAS}, y_{Frost} and y_{GSC} is the output of FAS-, Frost's LCMV- and GSC beamformers respectively. All signals are RMS normalized. Top: Full signals. Bottom left: SOI part signals. Bottom right: SONI part of signals zoomed in on the second SONI pulse after the SOI pulse in the top plot.

Table 4.7: PR and PCC between the Tx-signal and output signals for real experiment.

Signal	RCC	PR
Filtered Rx_1-signal	0.184	0.968
Output FAS	0.255	1.895
Output Frost	0.414	10.976
Output GSC	0.403	7.526

Chapter 5

Discussion

In this chapter a through investigation of the results, experiments, setup, algorithms and possible applications and improvements will be given. First, a broad overview over the performance of the different beamformers will be given. Afterwards, the results will be exhaustively dissected with the different fundamental theory in mind. Prolonging the discussion to the sources of errors and weaknesses in the conducted experiments before dwelling upon the applications within sea cage fish farming ending in suggestions for further work.

5.1 Beamformer performance

There are some reoccurring patterns in the results presented in chapter 4 for all the experiments conducted. The first and most obvious is that all of the experiments supports the theory that multiple array elements and beamforming are able to increase the SNR quality of noisy receiver signals compared to regular bandpass filtering of single element receiver signals. This observation is supported by the visual representation of the output signals as well as the PR and PCC measurements performed. It is also evident that the beamformers suppresses the interfering surface signal coming in from an angle different than the axisymmetrical listening direction, which is not the case for the single element measurement. Further interference suppression is apparent in the case of five virtual array elements compared to the virtual experiment with three elements. Which indicates that further improvements of the physical experiment could be the case if more hydrophones had been used.

Background noise levels are visually noticeable in the shape comparison plots (fig. 4.1, fig. 4.3 and fig. 4.12) as the amplitude between the pulses (both SOI and SONI). These levels decreases for all the processing techniques, but is most prominent for the beamformers.

There is also a difference in suppressing ability internally between the beamformers. The adaptive algorithms perform best in terms of interference suppression and SOI gain, which is most evident from the PR measurement where Frost's LCMV algorithm is superior in the physical experiment.There is therefore a significant advantage of using adaptive algorithms in environments similar to the experiment.

5.2 Array vs single element

In section 5.1 it was mentioned that that the array processing performed better than single element processing in terms of SNR improvement. By inspecting the the beampattern and directivity index plots in section 4.3.1 and section 4.3.2 it is clear that the ULAs will have largest gain in the look direction of 0° elevation (assuming no beam steering), and inherently lower gain in the other directions. This supports the theory, assuming isotropic receivers, that a single element respond equal in all directions whereas the ULA is assumed to inhabit increased axisymmetric directivity in the look direction and decreased directivity outside the main lobe. This phenomenon is supported by the virtual result in fig. 4.2 and fig. 4.4 where it is clear that the the SOI and SONI amplitudes are equal for the single element case while for the ULA the amplitude is higher for the SOI and lower for the SONI. The second thing deduced from the same figures is that the experiment with increased number of array elements inhabits increased directivity in the listening direction, and better ability to suppress the interference. Yet again, this is supported by the theory illustrated by the beampattern and DI plots in section 4.3.1 and section 4.3.2 showing slimmer mainlobe, higher listening direction directivity and lower sidelobe levels over all frequencies.

From the physical results presented, one can not say much about the mainlobe beamwidth of the array. To check the beamwidth, one could make a signal come in from different directions and observe where the gain from the signal increases its gain over a variety of frequencies.

Going back to the directivity analysis, two DI-frequency plots are shown in fig. 4.7 and fig. 4.10 for three and five array elements respectively. Notice how the DI in the look direction is quite steady in terms of array gain over all frequencies, while the DI in 30° elevation changes a lot for both arrays. In the case of three array elements and 30° angle, the DI decreases gradually until it drops fast around 10 kHz where it hits a null in between the the main-lobe and a side-lobe. The five element array has a DI that changes a lot more since the lobes are slimmer and the ULA will in general have more sidelobes with lower directivity level compared to three elements. Thus, the five element ULA's DI will hit the first null around 6 kHz at an angle of 30° elevation. Identifying this directional-frequency DI relationship allows for speculations about what happens in the time-domain for the two ULA cases as interference arrives at approximately 30° elevation while listening in 0°. The reduction of the interference contribution is linked with the decreasing lobe level for the three element case, and the null between the main lobe and and first side lobe for the five element case resulting in almost no SONI power left after beamforming.

As the array is designed for frequencies under 7.5 kHz one can expect grating lobes appearing for higher frequencies. This is not observed in the experiments mainly because the SOI operated within this limit with center frequency of 5 kHz and 0.5 FBW. Sidelobe formations that are equal or greater in gain compared to the main lobe, i.e. grating lobes, could cause deficient beamforming but this is not a problem in this thesis.

5.3 Fixed vs adaptive beamforming

By comparing the AG and PCC calculations for the virtual experiment with three array elements it is possible to see that the FAS beamformer performes inferior to Frost's LCMV and GSC beamformer, while it outperforms the adaptive beamformers in the five element case.

Inspecting the three elemental ULA results in fig. 4.2 it is apparent that the SONI amplitude of the FAS beamformer is larger than both LCMV and GSC beamformer. SOI amplitudes are relatively similar with a slightly higher amplitude in favor of the adaptive algorithms. Looking at the PR measurment verify this as well, where the adaptive beamformer PR is over twice the FAS beamformer PR. The inter-pulse noise level is quite similar for all the beamformers in this case as well. Thus, the correlation and AG measurements makes sense.

Why the adaptive algorithms are able to produce better results are tied to the adaptiveness of the algorithms provided by the LMS weight optimization problem embedded in the algorithms. Before elaborating on this, take a look at the experiment with five array elements. Here the SONI amplitude level for all the beamformers looks quite similar, see fig. 4.4, and the SOI level is in favour of the FAS beamformer. It is also possible to observe large amplitude fluctuations for the GSC beamformer just after the SOI pulse which is not seen for the other two beamformers. Since the GSC (and Frost's LCMV beamformer) updates the filter weights adaptively using the LMS algorithm utilizing gradient decent in search of optimal weights it is reason to believe that this amplitude fluctation is a result of the LMS learning rate choice which governs the "adaptiveness" of the weight update. This factor influences the responses of the adaptive beamformers in two different ways. Increasing the learning rate allows the adaptive beamformers to suppress the SONI better, as seen in the three element case, since it will authorizes the LMS algorithm to move toward a local minima faster and followingly be able to make fast changes. But then again, this makes the adaptive algorithms more sensitive to amplitude variations and can therefore use longer time to settle after a pulse is inflicted on it. This is more evident for the GSC than Frost's algorithm in the five element experiment. Remark that too large learning rate makes the adaptive weight adjustments unstable, and on the other hand, a too small learning rate makes the adaptive weight update slow and hence the beamformer will approximately act as a regular FAS beamformer with predefined filter response. Thus, in the case of five array elements it is natural to think that the GSC has been given a too large LMS step factor trading SONI suppression with local minima overstepping. Therefore, the GSC's amplitude fluctation behind the SOI pulse is the effect of local minima overstepping making the optimal weight recovery time long. The learning rate will also make the initial weight tuning process slow or fast depending on the initialization of the weights, start signal characteristics and the learning rate it self.

Comparing PR in the five element case indicates that the peak value ratio between the beamformers is largest for Frost's beamformer. Since there is not any major visual differences in the SOI and SONI region indicating that the PCC of the FAS beamformer should be better than the adaptive beamformers implies that the disparity must be between the pulses, i.e. the noisy region. It is hard to tell from the figures, but there is therefore reason to believe that the accumulated noise level for the adaptive algorithms in addition to the effects of overstepping makes the total PCC lesser compared with FAS, even though

the difference is quite small.

The beamformer performance in the five element case is, as elaborated on in section 5.2, tied to the incident angle of the SONI which indicate lobe-nulling in the frequency range of interest, thus making the beamformers inherently robust against noise/interference form that direction. Due to this ingrained feature, the adaptive algorithms will probably not be very affected by signal changes in this direction, and hence other directions may play a larger role in the LMS error.

Even though FAS do perform well in the five element ULA when the SONI arrives from $\sim 30^\circ$ it is not certain that it will perform equally good when the SONI arrives from a direction when a null is not present. This is why adaptive algorithms may be preferable unless one is designing a more fixed response array with known interference sources and directions and don't need the complexity of the adaptive algorithms.

When it comes to the distinction between Frost's LCMV and GSC, there are some points that were addressed when GSC was introduced in section 2.8.5 that should be confronted here. First of all, extensive learning rate tuning was neither performed or identified as a big concern in this thesis as the basic working methods and overall performance was of interest. Both algorithms may be exploited to perform even better if the right tuning is performed. When that is said, GSC inhabits a more general structure than the LCMV, thus playing with different structures, weights and blocking matrices may be profitable.

Filtering techniques in the most general form extracts information immersed in noise characterized by certain statistics, which has become evident in this thesis. Determination of the filter, fixed or adaptive, often follows two paradigms. (i) A priori knowledge of the noise field, or if this is unknown (ii) a priori knowledge about the characteristics of the SOI. LCMV and GSC mainly pursue the first paradigm, and are optimal for coherent noise by definition. In case of a high non-directional diffuse noise field it is observed that the performance of these methods do not differ greatly compared to FAS [36]. This fact can be tied to the interpulse noise suppression ability of all the beamformer looking quite similar in amplitude in the shape plots (fig. 4.1 and fig. 4.3) of the virtual experiments (with non-directional incoherent noise), and the fact that the SOI pulse is gained about the same for all the beamformers. The directional coherent noise part of the signal (i.e. SONI) is however suppressed better by the adaptive algorithms. As a sidenote: in highly reverberant environments with high diffuseness one could implement LCMV with adaptive constraints, called *linearly constrained adaptive beamforming with adaptive constraints*, by fusing the two paradigms together.

5.4 Physical vs virtual experiment

In the physical experiment, as for the three element virtual experiment, the adaptive beamformers outperforms the FAS beamformer. The correlation measurements in the physical environment are not as good as for the virtual analogue, which will be discussed later, but anyways, it is possible to see that the adaptive beamformers surpass the FAS performance to a greater extent than for the virtual experiment. Observe that the SOI and SONI amplitude level for the filtered single element receiver signal is roughly the same while the beamformers has higher SOI amplitude and lower SONI amplitude in fig. 4.13. All this supports the the results gathered in the virtual experiment with three elements as discussed

in section 5.2, thus referring to this section for elaboration on the results.

What differs the most about the results in the physical and virtual experiment is that the PCC and PR are quite different in favor of the adaptive algorithms. The disparity between the adaptive and fixed beamforming correlation results in the physical experiment is probably due to the poorly SONI suppression ability of the FAS. Since there are 11 clear interference pulses present in the audio signal of interest lasting for one second, see fig. 4.13, it is logical that the verdict of the SONI pulses plays a bigger role than in the virtual analog lasting for 0.3 seconds with only a single SONI pulse. The nature of the SONI in the physical experiment is also different than for the virtual case, and may contribute more to the signal variance/power. PR calculations suggests that the Frost beamformer is able to suppress the peak amplitude interference by a factor of 10.98 for the first SOI pulse and a factor of 1.90 for FAS. This is can be compared with a factor of 3.36 for Frost and 1.57 for FAS in the virtual case. This is comparable since the original peak values of SOI and SONI used to calculate PR in both cases is about the same, as seen by the PR of the filtered single element receiver signals being 0.97 and 1.11 in the physical and virtual experiment respectively. The reason why the physical experiment gives a higher PR is probably linked to the fact that the interference characteristics differs, the hydrophone radiation is not necessarily isotropic, the incident angles are not precisely the same due to the approximate position of the boat alongside sound refraction of the interference sound in the water. It can also be that the Frost's LCMV simply performs better in a combined diffuse environment.

The background noise present in the physical environment is most likely not white Gaussian noise as for the virtual experiments. The noise field is built up of surrounding noise sources such as water sloshing, boat and dock movement, distant and nearby motor noise, industrial noise from the nearby industrial area, wind etc. It is most likely a combined noise field with diffuse noise and coherent noise (such as motor noise) due to the nature of the marina and nearby noise sources. Therefore, the noise experienced by the different hydrophones might not be totally incoherent since the noise field might be somewhat directional. From the results obtained in this thesis, such directional noise effects has not been experienced other than motor noises from the motorboat and reverberations from multipath propagation of SOI and SONI. In fact one is able to observe the raise of reverberations as the power of the SOI increases. The incident angle of the reverberations can also be calculated from the phase delays between the receiver signals. This is an observation that will be present in an physical environment that allows for multipath signal propagation, but since this is not of great importance for the experiment in this thesis, no further calculations is conducted tied to this.

From the propagation velocity profile gathered by the CTD, it is clear that the velocity changes with depth. This is due to the salinity, temperature and consequently changes the density in the water making the propagation velocity not uniform with depth. The velocity changes most rapidly in the first few meters from the water surface before becoming more stable further down. This is probably due to the brackish water in the Trondheim fjord caused by freshwater inflow from rivers such as Nidelva which is backed by the CTD measurements. This factor is not taken into account in the code in this thesis, but because the SOI is arriving perpendicular to the array axis, i.e. propagating horizontally in the same depth, one can assume absence of large velocity changes in the propagation path.

5.5 Strengths and weaknesses with array processing

Looking at the results one is able to see various strengths that follows with array processing over single element processing. Array processing gives directionality to the signal reception which otherwise would not be present. Beam steering allows the operator to listen in different directions depending on the array structure and hardware specifications. Different kinds of array structures gives different types of directional performance, thus the design process is very important to obtain the desired performance. More array elements allows for slimmer mainlobe structures, increased directivity in the look direction, and decreased directivity in non-look-directions. Various types of beamformers exists for both narrowband- and wideband signals, and these are simple to implement in the software or directly in the hardware for increased calculation performance. The different types of beamformers can be chosen dependent on the wanted performance both computationally and output quality wise.

Even though there are several advantages with the use of arrays over single elements, there are some drawbacks and limitations that needs to be addressed. The first is that the the directivity is closely related to the design of the array. The spatial displacement between the array elements gives an upper bound on the frequencies of interest due to spatial aliasing described in section 2.4.2. The frequencies of interest are also lower bounded (to some extent) by the Fraunhofer distance stating the transition between the near field region and the far field region, unless the near field effects are accounted for. If it is desirable to listen to sounds not far from the array, the total extent of the array must be short enough to let the sound waves appear as plane waves at the array. Thus, both the interelemental spacing along side the total spatial extent of the array limits the available frequency interpretation in both ends of the spectrum.

Another disadvantage is the pricing. Underwater equipment is usually quite expensive. Processing power is also an factor that is multiplied by the number of array elements that is used. In an remote environment, where the array is supposed to autonomously log information, e.g. exposed fish farming sites, it can be desirable with system with low energy consumption for robustness. Synchronization is an factor that is extremely important to ensure good beamforming performance, thus a robust system needs to handle this properly. Certain array structures together with certain beamforming techniques are very sensitive changes in look-directions, thus, information about the listening direction is important in the design phase.

The design of the array used in this thesis ended up being an ULA which inhabits the axisymmetrical property following the design philosophy that this will give horizontal reception coverage within aquaculture sea cages with directional sound suppression from the surface and bottom. Increasing the number of array elements also allows for more flexibility in the array processing stage as multiple elements can be divided into subsets that are optimized for different kinds of signal frequencies, e.g. nested arrays. I.e. some elements may be used in multiple subsets, but the main difference between the subsets is the interelemental spacing allowing different types of beampatterns to emerge for the different subsets.

5.6 Sources of error and uncertainties

The hardware used in the physical environment experiments consisted of three hydrophones, where one of the hydrophones was of a newer model. Three different hydrophones means three different internal hardware. Even though the hydrophones most likely are tested and calibrated by the manufacturer differences in performance between the elements must be expected. An good example covered in section 3.2.4, described the flipped nature of the transducer signal for the newest hydrophone model compared to the two older ones. This discrepancy is accounted for in the data preparation, but the received signals will nevertheless differ due to the independent receiver hardware, self noise and surrounding noise field. As long as the signals are synced the SOI content will be strengthened as long as the distortion between the elements are within reasonable limits. The effects of multi elemental discrepancies is not measured and can therefor not be quantified. This effect is however presumed to be little due to the beamformer results acquired for both the physical and virtual experiment supporting each other.

Since the hydrophones have their own internal clock one should expect clock drift relative to each other. The physical experiment recorded data in 10 minutes intervals for three periods. Every recording contained 7 subsections of interest, and was split accordingly. Every recording subsection was then synchronised on sample level with the corresponding subsection gathered from the other hydrophones, leaving effect from clock drift minor and neglectable.

When it comes to the transmitter signals, there are several sources that can inflict distortion to the signal. The signal passes from the audio player through the amplifier and an impedance matcher before being radiated to the water by the underwater speaker. Unless the frequency response characteristics of the speaker is flat one can expect some signal distortion induced by the speaker since the signals contain a range of frequencies. The signal can also experience information loss in the signal transmission chain before arriving at the speaker. These are however considered minor. After launching into the water some signal distortion may occur due to propagation scattering, refraction, and attenuation. Sounds arriving at an angle relative to the hydrophone array axis will experience these effects more because of the longer travel distance between the elements and the differences in the propagation velocity profile. These effects are considerably important in the case of measuring sounds from fish since the motivation of this thesis is to understand the underlying content of fish sounds. These effects is seen as a limitation of the propagation channel since no signal reconstruction can be performed without a priori knowledge about the content of the original signal. In the case of the experiments in this thesis, it is easy to see that the signal content of the received audio reassembles the SOI but not completely. The reason for this is primarily tied to the character of the noise and how this propagates through the array processing. But, one cannot abandon the fact that distortion can happen in the medium and transmitter chain as well.

The most important sources of error in the physical environment experiment is liked to the experiment setup. Firstly, the transmitter source was lowered to the depth of the upper hydrophone element. It would probably be better to align it with the mid hydrophone to ensure the best possible perpendicular plane wave propagation to avoid phase delay between the array elements. This should however not be a big problem since the distance between the hydrophones and the speaker is much larger than the Fraunhofer distance, in

addition to the pre-processing synchronization.

Waves and water flow could make the distance and orientation of the speaker and hydrophones change during the experiment. These changes would be considered slow and would not contribute to large reception variations other than the neglectable time differences between the received pulses. If the weather/environment had been rougher, then this effect could be more observable.

The speaker and hydrophone array was not completely directed against each other due to the lack of rope length. This means that the incident azimuth angle of the incoming SOI was not optimal. But due to the axisymmetrical nature of the array this should not play a big deal. However, the array elements might not be isotropic as assumed in the experiment, and the sound reception might therefor be different in different directions. This non-isotropicness is probably the biggest flaw with the experiment, affecting the beampattern of the array. The effects of this is not explicitly apparent in the results, but could have greater consequences if beamsteering was involved. If beamsteering is needed, the single element beampattern and water column propagation velocity changes should be accounted for. No information about the icListen beampattern is found since the manufacturer do not provide this information in the data sheet.

Inaccurate measurements when building the hydrophone mounting frame is also an source of error. Having said that, since the array consists of only three elements, the effects of misplacement is not prominent.

5.7 Applications in fish farming and future work

The potential of this technology is considered high as the physical experiment have some promising results. Testing in a sea cage would be a natural next step. More knowledge about the sounds of salmon needs to be acquired, and the array should be designed thereafter. This could be an give and take interaction of recording and learning about the sounds over several iterations.

It would also be nice to make the code more robust with error handling and input parsing and cleaning the class structure. The code should also be tuned to explore its potential in different operational environments. Some suggestions for making the code more versatile is by adding more functionality such as:

- Audio partitioning functionality if large audio files are used. It is probably better to perform beamforming on smaller subsections of the data sequence at a time. This is also necessary in a real time application.

- Allowing for different hydrophone radiation diagrams, and embed this into the beamsteering fuctionality.

- Expanding the virtual environment using thermo-, halo- and pycnoclines to build propagation velocity profiles.

- Adding echoic simulation environments with reflective surface profile, bottom profile and objects in space.

- Testing with different (filter) weights, e.g. using Dolph-Chebyshev window to produce filter weights. This method is interesting because of its trait of lowering the highest sidelobes, at the expense of raising smaller sidelobes [10]. Hamming windows is currently used.
- Adding Cox's Regularization to the adaptive methods for robustness which allows for optimal solutions depending on the noise field type[4].
- Implementing frequency domain counterparts of the algorithms presented in this thesis.
- Implementing LCMV with adaptive constraints useful in high reverberation environments [36]. This could maybe be used in indoor fish tanks?
- Implementing null steering.

A good code library could potentially be merged with other sea cage models, e.g. fish schooling, that could potentially be used to simulate the soundscape within a sea cage which again could be used to predict behaviour.

Fish schooling recordings gathered from an axisymmetric ULA designs could potentially be very practical for determining group sound behaviour such as stress, hungriness, illness etc. and could followingly be used to make a detection or classification method treating the different situations. This could be used with traditional algorithms or using machine learning/deep learning. If one is able to isolate single fish sounds this could be used to detect and classify single fish behavior. Building a sound library could be used to train a *neural beamformer* (deep learning beamformer). Being able to listen to group sound and isolated fish sounds is a very good step towards automation following the precision fish farm and fish welfare mindset.

Combining beamforming with *voice activity detection* (VAD) can be useful to only record necessary audio. VAD is a technology commonly used in speech enhancement as an trigger functionality determining if a SOI is present or not. This can be a simple activation such as a amplitude threshold. Thus, VAD decides whether or not to process signals downstream (e.g. by a beamformer) silencing audio sections without SOI completely.

Exploration of the use of wideband beamforming using sensor delay-lines (spatial filters) [19] instead of temporal delay-lines such as FIR (temporal filters) could be interesting. This is an alternative to temporal filtering by expanding the physical array structure with more elements. Structures like this inhabits the same properties as regular temporal filtering based beamforming in addition that it overcome problems introduced by ultra-wideband systems requiring high speed circuits.

Further work on array design could also lead to some interesting applications. More elements could allow for slimmer main lobe and good beamsteering response that can follow the acoustical movement of the fish within the sea cage. Inspired by Norsonics Acoustic camera mentioned in the introduction, arrays generating onedirectional mainlobe with narrow beamwidth could potentially allow for acoustic tracking of single fish within the sea cage. Information such as swimming speed and direction could be readily available. Combining this with an camera could pinpoint what kind of fish makes what sounds, or pinpoint interference sources and block them using anti-jamming null-steering methodology. This technology could also be used in other areas for acoustical monitoring and

analysis. This could potentially include monitoring underwater structure sounds and detecting abnormalities, e.g. oil/gas lines, offshore windmills and other offshore equipment. detection of acoustic leakage or noise sources on ships that can indicate malfunctioning or to indicate unhealthy sound power levels with respect to aquatic animals. Bioacoustic applications such as listening to coral reefs or in precision fisheries listening for large schools of fish.

Chapter 6

Conclusion

In this thesis three hydrophones in a vertical ULA configuration submerged in water followed by wideband acoustic beamforming was tested to check its ability to passively listen in the horizontal plane while suppressing surface sounds in an marine environment. The beamforming methods used consisted of fixed FAS beamforming and adaptive LCMV- and GSC beamforming. Two virtual experiments was conducted using three and five array elements favoring the five element array for the purpose of increasing the the signal gain and suppressing directional noise/interference. A physical experiment validated the results from the virtual experiments, and proved that adaptive beamforming, and most notably Frost's LCMV beamformer, was able increase the SNR by gaining SOI coming from 0° elevation and suppressing SONI from $\sim 30^\circ$ elevation in addition to lowering the overall noise level compared to single element processing. Better SNR is expected using five hydrophones instead of three. The results suggests that this array processing design would be valuable and beneficial in acoustic fish school monitoring within sea cage aquaculture.

Bibliography

[1] Bertucci, F., Ruppé, L., Van Wassenbergh, S., Compère, P., Parmentier, E., 2014. New insights into the role of the pharyngeal jaw apparatus in the sound-producing mechanism of haemulon flavolineatum (haemulidae). Journal of Experimental Biology 217, 3862–3869.

[2] Charif, R.A., Clapham, P.J., Clark, C.W., 2001. Acoustic detections of singing humpback whales in deep waters off the british isles. Marine Mammal Science 17, 751–768.

[3] Clark, C.W., Charif, R., Mitchell, S., Colby, J., 1996. Distribution and behavior of the bowhead whale, balaena mysticetus, based on analysis of acoustic data collected during the 1993 spring migration off point barrow, alaska. Report-International Whaling Commission 46, 541–554.

[4] Cox, H., Zeskind, R., Owen, M., 1987. Robust adaptive beamforming. IEEE Transactions on Acoustics, Speech, and Signal Processing 35, 1365–1376.

[5] D'spain, G., Batchelor, H., 2006. Observations of biological choruses in the southern california bight: A chorus at midfrequencies. The Journal of the Acoustical Society of America 120, 1942–1955.

[6] Fish, M.P., Mowbray, W.H., 1970. Sounds of western North Atlantic fishes. A reference file of biological underwater sounds. Technical Report. RHODE ISLAND UNIV KINGSTON NARRAGANSETT MARINE LAB.

[7] Føre, M., Frank, K., Norton, T., Svendsen, E., Alfredsen, J.A., Dempster, T., Eguiraun, H., Watson, W., Stahl, A., Sunde, L.M., et al., 2018. Precision fish farming: A new framework to improve production in aquaculture. biosystems engineering 173, 176–193.

[8] Frost, O.L., 1972. An algorithm for linearly constrained adaptive array processing. Proceedings of the IEEE 60, 926–935.

[9] Griffiths, L., Jim, C., 1982. An alternative approach to linearly constrained adaptive beamforming. IEEE Transactions on antennas and propagation 30, 27–34.

[10] Harris, F.J., 1978. On the use of windows for harmonic analysis with the discrete fourier transform. Proceedings of the IEEE 66, 51–83.

[11] Hawkins, A., 1986. Underwater sound and fish behaviour. , 114–151URL: http://search.proquest.com/docview/14770578/.

[12] Holt, S.A., 2008. Distribution of red drum spawning sites identified by a towed hydrophone array. Transactions of the American Fisheries Society 137, 551–561.

[13] Kasumyan, A., 2008. Sounds and sound production in fishes. Journal of Ichthyology 48, 981–1030.

[14] Kasumyan, A., 2009. Acoustic signaling in fish. Journal of Ichthyology 49, 963–1020.

[15] Krim, H., Viberg, M., 1996. Two decades of array signal processing research: the parametric approach. IEEE Signal Processing Magazine 13, 67–94.

[16] Kuo, S.M., Morgan, D.R., 1996. Active noise control systems. New York: Wiley.

[17] Ladich, F., 2015. Sound communication in fishes. volume 4. Springer.

[18] Lichte, H., 1919. Über den einfluß horizontaler temperaturschichtung des seewassers auf die reichweite von unterwasserschallsignalen. 17, 385–389. Trans. by Wittenborn, A.F.: On the influence of horizontal temperature layers in the sea water on the range of underwater sound signals.

[19] Liu, W., Weiss, S., 2010. Wideband beamforming: concepts and techniques. volume 17. John Wiley & Sons.

[20] MathWorks, 2019a. Dsp system toolbox. URL: https://se.mathworks.com/products/dsp-system.html. (Accessed: 2020-05-19).

[21] MathWorks, 2019b. Phased array system toolbox. URL: https://se.mathworks.com/products/phased-array.html. (Accessed: 2020-05-19).

[22] MathWorks, 2019c. Signal processing toolbox. URL: https://se.mathworks.com/products/signal.html. (Accessed: 2020-05-19).

[23] Mungamuru, B., Aarabi, P., 2004. Enhanced sound localization. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 34, 1526–1540.

[24] Nebeker, N., 1998. Fifty years of signal processing: The ieee signal processing society and its technologies 1948-1998. The IEEE Signal Processing Society .

[25] Newton, I., 1803. Philosophiae naturalis principia mathematica: Vol 2. Knight & Compton. Trans. to English by Motte, Andrew: The Mathematical Principles of Natural Philosophy: Volume 2.

[26] Nocedal, J., Wright, S., 2006. Numerical optimization. Springer Science & Business Media.

[27] Norsonic, . Norsonic's acoustic camera. URL: https://web2.norsonic.com/product_single/324/. (Accessed: 2020-06-03).

[28] Nøttestad, L., 1998. Extensive gas bubble release in norwegian spring-spawning herring (clupea harengus) during predator avoidance. ICES Journal of Marine Science 55, 1133–1140.

[29] Piper, J.E., 2011. Beamforming narrowband and broadband signals. Sonar Systems .

[30] Saruwatari, H., Kawamura, T., Nishikawa, T., Lee, A., Shikano, K., 2006. Blind source separation based on a fast-convergence algorithm combining ica and beamforming. IEEE Transactions on Audio, speech, and language processing 14, 666–678.

[31] Shannon, C., 1949. Communication in the presence of noise. Proceedings of the IRE 37, 10–21.

[32] STADLER, J.H., 2002. Evidence for a hydrodynamic mechanism of sound production by courting males of the notchtongue goby, bathygobius curacao (metzelaar). Bioacoustics 13, 145–152.

[33] Statistisk sentralbyrå, a. 07326: Akvakultur. salg av slaktet matfisk, etter år. hele landet, laks, matfisk (1 000 kr). URL: https://www.ssb.no/statbank/table/07326/chartViewLine. (Accessed: 2019-10-24).

[34] Statistisk sentralbyrå, b. 07326: Akvakultur. salg av slaktet matfisk, etter år. hele landet, laks, matfisk (tonn). URL: https://www.ssb.no/statbank/table/07326/chartViewLine/. (Accessed: 2020-10-24).

[35] Takahashi, Y., Takatani, T., Osako, K., Saruwatari, H., Shikano, K., 2009. Blind spatial subtraction array for speech enhancement in noisy environment. IEEE transactions on audio, speech, and language processing 17, 650–664.

[36] Uncini, A., 2014. Fundamentals of Adaptive Signal Processing. Springer Publishing Company, Incorporated.

[37] University of Rhode Island, Inner Space Center, a. Between world war I and world war II: The 1920s and 1930s. URL: https://dosits.org/people-and-sound/history-of-underwater-acoustics/between-world-war-i-and-world-war-ii-the-1920s-and-1930s/. (Accessed: 2020-06-03).

[38] University of Rhode Island, Inner Space Center, b. World war ii: 1941-1945. URL: https://dosits.org/people-and-sound/history-of-underwater-acoustics/world-war-ii-1941-1945/. (Accessed: 2020-06-03).

[39] Urick, R.J., 1983. Principles of underwater sound. 3rd ed. ed., McGraw-Hill, New York.

[40] Valimaki, V., Laakso, T.I., 2000. Principles of fractional delay filters 6, 3870–3873.

[41] Valin, J.M., Michaud, F., Rouat, J., 2007. Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering. Robotics and Autonomous Systems 55, 216–228.

[42] Van Veen, B.D., Buckley, K.M., 1988. Beamforming: A versatile approach to spatial filtering. IEEE assp magazine 5, 4–24.

[43] Vester, H.I., Folkow, L.P., Blix, A., 2004. Click sounds produced by cod (gadus morhua). The Journal of the Acoustical Society of America 115, 914–919.

[44] Vøllestad, A., 2019. Fisk. URL: `https://snl.no/fisk`. (Accessed: 2020-05-31).

[45] Wahlberg, M., Westerberg, H., 2003. Sounds produced by herring (clupea harengus) bubble release. Aquatic Living Resources 16, 271–275.

[46] Webb, J.F., Fay, R.R., Popper, A.N., 2008. Fish bioacoustics. volume 32. Springer Science & Business Media.

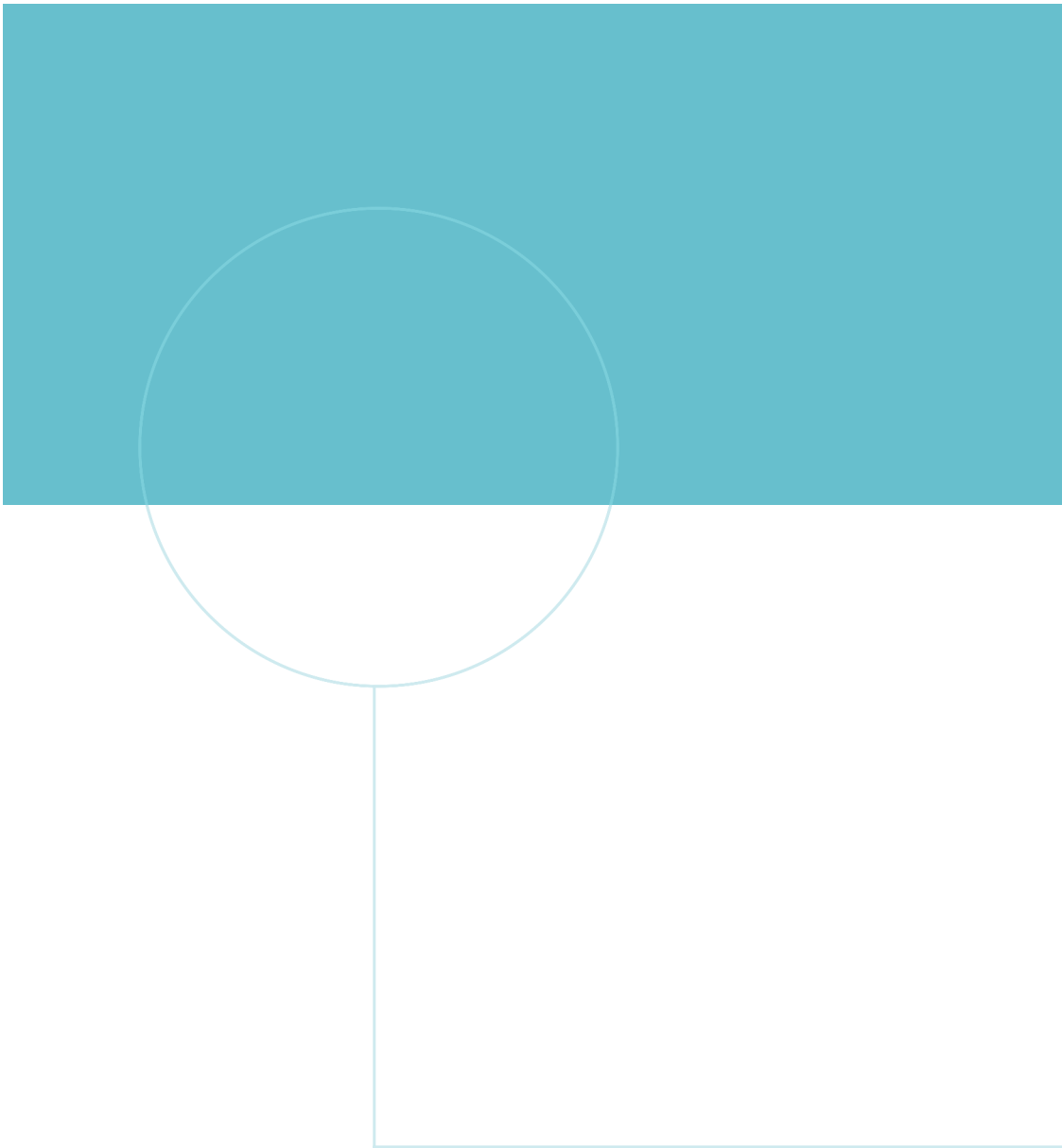
[47] Wenz, G.M., 1962. Acoustic ambient noise in the ocean: Spectra and sources. The Journal of the Acoustical Society of America 34, 1936–1956.

[48] Widrow, B., 1971. Adaptive filters, in: Aspects of network and system theory. Holt, Rinehart, and Winston. volume 687, p. 653.

[49] Widrow, B., Glover, J.R., McCool, J.M., Kaunitz, J., Williams, C.S., Hearn, R.H., Zeidler, J.R., Dong, J.E., Goodlin, R.C., 1975. Adaptive noise cancelling: Principles and applications. Proceedings of the IEEE 63, 1692–1716.

[50] Wilson, B., Batty, R.S., Dill, L.M., 2004. Pacific and atlantic herring produce burst pulse sounds. Proceedings of the Royal Society of London. Series B: Biological Sciences 271, S95–S97.

[51] Yan, S., 2019. Broadband Array Processing. volume 17 of *Springer Topics in Signal Processing*. Springer Singapore, Singapore.



NTNU

Norwegian University of
Science and Technology