An Assurance Metric and Robustness Evaluation of a Low-cost Acoustic Beamformer for Source Localization

Thomas C. Coleman

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Masters of Science
in
Mechanical Engineering

Steve C. Southward, Chair
Alfred L. Wicks
Cameron D. Patterson

April 27, 2018
Blacksburg, Virginia

Keywords: Metric Evaluation, Acoustic Beamforming, Source Localization
Copyright 2018, Thomas C. Coleman

An Assurance Metric and Robustness Evaluation of a Low-cost Acoustic Beamformer for Source Localization

Thomas C. Coleman

(ABSTRACT)

A rise in interest for service robotic rovers produces a need for a low-cost method for source localization in order for a prospective robotic unit to engage with a human operator. This study examines the use of the LMS algorithm for constructing a beamformer using an optimized Weiner filter solution for this source localization application and evaluates the robustness of a developed characterization method for assuring that a proper approximation for the desired signal is achieved. The method presented in this paper encompasses using a filter and sum method in which the sums are generated for a selected set of filter angles, and this set of sums are compared and characterized to produce a selection for an approximate arrival angle from the sound source to the microphone array. These filters are adaptively trained offline using a generated desired signal chirp to represent the average human whistle and a training data set for each of the four possible room configurations. This method was tested to determine if a selected filter configuration could still produce viable outputs for scenarios in which the testing room had been changed, whether noise was injected into the testing environment, if two or three microphones were used in testing process, and whether the filter angles are aligned with the arrival angles of the signal. Results on the robustness of the adaptive LMS beamformer are presented. Limitations of the system performance are discussed and possible solutions for results that have undesired performance are given in future work.

An Assurance Metric and Robustness Evaluation of a Low-cost Acoustic Beamformer for Source Localization

Thomas C. Coleman

(GENERAL AUDIENCE ABSTRACT)

A rise in interest for service robotic rovers produces a need for a low-cost method for locating a sound source so that a potential service robot can interact with a human operator. In this study, a beamformer is implemented to approximate a direction for the sound source. This beamformer is comprised of a set of trained filters for the designed microphone array. These filters were trained based on three training conditions of training room, the number of microphones used, and whether additive or ambient noise is used during training. The training signal for the filters consisted of a chirp from 1 to 2.5 kHz to mimic a portion of the human whistling spectrum. Once trained, these beamformers were then given data from separate tests to determine if a distinct and correct approximation could be determined. This paper suggests a method to use the correlation of each beamformer to the training signal to determine both the maximum correlated beamformer and whether correlation is distinct from greater than the other beamformers examined. These results are finally examined under an ANOVA and percent difference process to determine if the three training conditions improve the average prediction of the angle of arrival of the source signal for the generated beamformers.

(ACKNOWLEDGMENTS)

I would first like to thank my thesis advisor Prof. Steve Southward of the Mechanical Engineering Department at Virginia Polytechnic Institute and State University. Prof. Southward was always thrilled to speak with me when I ran into a trouble spot or had a question about my research or writing. He consistently allowed this paper to be my own work, but steered me in the right the direction whenever he thought I needed it.

I would also like to thank my committee members, Prof. Alfred Wicks and Prof. Cameron Patterson, for taking the time to read my thesis. Without the words of wisdom that you both provided, I would not be able to present this topic today. I am gratefully indebted to both of you for your valuable comments on this thesis.

With a special mention, I would also like to acknowledge my colleagues in the Mechatronics Lab at Virginia Polytechnic Institute and State University. Without you all, I would still be stuck trying to build, test, and explain a vast majority of this thesis. I thank you for your friendship and kindness that you have shown me over this trying time it took to complete this paper.

Finally, I owe my deepest gratitude to my wife, Kaitlyn, who has had more patience dealing with my skills as a writer and my enthusiasm for finding distractions to read my writing and give constructive feedback to improve the overall outcome of the paper. Thank you for giving me the space, the love, and the encouragement to make this paper understandable to those beyond myself.

Contents

List of Figures

List of Tables

Chapter 1

Introduction and Background

This section will cover the problem and motivation for this study. A brief literature review of the experiments and research relating to this topic will be examined to understand problems faced by other groups and strategies suggested for overcoming issues. At the end of this section, the thesis structure is detailed for the materials covered in this study.

1.1 Problem Statement and Motivation

This study is focused on producing results for a low cost microphone array for an unmanned ground vehicle (UGV) that would be used as a home sentry. The UGV would implement a microphone array beamformer based on results generated through this study. The microphone array beamformer would need to detect a given signal in order to alert the robotic unit and determine the direction from which this signal originated. This UGV would act as a home sentry and service rover for a human operator. Over the past 20 years, interest has arisen for service robotic units that are capable of interacting with a human operator through a given command [**?**]. This command is usually selected as either an impulse signal such as a snap of the fingers or a hand clap, a whistle, or speech signature that can be interpreted by the processing unit and can be easily performed by the average human operator [**?**][**?**]. For this study, the command was ultimately decided as a whistle of

a given frequency bandwidth to signal the UGV for attention.

Additionally, this experiment was requested to employ a method of detecting the desired whistle using inexpensive electret microphones. These types of microphones are implemented to show that a method exists to detect the desired whistle for equipment that is readily accessible to the public. The reasoning for this demonstration is to inspire other researchers to examine methods selected for this paper and implement those methods easily with readily available and cost effective materials. Normally, the equipment used in source localization and detection experiments are associated with a large cost, configurable microphones and computationally expensive processors and memory, which dissuade researchers from implementing real systems.

The final criteria for this study aims to evaluate the robustness of the chosen beamforming and selection methods to ensure that the microphone array beamformer can be utilized in a breadth of testing scenarios. The validation of the robustness of the selected beamforming method can be examined for two areas: the effectiveness of the beamformer to predict the correct source direction given scenarios outside of the trained data and the impact of varying one or multiple parameters of training and validation scenarios. The desired effect of this study is to determine if the selected method is both robust and cost-effective to implement for a realistic system.

1.2 Review of Microphone Arrays

A microphone array is defined as a set of acoustic sensors which are organized in an arrangement to sample the surrounding environment, with the microphones being separated into some geometry. As a result, the outputs of the array contain several signals, ranging from the signal of interest, to noise, to the propagation information from the sources to the microphones of the array. With these signals, a researcher can employ a microphone array for applications in localizing sound sources, suppressing ambient noise, separating different sound sources, and more [**?**]. A microphone array can be assigned to one of three configuration types used to search space depending on the desired viewing area of the operator.

These three types of microphone arrays can be described as a linear, planar, or volumetric array [**?**]. A linear array is when the microphones exist in a linear pattern and is primarily used for detecting sources that are directly in front of the array of microphones. Example applications for a linear array would be for amplifying sound in a singular point for a sports program or detecting a signal when a robot is facing an operator. In a planar array the microphones are arranged in a singular plane to detect a signal that is on one side of the microphone plane. Example applications for the planar arrays can be for teleconferencing or radar and sonar applications. The volumetric array is an array of microphones that is constructed in a three dimensional space. These arrays are used for detecting signals where not only the azimuth of the source changes but so does the altitude. For all of these cases, the arrays can be set to uniform, non-uniform, and random spacing. The spacing chosen can affect how the incoming signal is received, thereby changing how the array or beamformer can behave [**?**].

An array that implements a series of microphones in one of the described types of array configurations can be referred to as a differential microphone array (DMA). DMAs are microphone arrays that are designed to respond to spatial derivatives of an acoustic pressure field. The DMA was first used in the early 1930s, where omnidirectional pressure sensors and gradient ribbon microphones were combined to produce beam patterns such as dipole, cardioid, hypercardioid, and supercardioid [**?**]. One reason a DMA is used over a directional microphone is that directional microphones, when first created, were fixed to a specific directional response. This means that a directional microphone was not configurable, so a different microphone must be used in order to have a different directional response. Another favorable characteristic of a DMA is that a frequency-invariant beam pattern can be formed to process broadband signals. Additionally, DMAs are effective for both high and low frequencies with a small design package. Finally, the small area geometry of a DMA can have the potential to attain maximum directional gain for a given number of sensors. A DMA can be configured to respond for a specified set of directional responses depending on which array geometry and digital signal processing (DSP) technique is implemented for the microphone array [**?**]. This DSP technique creates what is referred to as a beamformer.

1.3 Review of Beamformers

A beamformer refers to the construction of spatial filters that are capable of receiving a signal from a desired direction, or set of directions, and attenuates the remaining signals from other directions. The concept for beamforming can be used in applications of both radiation and reception, but, this study focuses on reception [**?**]. These beamformers are the processing algorithm that microphone arrays use in order to complete the aforementioned applications [**?**]. This assertion means that beamformers can be used in a wide breadth of fields, ranging from radar and sonar, to telecommunications, to medical monitoring, depending on the parameters used to construct the beamformer [**?**][**?**].

These beamformers are capable of being used in at least two major categories of approach from the literature review of this study. The first and most popular category to implement is a time delay estimation (TDE), or time of delay, analysis on the incoming acoustic source. This approach involves finding the delay between when the signal was received from two or more microphones to determine the approximate angle of arrival [**?**][**?**]. When a signal source releases a signal that is desired, the sound wave, over a distance, turns into a plane wave if the range of the signal source is considered to be far field. This plane wave reaches the microphones on the array at different times, so a delay can be found between the microphones used [**?**]. This technique has multiple methods to both find said delay and process the delay to determine the angle of arrival. From the experiments observed, researchers have used a cross spectrum analysis to compare the relative delay between microphone signals [**?**], a cross correlation analysis with recurrent neural networks to train the beamformers based on received data [**?**], a model of sampling delays relating to the geometry of microphones on the array [**?**][**?**], and a digital comparative counter between microphones to determine the delay with an array model [**?**] to approximate the angle of arrival.

Using the TDE approach has the benefit that this method does not require the storage of filters. Additionally, this TDE concept is sufficient for passively determining where a signal originates, and the concept can be altered to an active process if the array is desired

to sweep certain areas. However, this method of determining the source location comes with several disadvantages. The first disadvantage is that most of the reviewed experiments required models of the microphone area geometry to be constructed to use the found delay to determine the angle of arrival. This limits the microphone arrays to be built using microphones that are uniformly spaced and similar in performance. As stated in Section 1.2, if a microphone in the array is unevenly spaced, the model becomes significantly more difficult to derive. Additionally, if a microphone in the array begins to fail and needs to be replaced, the new microphone will need to be checked against the currently used microphones to ensure a similar performance [**?**]. Another disadvantage is that this approach is highly susceptible to the reverberation inside the room where the system is implemented, as the system expects a singular plane wave to reach the array [**?**]. These methods for the TDE approach are effective at determining the angle of arrival of when finding a signal delay between microphones, but these methods do have obstacles with room reverberation and the limitation in array design.

The other dominant category used for beamforming can be described as a filter and sum method. The filter and sum method involves using a finite impulse-response filter for each sensor signal to train a filter response per channel. These filter responses are added into a single signal which is compared to a desired output signal [**?**]. The combination of the outputs from multiple microphones provides spatial selectivity, which can assist with suppressing noise and interference [**?**]. These filters are highly directional depending on the amount and quality of the microphones used and can be used for suppressing noise if the training signal does not include noise [**?**][**?**]. To train these filters, there are multiple methods one could implement find an optimal solution. We will mention two methods that use the least square error approach as the main proponent for training the filters.

The first training processes is referred to as the regressive least square (RLS) algorithm, which was first introduced in the mid 1950s following the release of a set of papers involving least square estimate [**?**]. This process involves implementing an estimate for filters and each data point collected to build an initial guess. This algorithm is able to update as more data is received and previous data that is uncorrelated to the current data can be re-

moved from the training of the filter weights by using a forgetting factor. The RLS algorithm allows for the tracking of a point of interest, similar to the Kalman filtering process [**?**]. The RLS algorithm performs best when a substantial amount of test data can be used to train and update the filters. Usually, the RLS algorithm requires at least 40 data samples and is process intensive as an inversion is required at the end each time a regression is taken. [**?**] While the RLS method of training filters can be robust given the right conditions, the RLS method is not computationally efficient which makes this method difficult to suggest for use on a platform where computational performance is a concern.

The next process is the least mean square (LMS) solution. This process takes a set of data and minimizes the error between the desired signal and the summed filtered responses. This minimization occurs because the received data is iteratively run over the weighted filter vector to produce a close approximation of the desired signal. Taking the LMS algorithm and adding some constraints for finding the minimum error can produce a different types of commonly used beamformers. If the LMS algorithm uses a linear constraint, then this process would construct a Frost beamformer which is one of the more commonly used beamformers [**?**][**?**]. The LMS algorithm approach is more computationally efficient compared to the RLS method of training filters. A drawback of using LMS comes from the rate of convergence on a minimum error due to the spread of eigenvalues and if multiple interference sources are present in the environment. [**?**]

Both the RLS and LMS algorithms use a data set and a gradient approach to solve for a minimum error between the received signals and the training signal. The Nokia Research Center performed an analysis using the filter and sum method and found that this approach resulted in an easily constructed and controllable beamformer. These results were found to resemble the theoretical responses well and achieved a great sensitivity for desired response [**?**]. From these results, a beamformer is described as a directional amplifier for a given given search space. When using the filter and sum approach, the generated filters will have two issues caused by the direction amplification. First, a high degree of spatial selection may affect the ability to monitor the broad scene within the search space of the array. For the second issue, the single output of the beamformer removes the cues from the array of when

signals arrive which can help with locating and segregating sound sources [**?**].

With these filter training algorithms, the beamformers can be constructed to be either fixed or adaptive in nature. A fixed beamformer is one that has time invariant coefficients, while an adaptive beamformer is continuously changing based on received array data. From a general standpoint, fixed beamformers are easier to construct as no data is required, but fixed beamformers are not as efficient as its adaptive counterpart when dealing with room reverberation, environmental noise, and competing sources [**?**]. The reason that the adaptive beamforming process is better suited than the fixed beamformer process is that the adaptive process uses real data which contains interference signals from noise and reverberation which be used to enhance the desired signal response of the beamformer. This observation means that using a filter and sum approach when performing an application with a microphone array can have a significant impact on performance when noise and interference are in the environment. The adaptive training process can also be performed as adaptive offline or adaptive online in real-time. Offline adaptive training refers to collecting training data prior to training the filters to create a beamformer for a specified scenario. Once these filters are trained, the filters are not updated unless another training data set is introduced to retrain the filters. Online adaptive filter training is the opposite where data is collected as the beamformer is used to consistently train the system. While the online adaptive method is more robust than the offline method, using the offline method of adaptively training the filters results in a more computationally efficient means of filter training.

1.4 Thesis Organization

This paper covers the construction, characterization, and evaluation of a low-cost beamformer response. For this study, suggestions will be made on how well the selected beamforming approach works for an assortment of test scenarios and what improvements can be made. This study is intended to examine only one beamforming algorithm and then perform a robustness evaluation and metric characterization of the beamforming outputs. Chapter 2 explains the formation of the desired excitation signal, the implemented micro-

phone array platform design, how the beamformer was constructed and evaluated on training data, and the hypothesis of what effects may be seen by the system during the review of the results. Chapter 3 introduces the the process designed to characterize the results of the beamformers to confirm if the desired signal was detected and in which direction if the signal was detected. Chapter 4 will cover the examination of the results for possible conclusions based on the tests performed.

Chapter 2

Empirical Testing and Design of the Beamformer

This section will discuss the decisions made with respect to the design, modification, and testing of the experiment and equipment used. Additionally, this section will also discuss how each of the beamformer conditions are varied in context to the microphone array and give the hypothesis for each variation of the beamformer conditions in the overall impact to the performance of the filter.

2.1 Experimental Test Configuration

For this experiment, a few constraints need to be specified in order to establish a means of determining the impact of changing filter variables on the performance of a filter bank given a specified excitation signal. Setting these constraints allows for a repeatable process where slight alterations can be made to the microphone array, the environment surrounding the array, or both during experimentation. These constraints include: what type of excitation signal is used, the frequency range to be processed by the microphone array, the design of the microphone array, and the types of tests that will be recorded by the microphone array during experimentation.

As stated in Section 1.1, the purpose of this experiment is to create and evaluate a microphone array beamformer that is capable of being used on a robotic service unit and home sentry that a human operator can signal with an audio command. For this reason, a common excitation signal, such as a whistle, is needed in order to signal to the mentioned unit for attention, like one would perform for a dog. There are many choices that one could use for the excitation signal such as a hand clap, the snapping of fingers, the speaking of a particular word or phrase, or a whistle. For this experiement, the excitation signal was requested be a whistle as this type signal would allow for a specific frequency range to be observed rather than an impulse input, such as the clapping of hands or snapping fingers, or a complex audio signature, such as speaking a word or phrase, for requesting attention. The excitation signal needs to be in a frequency range that is both capable of being produced naturally by a human and within a frequency band that is capable of being recorded by the inexpensive microphones selected for this study. The microphones used in this study were BOB-09964 microphones, which are inexpensive omnidirectional pressure foil microphones that are used mainly for hobby level activities and normally cost under five dollars per microphone [**?**]. These microphones were selected due to their low cost and simplicity of use. Unfortunately, this low cost and simplicity also result in a lower quality and the requirement for calibration. This requirement for calibration is the reason behind why a beamformer is required for this study. By training a beamformer, the microphones will need to be calibrated in order to form the beamformer correctly [**?**]. This experiment is to evaluate how robust a beamformer can be made when using lower quality electret microphones. Figure 2.1 illustrates the typical frequency response bounds of the BOB-09964 microphone used in this study from data provided by the the manufacturer [**?**]. The green highlighted area displays the normal magnitude responses that can be found from this model of microphone. These microphones have a consistent response for the input frequency ranges of 100 to 800 Hz and 3 to 5 kHz. These frequency bands have a magnitude response range of 6 dB and 11 dB for the input frequency ranges of 100 to 800 Hz and 3 to 5 kHz respectively.

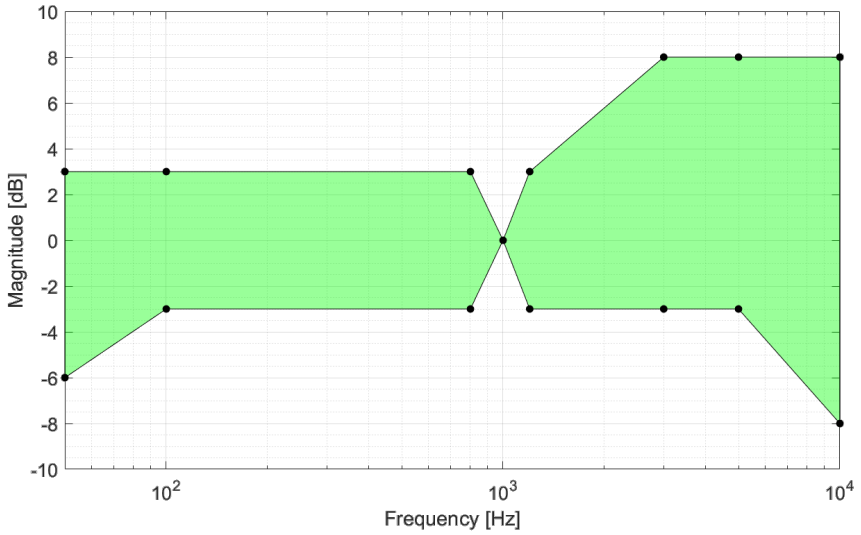
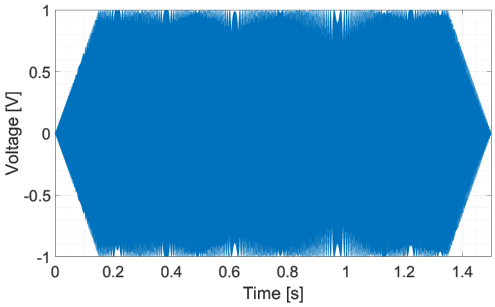


Figure 2.1: Typical microphone frequency response of the BOB-09964 microphones showing the upper and lower relative responses for a given input frequency

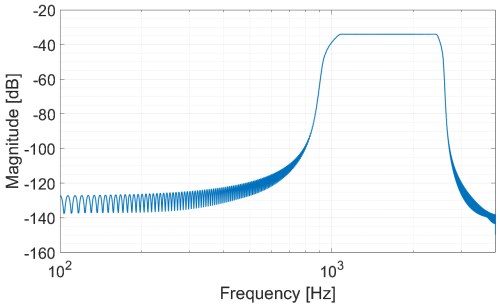
From a study performed by the Blekinge Institute of Technology, an average human was found to produce a whistle approximately between 500 to 5000 Hz [**?**]. The excitation signal should represent at minimum a portion of this frequency band and the represented frequency should have a majority of the response and a consistent magnitude over the frequency band. From the Blekinge study, the power of a human whistle appears to have magnitude response of -30 to -35 dB over a freqeuncy range of 1 to 2 kHz. After this frequency range, there is a decrease of the magnitude response to the noise floor, approximately -50 dB, at 3 kHz [**?**]. The frequency range from approximately 1 to 2.5 kHz was found to have a magnitude consistently above the -45 dB noise floor for a typical human whistle. This observation means that the generated excitation signal should be in the frequency range of 1 to 2.5 kHz. The 1 to 2.5 kHz frequency range was originally requested for the excitation signal, but the frequency was confirmed have both a consistent response and represent a noticeable portion of the average human whistle spectrum by the data reported from the Blekinge study. For these reasons, the excitation signal for this study was selected to be a chirp signal from 1 to 2.5 kHz.

To best represent this 1 to 2.5 kHz frequency range, a chirp signal was generated to fit this frequency band. Figure 2.2 illustrates a sample of the excitation signal with a tapering

window placed around the range from 1 to 2.5 kHz and the magnitude of the response in the frequency domain. Here, the excitation signal can be seen to have the full magnitude of -30 dB beginning at 1 kHz and decrease to the noise floor at 2.5 kHz which was desired. This excitation signal will be used for the training of the filters in Section 2.2 as the training signal for the algorithms.



(a) Time series data of chirp signal



(b) Frequency spectrum data of chirp signal

Figure 2.2: Plots of the of the proposed excitation signal for the (a) time series data for two seconds and (b) frequency series data

The microphone array was designed to represent a natural form (eg the head of a human or animal) with a configurable microphone placement. As mentioned in Section 1.2, the number of receivers (microphones or ears) used in an array can have a significant effect on the effectiveness of the array when implemented. Figure 2.3 displays the microphone array that was implemented for this study. The array was selected to have three microphones to determine the difference in filter performance between two and three channel filter configurations. A fourth channel was not implemented because the oscilloscope used for the experiment has four inputs where one input was used to trigger the oscilloscope to gather data. For consistency, when discussing which microphones are being observed, the microphones are numbered from one to three starting from the top left corner, moving clockwise as shown in Figure 2.3. Using the three channel configuration as the primary structure for the array, the microphones were placed with 150° of separation between M_1 and M_3 and M_2 and M_3. The front two microphones, M_1 and M_2, have a separation 60° represented as the 2ϕ in Figure 2.3. The center point of the microphones exists behind M_3 to create a non-radial array to determine if the difference between two microphone configurations plays a role in the

performance of the beamformers. Microphones one and two are separated by approximately nine inches in the illustrated configuration, while the distance between microphone three and microphones one and two was approximately four inches. This array is attached to a camera tripod with angle measurements in the yaw of the microphone array with a resolution of six degrees. Each microphone was surrounded by foam covered cardboard tubes with a plastic covering on the back. This implementation was used to assist in directing the sound input to each microphone and consists of easy to obtain materials. For data processing, the microphones for this array were connected to a 4000 series PicoScope oscilloscope that would collect data from the three microphone outputs and an external trigger signal. The signal from each microphone was gathered at 14 bits of resolution at a sample rate of 10 kHz.

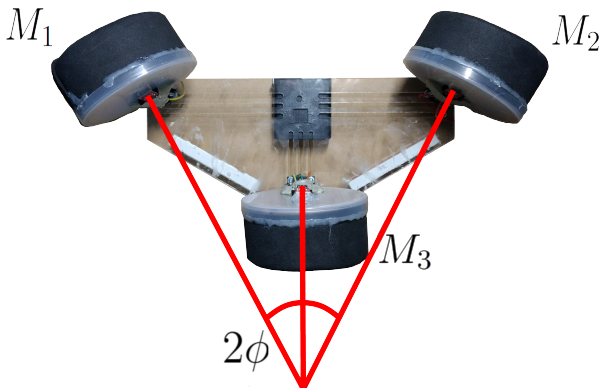


Figure 2.3: Top down representation of the microphone array used in this study

For testing, the unit was placed into both small and large classrooms on the Virginia Tech campus. An illustration of the placement of the speakers and microphone array, as well as a picture of the small testing room can be seen in Figure 2.4. In both rooms in which the system was tested, there were four speakers placed around the microphone array with three speakers, marked as NSpk1, NSpk2, and NSpk3, used to emit white noise from approximately 36 inches away from the center of the microphone array at 90°, 180°, and 270° from the starting position and a single speaker for the excitation signal, makred as SigSpk, located approximately 80 inches away from the center of the microphone array at zero degrees from the starting position. The zero position for the array in relation to the speakers can be seen in Figure 2.4b. The excitation speaker was placed next to the wall as this placement would reduce the amount of reverberation from the wall directly behind

the excitation speaker as the signal is sent [?]. This testing configuration allows noise to be injected into the environment surrounding the array using the remaining three speakers. The small room was approximately 20 by 45 feet while the large room was approximately 80 by 80 feet. The reason the rooms were selected was due to the differences in size and geometry. Both of these rooms can produce reverberations within the room that can affect the microphone array.

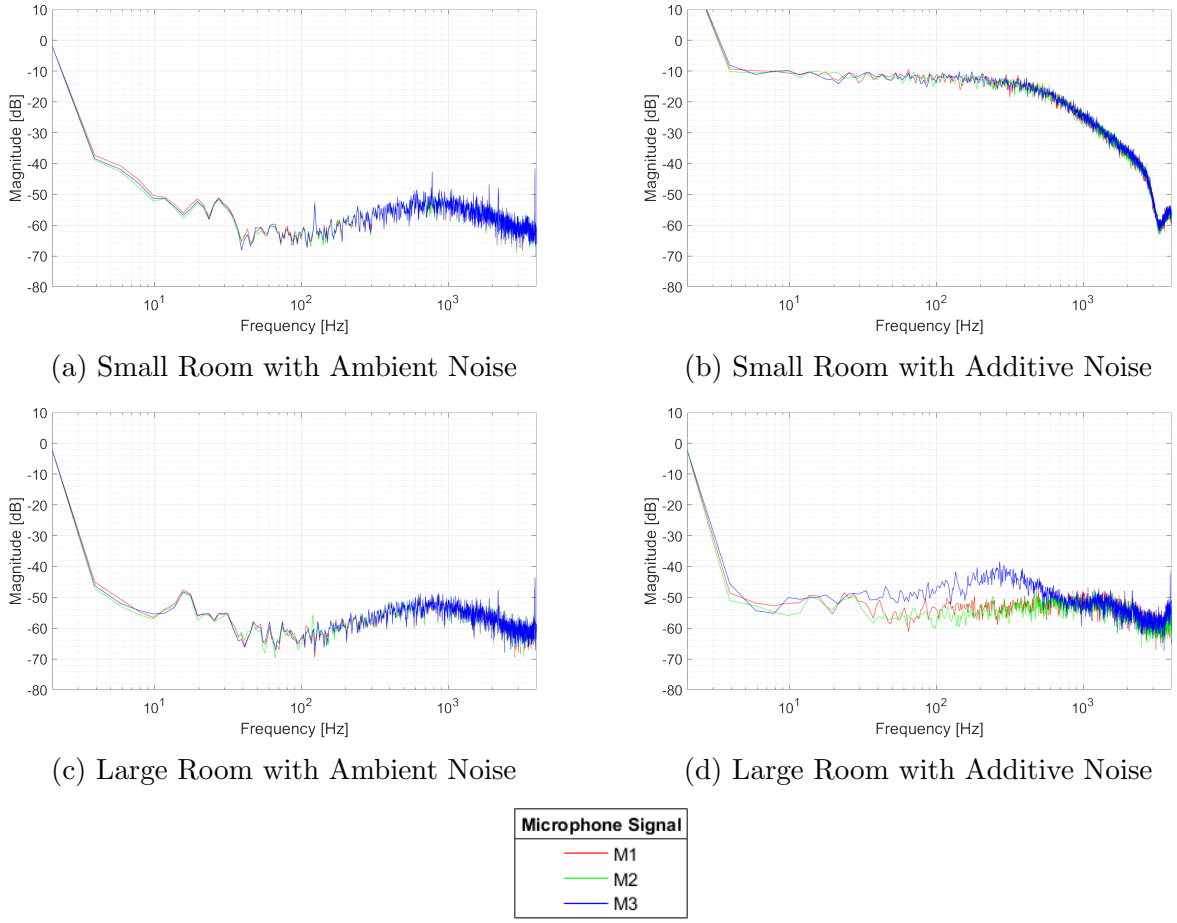
(a) Small Room Layout (b) Concept of Room Layout

Figure 2.4: Pictures of (a) the small room testing layout and (b) an illustration of the speaker and array layout at home position

When beginning a new test, the microphone array was set to zero degrees in which microphone three is facing directly away from from the excitation speaker. The array was set to collect data when the external trigger signal was received from the same computer that produces the excitation, or training, signal. When the trigger signal was received, two seconds of data, or 20000 sample points, was collected for either in training the filters of the beamformer or validating the beamformer filters. The array was rotated counterclockwise 45° until reaching the designated mark on the test stand. Then the excitation signal was replayed under the same environmental test conditions. This process was repeated until the array had been rotated 360° in 45° increments.

As mentioned previously, this experiment aimed to examine the differences between filters trained with and without noise. Noise was chosen as a condition to vary throughout testing as issues from noise propagation and room reverberation can cause complications

when determining the location of the signal source by masking the excitation. The training of the beamformer with or without noise being added can resolve some of these issues by suppressing noise outside a specified range depending on the beamformer training algorithm selected. [**?**] When a excitation signal is masked by interference from either another sound or noise source, this reults in a phenomenon in humans that can be referred to as the "cocktail party" effect. [**?**] The "cocktail" party effect is best described as trying to listen to a guest at a cocktail party while other guests are speaking to one another. The desired signal, the words from your guest, can become hard to understand or hear, but the brain filters out signals not corresponding to the particular signal of interest. This filtering makes the desired signal the focus which causes the signal to be heard and processed. As with humans, noise within a system can cause multiple problems for a beamformer, especially when the signal to noise ratio (SNR) for a desired range is below 10 dB [**?**]. The condition of whether noise is added or not to the environment during training was observed to determine if training with noise improved the performance of the beamformer. If no noise is added during training or validation, then the noise condition can be referred to as "ambient noise" since only the ambient noise of the testing room is used. When noise is added to the environment using the three noise speakers, the condition is referred to as "additive noise" for this study because noise has been added to ambient noise found in the testing rooms. For data taken for use in training or validating the filters using noise, the three designate noise speakers were given separate noise signals of pseudo white noise that is uncorrelated to the other two noise signals and the excitation signal. Figure 2.5 contains the frequency magnitude of the four different noise signals that the microphone array could be exposed to during testing. The data for both rooms ambient noise was collected prior to testing to observe the levels that existed within each respective room. The additive noise signals were specifically calibrated to produce a noise magnitude of -45 dB in the testing rooms. This noise magnitude is greater than the ambient noise level of approximately -60 dB for both rooms. The SNR for these noise cases was found to be on average 12.5 dB across the the excitation range of 1 to 2.5 kHz.



(a) Small Room with Ambient Noise

(b) Small Room with Additive Noise

(c) Large Room with Ambient Noise

(d) Large Room with Additive Noise

Figure 2.5: Spectral magnitude data for the noise that can be seen by the microphone array

The microphone array was also tested for a translation from the home position for both the ambient and additive noise conditions in the small testing room. These tests are added to determine microphone performance when the array was moved slightly to a position in the room which it had not been trained. For the translation length, the microphone array was moved one quarter of a wavelength for the home position in the X, Y, and Z directions independently. This translation was based on the average frequency of 1750 Hz of the 1 to 2.5 kHz range, which results in a translation of approximately 2 inches. Due to time constraints, a set of translation tests was unable to be taken in the large room. Both training and validation test data was used for performance comparisons between beamformer configurations. The microphone array was placed through a total of twenty-four tests under the following conditions:

- Three tests were conducted in the small room with ambient noise, where one of the three tests is used in the training set for the beamformers and the remaining tests were used for validation of the beamformer responses
- Three tests were conducted in the small room with noise added, where one of the three tests is used in the training set for the beamformers and the remaining tests were used for validation of the beamformer responses
- Three tests were conducted in the large room with ambient noise, where one of the three tests is used in the training set for the beamformers and the remaining tests were used for validation of the beamformer responses
- Three tests were conducted in the large room with noise added, where one of the three tests is used in the training set for the beamformers and the remaining tests were used for validation of the beamformer responses
- Three tests were conducted in the small room with ambient noise where the microphone array is translated 2 inches in the X direction with three tests for validation of the beamformer responses
- Three tests were conducted in the small room with ambient noise where the microphone array is translated 2 inches in the Y direction with three tests for validation of the beamformer responses
- Three tests were conducted in the small room with ambient noise where the microphone array is translated 2 inches in the Z direction with three tests for validation of the beamformer responses
- Three tests were conducted in the small room with noise added where the microphone array is translated 2 inches in the X direction with three tests for validation of the beamformer responses
- Three tests were conducted in the small room with noise added where the microphone array is translated 2 inches in the Y direction with three tests for validation of the beamformer responses

- Three tests were conducted in the small room with noise added where the microphone array is translated 2 inches in the Z direction with three tests for validation of the beamformer responses

These test conditions are used to determine how robust the trained beamformers are to deviations from the training conditions. The beamformer training algorithm will need to be examined to determine how these testing conditions were handled by the beamformers to produce acceptable results.

2.2 Beamformer Algorithm Development

The beamformer implemented in this study must meet certain expectations in order to be a valid solution. Since this experiment calls for a means of recording and processing data quickly, the selected beamformer would need to be efficient computationally. As mentioned before, the microphones used in this experiment are inexpensive and nominally used at the hobby level therefore, the beamformer should also incorporate a way of training the filters for the microphones. Some filter training methods mentioned in Section 1.3 involve using a specific model for the microphones implemented, meaning these microphones would need to be calibrated to properly represent the desired model and would need to be approximately identical to the other microphones implemented on the array.

To satisfy these constraints, a filter and sum adaptive offline training approach was selected. This type beamformer was targeted at a set of directions, or angles, that can be built from a desired output signal and a single set of received data. Because the filter was trained offline, the filters do not update as the system operates. Instead, the real-time implementation utilizes a set of fixed filter vectors that can be updated when the user chooses. Optimal Weiner filters were selected for the filter vectors of the trained beamformers. The intention is that the Wiener filter will remove data outside of the frequency set for the designed excitation chirp, which allows for a more accurate comparison between a received signal and the training signal.

These filters were trained through an adaptive offline algorithm, which uses a set of user designated parameters to solve for the weights of the Weiner filter vector. An LMS algorithm was used to solve for these weights. This process expects that the signal is stationary. While this conditions is expected, a stationary signal does not have to be satisfied in order to reach convergence for filter coefficients [**?**]. To illustrate the filter training process, Figure 2.6 displays the flow diagram for training a set of three filter vectors based on the received data and a generated desired signal. The flow diagram shows a three microphone system, which is representative of the maximum input that this study observed. The number of inputs can be expanded to the capability of the system, so this offline filter training process can occur with any number of microphones so long as the system can support the amount of inputs. The W_{ck} variable represents the column vector of filter coefficients for the given filter direction and set of training conditions, where c is the number of the microphone channels used and k is the current sample iteration value of the sample memory. The $M_{c,k}$ variable and S_k variable represent the microphone signal for microphone channel c and the training signal the k^{th} sample iteration respectively. If the filter vector is longer than the current sample memory, then W_{ck}, $M_{c,k}$, and S_k would all be column vectors that are the same length as the number of filter coefficients. The microphone signal starts at the current sample iteration value and includes the previous elements for the microphone signal in order to produce a vector the length of the filter vector. If a element does not exist, then the remaining elements are treated as zeros. Once the vectors are the same size, the vectors are convolved together to form the filter response due to the contributions of this and preceding data points. This complete process is repeated for all sample points. Once the sample is completed for training the filter vector, the process is repeated for a set amount of loops as defined by a user. A loop for this process is defined as taking the same sample and retraining the filter vectors using the output filter vector as the initial values for $W_{c,k}$. The loop iteration index is not shown in Figure 2.6 but repeats the process by updating the filter vectors between each loop.

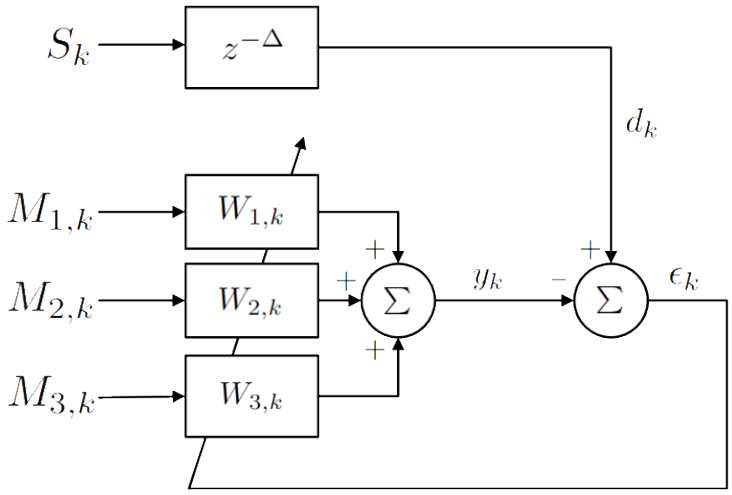


Figure 2.6: Flow diagram of offline adapter filter training when given a training signal and data from three microphones

The LMS process is designed to ensure that the error between the training signal and the summation of the filter responses will be minimized. The minimized error means that the filter response is a reasonable approximation of the training signal [**?**]. The error of filter response to the training signal is described in Equation 2.1:

$$\epsilon_k = d_k - y_k \tag{2.1}$$

where ϵ_k is the error seen by the filtering process, d_k is the desired signal, and y_k is the sum of the filter responses for the current sample point, which are all scalar values. If the y_k term is expanded to show the contribution for each filter response Equation 2.1 can be represented in a new form such as in Equation 2.2:

$$y_k = \sum_{c=1}^{C} M_{c,k}^T W_{c,n,k} \tag{2.2}$$

where the filter response signal, y_k, is the summation of the filter response for each microphone channel, which allows for any number of channels to be used in this process. N_C refers to the total number of channels used, where three channels are used for this study.

The initial weighted coefficients for the filter vector of each channel were set to a zero

vector as an initial guess. The length of these filter vectors was determined by the user prior to observing the training data. The number of filter coefficients for this experiment was initially selected empirically based on a binary base and was iteratively tested to determine what number of filter coefficients were necessary to produce both an acceptable filter response and a converging mean square error (MSE) below a selected MSE value. To calculate the MSE, the error, ϵ_k, is squared and divided by the number samples, N_k. Calculating the MSE of each loop of the filter training algorithm can be seen in Equation 2.3:

$$MSE_n = \frac{\epsilon_n^T \epsilon_n}{K} \tag{2.3}$$

where MSE_n refers to the mean square error at the n^{th} loop of the offline adaptive training process. The error vector of was then multiplied by itself and divided by the total number of sample points in the sample memory. It is desirable for the MSE to converge to a minimum value, which can be described as when the MSE has stabilized to a value that is at least -10 dB from the starting point. This minimum MSE value was selected to be -10 dB from the starting MSE as this would show a strong approximation of the filtered responses to the desired chirp signal. When observing the MSE of the filters as the filters are trained, the number of loops that the filter vectors were trained with was selected by the user on an empirical basis to examine if a filter had converged to a minimum error. For this experiment, the number of iterations was set to start at 750 loops and could be altered if a converging MSE could not be found. If this term is too small, then convergence of the MSE may not be seen. If this term is too large, then convergence can be found but iterations may be wasted.

The next step is to determine how the filter coefficients change per sample point iteration as this is how the weights are updated when moving through the data points. With the LMS algorithm, the change in filter coefficients can be described as seen in Equation 2.4:

$$W_{c,n,k+1} = W_{c,n,k} - \mu \hat{\nabla}_{c,k} \tag{2.4}$$

where μ is the step size taken by the filter for each loop and $\hat{\nabla}_{c,k}$ is the gradient estimate

for the change in filter coefficients for each channel per iteration [**?**]. The μ term governs the stability of the adaption and the speed at which convergence can be reached. Nominally, μ is defined as a positive scalar. If μ is selected as a large a number, then the filter can converge quicker. Unfortunately, if the selected μ is too large, then the steady state error of the filter training process will also be large which can lead to the filter adaptation becoming unstable [**?**].

The gradient for the Weiner filter can be estimated with the expected value of the squared error of the response to the training signal. This expected value can then be differentiated with respect to the filter vectors per data sample iteration. The results of this process can be seen in Equation 2.5:

$$\begin{aligned} J &= E[\epsilon_k^2] \\ \hat{\nabla}_{c,k} &= \frac{\partial J}{\partial W_{n,k}} \\ &= \frac{\partial \epsilon_k^2}{\partial W_{n,k}} \\ &= 2\epsilon_k \frac{\partial \epsilon_k}{\partial W_{n,k}} \\ &= -2 \sum^{C} \epsilon_k M_{c,k} \end{aligned} \tag{2.5}$$

where the gradient has been estimated in terms of the measured error and the input samples to the filter vector.

When combining the results of Equation 2.5 and 2.4, the change in filter coefficients can be expressed as seen in Equation 2.6:

$$W_{c,n,k+1} = W_{c,n,k} + 2\epsilon_k \mu M_{c,k} \tag{2.6}$$

This form of the equation will be used for each channel with the same error, so each filter vector per microphone can contribute to the training of the other filter vectors. This process of training the filter vectors means that as more microphone channels are added, the solving time becomes longer.

For the scenarios in which the MSE does not converge, Equation 2.6 can be updated to include a multiplier to the current filter vector, which helps drive the system to convergence. Adding this multiplier, γ_c changes the process to a Leaky LMS process. This update leads to the expression shown in Equation 2.7:

$$W_{c,n,k+1} = \gamma_c W_{c,n,k} + 2\epsilon_k \mu M_{c,k} \tag{2.7}$$

where γ_c is the leak coefficient per microphone channel. The leak coefficient is in a range of 0 to 1, and drives the MSE to converge as γ_c approaches zero. Additionally, as γ_c approaches zero, the change in MSE reaches is 0 dB, meaning that a signal cannot be constructed from the training signal and microphone inputs. The leak coefficient is normally suggested to be in a range of 0.9999 to 1 as going below this point drastically affects the minimum MSE that can be found [**?**]. For this experiment, the leak coefficient was set to unity unless convergence could not be found when altering any of the previous controllable parameters.

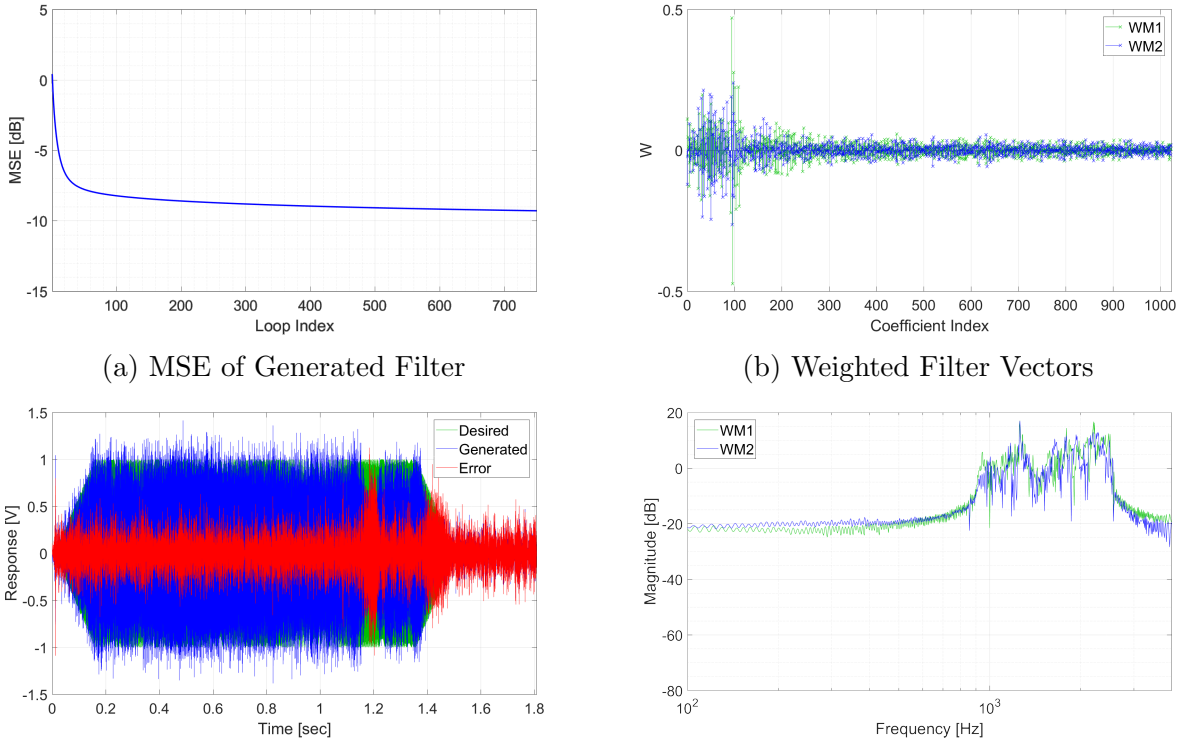
Using the filter and sum offline adaptive training process has a dramatic advantage when compared to TDE based methods for training the filters of beamformers. The main advantage is that using the filter and sum offline adaptive training process results in only four parameters needing to be altered to determine a solution as compared to adjusting models for array geometries as seen with TDE methods. These tunable parameters for this process can be summarized as:

- the number of selected loops that the filter will be trained
- the number of filter coefficients
- μ, the step size constant
- γ_c, the leaky LMS coefficient per microphone channel

and are selected prior to running the LMS process. To examine how the selected parameters produce an acceptable set of filter vectors, the responses of the generated filters will need to be examined.

When examining the filter outputs, there are four responses that need to be observed per filter. To observe these responses, the results of the zero degree beamformer for the small training room, no added noise, and the {1,2} microphone configuration will be used as an example. Figure 2.7 displays the responses from the filter training process using Leaky LMS algorithm with optimized Weiner filters on a zero degree direction filter for the small room with no training noise with the {1,2} microphone configuration. The first response for the filters is the MSE of the filters for each loop of the training of the filter vectors. The expectation is that the minimum MSE should be below -10 dB from the starting MSE with no slope on the final loop signifying that convergence has been reached. This limit is not required, but the lower the MSE, the better approximation is made from the received signal to the desired signal. This assumption means that the MSE for training sets that have training noise added during data collection should have a higher MSE compared to their no noise counterparts. From Figure 2.7a, the resulting MSE for this filter configuration is found to be below -8 dB. This finding means that the filter should have a strong magnitude frequency response which can be seen in Figure 2.7d as true where frequencies not in the 1 to 2.5 kHz bandwidth are suppressed to approximately -20 dB. The second response for the filter is the filter vector that was created. This plot details the impulse response of the filter vectors. The third response for the filters is the error between the training signal and generated filter response signal. The expectation is that the error between these two signals is as low as possible throughout a majority of the signal comparison. From Figure 2.7c, the error is low compared to the output signal for a majority of the signal comparison, meaning a reasonable approximation has been made. The final plotted response of the filters is the frequency response magnitude of each of the filter vectors. The highest magnitude of responses should be in the range of 1 to 2.5 kHz to match the desired chirp signal. The optimized Weiner filter is used to reject noisy signals outside of the chirp frequency range.

Unfortunately, with four training conditions, for the small and large rooms training with ambient and additive noise, and eight beamformer directions per training condition, this study had a total of 32 beamformer filter banks per microphone channel configuration used. If all of the possible two and three channel microphone configurations are used, this total is then 128 distinct beamformer filter banks that can be examined. For practical reasons,



(a) MSE of Generated Filter

(b) Weighted Filter Vectors

(c) Error of Generated Response

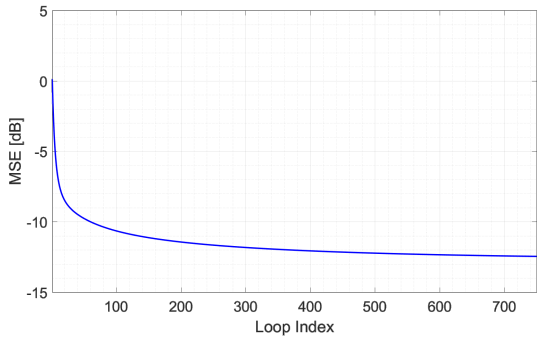
(d) Frequency Response Magnitude

Figure 2.7: Response components of the 0° direction beamformer using a {1,2} microphone configuration in the small room with ambient training noise

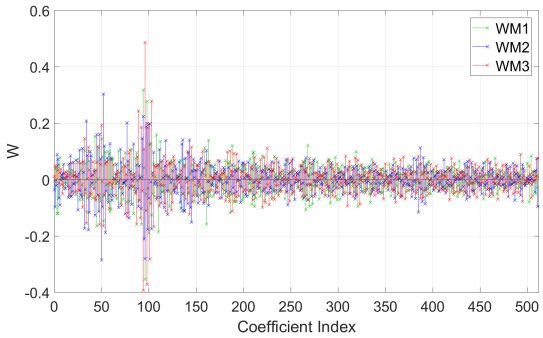
the responses for the generated filters have been limited to a set of examples that can be used to highlight the general response of the filters. The set of examples will come from the four testing conditions with similar microphone configurations and the 0° beamformer direction for small training room in which the {1,2} and {1,2,3} microphone configurations were trained with ambient and additive noise. These responses will be used as a basis for comparing the responses found.

The filter responses for the three microphone channel configuration for the small room with ambient noise can be examined in Figure 2.8. Figure 2.8 shows the responses from the filter training process using the Leaky LMS algorithm on a 0° direction beamformer for the small room with no training noise with the {1,2,3} microphone configuration. In this case, the resulting MSE was below -12 dB, so a significant improvement from the two channel filter configuration case. This observation means that adding another microphone reduces the MSE, therefore a closer response can be generated to approximate the desired excitation

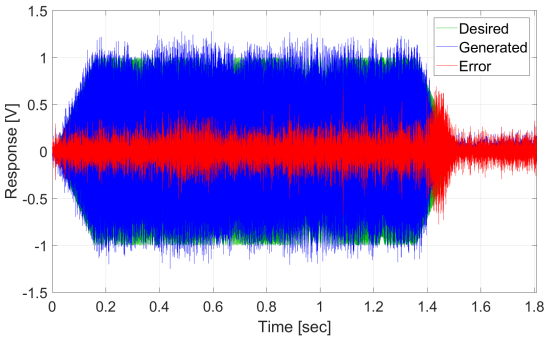
chirp. The frequency response for the three channel configuration has a similar response to two channel configuration. Both responses have the desired bandwidth of 1 to 2.5 kHz for the positive magnitude and have suppressed frequency bands outside the desired bandwidth.



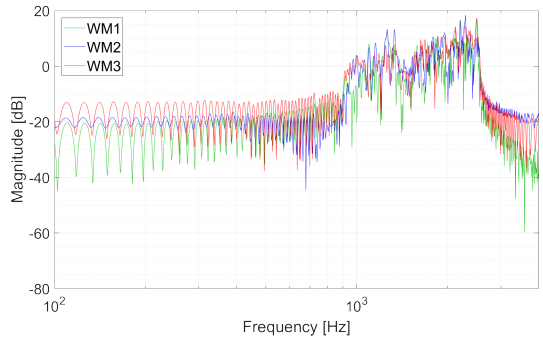
(a) MSE of Generated Filter



(b) Weighted Filter Vectors



(c) Error of Generated Response

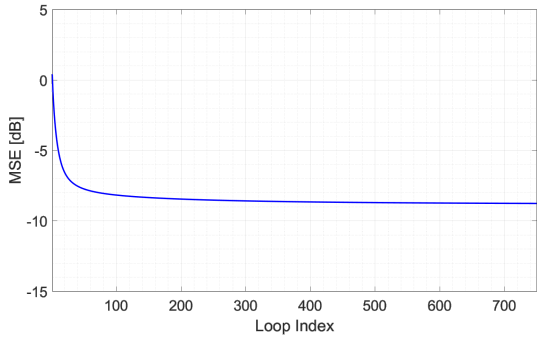


(d) Frequency Response Magnitude

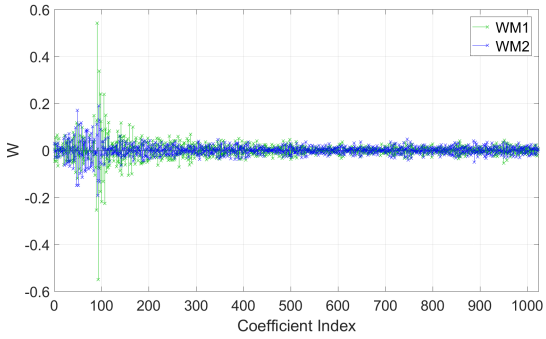
Figure 2.8: Response components of the 0° direction beamformer using a {1,2,3} microphone configuration in the small room with ambient training noise

Now the cases in which the filters are trained with additional noise will need to be examined to observe how adding training noise affects the responses for the constructed filters. Figure 2.9 presents the responses from the filter training process using the Leaky LMS algorithm on a 0° direction beamformer for the small room with additive noise for the {1,2} microphone configuration. The resulting MSE is below -8 dB, so the goal of -10 dB was not met. Remember that this limit for the MSE is not required, but the larger the MSE decreases by, the better the approximation is found between the generated filter response and the training signal. This observation validates the assumption that using additive noise to the training of the beamformer will decrease the received MSE for the filter configuration. Another result that can be seen is that by training with noise, the filter suppresses the

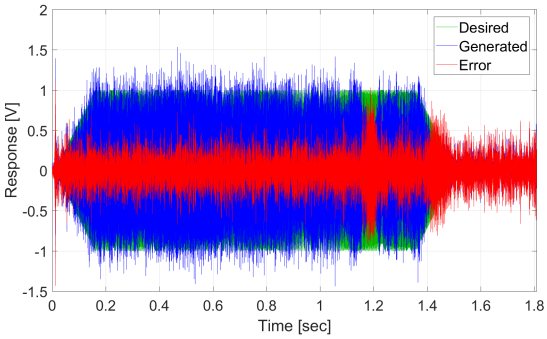
frequency bands outside the desired bandwidth to -40 dB, rather than the -20 dB that was seen in Figure 2.7d for the two channel beamformer that was trained with ambient noise. This observation means that by training with additive noise, the desired frequency band is enhanced as compared to frequency bands outside the 1 to 2.5 kHz frequency band. This trend agrees with the assumption stated in Section 2.1 that the addition of noise in training assists in enhancing the desired signal by further suppressing the signals outside the desired range.



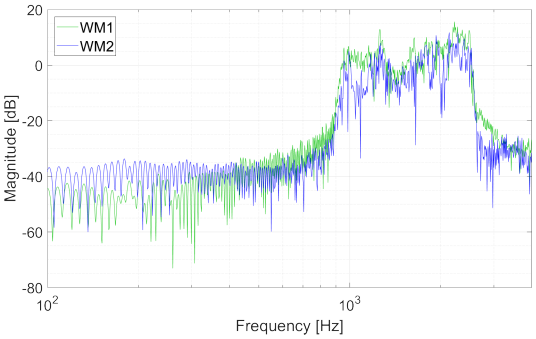
(a) MSE of Generated Filter



(b) Weighted Filter Vectors



(c) Error of Generated Response

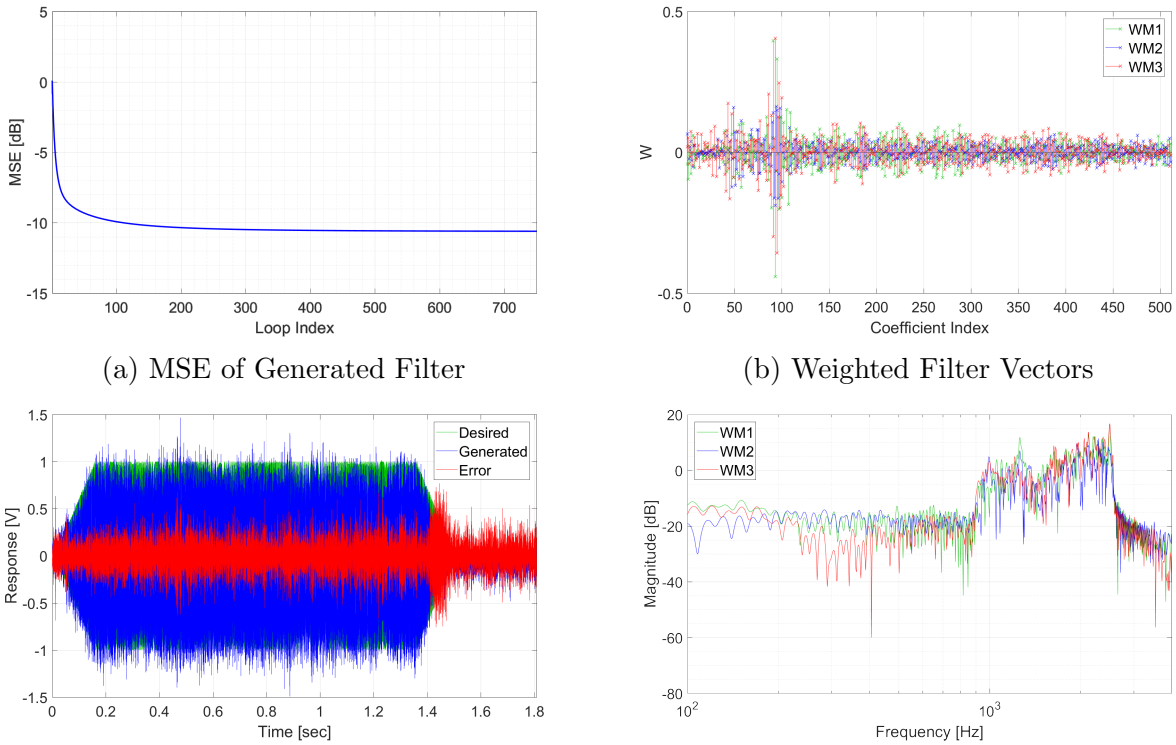


(d) Frequency Response Magnitude

Figure 2.9: Response components of the 0° direction beamformer using a {1,2} microphone configuration in the small room with the additive training noise

The final filter training responses are from the beamformer that was trained in the small room with additive noise for a three channel configuration. Figure 2.10 features the responses from the filter training process using the Leaky LMS algorithm on a 0° direction beamformer for the small room with additive noise for the {1,2,3} microphone configuration. For this beamformer, the resulting MSE was below -10 dB, which is lower than the two channel additive noise beamformer and larger than the three channel beamformer that was

trained ambient noise. This observation supports the claim that adding a third microphone channel decreases the received MSE, but adding noise during training increases the MSE for a beamformer. When examining the frequency response for this beamformer, the response is similar to the three channel case that was trained ambient noise.



(a) MSE of Generated Filter

(b) Weighted Filter Vectors

(c) Error of Generated Response

(d) Frequency Response Magnitude

Figure 2.10: Response components of the 0 degree filter using a {1,2,3} microphone configuration in the small room with the additive training noise

These examined filter responses give some insight into how these filters may perform when testing with similar validation data. Unfortunately, these responses do not convey how the responses for the beamformers operate for when the validation conditions differ from the training conditions. For these beamformers, the final user parameters were selected to be consistent across training group. All filters that were trained with two microphone channels had filter vectors with 1024 filter coefficients, and those filters trained with three microphone channels had filter vectors with 512 filter coefficients. All filters examined in this study had a μ of 0.0025 and each filter repeated the filter training process for 750 loops. No filter required the use of the leaky coefficient as all filters were able to converge during the training process

meaning that all filter vectors used a γ_c of unity during the LMS process. When these filters are used in a testing scenario as described in Section 2.1, a filter bank of eight beamformer directions are used with each beamformer direction consisting of either two or three filter vectors, depending on how many microphones are implemented for the microphone array. The results for how well each type of trained beamformer works for the described testing conditions will next need to be examined to frame how this study assumes the performance of a described filter will behave.

2.3 Hypothesis of Beamformer Performance Differences

Since the testing conditions have been detailed and the training of the beamformers has been discussed, hypothesis can be drawn on the significance that each beamformer condition can play on the responses of the beamformers. These beamformer conditions are the number of microphones used, the use of noise when training beamformer filters, and the room used to train the filters of the beamformers. These beamformer conditions will be the only sources of variation that are examined in this study, so a several part hypothesis must be drawn for this experiment for how these specific variations will affect performance. To best describe the hypothetical impact of each condition, these conditions will be broken down and discussed for the possible outcome in this experiment.

The first portion of the hypothesis is stated as the variation in the number of microphones implemented should have a significant impact on beamformer performance. As detailed in Section 1.2, the number of microphones directly aligns with how much data can be gathered in parallel. As the amount of parallel data from separate microphones increases, more detail can be discerned about the target signal. This is one of the fundamental principles of using a DMA, since a number of microphones can be used to form a highly directional beam as the number of microphones increases [**?**][**?**].

For the second portion of the hypothesis, the variation in using ambient or additive noise when training the beamformers should have a significant impact on beamformer performance, but this source of variation may not have as large of an impact on the beamformer

performance varying the number of microphones. As mentioned in Section 1.3, adding noise in training has been shown to assist in removing excess noise from the signal source when processed, meaning that for cases with noise, the beamformers trained with additive noise should perform better than beamformers trained with ambient noise. This condition works well for generating filters that have a stronger response to the training signal, but additive noise in training the filters does not directly add more detail data on the training signal.

For the final portion of the hypothesis, we speculate that the variation of the room in which the beamformers are trained should not have a significant impact on beamformer performance. From Section 1.3, the adaptive offline training process takes into account interference from factors such as noise and room reverberation. If this hypothesis is proven true, then using the LMS filter coefficient identification process would rule out the need for generating a specific model per room that the unit is desired to be implemented.

These hypotheses can be tested using the constructed beamformers that have been mentioned. These beamformers will be subjected to the discussed twenty-four tests to determine if any of these training conditions play a significant part in impacting the performance of the beamformers. For this evaluation, the output data from the filtered responses will be characterized and interpreted.

Chapter 3

Beamformer Bank Output Characterization

This section discusses the validation of the generated beamformer filter bank outputs for an explicit prediction of the direction of the sound source. The characterization process involves using statistical analysis of the correlation outputs of the beamformers as well as a proposed scoring system which can determine if the predicted direction of the beamformer filter bank should be accepted or rejected.

3.1 Interpretation of Beamformer Output Data

The first step in this analysis is establishing how the data was presented from the beamformer filter banks. In Figure 3.1 a sample output polar plot is given to illustrate the represented data. The theta values around the perimeter of this polar plot represent the beamformer angles (θ_B) of the eight trained beamformer directions. The radii values for the polar plot are the cross correlation values of the filtered and summed signal compared to the training signal. These radii have been normalized to the maximum correlated output from the beamformer bank to dictate which θ_B has the largest correlation to the training signal. For each test direction, the chirp signal arrived at a specified angle to the microphone array

giving an angle of arrival (θ_A) to be used in validation of the beamformer performance. In operational testing, θ_A is not known by the host system, so the angle must be approximated with θ_B in some manner.

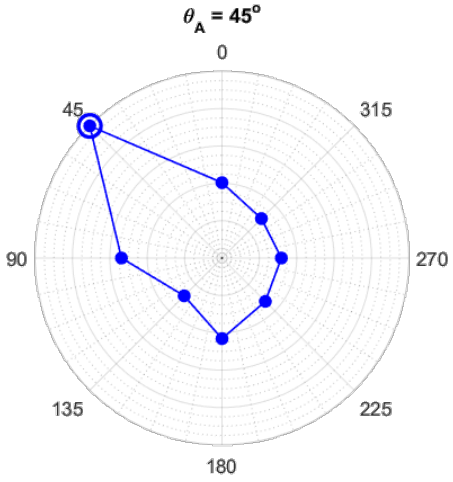


Figure 3.1: An example output polar plot of a θ_A of 45^o, θ_B from 0 to 315^o with a resolution of 45^o, and a maximum output correlation corresponding to 45^o

For this experiment, eight evenly spaced θ_B were selected which allowed for a 45^o resolution around the microphone array. The outputs for each θ_B are normalized to the maximum correlation value to remove the use of thresholds in the correlation responses as an indicator for accepting or rejecting an answer from the beamformer bank. Performing this normalization would assist in detecting a pure noise signal in which no desired excitation signal is received. Instead of using thresholds, this experiment employed a scoring mechanism to describe if the maximum correlated direction is a suitable candidate to predict the θ_A of the chirp signal.

The scoring system is broken into three metrics that describe not only the maximized correlation, but also how significant the maximum correlated output was compared to the other beamformer output correlations. These metrics are described as: scaled mean radii length, scaled maximum radius contribution, and scaled vector to maximum angle differential. These metrics are defined on a scale of 0 to 1 based on the maximum and minimum values desired and are used to generate an assurance score. This assurance score was then

compared to a threshold value to determine if the maximum correlation output direction will be accepted as a correct prediction. The process flowchart illustrating the general procedure once the audio signal is captured by the microphone array can be seen in Figure 3.2 for determining predicted θ_A for the chirp signal and whether to accept this estimate. In this flowchart, M_c represents the signal received by the c^{th} microphone, S refers to the chirp signal that is delayed before comparison, and $W_{c,h}$ is the weighted coefficient vector of the c^{th} microphone of the h^{th} beamformer. Each of the dashed red boxes represents one of the beamformers used in validation. This process allows for a streamlined procedure for determining a direction that has the largest correlation to the excitation direction as well determining if that prediction is valid.

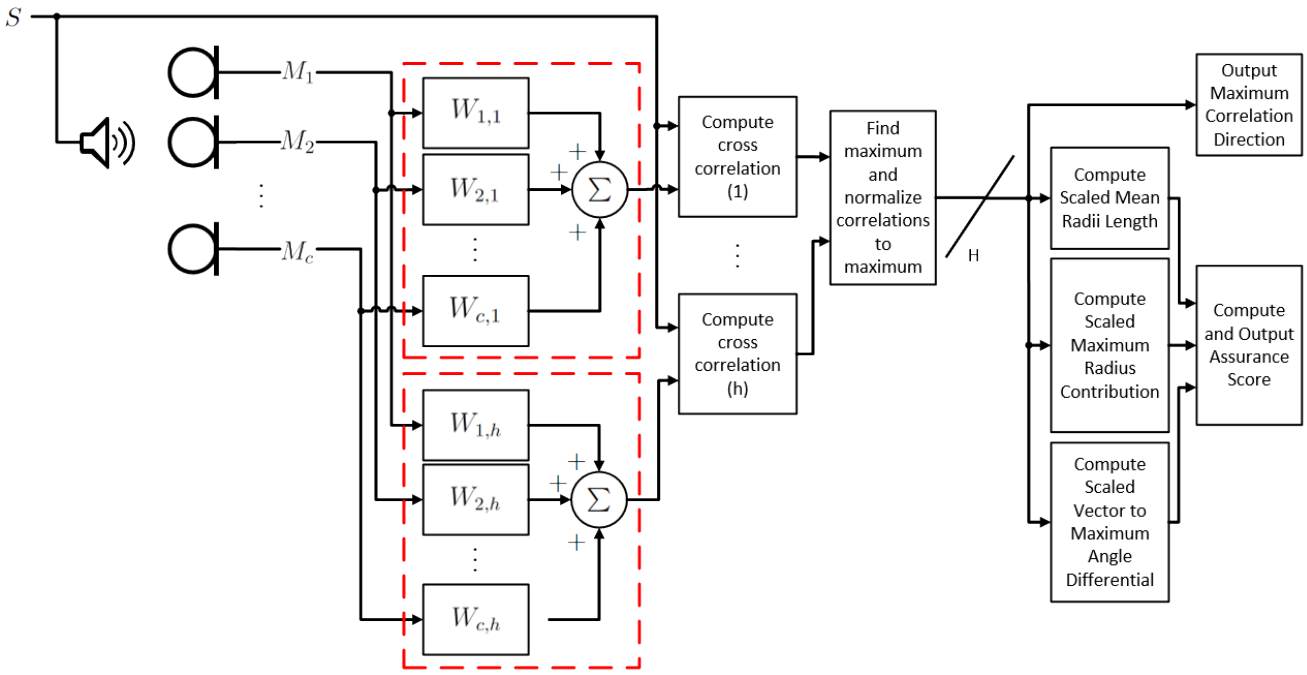
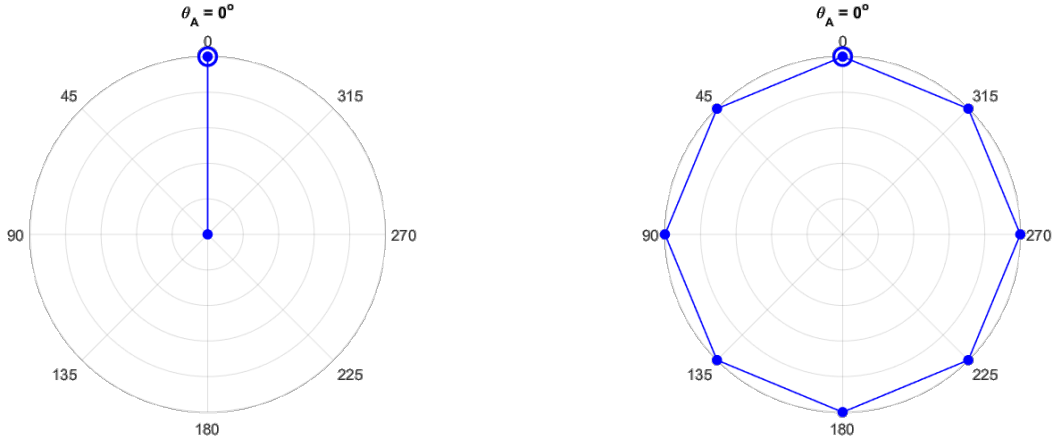


Figure 3.2: Flow diagram of procedure for gathering validation, filtering responses, comparing beamformer outputs, and approximating an angle of arrival

From the comparison of the correlation outputs, the generated metrics are assigned a relative weight then summed to generate an assurance score. When the metrics for describing the beamformer bank outputs were first discussed, the limits for the maximum and minimum beamformer bank outputs were suggested to come for responses were all information is know about the signal and a response where nothing is known about the signal. These cases have

been plotted as seen in Figure 3.3 on a polar plot similar to example beamformer bank output. Figure 3.3a shows the theoretical ideal sample output from the beamformer bank where all information is known about the signal and where only the correct beamformer direction receives a high correlation while the remaining directions have little to no correlation to the excitation signal. Figure 3.3b presents a maximum response in all θ_B meaning no information is known about the signal a proper θ_A cannot be predicted.



(a) Ideal sample

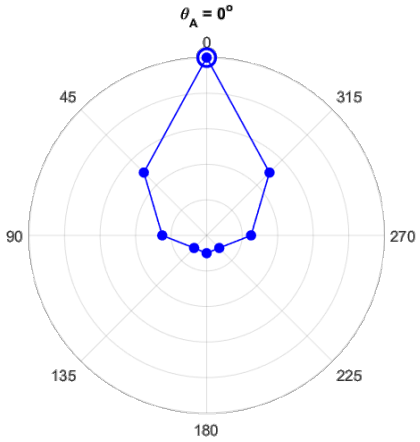
(b) Pure noise sample

Figure 3.3: Two theoretical sample plots in which the correct angle is guessed from the beamformer bank

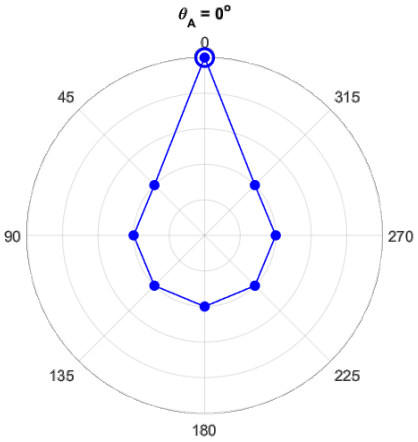
These plots represent the theoretical best and worst case scenarios for the beamformer bank outputs which may not be replicable when using real data. These cases may not be achievable with real data due to the beamformers still producing some correlation with the received filtered signal to the training signal even if the signal is not the desired chirp. For this reason, the beamformer bank outputs can be compared to the ideal and pure noise cases using the assurance score to determine if the predicted direction has a significant correlation to the chirp signal. The assurance score for each result will be designed to to indicate clearly whether a received beamformer bank output is more representative of the ideal or the worst case. For this experiment, the ideal and pure noise cases were designed to produce an assurance score of 1 and 0 respectively using the calculation method described in the flowchart procedure.

While the assurance score and the quantities involved in producing the assurance score are based on the theoretical maximum and minimums of the beamformer banks, the ideal and worst cases have been stated as only theoretical models, so these cases would be difficult or improbable to produce. Some auxiliary cases were generated to have characteristics that were seen across the responses for all tests. These cases can be seen in Figure 3.4 where four polar plots contain more realistic scenarios for data from correlating the beamformer responses to the chirp signal. Figure 3.4a shows a beamformer correlation output for a strong correct response where the normalized correlation is low in all directions except for the dominant predicted direction. Figure 3.4b illustrates a moderate response which is similar to the strong response but has less cancellation in the directions that are not the predicted direction meaning the correlated beamformer outputs are larger than the previous figure. Figure 3.4c describes an output in which there is a much larger response of the beamformer direction correlations near the predicted beamformer direction. This type of response leads to a diverging response in which two directions that are close to one another have similar correlated outputs to the chirp signal. Lastly, Figure 3.4d is a step beyond the diverging case in which the two strong predicted beamformer directions have separated further apart creating a split response. All of the responses were set to have the chirp come from θ_A of zero degrees, but each response represents different characteristics that can be seen within the received beamformer bank data. Again, these responses were generated to better illustrate the characterization method implemented in this study and were not gathered using real microphone signals.

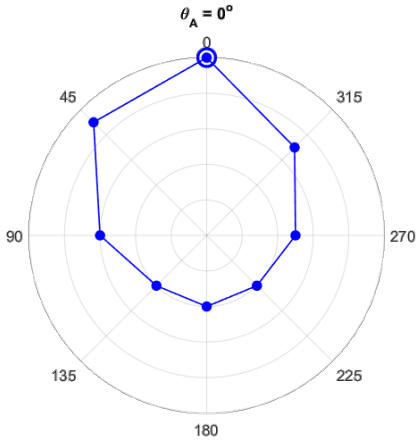
These responses, along with the three previously mentioned cases, will be used as metrics to verify how well the scaled mean radii length, scaled maximum radius contribution, scaled vector to maximum angle differential, and assurance score represent the desired and undesired responses. From these values, there are two primary outputs that will be received each time a comparison is made. The first output will be the predicted direction which has the highest correlation to the chirp signal compared to the remaining beamformer directions. The second output will be the assurance score that gives an assessment of how evident the prediction is as compared to the other beamformer directions. Each of the categories that represent a desired characteristic of the received filter data will be examined to verify both



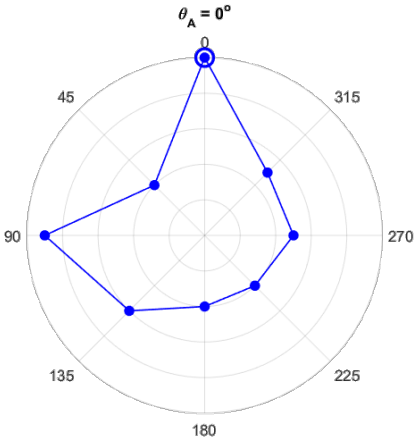
(a) Strong response



(b) Moderate response



(c) Diverging response



(d) Split response

Figure 3.4: Four sample plots of possible output radii length configurations for $\theta_A = 0$ where all beamformer banks have predicted correctly

the implementation of this calculation based method and the portrayal each response receives when observed in each category.

3.2 Scaled Mean Radii Length Metric

The Scaled Mean Radii Length (SMRL) characterizes the average of the radii lengths of the beamformer responses into a relative value. This operation is done by converting the mean of those radii lengths to a scale value from 0 to 1 as seen in Equation 3.1:

$$S_{MRL} = -p\frac{\sum r_h}{H} + p$$
$$p = \frac{H}{H-1} \tag{3.1}$$

where S_{MRL} is the SMRL, N_{θ_B} is the number of beamformer directions, and r_i is the radius length of the i^{th} element of the beamformer directions up to N_{θ_B}. The reason the average for the beamformer responses was used was because this average illustrates the spread of highly correlated or similar beamformer responses. This spread indicates the amount of significance the predicted direction has as the maximum correlated direction. For the conversion to a 0 to 1 scale value, the equation is designed to set the worst theoretical case where the average normalized radii length is the maximum value of 1 to correspond to an SMRL of 0. The ideal case would result in an average of one ${N_{\theta_B}}^{\text{th}}$ of a radius length as all other directions have no correlation to the chirp signalfor the respective beamformer response. For the ideal case, the SMRL of this case would result as unity according to the conversion.

The SMRL is the primary metric that is used in developing the assurance score. This response would have a similar correlation response at a majority of the beamformer directions which can be caused by receiving a signal that does not contain the chirp signal or the signal could not be limited to a single direction. For these responses, Table 3.1 contains the results for the theoretical best and worst cases, along with the characteristic and real sample, beamformer bank responses. Both the diverging and split cases have a lower response to the SMRL metric as compared to the strong, moderate, and real sample responses. From this observation, the diverging and split cases would have beamformer responses would have a higher amount of correlation for a larger amount of the beamformer directions meaning that a distinct prediction would be harder to achieve.

Table 3.1: SMRL results for the sample responses

Result Category	Ideal	Strong	Moderate	Diverging	Split	Worst	Real Sample
SMRL Score	1	0.743	0.6	0.443	0.471	0	0.628

However if only the SMRL is considered, some information is missing from understanding the shape of the beamformer bank output. Only the average beamformer response

can be represented using the SMRL. While the diverging case is not the most distinct prediction of a beamformer direction, the diverging case should be able to be included as an acceptable beamformer response. The reason behind making this case acceptable is that the diverging case response has an overall response that points within one beamformer resolution of the predicted direction. This shape of the beamformer output would indicate that while a single direction cannot be predicted, two directions in close proximity to one another means the excitation response could be expected to lie between the two predicted beamformer directions. To better represent the remaining attributes for an acceptable beamformer bank response, the contribution of the predicted beamformer direction and the vector angle differences will need to be examined to see how these metrics would subsidize the generation of the assurance score.

3.3 Scaled Maximum Radius Contribution Metric

The second property that will be discussed is that of the Scaled Maximum Radius Contribution (SMRC). This metric for the assurance score utilizes the maximum radius length and the radii lengths of the beamformer directions on both sides of the maximum output to describe how much of the data is directed towards the predicted beamformer direction. Like SMRL, the SMRC was defined on a 0 to 1 scale using a linear mapping as seen in Equation 3.2:

$$S_{MRC} = \begin{cases} -q\frac{\sum r_g}{\sum r_h} - \frac{q}{H} & \text{if } \frac{\sum r_g}{\sum r_h} > \frac{3}{H} \\ 0 & \text{if } \frac{\sum r_g}{\sum r_h} \leq \frac{3}{H} \end{cases} \tag{3.2}$$
$$q = \frac{H}{H-3}$$

where S_{MRC} is the SMRC received from the beamformer bank and r_k are the radius lengths of beamformer direction k. The values of k include only three beamformer directions covering the predicted beamformer direction and the two nearest neighbors. The index i ranges over all eight beamformer directions. As an example, if the predicted beamformer direction was

0°, then the correlation values from beamformer directions of 315°, 0°, and 45° would be used to calculate the SMRC value. This metric uses three beamformer directions in the calculation rather than just the predicted direction to allow responses in which the beamformer bank response is diverging to be considered acceptable. Since the two neighboring beamformer directions are used in conjunction with the maximum beamformer direction, using this metric limits the system to using a minimum of four beamformer quantization directions to identify the direction of a sound source. The reason for the four minimum beamformer direction requirement is that if only four directions are used, then a single hemisphere is considered each time the SMRC is calculated. For a four direction beamformer bank, if the 0° direction is predicted, then the 270°, 0°, and 90° outputs will be used for the SMRC. For a three or less direction beamformer bank, the SMRC would always result in a zero since the condition for a zero output would always be satisfied.

While the SMRL metric indicates the general correlation results across all beamformer directions, the SMRC is focused on representing the contribution of the predicted direction along with neighboring outputs. This observation means SMRC value is designed specifically to represent the predicted direction and give more targeted information on how defined the results are compared to the remaining beamformer directions. For these responses, Table 3.2 contains the results all of the described cases for the metrics of SMRL and SMRC output responses. This table shows that the diverging response performs significantly better using the SMRC than the SMRL metric. Along with this improvement for the diverging response, the strong and moderate response also have acceptable responses. These observations mean that the split case is primarily targeted to be deemed unsuccessful.

Table 3.2: SMRL and SMRC results for the sample responses

Result Category	Ideal	Strong	Moderate	Diverging	Split	Worst	Real Sample
SMRL Score	1	0.743	0.600	0.443	0.471	0	0.628
SMRC Score	1	0.943	0.558	0.649	0.447	0	0.662

This metric identifies if the θ_A is between two quantized beamformer directions or if two beamformer directions that are next to one another share a similar correlation response. As can be seen when observing the sample polar plots, as the SMRC increases, the more

defined the predicted direction becomes. This trend is evident in the provided samples, but only refers a small portion of the collected data to the predicted direction. From the observing the SMRL and the SMRC, the average response of the beamformer outputs have been described, as well as the strength of the predicted direction. Unfortunately, neither of these metrics take into account for the predicted direction of the beamformer points for a possible predicted direction. The need for a vector angle of the beamformer bank outputs leads into the final mentioned metric where the vector angle difference generated by the beamformer outputs will be compared to the predicted beamformer direction.

3.4 Scaled Vector to Max Angle Difference Metric

The third and final metric that will be discussed is the Scaled Vector to Maximum Angle Difference (SVMAD). This metric provides information about the vector angle of the beamformer bank outputs and was converted to a 0 to 1 scale to incorporate the value into the assurance score calculation. The purpose behind the inclusion of this metric is to illustrate the difference between the predicted beamformer direction and the vector direction of the beamformer bank response. The linear mapping for the SVMAD can be seen in Equation 3.3:

$$S_{VMAD} = \begin{cases} -\frac{H}{360} D_{VM} + 1 & \text{if } D_{VM} < \frac{360}{H} \\ 0 & \text{if } D_{VM} \geq \frac{360}{H} \end{cases} \tag{3.3}$$

where S_{VMAD} is the SVMAD received from the beamformer bank and D_{VM} is the magnitude of the difference between angles of the predicted beamformer direction and the vector direction of the beamformer bank results in degrees. The importance of using the SVMAD as a metric to characterize the beamformer bank output is that more data is provided to describe if the beamformer bank has a diverging or split response.

In Equation 3.3 the maximum acceptable D_{VM} is a difference of one resolution of the beamformer directions or 45° for this research. If the D_{VM} is found to be above one

resolution, then the SVMAD is converted to a minimum score of 0 stating that the vector angle is more than one resolution from the predicted beamformer direction. The calculation for D_{VM} can be seen in Equation 3.4:

$$D_{VM} = \begin{cases} |\theta_{com} - \theta_{max} - 360| & \text{if } \theta_{com} - \theta_{max} > 180 \\ |\theta_{com} - \theta_{max} + 360| & \text{if } \theta_{com} - \theta_{max} < -180 \\ |\theta_{com} - \theta_{max}| & \text{if } |\theta_{com} - \theta_{max}| \leq 180 \end{cases} \tag{3.4}$$

where θ_{com} is the vector angle from the beamformer bank outputs, θ_{max} is the angle of the predicted beamformer direction. The D_{VM} is set as a magnitude because only the angle difference is required rather than knowing where the angle is in relation to the predicted beamformer direction. The final portion to calculate is the θ_{com} value of the beamformer bank.

To calculate the θ_{com} value, the beamformer outputs are converted to a cartesian mapping and the x and y components of each beamformer direction are then summed in each axis. The x and y axis summations are then used to produce a vector angle for the beamformer bank. This process can be seen in Equation 3.5:

$$\begin{gathered} \theta_{com} = \begin{cases} atan2(y_{com}, x_{com}) & \text{if } atan2(y_{com}, x_{com}) \geq 0 \\ atan2(y_{com}, x_{com}) + 360 & \text{if } atan2(y_{com}, x_{com}) < 0 \end{cases} \\ x_{com} = \sum r_h cos(\theta_h) \\ y_{com} = \sum r_h sin(\theta_h) \end{gathered} \tag{3.5}$$

where x_{com} and y_{com} are the x and y components of the beamformer bank outputs, r is an N_{θ_B} length column vector of the beamformer bank radii lengths, and θ_B is an N_{θ_B} row vector of the beamformer angles. For this development, the SVMAD is set to the minimum value as no difference between the predicted beamformer and vector angles could be found.

As mentioned in Section 3.1, the beamformer bank output directions must be considered evenly spaced to use this method for finding the SVMAD value. This requirement

allows cancellation in both the x and y axis of the beamformer responses with symmetrical beamformer angles over the plane. The plan for the SVMAD score is to provide a clearer distinction for responses that are either diverging or split for the beamformer banks. For these responses, Table 3.3 contains the results all of the described cases for the metrics of SMRL, SMRC, and SVMAD output responses. These results show that when the response has split, the SVMAD score will be low. For the higher scoring cases, the strong and moderate responses have no distinction between responses according to the SVMAD metric.

Table 3.3: SMRL, SMRC, and SVMAD results for the sample responses

Result Category	Ideal	Strong	Moderate	Diverging	Split	Worst	Real Sample
SMRL Score	1	0.743	0.6	0.443	0.471	0	0.628
SMRC Score	1	0.943	0.558	0.649	0.447	0	0.662
SVMAD Score	1	1	1	0.740	0.075	0	0.7472

These results provide a distinct illustration for how this method distinguishes between diverging and split responses. While this metric describes the vector output of the beamformer bank response as compared to the predicted beamformer direction, this method alone does not characterize the contributions of the predicted beamformer direction specifically or the average beamformer response. This knowledge leads to each of the metrics being used to produce a singular score that can be used to describe the contribution of the predicted beamformer direction, the average beamformer response, and the vector angle response of the beamformer bank. A scoring system will next be developed to use the contributions of each of the metrics to describe if the predicted beamformer direction is an acceptable approximation for the originating direction of the chirp signal.

3.5 Assurance Score

As mentioned in Section 3.1, the correlation value threshold of the beamformer outputs was not used in determining if the maximum beamformer direction was an acceptable approximation for the chirp arrival direction. Instead, a scoring method was developed to provide a score to determine if a predicted beamformer direction is an acceptable approxima-

tion of the chirp arrival direction. This score was used to determine how viable the predicted beamformer direction is over other possible beamformer directions. The assurance score is calculated using Equation 3.6:

$$AS = 0.4S_{MRL} + 0.3S_{MRC} + 0.3S_{VMAD} \tag{3.6}$$

where AS is the assurance score, which is on a 0 to 1 scale. The coefficients were predicted to have near similar contribution to the assurance score for each metric. Only the SMRL was given a larger weighting. The SMRL was given a larger weighting because as the average radii length increases the less likely a distinct solution can be found.

As mentioned in Section 3.1, the assurance score has been assigned a threshold of 0.5 for a beamformer bank response to be considered acceptable for the predicted beamformer angle to be used. This threshold was chosen as 0.5 for the assurance score as this score would mean that the beamformer output had at least as the median assurance between the ideal and worst theoretical cases. The results can be seen in Table 3.4 for all of the examined sample and theoretical cases. From the results of these cases, the strong and moderate responses have an assurance score that is well above the 0.5 threshold. This was desired as both responses had a distinct direction in which the beamformer predicted. The diverging case also has a beamformer response that was deemed acceptable, and the split response was flagged as unacceptable due to several conditions. This outcome means for beamformer responses that have two predicted beamformer directions that are neighboring to one another, then the assurance score is designed to expect these responses to be acceptable.

Table 3.4: SMRL, SMRC, SVMAD, and Assurance Score results for the sample responses

Result Category	Ideal	Strong	Moderate	Diverging	Split	Worst	Real Sample
SMRL Score	1	0.743	0.6	0.443	0.471	0	0.628
SMRC Score	1	0.943	0.558	0.649	0.447	0	0.662
SVMAD Score	1	1	1	0.740	0.075	0	0.7472
Assurance Score	1	0.897	0.707	0.594	0.345	0	0.6741

From these results, all cases that received a score of 0.7 or higher had a strong to

moderate response. Out of the cases presented, only one case beyond the worst theoretical case failed to receive at least a passing score. Of those cases that passed, the two cases of the real sample and the diverging responses were below the observable secondary threshold for a distinct response. Because of the observable threshold, the statement can be made that if a beamformer bank responds with an assurance score above 0.7, then that beamformer response should be accepted automatically. If a case arises where the assurance score is below 0.7 yet above 0.5 and there is time for another signal to be taken, then the system could retake another set of data to verify the direction of the excitation source. The next step is to take the assurance score and use the value in a way to identify how changing filter training and validation conditions can affect beamformer performance.

Chapter 4

Results of Empirical Testing and Output Characterization

This section will review the analysis of variance (ANOVA) process and how this process is used as a comparison tool for the test scenarios presented in Section 2 in relation to how changing the beamformer conditions affect the performance of the beamformers. The best beamformer setting for the testing conditions was determined for the observed testing conditions using tests where the arrival angle of the chirp signal is aligned with a beamformer angle along with half angle resolution and pure noise cases.

4.1 Data Comparison Methods

An analysis of variance process was selected in order to determine how beamformer conditions affect beamformer performance. This process requires that the data is drawn from a normally distributed population, that the selected sample population has a common variance, and that all samples are taken independently of each other. Since each test was conducted separately in a short period in order to preserve environmental conditions between tests, the requirement of independent sampling can be assumed as satisfied. The remaining requirements of a normally distributed set of data and common variance will be need to be

verified when examining the data from the beamformer bank results. Since in this experiment there are three sources of variation, a traditional single ANOVA cannot be used for this experiment. Instead a three factor ANOVA was used to determine how the variation in the three testing conditions and their interaction impact beamformer performance.

The primary output of the ANOVA process is the f-value, which indicates if there is a significance in the given variation source. This f value is then compared to the critical f-value [**?**]. The critical f-value for the ANOVA processes are set with a 95 percent confidence interval. This interval contains two standard deviations from the normally distributed mean. If the f-value is greater than the critical f-value, then the change in that source or sources has a significant impact on the performance of the beamformers. When comparing the significance between difference variances, the larger the f-value is, the more significant the change in that factor is considered [**?**].

There are three principle sources of variation in this study: the number of microphones used, the use of ambient or additive noise during training, and the size of the room where the beamformer was trained. These sources of variation will be used to divide the data into groups of similar conditions, and an assurance score was calculated for all data collected. The assurance score is the primary metric used for observing general beamformer performance, and can be used to determine if an acceptable approximation of the chirp direction was found. For these ANOVA tests, the amount of variation will remain constant with respect to the type of noise used during training and the room where the beamformer was trained. Each of these sources had only two choices for possible configurations. The source of variation for the number of microphones was considered the combination of the microphones used for training the beamformers. For example, when two microphones were used the microphone configuration could be either {1,2}, {1,3}, or {2,3} giving three possible degrees of freedom for this source of variation.

The ANOVA process only indicates if a variation of a source condition is significant by stating if the average response has changed by comparing data that differs based a condition or set of conditions [**?**]. The ANOVA process does not explain how the variation affects the beamformer response in either a positive or negative manner, only the magnitude of

the change. To remedy this issue, the assurance score can be examined to determine how variations in conditions positively (or negatively) affected the beamformer performance between variation groups. The percent difference between the mean assurance score of different groups, as well as the standard error, can be calculated and compared over all beamformer directions. Observing these results per tested chirp direction can illustrate if the significance from the ANOVA process can be attributed to a given chirp direction or set of chirp directions. The actual chirp directions are used rather than the predicted beamformer directions as the excitation that was predicted may not always be aligned with the chirp direction. Using the chirp direction instead allows the number of true and false positives and negatives from the beamformer prediction results to be observed. These results give insight of how often the beamformers predict a correct excitation direction for the chirp signal. The number of false positives, or when the assurance score determines there was enough information to predict a direction and estimates the wrong excitation direction, should be as close to zero percent as possible. Additionally, the percentage of positive responses, or when the assurance score determines the amount of information from the beamformer bank was enough to predict a possible excitation direction, should be higher than the percentage of negative results. This requirement would mean that the beamformer predicts an acceptable estimate more often than an unacceptable one. With these tools, the performance of the beamformers can be described concisely, which assists in determining which set of beamformer training conditions can yield the better results.

4.2 Beamformer Aligned Tests

The ANOVA results of the data collected from the twenty four tests are reviewed. These tests include the training, validation, and translation data sets which align with the beamformer directions. This data should give the largest correlation response because the excitation signal is directly aligned with the trained directions of the beamformers for this scenario.

In order to validate if the ANOVA process generates a reasonable result, a configura-

tion that should have a similar result will be observed. If the two side channel configurations of {1,3} and {2,3} are observed, then a nonsignificant response should be received. The two side channels were selected for comparison, rather than all three two-channel configurations because the {1,3} and {2,3} configurations share the same geometric layout; {1,2} does not. With a similar geometric layout, the response from the beamformer banks of these two side channel configurations should be similar because they should receive and process data in the same manner. The chirp projected from the excitation speaker will approach these side channel configurations in the same path when the these configurations are flipped along the y-axis of the microphone array. This assumption means that the {1,3} response at the 45° arrival angle is assumed to be similar to the {2,3} response at the 315° arrival angle. The microphone configurations would be comparably aligned which means the beamformers should respond in an almost identical manner. This similarity in response should result in an f-value for the variation in the microphone configurations below the critical f-value. The other forms of variation for the beamformer can still receive a significant impact, but, based on our assumptions, the variation of two microphone configurations with a similar layout and the variation in the testing room should both be nonsignificant for this scenario. Table 4.1 provides the ANOVA results for the assurance score of the data sets for the {1,3} and {2,3} microphone configurations.

The variation in the number of microphones and their configuration has been labelled with the letter "A" for reference. The variation in the type of noise used during training and the variation in which training room was used have been labelled with the letters "B" and "C" respectively. This referencing is done because, as mentioned earlier, a three-way ANOVA can also provide the significance between the interaction of two or three conditions being changed. So the variation of "A x B" is the significance for when both the microphones and type of training noise are varied. While the critical f-value (labelled as "FCrit") and found f-value (labelled as "FFound") are the most important calculations to observe from the ANOVA process, the remaining reported calculations in the ANOVA result tables are the degrees of freedom, sum of squares, and the mean of squares of the beamformer variation conditions or a set of variation conditions that have been labelled as "dF", "SumSq", and "MeanSq" respectively. The sum of squares is the summation of the deviations of all observations

in a group to the group mean. The sum squares relates back to the total variance of the observation per group, where the groups are the conditions that have been varied for the beamformer. The mean squares is the sum of squares divided by the number of the degrees of freedom. This value is the sample variance for each source of variation. The "Error" given in the ANOVA result table is the residual error for the variability in the responses that cannot be explained directly by the ANOVA process [**?**].

When examining the results of Table 4.1, the assumption of an insignificant impact from the variation in microphone configuration and room geometry was found to be correct as indicated by the found f-value being lower than the critical f-value for the A and C variations. Along with these results, the variation in noise and the variation in the interaction between the microphone configuration and the training noise inclusion for the significant variations. All sources of variations that are deemed significant have been highlighted in the resulting table to ease observation of significant contributions to beamformer performance.

Table 4.1: Assurance Score ANOVA results for beamformers of the two microphone configurations on the sides of the microphone array ({1,3} and {2,3}).

Source of Variation	dF	SumSq	MeanSq	FCrit	FFound
Mics (A)	1	$2.76e^{-4}$	$2.76e^{-4}$	3.848	0.0165
Noise (B)	1	1.631	1.631	3.848	97.79
Room (C)	1	0.0147	0.0147	3.848	0.8806
A x B	1	0.0721	0.0721	3.848	4.324
B x C	1	0.0075	0.0075	3.848	0.4472
A x C	1	0.0107	0.0107	3.848	0.6439
A x B x C	1	0.0022	0.0022	3.848	0.1325
Error	1528	25.49	0.0167		
Total	1535	27.23	0.0177		

Feedback for the significance of the training room are desired as these results confirm the assumptions made in Section 2.3. As stated in Section 2.2 when discussing the formation of beamformers for the microphone array, the offline adaptive beamformer training process should take into account different system dynamics of the array and of environmental factors surrounding the array during training. Of these dynamics, the room dynamics in which the system is trained would be a concern if a significant impact was found. This would be a concern because if significant change in performance is found, then the system operation

would be limited to the room, or type of room, in which the system was trained. To better understand how these variations affect the performance of the beamformer banks, the percent difference for the different variations in the sources will be observed.

The next step is to observe the percent differences of the mean assurance scores for each variation source that was deemed significant. For this initial analysis, the nonsignificant variation for the source of the microphone configurations of the two side channels are included in this observation of the differences in beamformer performance. Including the variation of microphone configurations can illustrate how the percent difference of a nonsignificant variation compares to that of a significant variation. The results for the mean assurance score per arrival angle of the two side channel microphone configurations has been provided in Figure 4.1 for the sets of data where the chirp signal was aligned with one of the beamformer directions. For this variation, the results for the arrival angles show that there was a similar response across all directions when comparing the {1,3} and {2,3} microphone configurations. The {1,3} microphone configuration had an under-performing response from the 45° and 90° arrival angles but had an over-performing response at 270° and 315°. Both responses also had a similar variance which satisfied the ANOVA conditions. Since all ANOVA conditions had been satisfied, the variation due to the selected microphone configuration can be taken as nonsignificant when considering only the side two-channel configurations.

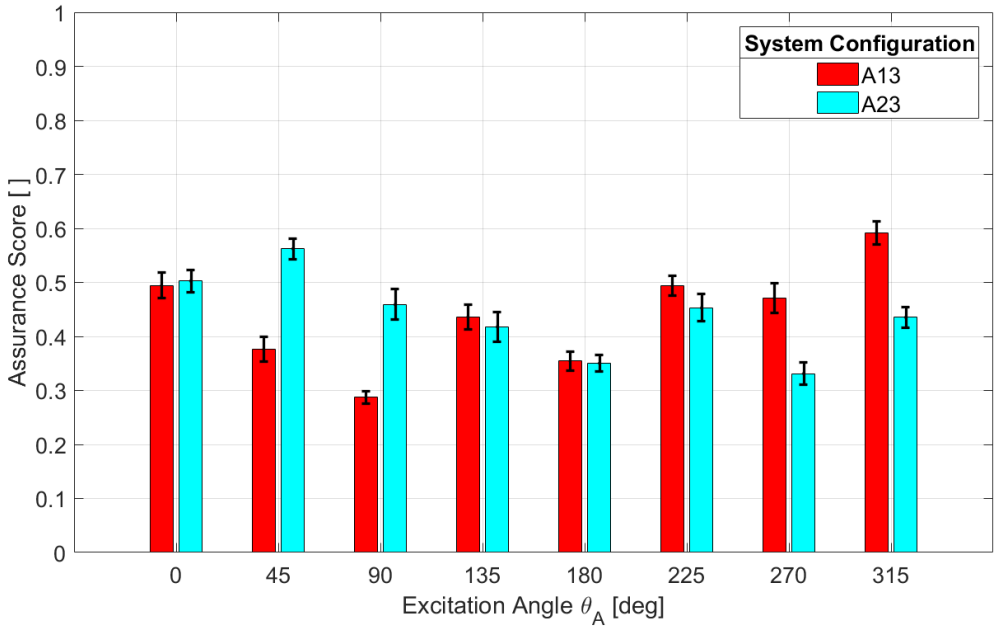


Figure 4.1: Output of mean assurance score with standard error for the eight excitation directions that align with the beamformer directions, where microphone configurations are varied from {1,3} to {2,3}

Figure 4.1 has both groups including the use of beamformers trained in the large and small rooms and beamformers that were trained with ambient and additive noise. The red (A13) series uses the {1,3} microphone configuration, while the blue (A23) series uses the {2,3} microphone configuration. The numbers in the series notation represent the channels used to represent the group shown, while the "A" notes that all of the noise and room conditions for the beamformer are included as well. The mean value across all arrival angles for {1,3} and {2,3} were found to be 0.4379 and 0.4388 respectively. The mean assurance scores result in a percent difference of 0.21% increase when comparing the {2,3} to the {1,3} microphone configuration. The performance of the beamformers were found to be similar.

The positive and negative response response rates of the compared series were then observed. The true and false connotation had been used to help decipher if the approximation made by the assurance score generation process was correct in the decision of accepting or rejecting the beamformer bank response. A result was considered positive if the beamformer bank response received an assurance score greater than or equal to 0.50. A true

positive result was considered if the beamformer bank output resulted in the known arrival angle was within one beamformer quantization direction, or 45° for this study, on either side of the predicted beamformer direction. The desired output for these tables was a beamformer response that has a near zero result for false positives and had a positive rating for a majority of the responses. Table 4.2 illustrates the results for the variation in microphone configurations for the {1,3} and {2,3} configurations. Unfortunately, neither of these system configurations resulted in acceptable response. Both configurations received responses that had a positive response percentage below fifty percent, but these configurations had a false positive response rate near zero percent. A false negative means that the beamformer bank may have predicted the correct excitation direction, but the beamformer bank may not have had enough information to do so conclusively. The results from Table 4.2 mean that the side two-channel configurations are not acceptable for the stipulations made in Section 4.1 for an acceptable beamformer bank response. The variation of the remaining significant sources will be observed to understand how these variations affect the response rates of the beamformer banks.

Table 4.2: Positive and negative response rates for the microphone configurations of {1,3} and {2,3}

A13	True	False
Positive	34.24%	2.34%
Negative	26.43%	36.98%

(a) Microphone Configuration {1,3}

A23	True	False
Positive	35.42%	1.17%
Negative	27.21%	36.20%

(b) Microphone Configuration {2,3}

The next source that was observed was the variation of using ambient or additive noise during training. The mean assurance score per arrival angle can be found in Figure 4.2 for the variation in the use of noise during training for tests in which the excitation signal arrival angle is aligned with the predicted beamformer angles. From these results, the beamformers trained with ambient noise under-performed against the beamformers trained with noise in all arrival angles except for one. The mean of each configuration and the standard error are similar for all arrival angles. This observation means the data has a similar variance for the received data so the conditions for the ANOVA process has been satisfied. With the ANOVA conditions satisfied, the variation of the training noise condition

for the beamformers is shown to have a significant impact on beamformer performance.

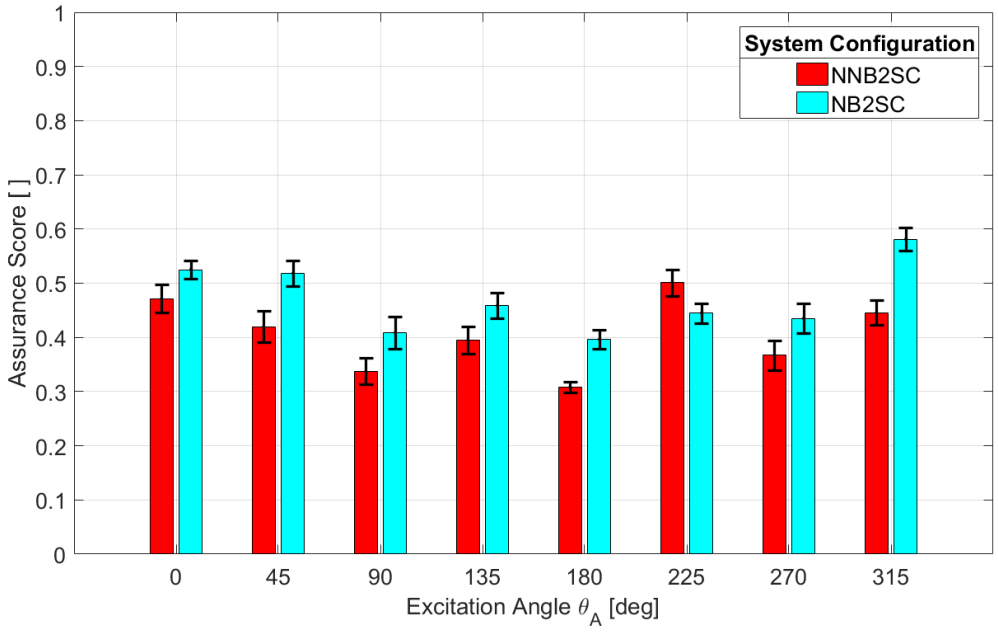


Figure 4.2: Output of mean assurance score with standard error for the eight excitation directions that align with the beamformer directions, where noise is varied from ambient noise to ambient and injected noise

Both groups in Figure 4.2 used beamformers trained from both the large and small room data and using the two side channel microphone configurations for the comparison. The red (NNB2SC) series represents the beamformer configurations where ambient noise was used during training, the blue (NB2SC) where additive noise was added used during training. The "NN" and "N" of the series notes that the group used either ambient or additive noise during training respectively. Along with this notation, the "S2C" notes that the {1,3} and {2,3} microphone configurations were used in the group. The "B" in this notation refers to both training room conditions being used for the groups. When comparing the mean assurance score across all beamformer directions, the mean assurance score for the beamformers that were trained with ambient noise was found to be 0.4058, while the assurance score for the those beamformers that were trained with additive noise were found to have an average of 0.4710 across all beamformer directions. This variation in the types of noise used in training results in a 14.87% increase in performance for the percent difference between these two

beamformer configurations. As stated in Section 2.3, adding noise to the environment while the beamformer was training can improve the performance of the beamformers, which results in higher values for the assurance scores.

The positive and negative response rates of the noise variation configurations are presented in Table 4.3. These rates are presented for beamformers where ambient and additive noise were used during training. The first condition of a positive response rate in the majority was not met by either beamformer configuration. The maximum positive response rate for the configurations was seen by the noise trained beamformers at approximately 44%, which is still below the threshold for an acceptable response. Additionally, while the positive response rate increased overall when training with additive noise, so did both the false positive and false negative rate. While both configurations still had near zero false positive response rates, an increase in the false positive rating was noted. This result means that using the beamformers trained with additive noise may result in higher average assurance scores, but using these beamformer configurations may also result in a slightly higher false report rating.

Table 4.3: Positive and negative response percentages for beamformers trained with and without noise

NNA2SC	True	False
Positive	29.04%	2.6%
Negative	36.20%	34.51%

(a) beamformer not trained with noise

NA2SC	True	False
Positive	40.63%	3.26%
Negative	17.45%	38.67%

(b) beamformer trained with noise

The variations in both microphone configuration and the addition of noise in training affect beamformer performance when varied simultaneously can be observed for the effects on beamformer performance. The output mean assurance scores for each of the eight arrival directions can be found in Figure 4.3 where the beamformers are the variations of the microphone configuration and the addition of noise in training. From the beamformer bank outputs, a single arrival direction did not directly contribute to any of the change in performance by drastically over or under performing as compare to the other arrival angles. Additionally, the addition of noise in training increased the mean assurance score for a

majority of excitation angles.

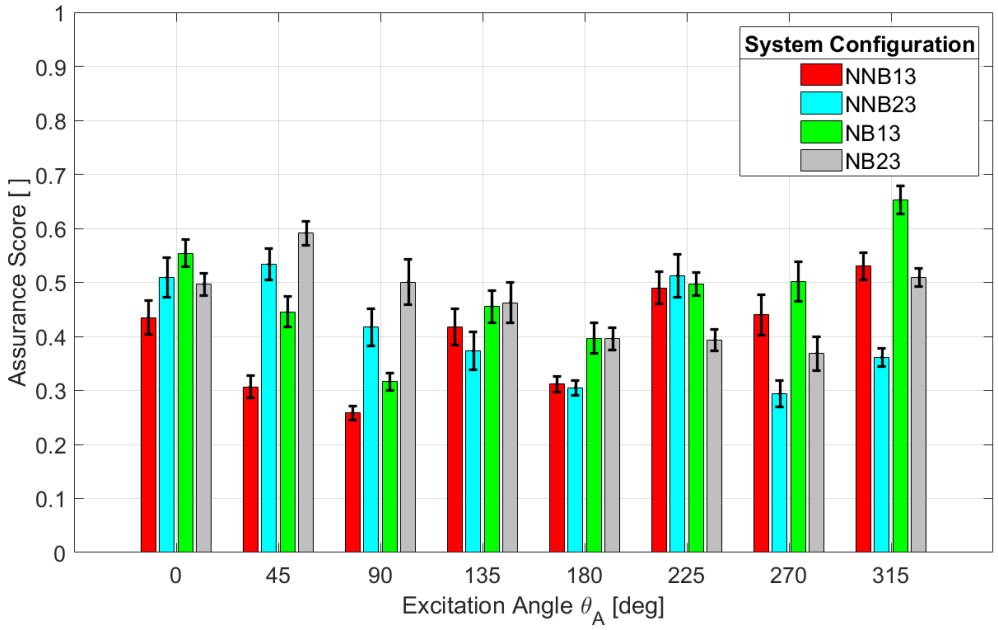


Figure 4.3: Output of mean assurance score with standard error for the eight excitation directions that align with the beamformer directions, where both the noise condition and the microphone condition for the side two-channel configurations

For this figure, the red (NNB13) and blue (NNB23) groups represent beamformers that were trained without noise for microphone configurations of {1,3} and {2,3} respectively while the green (NB13) and gray (NB23) series represent beamformers that were trained with noise for microphone configurations of {1,3} and {2,3} respectively. Contrary to the conclusion that was drawn when solely examining the variation in the training noise, the NNB23 beamformer performed as well if not better than the NB23 beamformer for the excitation directions of 0° and 270° arrival angles. For these cases, the assumption was made that adding noise during training would increase beamformer performance meaning a higher average assurance score should be attained. Since all of these cases do not match the stated assumption, then this finding may mean that using additive noise during training increases the likelihood of, but may not guarantee a higher assurance score.

The performance difference of the beamformer configurations were observed to describe which beamformer condition has an impact on beamformer performance. Table 4.4

displays the average assurance score for the beamformers configurations and their respective percent differences for when the excitation arrival angle is aligned with the designated beamformer directions. Table 4.4a shows the average assurance score results across all the excitation directions where the noise trained beamformers performed better than the beamformers trained without noise. Unfortunately, none of these group configurations have an average passing score across all of the arrival directions, but the noise-trained beamformers come close by having a minimum mean assurance score of 0.465, only 0.035 below an acceptable mean assurance score. Table 4.4b demonstrates the percent differences for the different combinations of system configurations. The comparisons in this table are how the column group compares to the row group. All instances where the column had a better performance than the row group have been highlighted for convenience. For these results, the highest differences are seen when comparing the beamformers trained with ambient noise to those trained with additive noise. Results for the percent differences were noted with a plus or minus to show the interaction between combinations to clarify if the change in performance was either positive or negative respectively. This table shows that the greatest difference is 18.01% between the NNC13 and NC13 beamformer configurations. Because the size of the performance difference between the beamformers trained with and without noise, the variation of training noise will be observed to understand if this source of variation has an impact to beamformer performance versus microphone configurations that do not have a similar geometric layout.

Table 4.4: Results from the variation interaction between microphone configurations and the type of training noise for the side two-channel configurations

Config	AS
NNB13	0.3985
NNB23	0.4131
NB13	0.4774
NB23	0.4645

(a) Mean Assurance Score

Config	NB23	NNB23	NB13
NNB13	(+)15.30%	(+)3.60%	(+)18.01%
NB13	(-)2.74%	(-)14.44%	
NNB23	(+)11.71%		

(b) Percent Differences of System Combinations

Now that the general trends for the beamformer performance have been observed over more specific cases, the results for the prediction of an acceptable estimate of the chirp

direction when varying both the type of noise during training and microphone configuration can be observed. Table 4.5 displays the results for the different possible combinations of these variations. As was seen with the earlier variation response results, all compositions have not passed the first requirement of an acceptable response rate. All beamformer configurations have a positive response rate below fifty percent for the tested data. Additionally, the false positive rating increased when switching from beamformers trained without noise to beamformers trained with noise. Caution should be used if the false positive response rate is near five percent for a given beamformer configuration because, as stated in Section 4.1, a negative response was assumed to be vastly more desired than a false positive response. From these response rates, the beamformers of the side two-channel microphone configurations cannot be accepted as none of the beamformer configurations meet the the two requirements of an acceptable response rate. This finding means the remaining configurations must be observed to see if an acceptable response can be found or if the requirements need to be altered, but the observations discussed have culminated into several trends that may assist in discovering what kind of configuration produces an acceptable beamformer.

Table 4.5: Positive and negative response percentages for the microphone configurations of {1,3} and {2,3} and the beamformers trained with ambient and additive noise

NNB13	True	False
Positive	25.52%	0.00%
Negative	34.38%	40.10%

(a) {1,3} trained without noise

NB13	True	False
Positive	42.97%	4.69%
Negative	18.49%	33.85%

(b) {1,3} trained with noise

NNB23	True	False
Positive	32.55%	0.52%
Negative	38.02%	28.91%

(c) {2,3} trained without noise

NB23	True	False
Positive	38.28%	1.82%
Negative	16.41%	43.49%

(d) {2,3} trained with noise

Since the {1,3} and {2,3} microphone configurations were observed for performance qualities, the remaining configurations for the microphones will need to be reviewed to see how the beamformer performance was impacted. The {1,2} microphone configuration was observed in conjunction with the side two-channel configurations from the previous comparison. Including the {1,2} channel configuration allows for the observation of any effects due to a slight change in geometric layout. As discussed in Section 2.1, these microphones

were assumed to be similar quality to which any differences between the microphones would be compensated for in the beamformer training process. The result of this ANOVA process for all of the two-channel configurations can be seen in Table 4.6 and should illustrate the impact of varying the microphones geometries on each of the two-channel microphone configurations. A primary difference from what was seen in Table 4.1 is the change in significance for the variation in microphone configuration. In this comparison, the variation for the microphone configuration was considered to be significant, as well as the variation in the type of noise and the interaction between microphone configurations and noise. This table shows that the significance for the variation in the room used for training is still not a significant impact on beamformer performance. The variation in noise for this comparison was still the dominant source of variations in terms of significance, but the significance of the variation in microphone configuration affirms the assumption that geometric layout of the microphone array has an impact on the beamformer performance. This significance means that the {1,2} microphone beamformer configuration should be observed to compare the performance difference between this beamformer and the two side channel beamformers.

Table 4.6: Assurance Score ANOVA results for the beamformers for all of the two microphone configurations of the microphone array ({1,2}, {1,3}, and {2,3})

Source of Variation	dF	SumSq	MeanSq	FCrit	FFound
Mics (A)	2	0.6505	0.3253	3.000	17.40
Noise (B)	1	1.539	1.539	3.848	82.31
Room (C)	1	0.0035	0.0035	3.848	0.1885
A x B	2	0.2816	0.1408	3.000	7.531
B x C	1	0.0263	0.0263	3.848	1.409
A x C	2	0.0971	0.0486	3.000	2.5967
A x B x C	2	0.0086	0.0043	3.000	0.2298
Error	2292	42.86	0.0187		
Total	2303	45.46	0.0197		

The results from previous the ANOVA process in Table 4.1 demonstrated that examining the interaction of variation sources rather than inspecting each source individually would yield the same result, so only the interaction between the variation of the microphone configuration and the type of noise used during training will be observed for this analysis. To clarify the information shown in each figure, the interaction of the two sources of varia-

tion are split into two separate figures. Figure 4.4 illustrates the mean assurance score per excitation direction for beamformers trained with ambient noise in both training rooms. The beamformer for the {1,2} microphone configuration performed as well as the {1,3} and {2,3} beamformer configurations, but the {1,2} configuration outperforms both of the other series only in the excitation direction of 180 degrees.

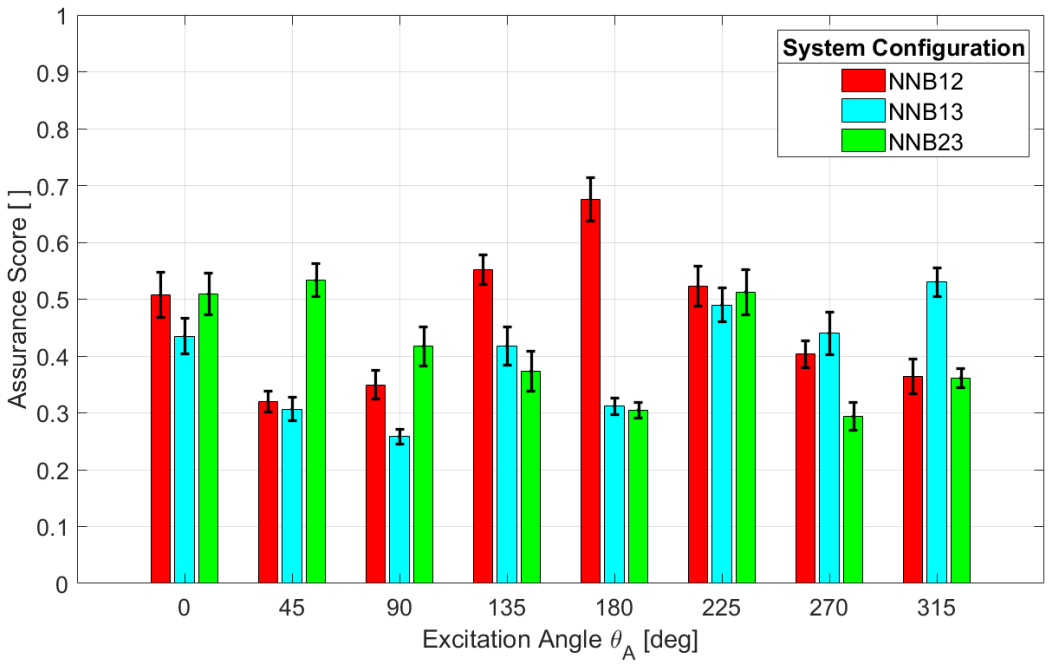


Figure 4.4: Output of mean assurance score with standard error for the eight excitation directions that align with the beamformer directions, where the microphone condition for all of the two-channel configurations for ambient room noise

For the beamformers trained with additive noise, the response was more consistent over the chirp directions. Figure 4.5 shows the mean assurance score per excitation direction for beamformers trained with additive noise in both training rooms. The {1,2} beamformer had a more consistent response over the excitation directions compared to the {1,2} beamformer trained with ambient noise. Additionally, the variation in the standard error for each excitation direction for the {1,2} beamformer configuration was larger than the standard error of either the {1,3} or {2,3} microphone configurations for beamformers trained with additive noise. These results show that the beamformer trained with additive noise in both training rooms have the tendency to perform better than either of the other configurations.

To understand more about how these additional beamformer configurations impact beamformer performance, the percent differences for the {1,2} beamformer configurations will be observed.

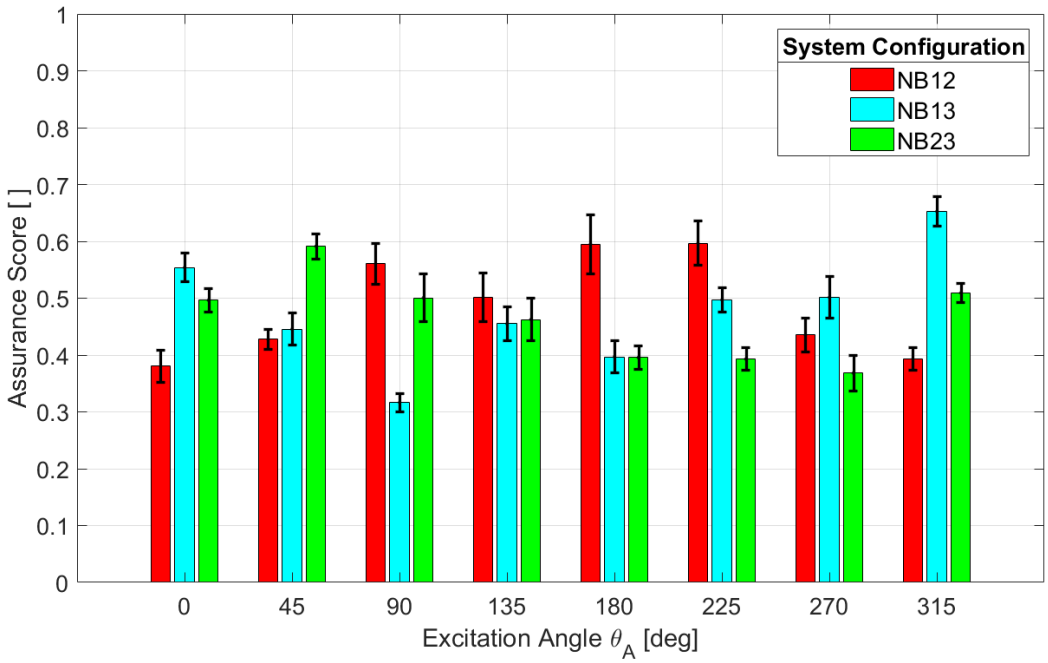


Figure 4.5: Output of mean assurance score with standard error for the eight excitation directions that align with the beamformer directions, using the microphone condition for all of the two-channel configurations for injected room noise

The assurance score for the {1,2} beamformer configurations were compared to the previous assurance scores of the {1,3} and {2,3} beamformer configurations. Table 4.7a shows the mean assurance score across all excitation directions that align with the designated beamformer directions. Table 4.7b displays the percent differences of the {1,2} beamformer configurations as compared to all of the two-channel beamformer combinations. For the mean assurance scores, none of the mean assurance score values were above the passing assurance score of 0.50 for any of the configurations. This means that while the beamformer performance may have improved when shifting from one of the other two-channel beamformer configurations, the improvement was not large enough to produce a positive result more than half of the time the beamformer was used. Looking at the percent differences, the main difference in the mean assurance score values comes from the comparing the beamformers

trained with ambient noise to those trained with additive noise. These beamformers had a percent difference of 14.67% and 11.09% compared to the NNC12 to the NNC13 and NNC23 beamformers respectively. This percent difference was similar to that of the {1,3} and {2,3} beamformers when compared in Table 4.4b in the side two-channel configuration ANOVA process.

Table 4.7: Results from the variation interaction between microphone conditions and addition of training noise for all of the two-channel configurations

Config	AS
NNB12	0.4616
NNB13	0.3985
NNB23	0.4131
NB12	0.4864
NB13	0.4774
NB23	0.4645

(a) Mean Assurance Score

Config	NB12	NNB12
NNB13	(+)19.87%	(+)14.67%
NB13	(+)1.87%	(-)3.37%
NNB23	(+)16.30%	(+)11.09%
NB23	(+)4.61%	(-)0.63%
NNB12	(+)5.23%	

(b) Percent Differences of System Combinations

The results for mean assurance scores, while under the desired 0.50 assurance score rating, improved from the previous observation. The NB12 beamformer resulted in a mean assurance score of 0.4864 across the eight excitation directions that were aligned with the designated beamformer directions, which is only 0.01 lower than the desired 0.50 assurance score.

To understand how well the assurance scores indicate the approximate chirp direction, the positive and negative response rates were examined. Table 4.8 shows the positive and negative response rates for the {1,2} beamformers. As was predicted, neither beamformer configuration was able to satisfy the first requirement for the results of a greater than fifty percent positive response rate. This response rate means that these beamformers were not acceptable to use over all cases in which the excitation signal is aligned with one of the beamformer directions. As can be seen from these results, the use of additive noise in training for the {1,2} beamformers both increased the overall positive response rate and decreased the false positive response rate from the beamformers trained with ambient noise. The decrease in the false positive rating differs from what was found in Table 4.5 for the response rate where by adding noise during training tended to increase the false positive response rate.

Table 4.8: Positive and negative response percentages for the {1,2} microphone configuration using beamformers trained with and without noise

NNB12	True	False
Positive	30.99%	7.55%
Negative	15.36%	46.09%

(a) {1,2} beamformer trained without noise

NB12	True	False
Positive	39.84%	1.82%
Negative	6.51%	51.82%

(b) {1,2} beamformer trained with noise

Finally, the case of all microphone configurations was observed to determine if there were any impacts to beamformer performance using the {1,2,3} microphone configuration. This three-channel configuration was examined because none of the two microphone configurations met the expectations for an acceptable response rate for any beamformer configuration and was apart of the original testing conditions. The results from the ANOVA process for all of the microphone configurations of the array can be found in Table 4.9 for all tests where excitation signals directly aligned with one of the beamformer directions. In this table, the significance for the variation in microphone configurations dramatically increased to become the dominant source of significance on beamformer performance. This result means that by adding the {1,2,3} microphone configuration to the comparison, a major shift in the mean assurance score was observed. The reason behind this change in mean assurance score is that the addition of microphones assists in the gathering of data in parallel as Section 2.2 discussed previously. This data illustrates a more accurate beamformer prediction can be made using a three-channel configuration over a two-channel confiugration, which means a higher assurance score can be achieved. As was seen in the previous comparison, the variation in the type of noise used in training and the interaction between microphones and noise are deemed significant, while the variation in the training room was still insignificant.

Table 4.9: Assurance Score ANOVA results for beamformers of all microphone configurations of the microphone array ({1,2}, {1,3}, {2,3}, and {1,2,3})

Source of Variation	dF	SumSq	MeanSq	FCrit	FFound
Mics (A)	3	13.77	4.590	2.608	272.3
Noise (B)	1	2.055	2.055	3.848	121.9
Room (C)	1	0.0286	0.0286	3.848	1.699
A x B	3	0.2816	0.0939	2.608	5.568
B x C	1	0.0101	0.0101	3.848	0.6010
A x C	3	0.1275	0.0425	2.608	2.521
A x B x C	3	0.0312	0.0104	2.608	0.6161
Error	3056	51.52	0.0169		
Total	3071	67.83	0.0221		

The mean assurance scores for the three-channel beamformer configurations was examined and compared to the two-channel counterparts. To observe the results of the {1,2,3} beamformer configuration, Figure 4.6 illustrates the mean assurance score for all of the microphone configurations per excitation direction for beamformers trained with ambient noise in both training rooms. The results for the NNB123 beamformer were consistent across the eight excitation directions and six of the eight directions had a mean assurance score above the 0.50 threshold required for a positive score. This observation means that the NNB123 had vastly improved compared to the two-channel configurations, which affirms the assumption made that the three-channel configuration would perform better than the two-channel configuration.

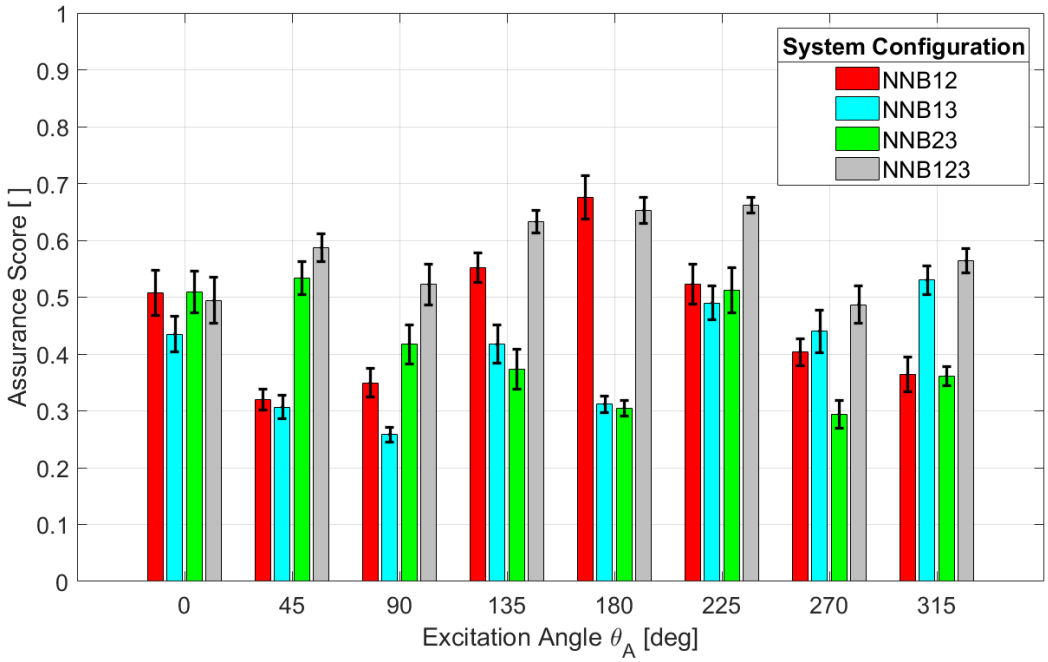


Figure 4.6: Output of mean assurance score with standard error for the eight excitation directions that align with the beamformer directions, where the microphone condition for all of the channel configurations for ambient room noise

These results show that the three-channel microphone configuration was superior to the two-channel configurations. For those beamformers trained with additive noise, Figure 4.7 shows the mean assurance score per excitation direction for all of the microphone configurations that were trained with noise and in both rooms. Figure 4.7 also demonstrates that the {1,2,3} microphone configuration had a consistent output for the assurance score over the eight excitation directions and results in an assurance score for all of the eight directions being above the required 0.50 threshold for a positive result. The improved performance in the {1,2,3} beamformer after changing to additive noise in training holds with the trends that were observed in the earlier comparisons. From the results of the beamformers trained with additive noise, the NB123 beamformer performed as well or better than any of the other noise trained beamformers. Six of the eight excitation directions resulted in above 0.60 for an average mean score over the six arrival angles.

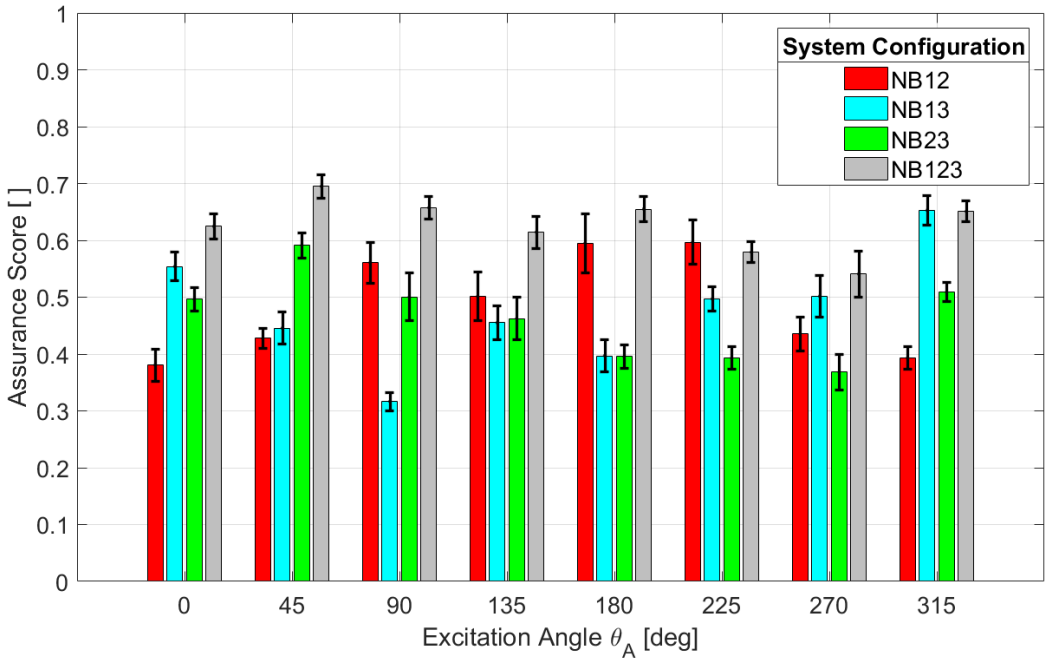


Figure 4.7: Output of mean assurance score with standard error for the eight excitation directions that align with the beamformer directions, where the microphone condition for all of the channel configurations for injected noise

The mean assurance scores for the three-channel beamformer configurations were next examined to see how including a third microphone changes the beamformer performance. To show the difference between the discussed beamformers, Table 4.10a displays the mean assurance score for the beamformer configurations for all of the excitation directions when the excitation signal was aligned with one of the beamformer directions. Table 4.10b shows the percent differences between the three-channel beamformer configuration mean assurance scores as compared to all of the observed beamformer configurations. From these results, the largest percent difference comes when comparing the two-channel beamformer configuration to the three-channel counterpart, which was predicted by the ANOVA process. When examining the resulting percent differences for the discussed beamformer configurations, the average difference is approximately 29% when observing beamformers that share the same noise condition for training, or comparing NNB123 to NNB12, NNB13, and NNB23. This observation means that using the three-channel configuration rather than any of the two-channel combinations for the beamformer configuration resulted in an increase on average

by 29% in performance, which is a noteworthy improvement. Additionally, the mean assurance scores of the {1,2,3} beamformers are above the required 0.50 positive response mark, meaning the implementing a three microphone system will acquire sufficient information for predicting the excitation direction regardless if the system is trained with or without noise.

Table 4.10: Results from the variation interaction between microphone configurations and addition of training noise for all of the channel configurations

Config	AS
NNB12	0.4616
NNB13	0.3985
NNB23	0.4131
NNB123	0.5752
NB12	0.4864
NB13	0.4774
NB23	0.4645
NB123	0.6271

(a) Mean Assurance Score

Config	NB123	NNB123
NNB13	(+)44.58%	(+)36.29%
NB13	(+)27.11%	(+)18.58%
NNB23	(+)41.15%	(+)32.80%
NB23	(+)29.79%	(+)21.29%
NNB12	(+)30.40%	(+)21.91%
NB12	(+)25.27%	(+)16.73%
NNB123	(+)8.63%	

(b) Percent Differences of System Combinations

The results from Table 4.10b show that using additive noise during training was also significant by the increase in the assurance score when switching between NNB123 to NB123. The difference between these beamformers results in an 8.63 percent increase in the mean assurance score. This observation builds on the general trends observed in Tables 4.4b and 4.7b for the two-channel configurations. In these cases, using additive noise during training would increase beamformer performance around 5% for all microphone configurations. So the assumptions made in Section 4.1 that the variation in microphone configurations and the variation in the type of noise in training would be significant have been confirmed, with the microphone configuration playing a larger role when the three-channel configuration was included as shown by the increased f-value and assurance score.

The positive and negative response rates for the three-channel beamformer configurations were also observed to indicate whether the three-channel configurations can accurately predict the arrival angle of the excitation signal. Table 4.11 contains the response rates of the three-channel configurations for the tests in which the excitation signal was aligned with one of the beamformer directions. From these results, both beamformers surpass both

requirements for an acceptable beamformer in general use. Both beamformers had a positive response rate over 79%, meaning the beamformers have satisfied the first condition of an acceptable response rate. The NB123 beamformer had a positive response rate slightly over 90% for the excitation signals that were aligned with one of the beamformer angles. Additionally, the false positive rates are both near zero where both beamformers have a response rate under one percent. These findings for the three-channel configurations are a considerable improvement from the two-channel configuration beamformers, and the findings mean that a viable beamformer or set of beamformers were been found.

Table 4.11: Positive and negative response percentages for the {1,2,3} microphone configuration using beamformers trained with and without noise

NNB123	True	False
Positive	78.91%	0.52%
Negative	2.34%	18.23%

(a) {1,2,3} beamformer trained without noise

NB123	True	False
Positive	89.84%	0.26%
Negative	1.30%	8.59%

(b) {1,2,3} beamformer trained without noise

Based on these results, the NB123 beamformer was more recommended for general use as this beamformer configuration yielded a positive response rate above 90% overall and had a near zero false positive rate. This beamformer also had a consistent assurance score over the eight beamformer directions with a low standard error for the given excitations. While either three-channel beamformer configuration can be used, the NB123 was recommended because this beamformer performs better than ambient noise-trained counterpart in all desired characteristics. In order to ensure that these configurations work for overall use regardless of the signal alignment, the functionality of the three-channel beamformers will be tested on cases in which the excitation signal is not aligned with one of the beamformer directions as this recommendation can only be made for when the excitation signal was aligned with one of the eight beamformer angles.

4.3 Supplementary Tests

This section will cover the supplementary tests performed to analyze how well the three-channel beamformer performed when receiving data for which the beamformer had not been trained. These supplemental tests include those where the excitation signal was not aligned with one of the beamformer directions, as well as a pure noise test. These scenarios are meant to represent the worst cases that these beamformers can receive, which allows for a general view of how the beamformers perform overall, rather than just testing with trained data sets. The acceptable response requirements for the beamformers remain the same as with the previous tests. These requirements are that the positive response rate should be the majority of the responses received with a false positive rate that is near zero. Along with these criteria, the pure noise tests should result in an assurance score below 0.50 for all of the tested beamformers. This requirement means the beamformers would not determine that a prediction for the chirp direction was acceptable when the chirp signal was not received.

To receive the worst possible results for the cross correlation when the chirp signal is used, the excitation will originate from between two beamformer quantized directions. The selected excitation directions were limited to four angles spread over the four quadrants in the two dimensional plane of the microphone array. Assuming that the responses for the beamformers are consistent for angles that are separated by 180°, the excitation angles were selected as 67.5, 112.5, 202.5, and 337.5 degrees for the sample excitation directions. These excitation directions were gathered in all tests that were not used for training, for a total of twenty tests or a total of eighty data points.

For these chirp signals, only the {1,2,3} microphone configuration beamformer was used to test for acceptable responses. This decision was made because only the three-channel configuration had beamformers that were capable of meeting the requirements of an acceptable beamformer when the chirp signal was aligned with one of the beamformer directions. Using only this beamformer configuration, there are a possible four different beamformers that can be used. The four three-channel beamformers are the beamformers trained with additive noise in small and large training rooms and the beamformers trained

with ambient noise in the small and large training rooms. This low number of variations means the ANOVA process is not directly needed and the results can be observed directly to see the differences in the performance between the beamformers. To see how the possible variations differ in performance, Figure 4.8 contains the mean assurance score for the four excitation directions of the four three-channel beamformer configurations. In these results, the beamformers output a comparable mean assurance score for each excitation direction. These results only differ when the NNS123 beamformer over-performs and the NNL123 beamformer under performs at the 337.5 degree excitation direction. This observation means that the performance difference should be nominal across all configuration comparisons.

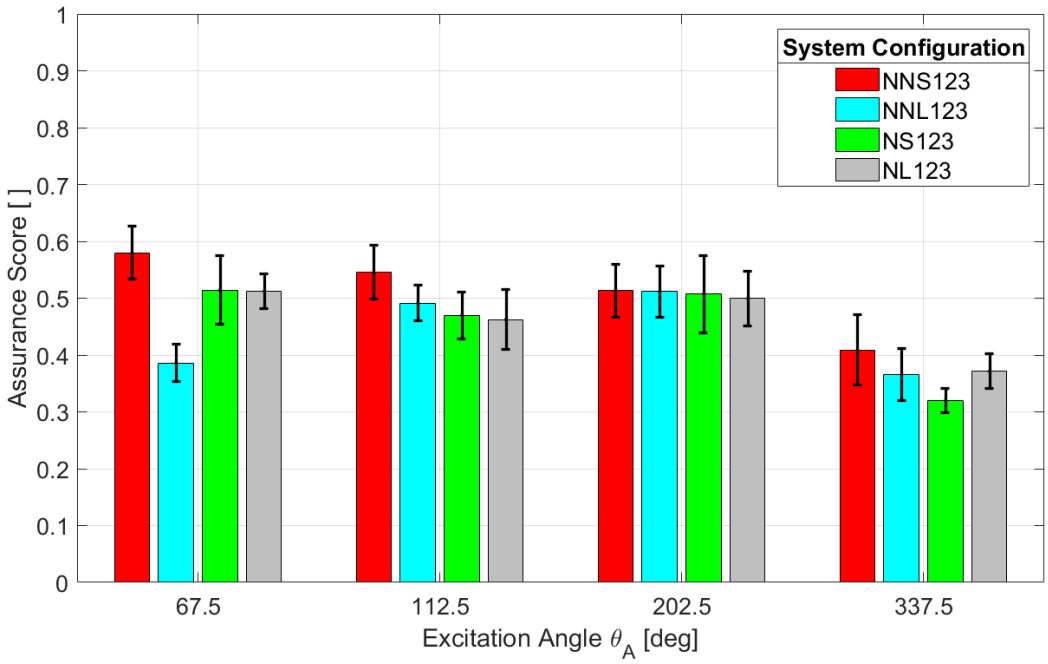


Figure 4.8: Output of mean assurance score with standard error for the four tested excitation directions across all tests using both room settings, where the excitation signal is not aligned with a beamformer direction

These results were not ideal for an acceptable beamformer due to the criteria that was used in this experiment. Out of the responses, all beamformers produced a mean assurance score below 0.50 for the 202.5° excitation angle. This result was similar to the responses of the two-channel beamformer responses from Section 4.2 where the mean assurance scores were under 0.50, which would produce a positive response rate below fifty percent. To see

how well the beamformers performed overall for the nonaligned tests, Table 4.12a shows the mean assurance score across the nonaligned excitation angles, and Table 4.12b depicts the percent differences between all of the beamformer combinations. The notation for these groups of beamformer follow the previous notation from Section 4.2, but now includes the notation for "S" and "L" that refers to training in the small or large room respectively. Only the NNS123 beamformer configuration generated a response that had an assurance score over the threshold for a positive response with a 0.5119 assurance score over the four excitation directions. From past experiences, this observation means that most likely the NNS123 beamformer will be the only beamformer configuration to pass the first criteria of an acceptable beamformer response for this experiment, but the second criteria for a near zero percent false positive response rate will need to be verified. The results in the table show that there is not a large discernible difference between the beamformer responses, but the NNS123 beamformer did still generate a response that was above the specified positive threshold. This makes the NNS123 the only candidate likely to pass both criteria of an acceptable beamformer response.

Table 4.12: Results from the variation interaction between room and noise conditions for the three-channel configuration

Config	AS
NNS123	0.5119
NNL123	0.4386
NS123	0.4529
NL123	0.4612

(a) Mean Assurance Score

Config	NNS123	NNL123	NS123
NL123	(+)10.42%	(-)5.02%	(-)1.82%
NS123	(+)12.23%	(-)3.21%	
NNL123	(+)15.42%		

(b) Percent Differences of System Combinations

The positive and negative response rates for the four beamformer configurations were observed to determine if any configuration could be deemed acceptable according the beamformer response criteria. Table 4.13 contains the response rates of the four beamformer configurations for the four non-aligned excitation angles. As was predicted, the only beamformer configuration to satisfy the first condition of majority positive responses was the NN123 beamformer, but this beamformer fails the second condition of a near zero false positive response rate by having approximately 30% false positive response rate for the non-

aligned excitation signals. This evidence means that none of the current beamformer configurations are acceptable to use when the excitation angle is not aligned with the beamformer directions with the available options for configurations.

Table 4.13: Positive and negative response percentages for the {1,2,3} microphone configuration where the noise and room conditions are varied when excitation is not aligned with a specific beamformer direction

NNS123	True	False
Positive	40.00%	28.75%
Negative	18.75%	12.50%

(a) NNS123 Beamformer Configuration

NNL123	True	False
Positive	17.50%	16.25%
Negative	40.00%	26.25%

(b) NNL123 Beamformer Configuration

NS123	True	False
Positive	33.75%	5.00%
Negative	37.50%	23.75%

(c) NS123 Beamformer Configuration

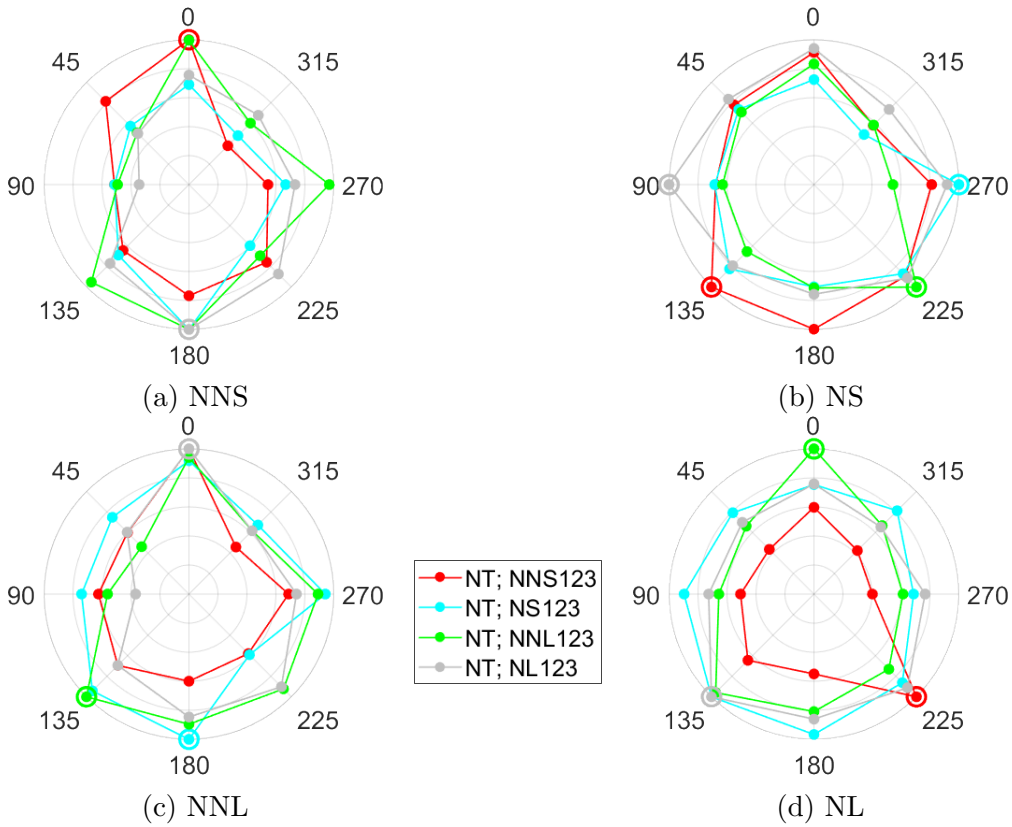
NL123	True	False
Positive	27.50%	12.50%
Negative	30.00%	30.00%

(d) NL123 Beamformer Configuration

These results indicate that a three-channel microphone array was not sufficient using the discrete beamforming technique with the current variations available. From the responses, the NNS123 beamformer generated responses with over a 60% positive response rate, but this beamformer also generated a false positive rate of almost 30%. This observation means that approximately one-third of the responses would conclude an acceptable prediction for arrival angle when the true chirp direction was beyond the acceptable threshold of 45° of the predicted excitation direction. For this system, the three-channel configuration was verified to fulfill the acceptable beamformer response conditions for when the excitation signals are aligned directly with the beamformer directions, but these beamformers were unable to satisfy these same conditions when the excitation was not aligned with one of the beamformer directions. The system would need to be altered and retested to satisfy these conditions.

Beyond the observations for excitations not aligned with one of the beamformer directions, another test set for the beamformers was the response the beamformers had to a pure noise signal. These noise signals were gathered by placing the microphone array in the orientation that would be used for an excitation signal coming from zero degrees. From this orientation, the array would gather two different sources of noise for the train-

ing rooms: ambient noise and additive noise. The initial responses for each of the rooms was provided in Section 2.1 for the SNR of the ambient and additive noise. For the actual beamformer bank responses to the pure noise base excitations, Figure 4.9 contains the beamformer responses to the small and large rooms with ambient and additive noise for the three-channel beamformer configuration. A majority of these responses had a spread similar to the pure noise expectated response as seen in Figure 3.3b for the beamformer responses for the three-channel beamformer configurations. From these responses, the expectation was that assurance scores for each of the responses should be below the threshold for a positive result. This outcome was desired because the beamformers should be able to neglect signals in which the chirp signal does not exist.



(a) NNS (b) NS

(c) NNL (d) NL

Figure 4.9: Beamformer responses for the excitations that are pure noise with no desired signal given to test noise handling capabilities of the three-channel beamformer configuration

In order to best examine the results of the beamformer responses due to a pure noise input, Table 4.14 has the response per room and noise condition used for the excitations with

the beamformer configuration used to generate the response. These results show that fifteen of the sixteen responses yielded a negative response with only the NNL123 beamformer in the small room with no added noise scenario generating a positive response. Along with the near positive responses, this finding means that more noise testing is required to verify how well the system can interpret a pure noise signal.

Table 4.14: Noise testing of the three-channel beamformer configurations in the small and large testing rooms with the three-channel beamformer configurations

		Room and Noise Settings			
		NNS	NS	NNL	NL Noise
Config.	NNS123	0.2816	0.2780	0.4157	0.4450
	NNL123	0.5916	0.2402	0.2946	0.4560
	NS123	0.2600	0.4453	0.2253	0.2320
	NL123	0.2989	0.1867	0.2520	0.2693

As can be seen in Table 4.14, the beamformers trained with additive noise on average produced a lower assurance score over all of the room and noise setting combinations. From this observation, the beamformers trained with additive noise were considered to perform better than the beamformers not trained with noise since the assurance score is closer to zero for a pure noise excitation. This result follows the general trend that the beamformers trained with additive noise saw an increase in performance compared to their counterparts that were not trained with noise.

From these tests, two conclusions can be drawn. The first was that while the three-channel beamformer configuration was well suited for predicting excitation directions when the excitation direction was aligned with a beamformer direction, that same beamformer was not capable of reliably predicting the direction of an excitation source when the excitation was not aligned. To correct this issue, more microphones could be added and tested with the array, as the addition of microphones greatly improved the acceptability of the beamformer performance as observed in the aligned excitation tests. The second conclusion was the response of the three-channel beamformers to a pure noise excitation where the beamformers tend to have a better response when the beamformer had been trained with additive noise. To verify how well the beamformers handle the noise excitation, the beamformers should be

tested multiple times for each noise and room condition combination as the initial results for a single test indicate that the beamformers are capable of generating a negative response for a signal in which the chirp signal does not exist.

Chapter 5

Summary Remarks and Conclusions

In this study, an experiment to implement and evaluate a microphone array beamformer utilizing inexpensive electret microphones for a service robotic rover and home sentry was performed. A literature review of the design and implementation of microphone array beamformers was examined to understand the desired performance characteristics and potential problems which can arise during implementation. As shown in this review, a microphone array can exist as either a linear, planar, or volumetric array with uniform, nonuniform, or random spacing between microphones on the microphones. The beamforming method for this study was selected as a filter and sum offline adaptive training process due to this approach allowing for a greater deal of suppression of interference from noise and signal reverberation within the testing room. Additionally, the filter and sum offline adaptive training process reduces the number of parameters that a user is required to determined while producing beamformers. A microphone array was utilized with a three-microphone configuration inspired design with {1,3} and {2,3} microphone configurations having a similar spacing but smaller than the spacing of the {1,2} microphone configuration. The excitation signal was selected to be a chirp as this could mimic the human whistling spectrum. This chirp would need to be both able to be generated by the average human and have a near consistent frequency response range across the microphone. From these constraints, the excitation whistle was chosen to be a chirp signal from 1 to 2.5 kHz for the unit to detect and localize.

The beamformer training process was selected as an LMS algorithm to determine using the optimal Weiner solution per the selection of the gradient step size, the number of filter weights, and the number of loops the filter training occurred. A Leaky LMS was included in case convergence was not obtained for a given filter, but all filters converged without the need for the leak coefficient. This beamforming training approach was selected since this method minimized the error between the chirp training signal and the received filtered response, meaning a close approximation could be made from the received data to the desired signal. Beamformers were trained for eight evenly spaced angles around the microphone array to give a resolution of 45^o for the implemented system. These beamformers were trained adaptively offline to generate a response based off a training data set for the specified filter angles that were of interest in this research.

Training data was collected for scenarios where the system was subjected to ambient and additive noise. Additionally, the system was tested in two training rooms that differed in size to determine if any differences existed in the responses from the filtered data. The beamformers were trained empirically for these four room conditions, as well for the {1,2}, {1,3}, {2,3}, and {1,2,3} microphone configurations, giving a total of sixteen possible beamformer arrays for this experiment. Along with the training data sets, two sets of validation data was also collected for each of the environmental training parameters, giving eight more data sets. Additionally in the small testing room, three tests of X, Y, and Z axis translations were taken as well, resulting in a total of twenty-four tests for the experiment.

Once the data was collected and the beamformers were trained, the output responses for the beamformer banks were characterized to determine if an acceptable prediction for the chirp direction could be estimated. This validation method resulted in the creation of the assurance score, which used three metrics of the cross-correlation between the training signal and the output responses from the beamformers to determine if the predicted beamformer direction was a suitable approximation of the excitation angle. The assurance score outputs for the training and validation data was tested with a three factor ANOVA to ascertain how the variations in the configuration of microphones in the array, the presence or absence of noise in training, and the room in which the system was trained impacted the performance of

the beamformers. These results indicated the significance of each source of variation, which was then used to inform on what training conditions warranted further investigation. From these reports on significant sources of variation, the positive and negative response rates were observed to give the final approval of the use of a given beamformer configuration for a certain scenario of operation.

Of these scenarios, the tests where the excitation chirp was aligned with one of the beamformer angles resulted in the conclusion that only the three-channel beamformer configuration was found to have an acceptable response rate. The NNB123 and NB123 beamformers produced results that had an overall positive response rate of 79.5% and 90.1% respectively. Both of these filters also produced a low false positive response rate of 0.52% and 0.26% for the NNB123 and NB123 beamformers respectively. The beamformers trained with additive noise tended to produce higher assurance scores, lower false positive response rates, and higher overall positive response rates. This observation validated the hypothesis that training with noise would produce more acceptable responses for the beamformer responses. When examining tests where the arrival angle for the chirp signal was between two of the trained beamformer angles, no filter configuration that was examined in this study satisfied the conditions of an acceptable response rate. These results represent the worst case scenario, where the chirp exists but is not aligned closely with any one specific beamformer direction. Suggestions for improvement would be to incorporate either more microphones on the array or more trained beamformer angles. The addition of microphones would allow more data to be used for the comparison of the directions examined. Unfortunately, this option would mean that the system would need to be capable of processing a minimum four signals, which may not be possible for some systems if the amount of memory is not available for allocation or the system does not have the capability to receive more than three signals at a time. Adding more trained beamformer angles would allow the system to have more points of reference, allowing for a higher probability of receiving a high cross correlation with the excitation chirp. This result does not guarantee that all angles can be represented accurately for the microphone array. Additionally, the inclusion of more beamformer quantization angles results in greater processing time for producing a beamformer bank response. The final tests where no excitation chirp was provided to the beamformer, and only the

ambient and additive noise signals were recorded found only one undesired response from the sixteen comparisons. To affirm that the observed trend was correct, more validation data is suggested for the room conditions of ambient and additive noise to examine if the trend continues.

From this study, a batch of beamformers were trained using a three-microphone array to detect and localize a human whistle in two testing rooms. These beamformers were trained with data from the two rooms of differing size with the use ambient and additive noise for this array. After training, these beamformer outputs were characterized using a scoring metric deemed the assurance score to inform the user which direction was the best approximation of the arrival angle for the source signal and whether this prediction was valid enough to accept. The results of the research indicate that training with additive noise and more microphones significantly improves the output response of the beamformers for the produced assurance scores. Of the examined beamformers, the beamformers that implemented three microphones were able to be deemed acceptable based on the criteria of a majority of valid predictions with a near zero false positive response rate for the tests in which the excitation arrival angle was aligned with one of the trained beamformer angles. The tests in which the excitation arrival angle was not aligned with one of the trained beamformer angles no examined beamformer configuration was found to have an acceptable response. Finally, tests in which the excitation chirp was not used and only the interference signals from the noise were provided to the microphone array indicated that the trained three-channel beamformers were capable of rejecting an approximation for an arrival angle for fifteen of the sixteen trials.