Computationally Efficient Methods for Detection and Localization of a Chirp Signal

Aditya Kashyap

Thesis submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Master of Science

in

Electrical Engineering

Amos L. Abbott, Chair

Steve C. Southward, Co-chair

Pratap Tokekar

December 05, 2018

Blacksburg, Virginia

Keywords: Microphone Array, Chirp Detection and Localization, Computationally Efficient

Computationally Efficient Methods for Detection and Localization of a Chirp Signal

Aditya Kashyap

(ABSTRACT)

In this thesis, a computationally efficient method for detecting a whistle and capturing it using a 4 microphone array is proposed. Furthermore, methods are developed to efficiently process the data captured from all the microphones to estimate the direction of the sound source. The accuracy, the shortcoming and the constraints of the method proposed are also discussed. There is an emphasis placed on being computationally efficient so that the methods may be implemented on a low cost microcontroller and be used to provide a heading to an Unmanned Ground Vehicle.

Computationally Efficient Methods for Detection and Localization of a Chirp Signal

Aditya Kashyap

(GENERAL AUDIENCE ABSTRACT)

As humans, we rely on our sense of hearing to help us interact with the outside world. It helps us to listen not just to other people but also for sounds that maybe a warning for us. It can often be the first warning we get of an impending danger as we might hear a predator before we see it or we might hear a car brake and slip before we turn to look at it.

However, it is not merely the ability to hear a sound that makes hearing so useful. It is the fact that we can tell which direction the sound is coming from that makes it so important. That is what allows us to know which direction to turn towards to respond to someone or from which direction the sound warning us of danger is coming. We may not be able to pinpoint the location of the source with complete accuracy but we can discern the general heading.

It was this idea that inspired this research work. We wanted to be capable of estimating where a sound is coming from while being computationally efficient so that it may be implemented in real time with the help of a low cost microcontroller. This would then be used to provide a heading to an Unmanned Ground Vehicle while keeping the costs down.

Acknowledgments

First and foremost, I would like to thank my advisor Dr. Steve Southward for his support and guidance througout this research. He was instrumental in helping me clearly define the problem, keeping me focused when I would go off track and guiding me when I got stuck or overwhelmed with a problem. I would also like to thank my committee members, Prof. A. Lynn Abbott and Prof. Pratap Tokekar for taking the time to read my thesis and providing me with valuable inputs to help improve my work. Finally, I would also like to thank my parents and my friends for their support and encouragement.

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 prior research relating to this topic will be examined to understand problems faced by previous groups and hence develop strategies for overcoming them. 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 research is focused on the detection and localization of a sound source generating a chirp while being computationally efficient so that it may be implemented in real time on a low cost microcontroller using a cost efficient sensing platform. The sound source may be a computer generated chirp or a human whistling in a chirp like manner.

The motivation was the desire to enable an Unmanned Ground Vehicle to estimate the direction of a sound source just as humans can while staying within a low cost budget. This could be used to provide it with an accurate heading or as a first estimate of where to go

while keeping the cost down.

For almost 20 years now, there has been research in enabling robots to respond to human operators through spoken commands. [13] [12]. There has also been extensive research towards the recognition of speech commands [11] but the current research relaxes the requirement to recognize different commands and focuses on detecting only one command, namely a chirp.

In the past, different methods for the detection of whistles have been proposed. One method focus on extracting features characterizing the incoming sound using different methods and analyzing the feature vector against known characteristics of a whistle for detection [14]. This method has the drawback of being computationally expensive.

Another method relies on passing the incoming audio through 3 band pass filters, converting the AC waves to a DC voltages and comparing the output of the band pass tuned to the whistle to the output of the other two filters [9]. This method needed not only a fast microcontroller but also dedicated hardware filters.

Thus there was a need to develop a method that could detect a whistle in a computationally efficient manner. To achieve this, a constraint was put on the whistle. Instead of trying to detect any and all human whistles, the developed method looked at detecting only certain kinds of whistles. This allowed the development of a method that could be computationally and memory efficient. The whistle being analyzed will be discussed in detail in Section 3.3.

The idea for using a sound source instead of a vision based system was motivated by the fact that a vision based system would need cameras that are invariably costlier than simple microphones and that would also require substantially more computational power which would further increase the cost. Vision based systems are also limited to responding to signals that are within their line of sight whereas a sound based system does not have that limitation.

The idea for localization of the whistle was inspired by the observation that humans often hear a sound or disturbance before they see the cause of it and can instinctively turn in the general direction of the source to pinpoint its location with vision. Hence, the aim was to produce results that could be used to provide a general heading and if need be, other sensors could be used to further discern the location to a better accuracy.

The commonly used methods for the localization of a sound source are Interaural Time Difference, Interaural Phase Difference or Interaural Level Difference [4] [3]. These techniques rely exclusively on the fact that signals received by the microphones are shifted versions of one another and this shift depends on the location of the microphones relative to the source. There are different methods developed that rely on this relative shift of the audio signal. These include beamforming [17] and time delay estimation methods [6][10]. However, these methods have the drawback of being computationally intensive.

Another method that was shown to be effective was Bayesian Acoustic Localization [5] which was a generalization of the TDE method by applying Bayes rule to get a probabilistic estimate of the location. However, this system was found to be effective only up to an SNR of 3dB while also being computationally expensive.

Recent developments have explored the use of Recurrent Neural Networks[13]. These techniques use a multilayer perceptron feedforward neural network structure and work by forming a feature vector after computing the cross-correlation power between samples from adjacent pairs of sensors. This method was able to locate a source up to an accuracy of 3.5° but it further drove up the computational cost required.

Since, this system was targeted for use on an Unmanned Ground Vehicle acting as a home sentry, it was considered acceptable to have a tolerance of $\pm 10°$. The vehicle could repeatedly listen to the source and improve its overall accuracy by reestimating the location as it moved

towards the source.

1.2 Review of Microphone Arrays

A microphone array is a group of microphones operating in tandem for a common purpose. Microphone arrays can be linear, such that all microphones are arranged in one plane [15] or spread out spherically in 3D space [2]. The number of microphones used can be as low as two or as even as many as 10 [16]. The arrangement of microphones depends on the intended use case. A spherical array would be used when the calculation of both the altitude and azimuth angle are required whereas a planar array would be more appropriate when only the azimuth angle needs to be determined.

The number of microphones used also has a direct impact of the cost of the system and the resolution of the estimated angle that can be expected. [1][15].

In this research, an array of four microphones arranged in a plane perpendicular to each other was used to analyze the sound coming in from different directions. The aim was to estimate the angle of arrival on a 2D plane and hence a planar array was used. The four microphones were separated by a physical barrier to give them a sense of directionality. These would act as the 'ears' of the sentry and by calibrating the microphones and physically separating them we were able to cut down on the computational needs by removing the need to build an adaptive filter or an acoustic beamformer. Effectively, a clever physical design helped offset the computational cost that would have been required for beamforming or adaptive filtering.

1.3 Review of Chirp

A chirp is a signal in which the frequency increases or decreases with time. It is also known as a sweep signal.

Prior research on Human whistles indicate that the typical frequency range of whistles tends to be from 500-5000Hz [14] [9] and it can be observed in Section 3.3 that it is easy for humans to whistle in a chirp like manner. That is, the frequency of a whistle can be made to increase over time.

Hence it was decided to use a linear up-chirp whose frequency increases from 1.2kHz to 1.7kHz over a duration of 0.1 seconds as the ideal sound source and this was later generalized to using a chirp like human whistle. The reasons for selecting these frequencies and this duration are explained in further detail in Section 3.3.

Other common commands that are easily reproducible include an impulse such as a clap or snap but a whistle could be louder and be spread across a wider bandwidth.

1.4 Thesis Organization

This thesis covers the construction of the sensing platform, the source signal, the detection of the source signal and the localization of its position.

It focuses on being computationally efficient to get a general sense of heading. The aim was not to pinpoint a sound source but rather to be able to tell the general direction of the sound source in a timely manner despite using only a low cost microcontroller and sensing platform. The system was tested under a variety of different scenarios to determine the conditions under which it was successful in estimating the direction of the sound source as

well as conditions that caused an incorrect estimate of the sound source.

Chapter 2 gives details about the sensing platform and its design and calibration. It defines the coordinate system and the conventions used in this thesis. It also discusses the setup that was used for running the experimental tests.

Chapter 3 goes into detail about the method used for detection of the command signal.

Chapter 4 discusses the methodologies developed to estimate the direction of the sound source.

Chapter 5 discusses the results obtained for the detection of the command signal in simulation and real world testing. It details the different scenarios under which the detection methodology was tested. It also discusses the results obtained for the localization of the source under different scenarios. Finally it states the appropriate working scenarios for the proposed methods based on the results obtained.

Chapter 6 summarizes the work done, presents the conclusions reached about the detection and localization methodologies that were developed and lays out the appropriate working conditions for the methodologies.

1.5 Contributions

This thesis will detail the sensing platform and the methodologies developed for the detection and localization of a source signal which could be a computer generated Chirp or a human whistle.

Chapter 2

Hardware and Experiment Design

2.1 Sensing Platform

The sensing platform consists of a four microphone array arranged in a plane perpendicular to each other so as to point in 4 different directions.

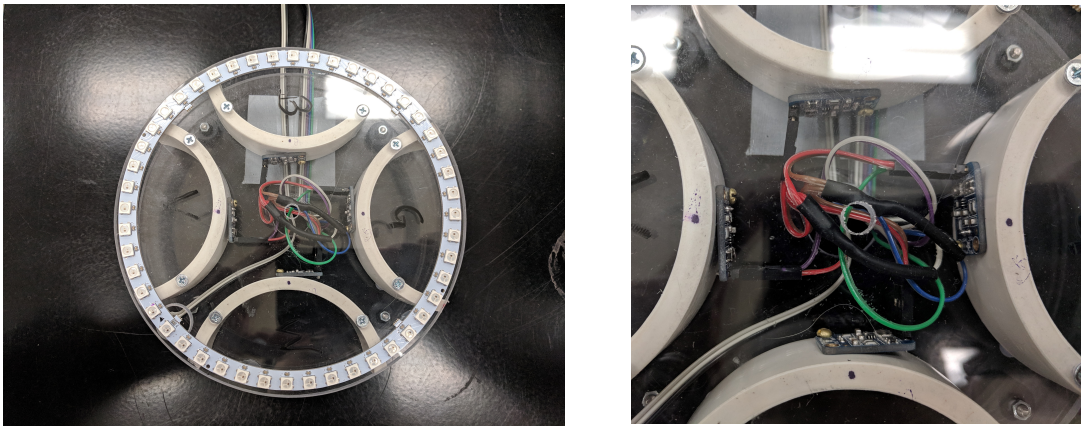
The microphones were sandwiched between two polycarbonate planes that served as the mounting base with holes cut out for the wires. An LED ring that could be used to visually indicate the final direction was also implemented. The microphones were separated by 4 C-shaped sections cut from a PVC pipe. Effectively, each microphone encased within the enclosure acted as a directional 'ear' and hence there were four 'ears' listening in different directions.

The microphones used were low cost electret microphones [1], each connected to a separate integrated MAX4466 OP-AMP. The microphones were calibrated by adjusting the gain to ensure that the peak to peak Voltage level was the same for all four microphones when exposed to the same sound source. The calibration method is discussed in the upcoming

section 2.2.

This physical design and the calibration step were critical to the success of this methodology. Since the microphones were calibrated, a microphone nearer to the sound source could be trusted to produce a larger output and that was used as the underlying basis for our calculations. The details of the methodology are discussed in the upcoming chapters.

The data was collected using a four channel oscilloscope and analyzed on a computer later.



(a) Overall structure (b) Close up of microphones

Figure 2.1: Picture shows the platform used for the microphone array

2.2 Calibration of The Microphones

A batch of 10 electret microphones were bought and one of the microphones was randomly selected as the reference microphone. The gains on the remaining 9 microphones were adjusted to ensure that the output of the microphones when exposed to a 1kHz sine wave from the same distance was identical.

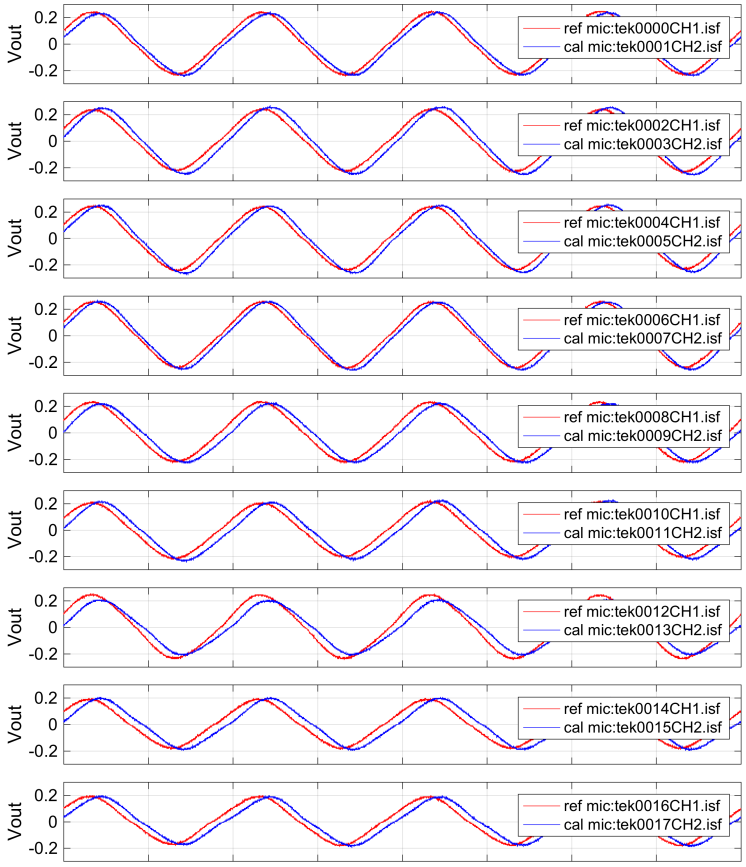


Figure 2.2: Figure shows the calibration of the 9 microphones against a reference microphone

2.3 Defining the Channels and Angles

As shown in Figure 2.3 the microphones were labeled as Channels 1 to 4 in a counter clockwise manner with Channel 1 placed along the positive x-axis and the angles defined to be increasing in a counter clockwise manner with 0 degrees defined along the positive x-axis.

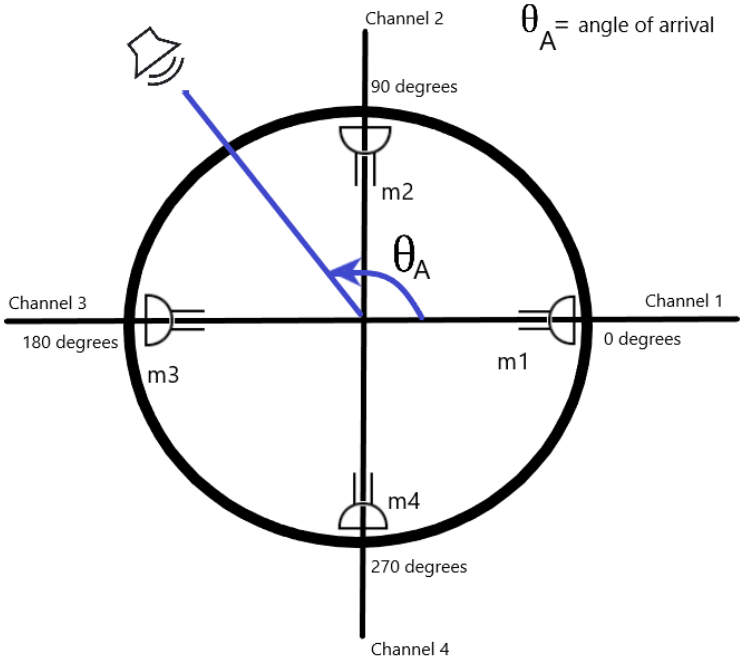


Figure 2.3: Picture shows how the convention used for the naming of the microphones and the definition of the angles

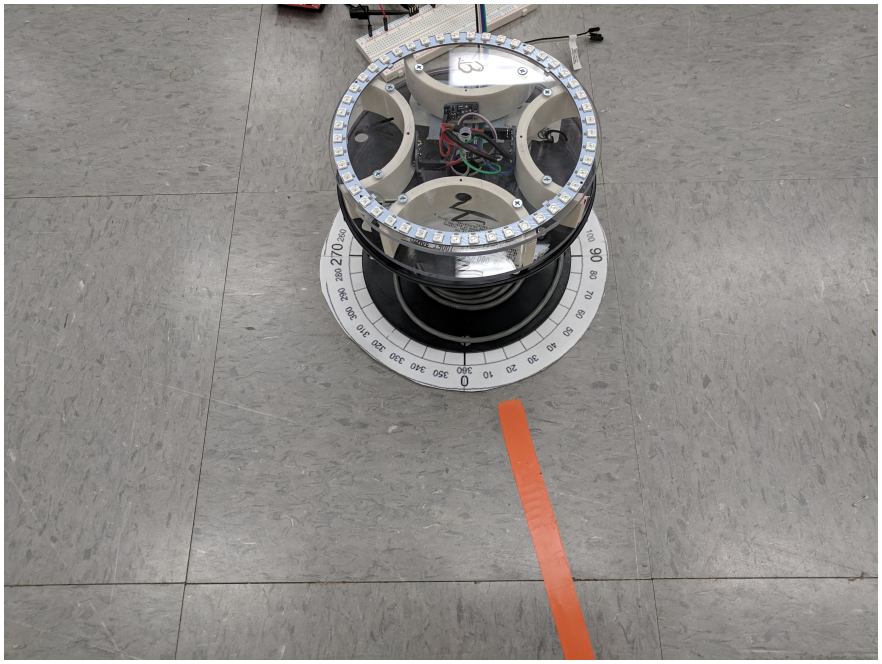
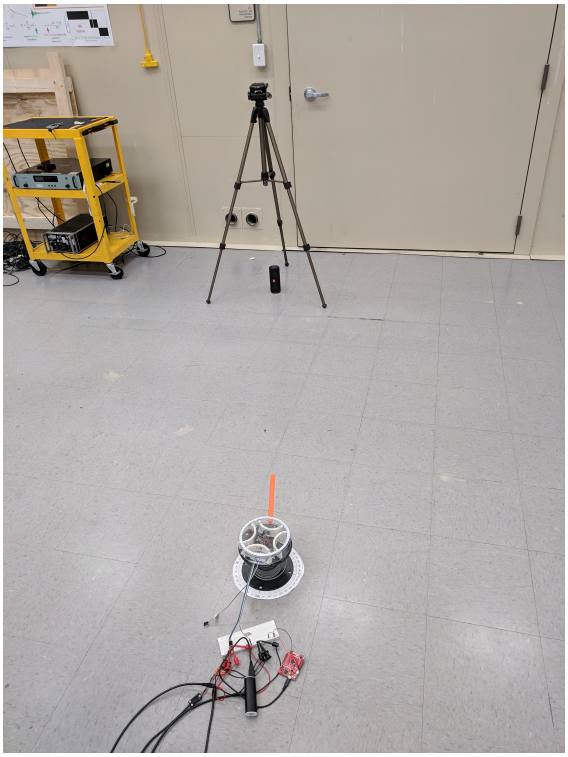
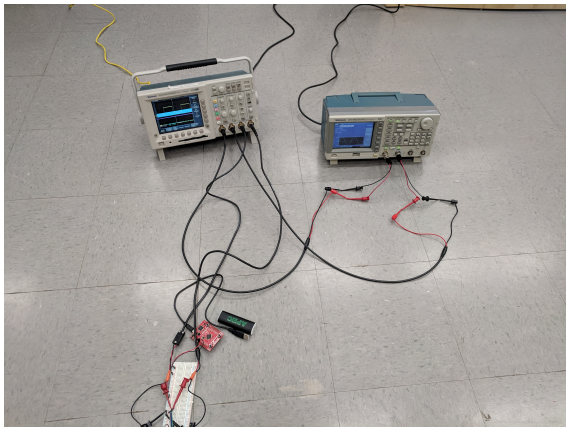


Figure 2.4: Figure shows how the actual angle to the source was calculated.

2.4 Experimental Test Setup



(a) Audio Source



(b) Collecting the data

Figure 2.5: Pictures show the setup used for running the experiments. They also emphasize why it was difficult to measure the angle of arrival to a high accuracy. While more sophisticated methods for measuring could have been used, an accuracy of $\pm$ 10 degrees was considered sufficient for the intended use case.

2.5 Variables of Focus

The following variables were varied to evaluate the performance of the detection and localization methods developed.

The following scenarios were considered for testing:

- **Source Signals:** Computer Generated Chirp, Human Whistle

- **SNRs of Chirp with respect to Noise:** [$+\infty$, 5, 0, -2, -5] dB
- **Height between source and microphone array:** [0, 1.5]m
- **Angle of Arrival θ_A:** [0 15 22.5 30 45 67.5 90 112.5 135 157.5 180 270] degrees
- **Room Conditions:** Room with and without multiple reflections of the sound source before it reaches the sensing platform
- **Detection Template Length:** [16 32 64 128 256 512] samples

2.6 Source Signals

Two source signals were evaluated for the detection and localization methodologies.

The first was a Human Whistle imitating a chirp and is shown in the following figure.

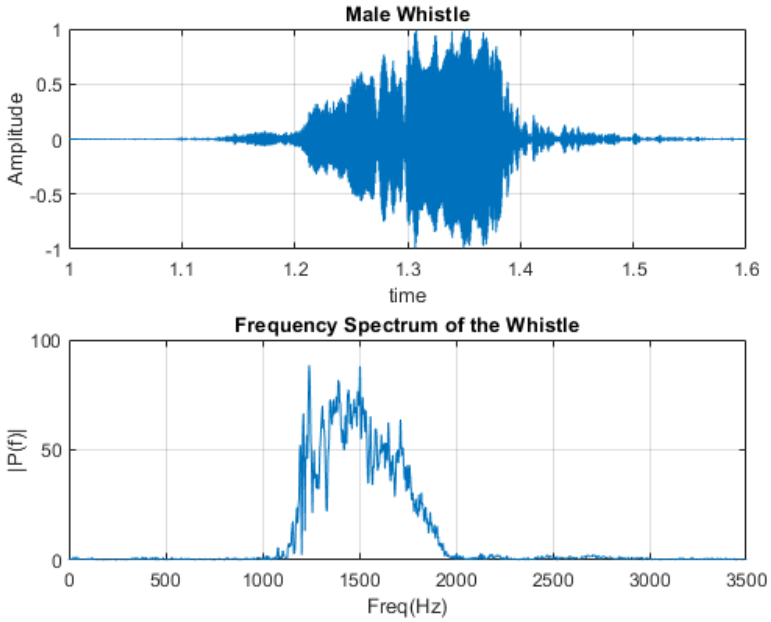


Figure 2.6: Figure shows the time and frequency content of the human whistle being detected.

The second was a Computer generated Ideal Chirp with similar frequency characteristics and duration as the expected Human Whistle. This ideal chirp is shown in the following figure.

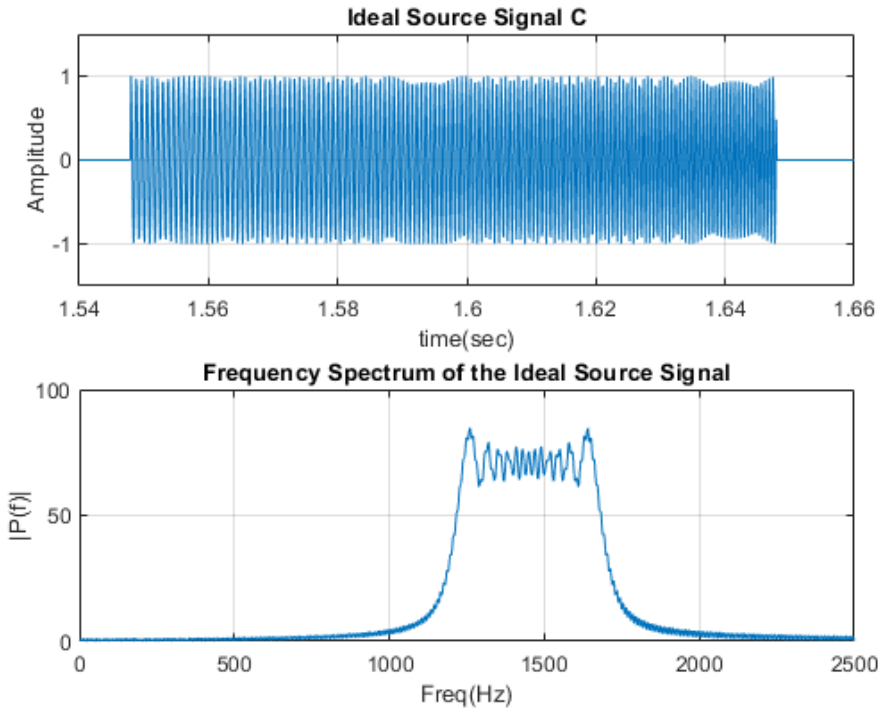


Figure 2.7: Figure shows the time and frequency content of the Ideal Source being detected.

2.7 Signal Processing Flow

This section shows a high level overview of the Signal Processing Flow with each step being discussed in further details in the upcoming chapters.

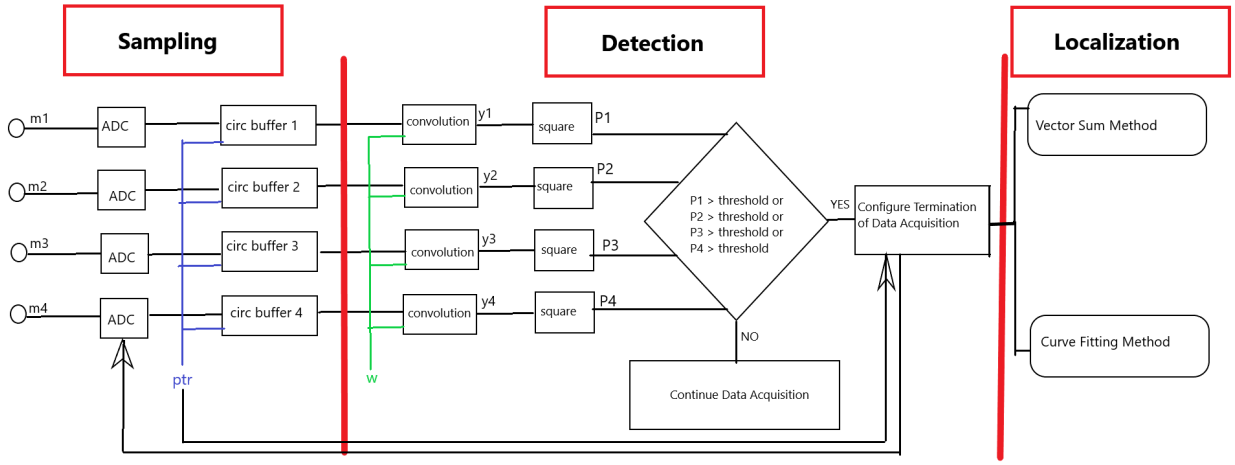


Figure 2.8: Figure shows the Signal Processing Flow on a high level

2.8 Intended Future Use of the Sensing Platform

As mentioned in 1.1, the motivation for this work is to provide an unmanned ground vehicle an estimate of the direction of the sound signal so that it may navigate towards it.

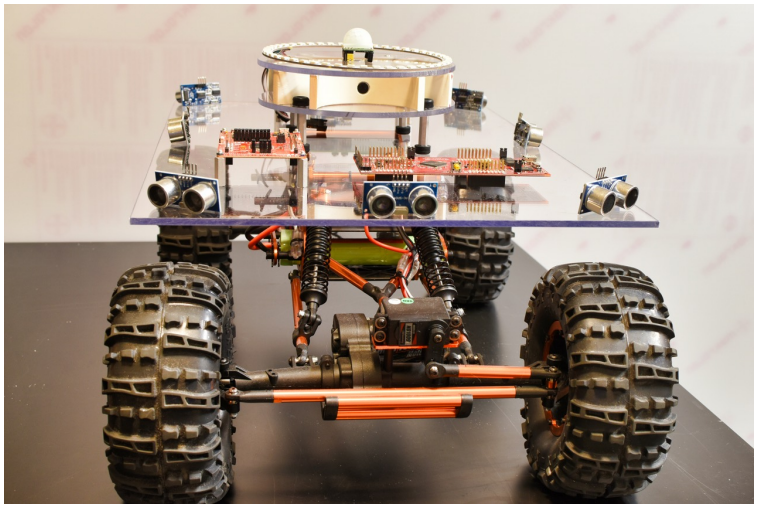


Figure 2.9: Picture shows the UGV on which the sensing platform is intended to be used

Chapter 3

Sampling and Detection

3.1 Introduction

This section looks at the method used for the sampling and detection of the source signal and examines its effectiveness under a variety of different possible scenarios. Since the aim was to make this system cost effective and target low cost microcontrollers which generally run around 80-200 Mhz and have a Flash memory of around 256-1024 kB [7] [8], it was important to sample and analyze the audio in an efficient manner so that the start of the source signal could be detected in real time. Once the start had been detected, the rest of the source signal from all four channels could be saved in memory and then further processed for the localization of the sound source. The amount of samples to be saved depended on the sampling frequency and the expected duration of the source signal which will be studied in the upcoming section 3.3.

3.2 Motivation for using a Human Whistle

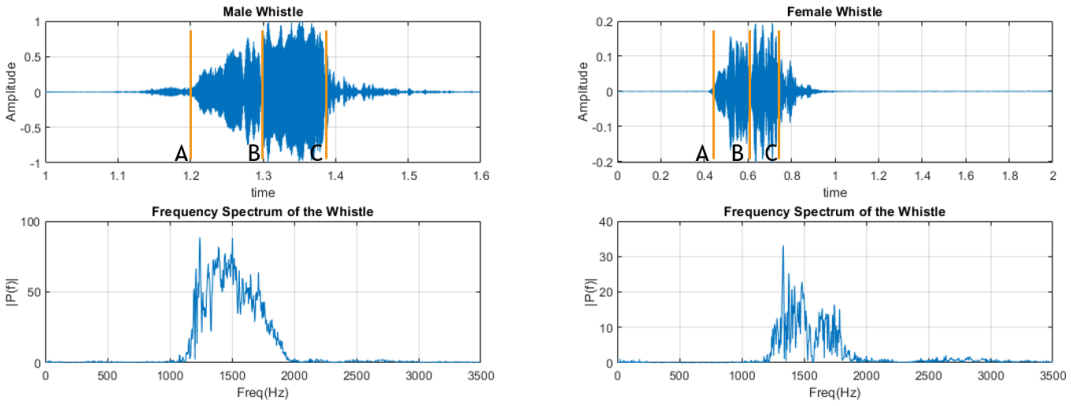
A linear chirp signal is defined as a signal whose frequency increases linearly with time. It is also commonly known as a sine sweep signal. The starting frequency, ending frequency and the time over which this frequency increases can be defined according to our needs.

It was observed that it is easy for humans to whistle in this 'chirp like' manner. While other signals such as a clap or snap are also easily reproducible, a chirp had the advantage of having a higher bandwidth and being spread over a larger duration. Hence a 'chirp like' whistle was used as the signal being detected and will be referred to as the source signal.

Before describing the detection methodology, the human whistle that is being detected must be analyzed and understood.

3.3 Analyzing Human Whistles

Both male and female chirp like whistles were analyzed to better understand the expected source signal.



(a) Male Whistle (b) Female Whistle

Figure 3.1: Figure shows male and female chirp-like whistles

It was observed that it was easy to whistle in this manner. The duration of such whistles (A to C) ranged from 0.2 - 0.3 seconds. The first 0.05 - 0.15 seconds of the whistle tends to be a sibilance (A to B). The duration of the part of the whistle that was captured was 0.1 seconds (B to C). Male whistles tended to sweep from 1.1kHz to 2kHz and female whistles tended to be slightly higher in frequency.

This analysis also helped set the sampling rate. The sampling rate had to be high enough to satisfy Nyquist's criteria while being low enough to be supported by most microcontrollers [7] [8]. It was set to 10kHz since this satisfied both requirements.

Setting the sampling rate to 10kHz meant that 1024 samples had to be captured to save approximately 0.1 seconds (B to C) of the whistle.

3.4 Generating the Ideal Chirp

After the analysis of human whistles in Section 3.3 it was observed that human whistles commonly have a frequency range between 1.2kHz - 1.7kHz and since the whistle was in a 'chirp-like' manner, the whistle could be approximated by a linearly increasing chirp signal. Based on this analysis, an Ideal Chirp C was generated using a computer with a starting frequency of 1.2kHz, ending frequency of 1.7kHz and duration of 0.1 seconds. This Ideal Chirp also served as one of the sources signals to be detected.

The time and frequency plot of such a chirp is shown below.

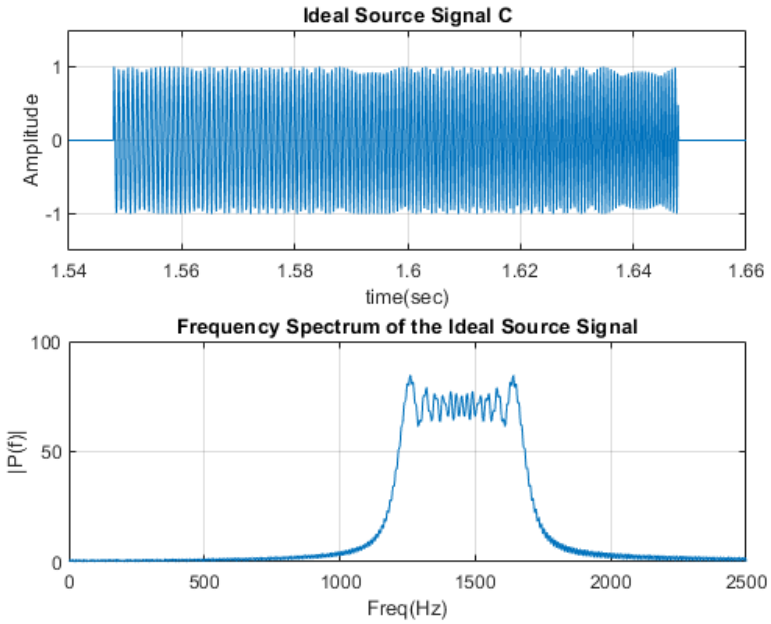


Figure 3.2: Figure shows the time and frequency content of the Ideal Source being detected.

Let,

$$f_{Start} = \text{starting frequency to be detected} = 1.2\text{kHz},$$

$$f_{End} = \text{ending frequency} = 1.7\text{kHz},$$

$$\text{duration of Ideal Chirp} = 0.1 \text{ sec, hence}$$

$$\text{length n} = 1024 \text{ samples at sampling frequency } f_S = 10\text{kHz}.$$

Thus the ideal chirp was generated as

$$\text{C} = \text{chirp}(f_{Start}, f_{End}, \text{n}), \text{ where}$$

$$\text{C} = [C_0\ C_1\ ...\ C_{n-1}]$$

$$= [C_0\ C_1\ ...\ C_{1023}]$$

3.5 Extract Template for Detection

From the Ideal Chirp C, a portion W is extracted and will be used to detect the presence of a source signal. This portion is henceforth referred to as the Signature Template W. The length L of the extracted portion and where this portion is extracted from dictates the frequency range present in it.

The ideal chirp C has length $n = 1024$ samples with $f_{Start} = 1200$ and $f_{End} = 1700$ Hz.

Let j be the index of the first sample from where the Signature Template W is extracted from the Ideal Chirp C. The offset j dictates the starting frequency in the signature template. If $j = 0$, the starting frequency would be 1200Hz.

Let the length of the Signature Template W be L.

Then, the starting frequency present in the signature template is:

$$1200 + j * \frac{500}{1024} Hz$$

The ending frequency is:

$$1200 + (j - L) * \frac{500}{1024} Hz$$

And the frequency spread of the Signature Template is given by:

$$1200 + L * \frac{500}{1024} Hz$$

The incoming signal is matched to this extracted portion W to detect the start of the source signal. The offset j was set to 256 and hence the starting frequency was 1325Hz. The offset was set after analysis of the human whistle 3.3 and recognizing the frequency range present in the start of the whistle.

$$1200 + j * \frac{500}{1024} Hz = 1200 + 256 * \frac{500}{1024} Hz = 1325 Hz$$

To accomplish the goal of being computationally efficient, it was desired to have as small a template length as possible while still being able to detect the source signal.

The frequency range present in the Signature Template as a function of the template length L is:

Table 3.1: Effect of Template Length on Frequency Range

Length L	Starting Frequency	Ending Frequency	Frequency Spread
32	1325	1335.625	015.625
64	1325	1356.250	031.250
128	1325	1387.500	062.500
256	1325	1450.000	125.000
512	1325	1575.000	250.000

A larger template length meant that the incoming audio signal was compared to a signature template containing a wider range of frequency. However, a larger signature template came with a larger memory and computational cost which was critical on a microcontroller. Hence the signature template length needed to be optimized to represent the best balance between accuracy, memory and speed requirements.

The length L was set as 128 . The method to obtain the template length and its effect is discussed in Section 3.8.

Thus the Signature Template can be given by,

$$\mathrm{W} = [C_j \; C_{j+1} \; ... \; C_{j+L}]$$

$$= [C_{256} \; C_{257} \; ... \; C_{383}]$$

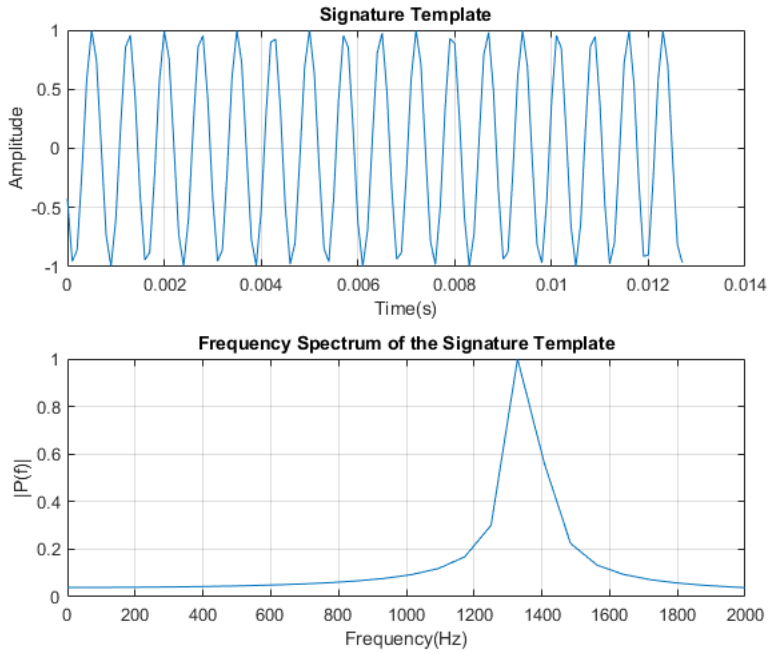


Figure 3.3: Picture shows the time and frequency content of the L = 128 Signature Template.

3.6 Sampling the Incoming Audio

As detailed earlier, 4 channels were being used to record the audio and the data from all 4 channels were used for the localization of the sound source as shown in Chapter 4.

Since a microcontroller was being targeted, it was crucial to optimize the amount of memory required to save the data coming from all 4 channels. It has been shown that the source signal was expected to last 0.1 seconds and hence each channel needed 1024 samples when using a sampling frequency of 10kHz to capture the source signal. The data streaming in from the 4 channels were saved in 4 circular buffers each of length N = 1024.

A circular buffer is an array of length N with a pointer *ptr* that contains the index of the location to which the data will be saved next. This assumes that data is first written to the

location ptr is pointing to and then ptr is incremented.

Thus, the oldest sample in a circular buffer is at ptr and is the sample that will be overwritten when a new sample comes in. The latest sample that was saved is at $ptr - 1$ and sample taken before that is saved at $ptr - 2$.

To save the audio, each channel is sampled at time step k:

$$m_{1k} = \text{sample from Channel 1 at time step } k$$

$$m_{2k} = \text{sample from Channel 2 at time step } k$$

$$m_{3k} = \text{sample from Channel 3 at time step } k$$

$$m_{4k} = \text{sample from Channel 4 at time step } k$$

These samples are then stored in the circular buffers of Length N = 1024.

$$b_1(ptr) = m_{1k}$$

$$b_2(ptr) = m_{2k}$$

$$b_3(ptr) = m_{3k}$$

$$b_4(ptr) = m_{4k}$$

and the pointer is incremented as follow:

$$ptr = (ptr + 1)\%N$$

Circular buffers ensure that there is always 1024 samples available in memory at any time after the initial filling of the buffer.

3.7 Detection Methodology

3.7.1 Overview

The detection methodology looked at correlating the audio streaming in with the predefined Signature Template by convolving them to match the temporal and spectral content of the signals. The instantaneous power of this match was considered as a measure of the match and when a predefined threshold had been exceeded, the start of the source signal was considered to be detected.

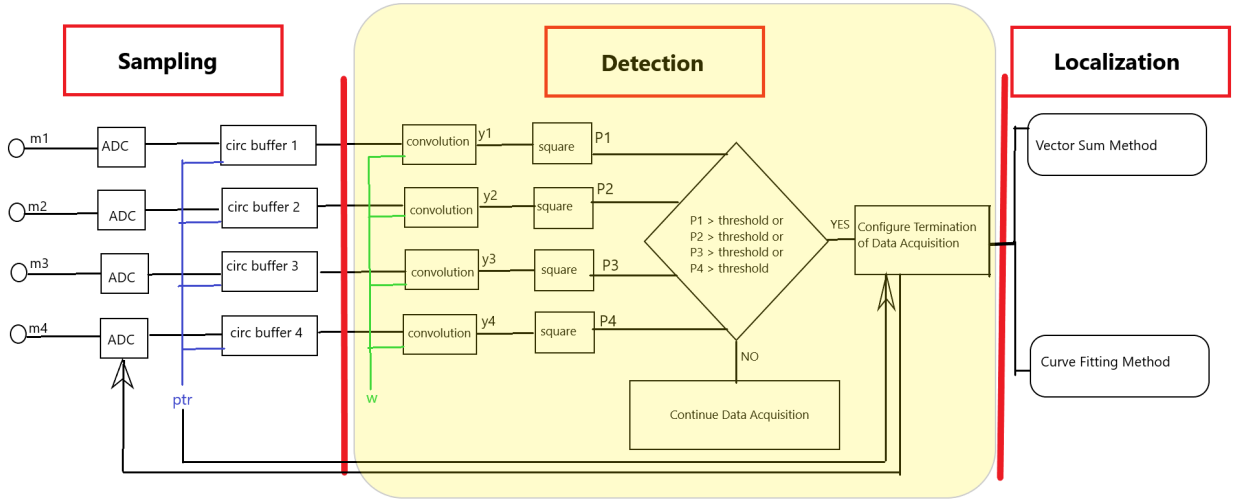


Figure 3.4: Figure shows an overview of the detection methodology

3.7.2 Correlation and Power Step

This subsection details the method to correlate the audio streaming in from the four channels with the predefined template to look for a match for the detection of a whistle.

The audio streaming in from Channel m was saved in a circular buffer b_m and then convolved with the Signature Template, which is essentially equivalent to filtering the audio coming

from each microphone through the Signature Template, and the output of this convolution was then squared to get an estimate of the convolution power, and compared to a threshold to detect a match.

The output of the convolution is given by,

$$y_{1k} = \sum_{j=0}^{L-1} \mathrm{W(j)} \cdot \mathrm{b_1((ptr\text{-}L+j)\ \%\ N)}$$

$$y_{2k} = \sum_{j=0}^{L-1} \mathrm{W(j)} \cdot \mathrm{b_2((ptr\text{-}L+j)\ \%\ N)}$$

$$y_{3k} = \sum_{j=0}^{L-1} \mathrm{W(j)} \cdot \mathrm{b_3((ptr\text{-}L+j)\ \%\ N)}$$

$$y_{4k} = \sum_{j=0}^{L-1} \mathrm{W(j)} \cdot \mathrm{b_4((ptr\text{-}L+j)\ \%\ N)}$$

For any channel m, when j = L - 1, W(127) is multiplied with b_m(ptr-1) which is equivalent to multiplying the 128^{th} sample of the Signature Template with the newest sample saved in b_m and pointed to by $ptr - 1$.

Similarly, when j = 0, W(0) is multiplied with b_m(ptr-128) which is equivalent to multiplying the 1^{st} sample of the Signature Template with the 128^{th} oldest sample saved in b_m and pointed to by $ptr - 128$.

The modulus of the pointer value *ptr* is taken with the length of the circular buffer N, so that the pointer wraps around to the start of the buffer when it reaches the end.

The convolution output from each channel is then squared to generate an estimated power corresponding to the cross-correlation of the audio from that channel and the signature template,

$$\mathrm{p_{1k} = (y_{1k})^2},\ \mathrm{p_{2k} = (y_{2k})^2}$$

$$\mathrm{p_{3k} = (y_{3k})^2},\ \mathrm{p_{4k} = (y_{4k})^2}$$

Lastly, the power thus obtained after every sample from each channel is compared to a threshold and when it exceeds a preset value, a whistle is assumed to be detected.

When,

$$p_{mk} > \text{threshold}$$

a whistle has been detected. The method to determine the threshold is discussed in the upcoming Section 3.8

3.7.3 Terminating Data Acquisition

Once the power output of any channel exceeds the threshold, the source signal has been detected. Since the sampling frequency was set to be 10kHz and the expected duration of the whistle from the analysis in Section 3.3 was 0.1 seconds, 1024 samples were required to capture the whistle from each channel. These values must be stored in the circular buffers defined earlier.

An offset value is determined and is set to be the number of samples of the whistle that is expected to already be saved in the circular buffers by the time the whistle is detected. The offset value was taken to be equal to the length of the Signature Template L as those many samples from the whistle were already in the buffer.

An index *start* is defined as follows,

$$\text{start} = (\text{ptr} - \text{offset}) \ \% \ \text{N}$$

and all 4 circular buffers are now filled up until the ptr reaches *start* again and the data acquisition is then stopped.

3.8 Determination of Template Length and Threshold

This section discusses the effect of varying the length of the Signature Template on the output of the convolution power and its effect on setting the threshold for detection. It was important to optimize the length because of the memory and speed constraints of the microcontroller. A larger template length required more memory to be saved and needed more operations to be performed to get the convolution power output which in turn would need a faster microcontroller. Hence the goal was to take as small a template length as possible while still being effective in detecting the whistle.

A larger template length will cause a larger overall output of the convolution power hence the threshold had to be decided as a function of the template length.

For determination of the threshold and template length, the output is observed when the Source Signal is present without any noise and then again with worse values of SNR to understand the effect of the length as SNR decreases.

3.8.1 Human Whistle as the Source Signal

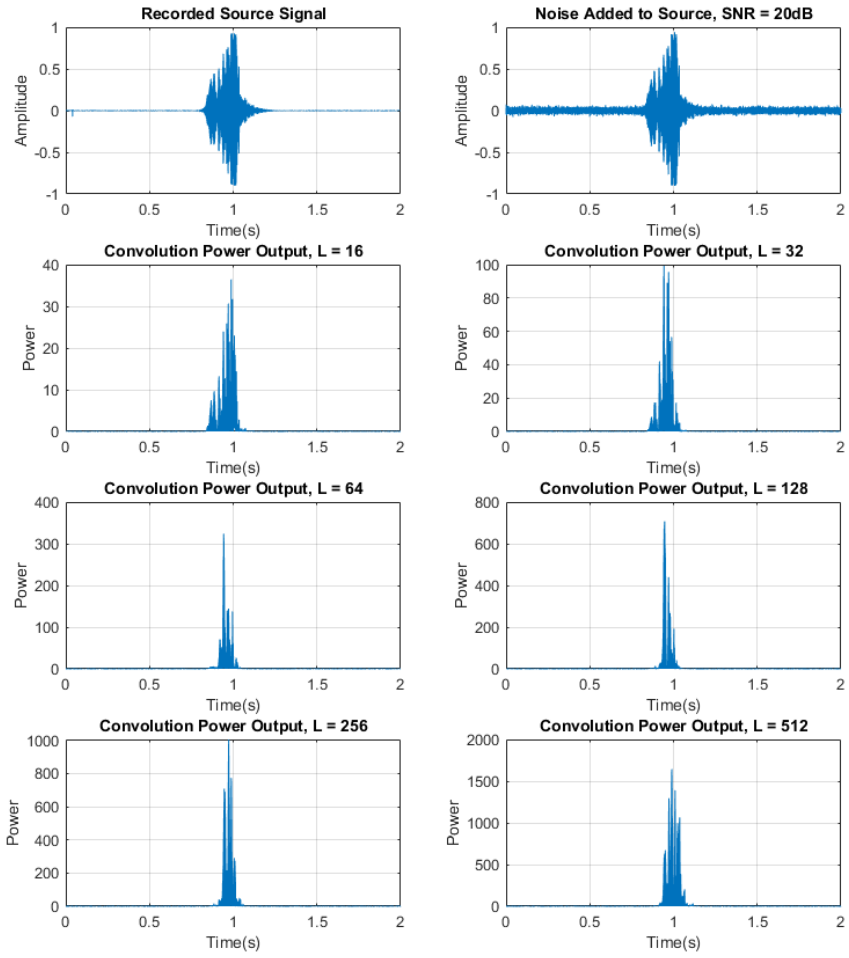


Figure 3.5: The convolution power output in the ideal case (SNR = 20dB)

This figure shows a human whistle corrupted with Noise with an SNR of 20dB, almost the ideal case, and the convolution power output with different template lengths. It can be observed that in this ideal case, almost any template length could be chosen and the detection algorithm would work.

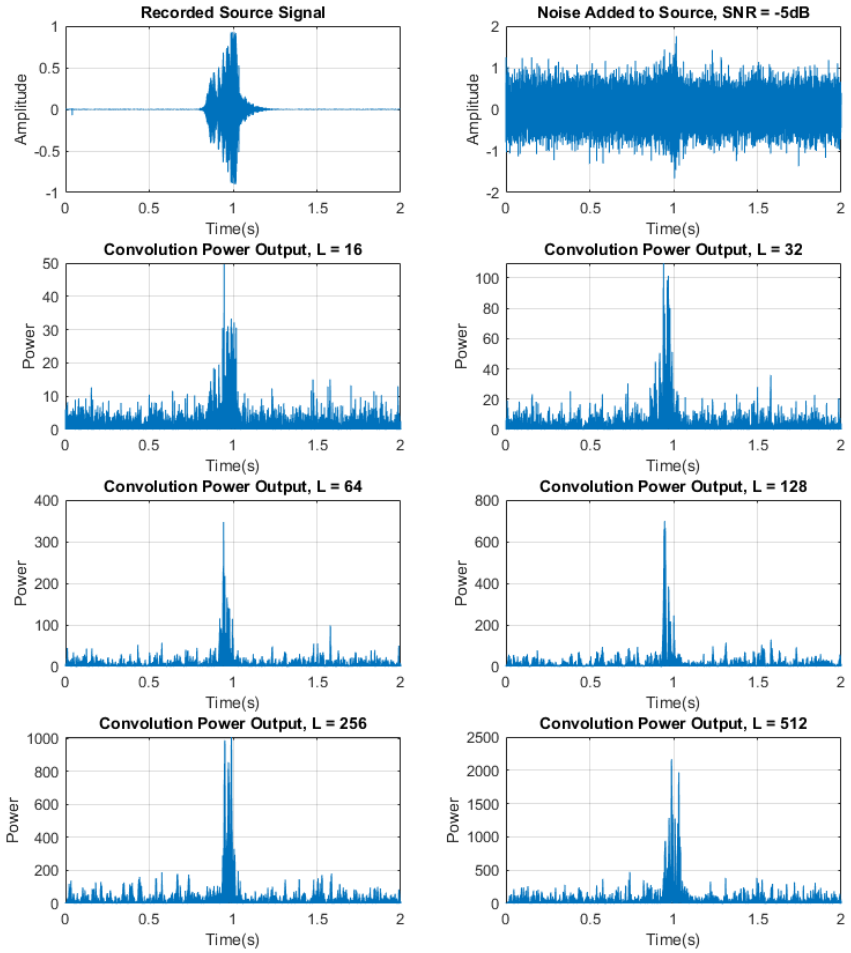


Figure 3.6: The convolution power output when SNR is -5dB

In this case, the human whistle was corrupted with noise with an SNR of -5dB. Once again, it can be observed that any template length would work with this SNR.

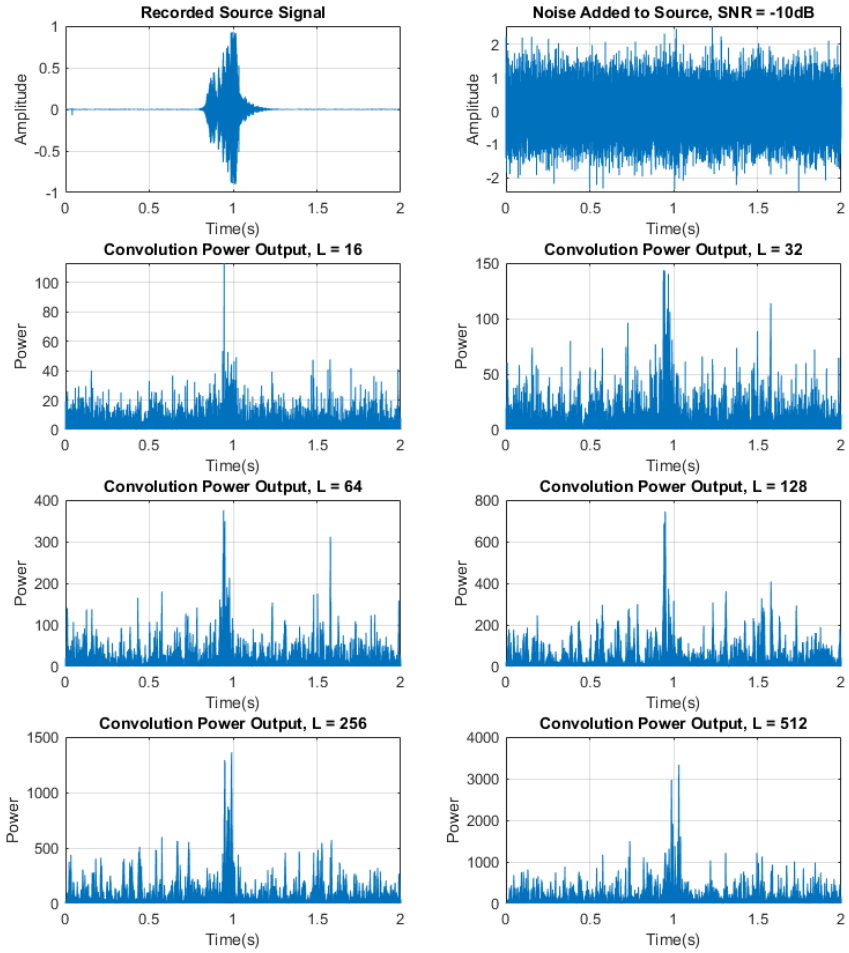


Figure 3.7: The convolution power output when SNR is -10dB

In this case, the human whistle was corrupted with noise with an SNR of -10dB. Now it can be observed that taking a template length of 16, 32 or 64 can result in faulty detections. This is deduced by the fact that the output from the noise (at around 1.6s) is very close in magnitude to the output due to the whistle.

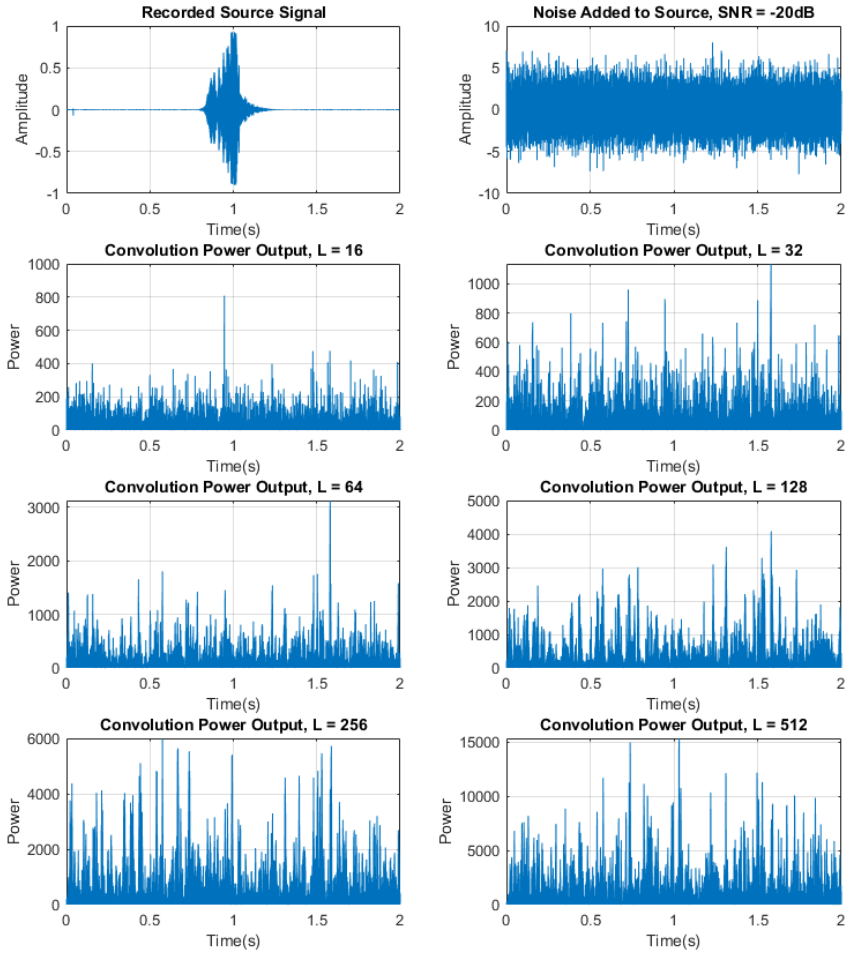


Figure 3.8: The convolution power output when SNR is -20dB

In this case, the human whistle was corrupted with noise with an SNR of -20dB. In this case, it can be observed that no template length would work and the detection algorithm fails.

Hence, the detection methodology can be rated to work up to an SNR of -10dB if a template length of 128 is chosen. If the SNR is not expected to be worse than -5dB then a template length of 64 can be chosen which would decrease the computational cost.

3.8.2 Ideal Chirp as the Source Signal

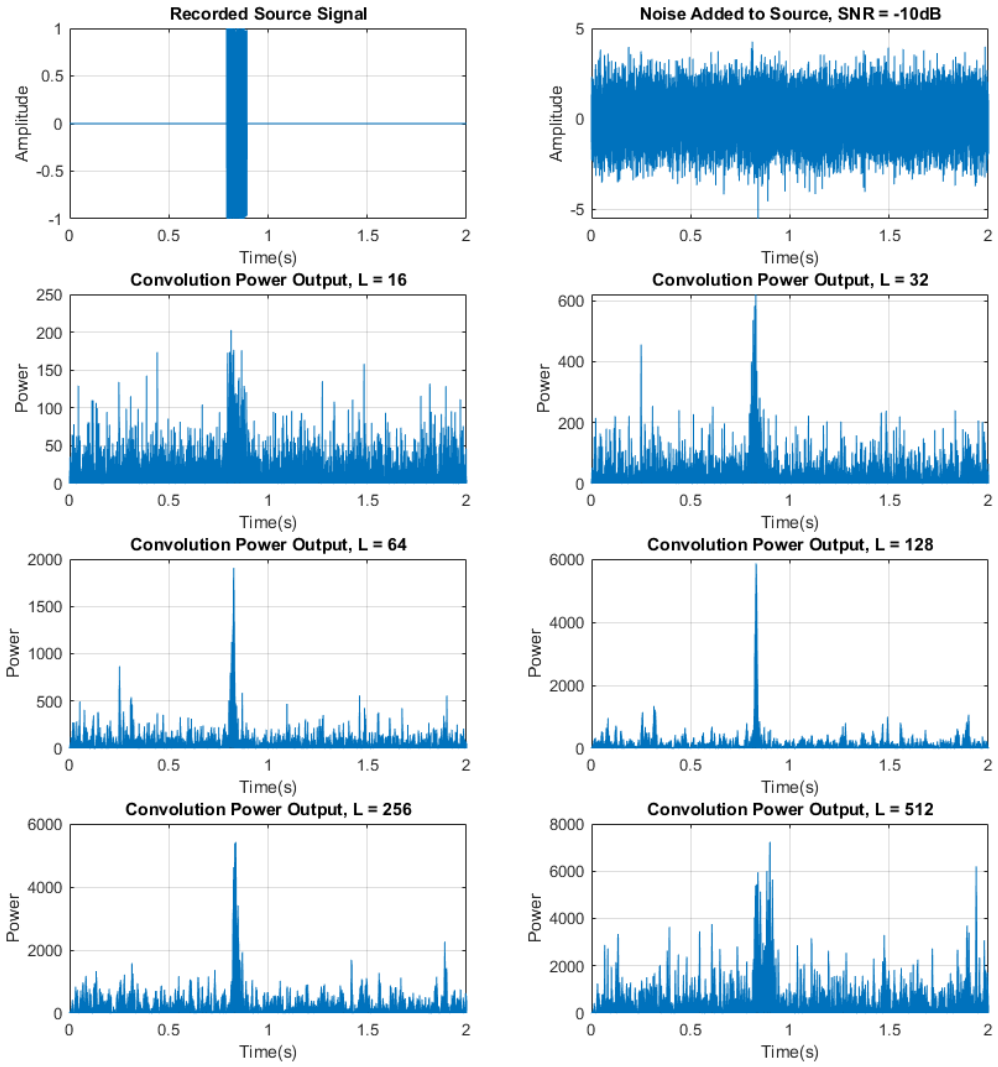


Figure 3.9: The convolution power output when using an Ideal Chirp with SNR = -10dB

This figure shows an Ideal Chirp corrupted with Noise with an SNR of -10dB and the convolution power output with different template lengths. It can be observed that in this case, template lengths of 128 or longer would work best.

After this analysis, when a human whistle was being used, the template length L was chosen as 128, the threshold was set as 500 and the detection methodology was rated to work up to an SNR of -5dB to be conservative.

When an Ideal Chirp was being used as the source signal, the template length was set as 128 and the threshold was set as 4000. This is in contrast to the case where the source signal was a human whistle and the threshold was 500 despite having the same template length.

3.9 Example of the Detection Methodology

Figure 3.10 shows the detection methodology successfully detecting the start of a human whistle. Since the duration of the whistle was determined to be around 0.1 seconds in Section 3.3, 1024 samples were saved for further analysis.

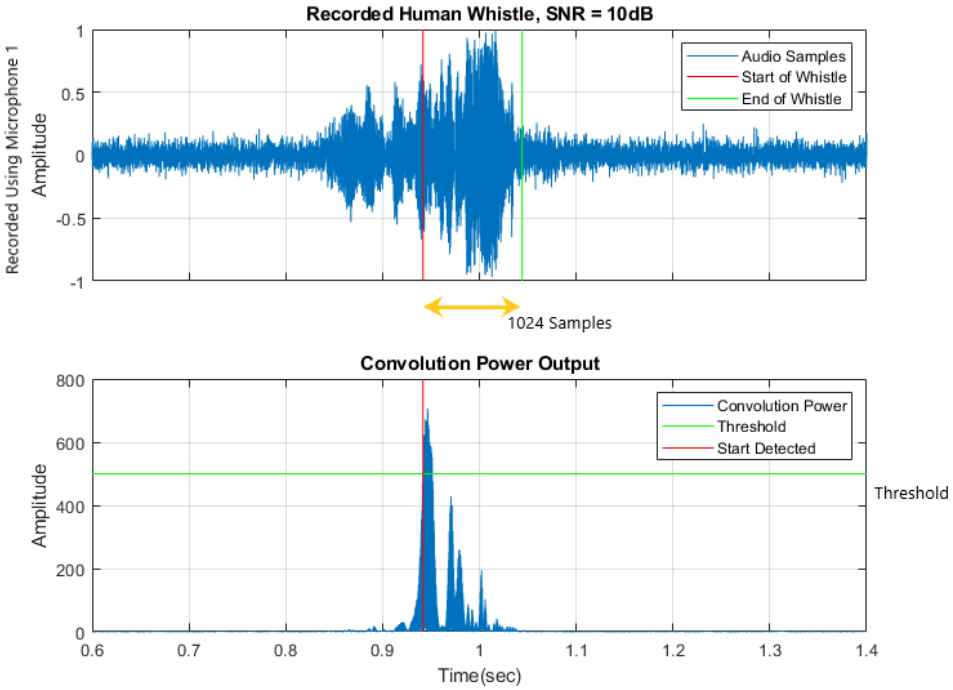


Figure 3.10: Figure shows the detected start and expected end of the whistle

Chapter 4

Localization

4.1 Introduction

This section discusses the methods used for the localization of the source after a command signal has been detected. As discussed in Section 3.7.3, once the whistle had been detected, 1024 samples from each channel were saved in 4 circular buffers called b_1, b_2, b_3 and b_4.

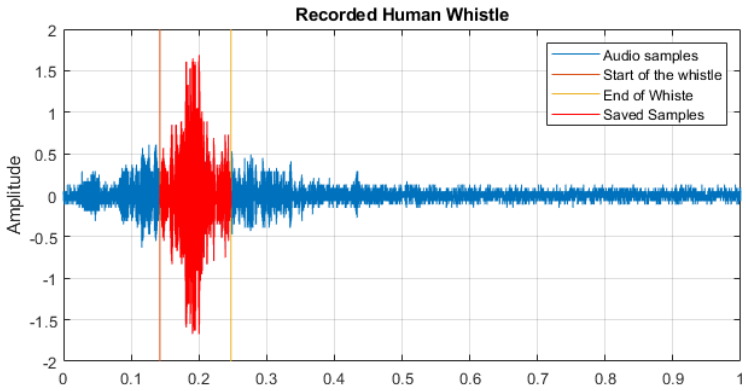


Figure 4.1: Figure shows the samples saved from one of the channels after detection of the whistle

4.2 Accumulated Power Calculations

Two different methodologies were considered for the calculation of the direction of the source. These are the Vector Sum Method and the Quadratic Curve Method, and they are explained in the upcoming sections.

For both these methods, the audio samples saved in the 4 circular buffers after the whistle was detected are first convolved with the Ideal Chirp C.

Convolving a signal with this ideal chirp C may be considered equivalent to passing the signal through a band pass filter with the lower cutoff frequency of 1.2kHz and a higher cutoff frequency of 1.7kHz as can be observed in the upcoming sections.

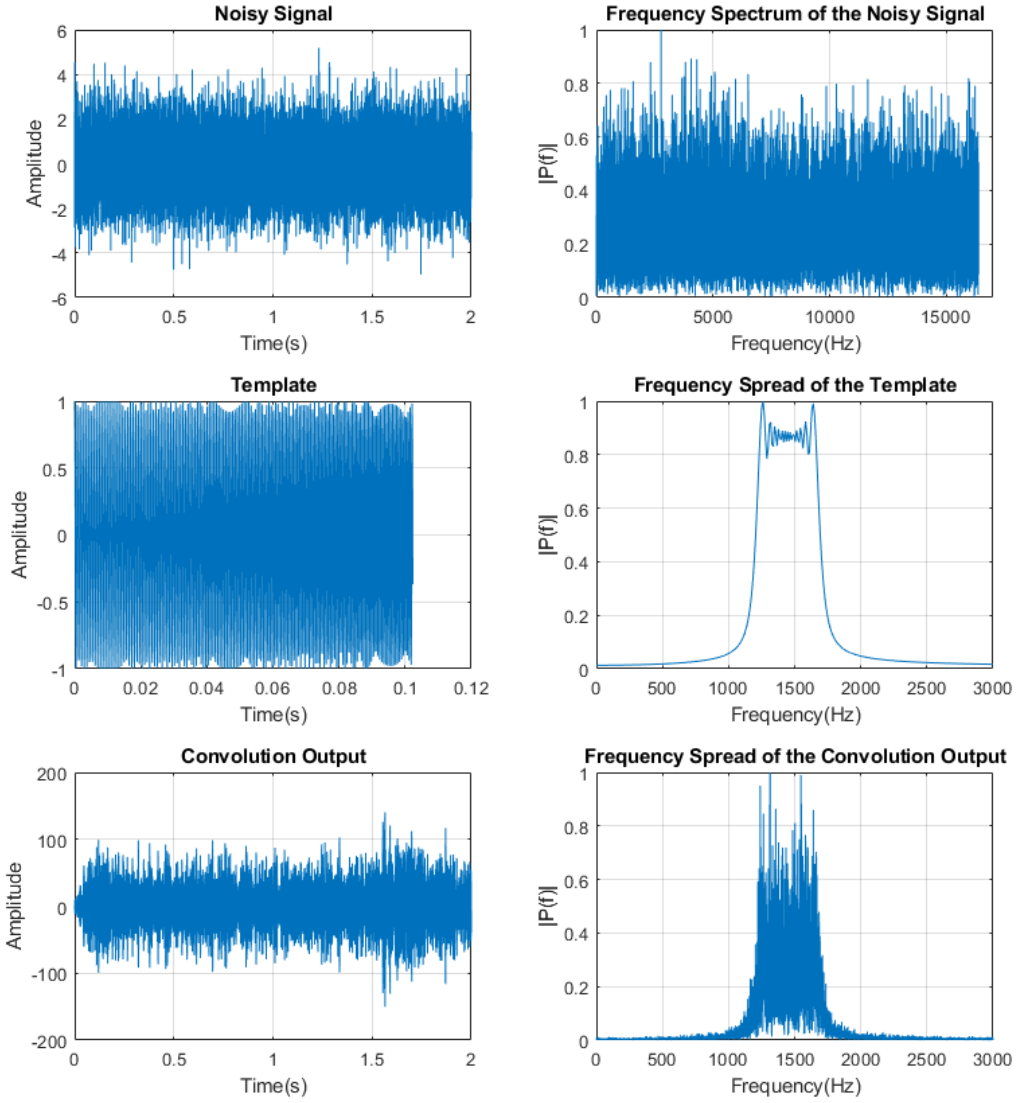


Figure 4.2: Figure shows the band pass effect of convolving a signal with the Ideal Chirp C

The band pass filtering effect can be seen in the figure 4.2. A noisy signal containing power across a wide spectrum is convolved with the chirp and frequencies below 1.2kHz and above 1.7kHz have been attenuated.

Recall that after the whistle was detected, all 4 circular buffers were filled up and the *ptr* is

currently pointing to *start*.

Thus, the output of the convolution is given by,

$$\text{For } k = 1 : 1024$$

$$y_{1k} = \sum_{j=0}^{L-1} \text{C(j)} \cdot \text{b}_1\text{((ptr-L+j) \% N)}$$

$$y_{2k} = \sum_{j=0}^{L-1} \text{C(j)} \cdot \text{b}_2\text{((ptr-L+j) \% N)}$$

$$y_{3k} = \sum_{j=0}^{L-1} \text{C(j)} \cdot \text{b}_3\text{((ptr-L+j) \% N)}$$

$$y_{4k} = \sum_{j=0}^{L-1} \text{C(j)} \cdot \text{b}_4\text{((ptr-L+j) \% N)}$$

$$ptr = (ptr + 1)\%1024$$

And the accumulated power in each channel can be obtained as:

$$\text{The power in channel m, } P_m = \sum_{i=1}^{1024} (y_{mi})^2$$

These accumulated powers will be used in both the upcoming methods.

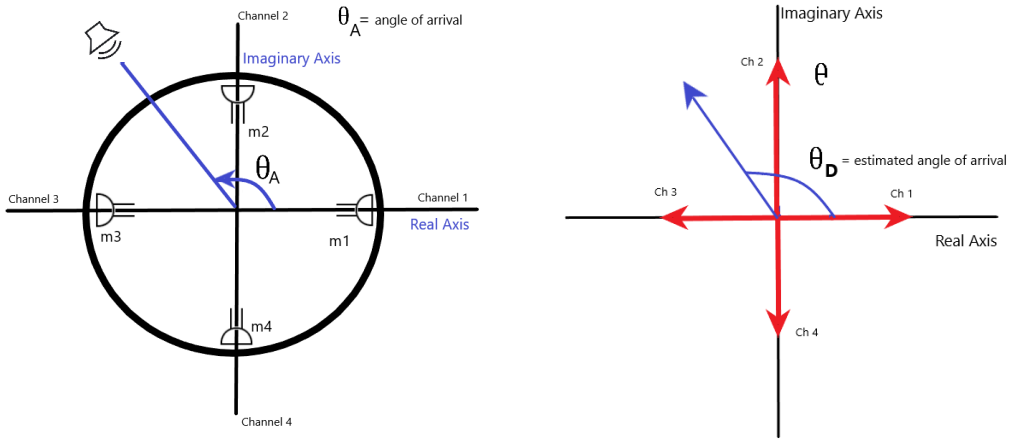
4.3 Vector Sum Method

4.3.1 Overview

This method relied on the fact that the microphones were calibrated, as discussed in Chapter 2, and hence it could be expected that the microphone nearer to the source would produce a larger output than one farther away.

In this method, the power contained in each of the 4 channels, after they were filtered through the Ideal Chirp C, was calculated by summing the squares of the output. The power values

were treated as a vector and added together to get the net power output. The angle of this resulting vector was calculated and taken to be the estimated direction of the source. The directions of each channel were as defined in Figure 2.3 and is shown again in Figure 4.3



(a) Setup

(b) Power Vectors

Figure 4.3: Figure shows the conventions used for the direction

4.3.2 Calculations

Define,

$$\sum P_x = P_1 - P_3, \text{ and}$$

$$\sum P_y = P_2 - P_4$$

then the estimated angle of the source =

$$\theta_A = \arctan(\frac{\sum P_y}{\sum P_x})$$

The following Figure 4.4 shows an example of the polar plot with the 4 channels, the resulting

vector sum and the calculated angle. The magnitude of the resulting vector sum has been normalized for better visualization.

As will be shown in Chapter 5, this method works well in a room without reflections. A safety mechanism is set that informs us not to trust this calculation when the length of the resulting vector sum is less than 1/10th of the smallest power vector. This is done because the angle calculation can become extremely sensitive to small changes when the net length is very small.

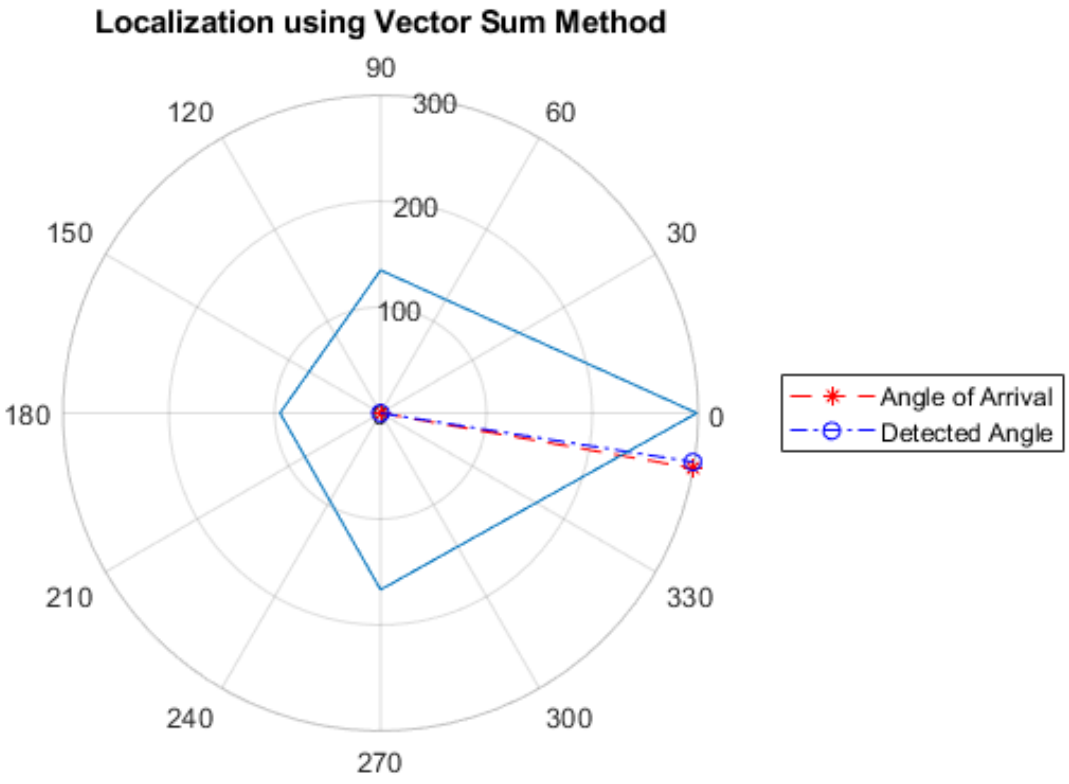


Figure 4.4: Figure shows the power output of the 4 channels, the actual angle of arrival and the detected angle

4.4 Curve fitting method

4.4.1 Overview

In this method, a quadratic curve is fit through three of the calculated power values. The channel with the highest power is selected as the primary channel, P_{main} and the two nearest neighbors to it are taken as the support channels P_{N1} and P_{N2}. The quadratic curve is fit through these values and the location of the peak of this curve is taken as the estimated angle of arrival. A quadratic curve is used since it has only one maxima or minima.

4.4.2 Calculations

There are 4 possible scenarios when deciding the channels:

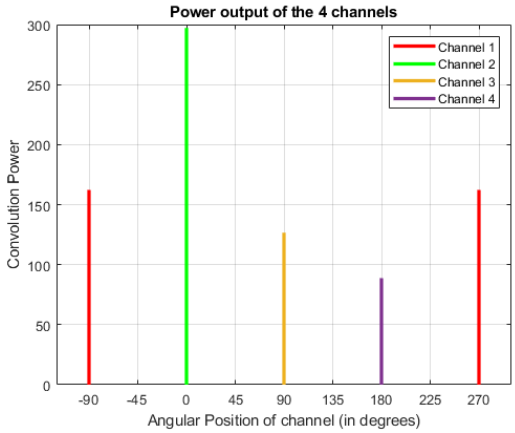
$$\text{If } P_1 \text{ is max: } P_{N1} = P_4,\ P_{main} = P_1,\ P_{N2} = P_2$$

$$\text{Else if } P_2 \text{ is max: } P_{N2} = P_1,\ P_{main} = P_2,\ P_{N2} = P_3$$

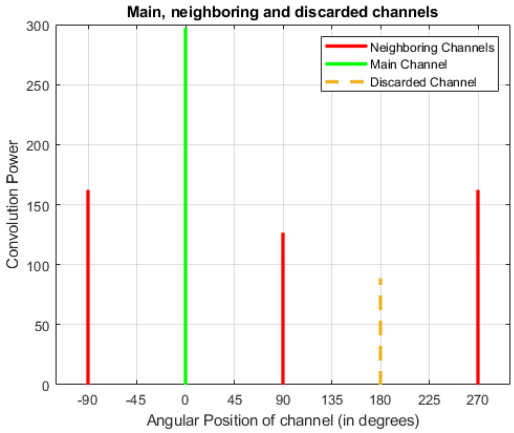
$$\text{Else if } P_3 \text{ is max: } P_{N3} = P_2,\ P_{main} = P_3,\ P_{N2} = P_4$$

$$\text{Else if } P_3 \text{ is max: } P_{N4} = P_3,\ P_{main} = P_4,\ P_{N2} = P_1$$

Hence, *main* is the index of the channel with the maximum power.



(a) Power Output of Each Channel



(b) Selecting the Channels

Figure 4.5: Figure shows the method used for selecting the main and neighboring channels based on power output

In Figure 4.5a it can be observed that Channel 1 has the maximum output. The 2 channels neighboring Channel 1 are Channel 2 and Channel 4. Figure 4.5b shows the main channels, the two neighboring channels and the discarded channels.

The angular position θ of these channels is scaled to a new axis called the ϕ axis. The scaling is given by the transformation:

$$\phi = \frac{(\theta - \theta_{main})}{90}$$

$$\implies \phi_{main} = 0$$

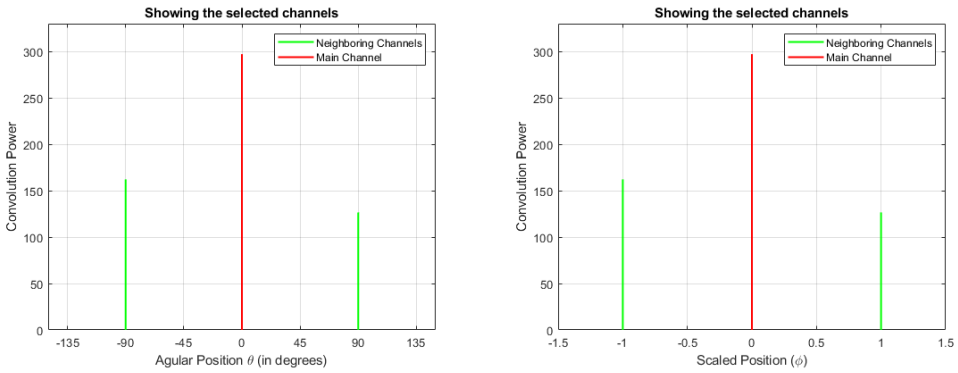
$$\implies \phi_{N1} = \text{-}1$$

$$\implies \phi_{N2} = 1$$

This gives the following points on the ϕ axis:

$$(\phi_m, P_m) = \{(\phi_{N1}, P_1), (\phi_{main}, P_2), (\phi_{N2}, P_3)\}$$

This can be observed in Figure 4.6 where the actual angles of the channels on the θ axis, as shown in Figure 4.6a, have been transformed to the scaled positions on the ϕ axis, as shown in Figure 4.6b.



(a) Power Output of Each Channel (b) Selecting the Channels

Figure 4.6: Figure shows the method used for selecting the main and neighboring channels based on power output

Thus, a quadratic curve is being fit through 3 points:

$$(\text{-}1, P_{N1}) \ , \ (0, \ P_{main}) \ , \ (1, \ P_{N2})$$

This leads to the following equations:

$$P_{N1} = a \cdot \phi_1^2 + b \cdot \phi_1 + c$$

$$P_{main} = a \cdot \phi_2^2 + b \cdot \phi_2 + c$$

$$P_{N2} = a \cdot \phi_3^2 + b \cdot \phi_3 + c$$

Which can be simplified as:

$$\begin{bmatrix} P_{N1} \\ P_{main} \\ P_{N2} \end{bmatrix} = \begin{bmatrix} \phi_1^2 & \phi_1 & 1 \\ \phi_2^2 & \phi_2 & 1 \\ \phi_3^2 & \phi_3 & 1 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

since ϕ_1 = -1, $\phi_2 = 0$, $\phi_3 = 1$, this equals:

$$\begin{bmatrix} P_{N1} \\ P_{main} \\ P_{N2} \end{bmatrix} = \begin{bmatrix} 1 & -1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

Hence, the coefficients can be calculated as:

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 0.5 & -1 & 0.5 \\ 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} P_{N1} \\ P_{main} \\ P_{N2} \end{bmatrix}$$

This gives a simple recipe for the calculation of the coefficients. The matrix of constants can be readily saved on a microcontroller and multiplied by the matrix containing the power of the main channel and the neighboring channels to obtain the coefficients of the quadratic curve.

The quadratic curve is given by: $a \cdot \phi^2 + b \cdot \phi + c$ and once the coefficients have been calculated, the location of the peak of this curve can be found by taking its derivative and setting it to 0.

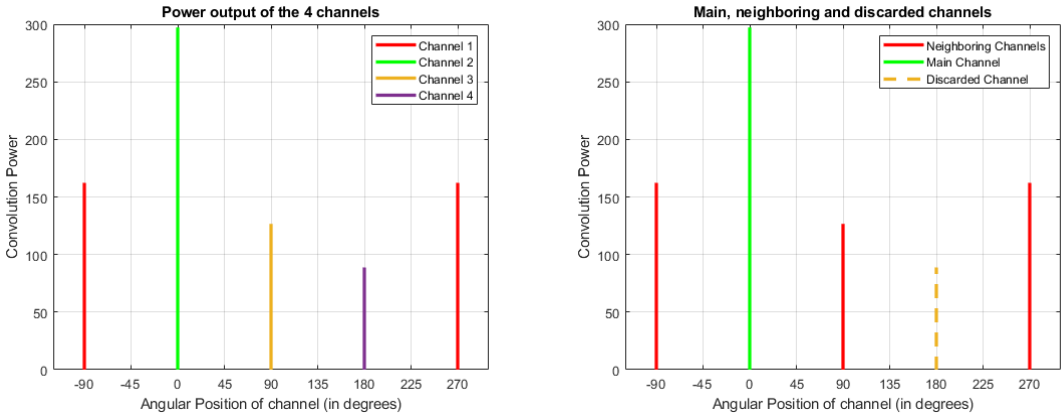
$$\frac{d}{d\phi}(a \cdot \phi^2 + b \cdot \phi + c) = (2 \cdot a \cdot \phi + b) = 0$$

$$\implies \phi_{peak} = \frac{-b}{2 \cdot a}$$

This value is then used to estimate the angle of arrival in degrees by the linear transformation:

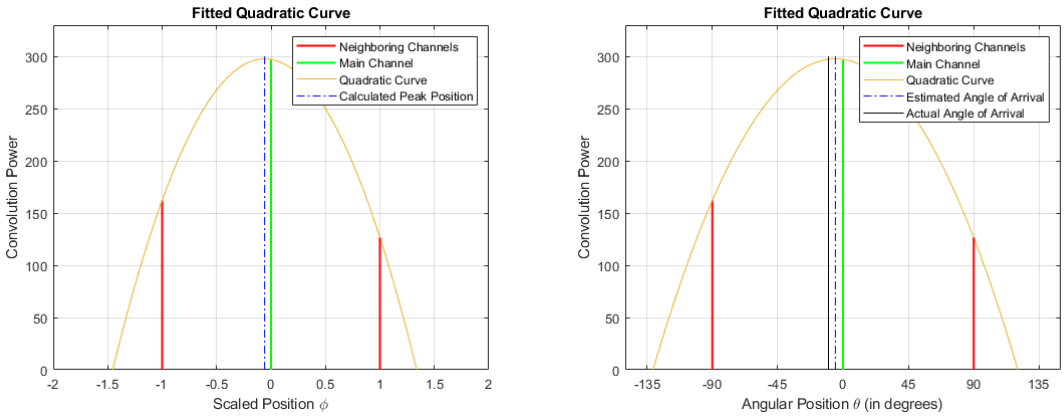
$$\theta_{arrival} = (\phi_{peak} + (main - 1)) \cdot 90$$

where, main = index of the channel with maximum power as stated earlier.



(a) Power Output of Each Channel (b) Selecting the channels

Figure 4.7: Figure shows the method used for selecting the main and neighboring channels based on power output



(a) Fitting Quadratic Curve to the scaled positions (b) Transforming back to get the angle in degrees

Figure 4.8: Figure shows the method used for estimating the angle of arrival using a fitted quadratic curve

Chapter 5

Results

5.1 Introduction

This section discusses the results obtained via simulation and real world experiments for the detection and localization of the Source Signal under a variety of different scenarios.

5.2 Simulation Results for Detection

The detection methodology discussed in Chapter 3 was simulated in Matlab to test its effectiveness under different SNRs. Both the computer generated ideal chirp and the human whistle were simulated under different SNRs to ensure that the methodology worked as expected.

- **SNRs:** [-5, -20] dB
- **Source Signals:** [Computer Generated Ideal Chirp, Human Whistle]

5.2.1 Detection of a Computer Generated Ideal Chirp

This section shows the results obtained during the simulation of detection of the start of a noisy computer generated ideal chirp padded with a random number of zeros at the start and end. It can be observed that the start and end of the whistle is accurately detected when the SNR = -5dB.

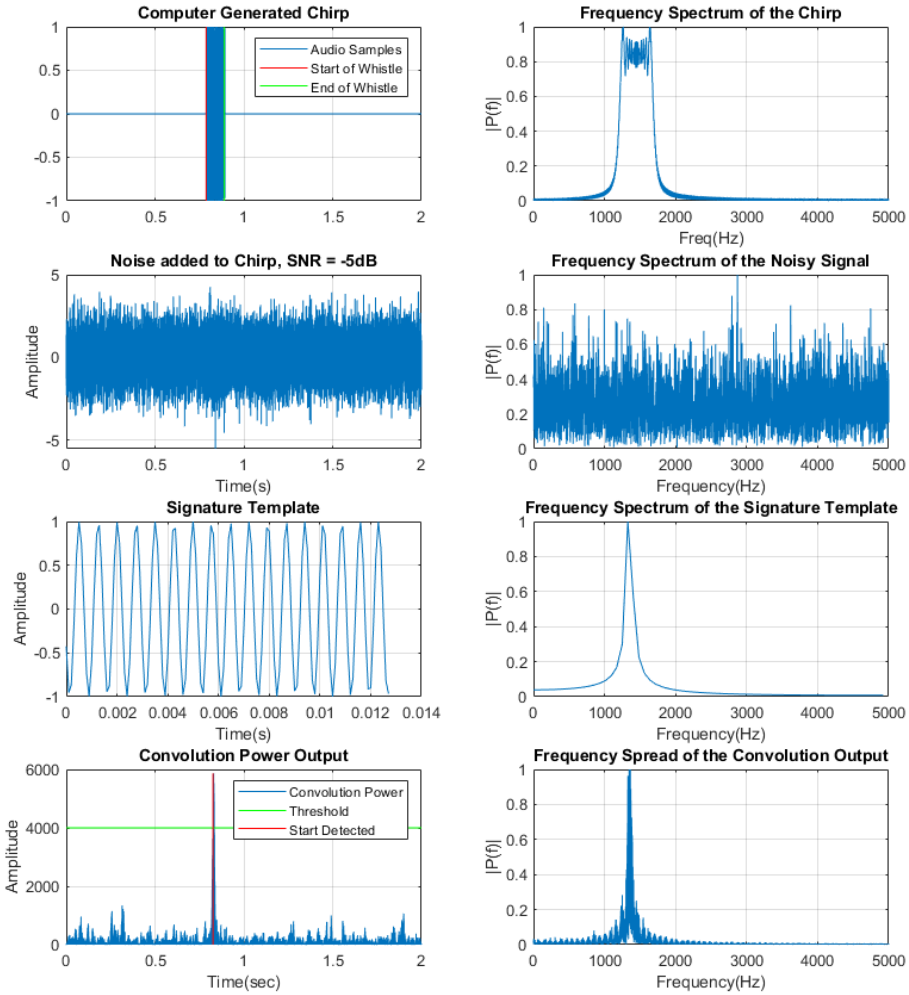


Figure 5.1: Figure shows a simulation study for the detection of Ideal Chirp with an SNR of -5dB

Figure 5.2 shows the case when the SNR = -20dB. It can be observed that the noise completely overwhelms the system and leads to an incorrect detection of the start.

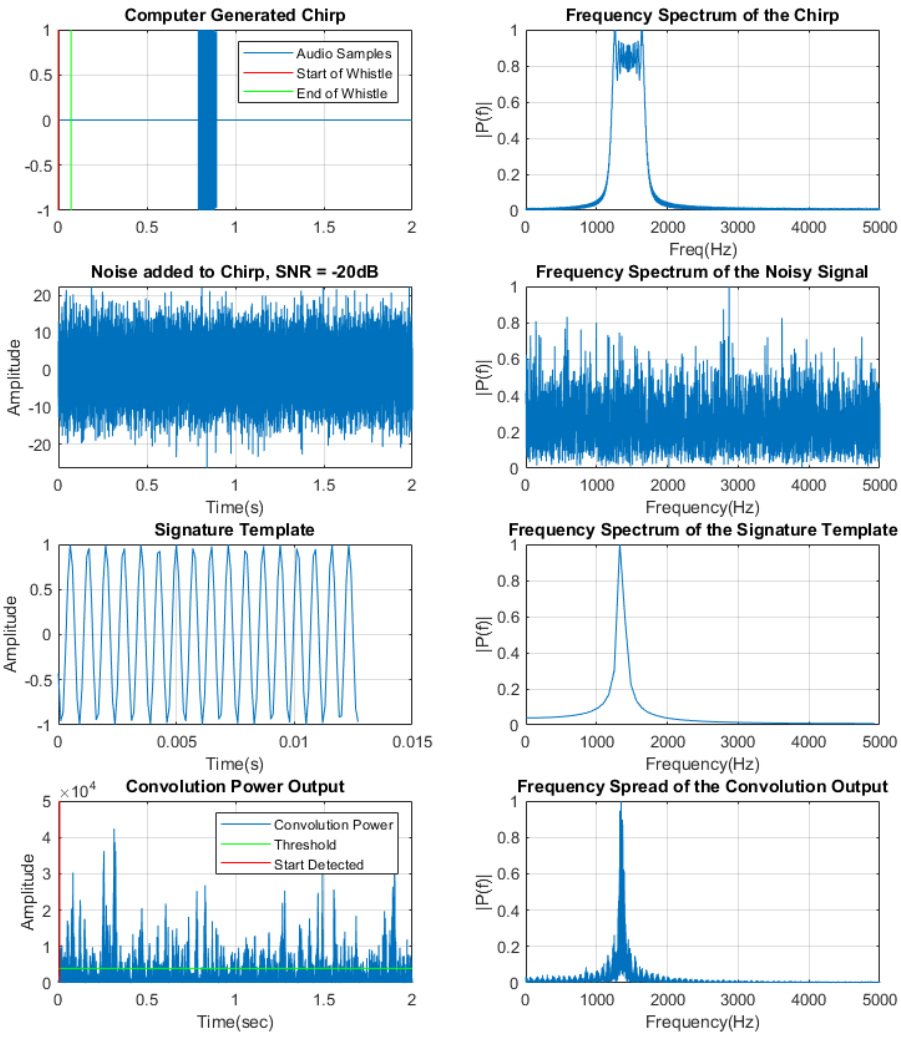


Figure 5.2: Figure shows a simulation study for the detection of Ideal Chirp with an SNR of -20dB

These simulation studies helped impose a constraint on the detection methodology. The methodology can only be trusted to detect Ideal Chirps when the SNR is better than -5dB.

5.2.2 Detection of a Human Whistle

This section shows the results obtained during the simulation of detection of a human whistle. It can be observed that the start has been detected accurately when SNR = -5dB.

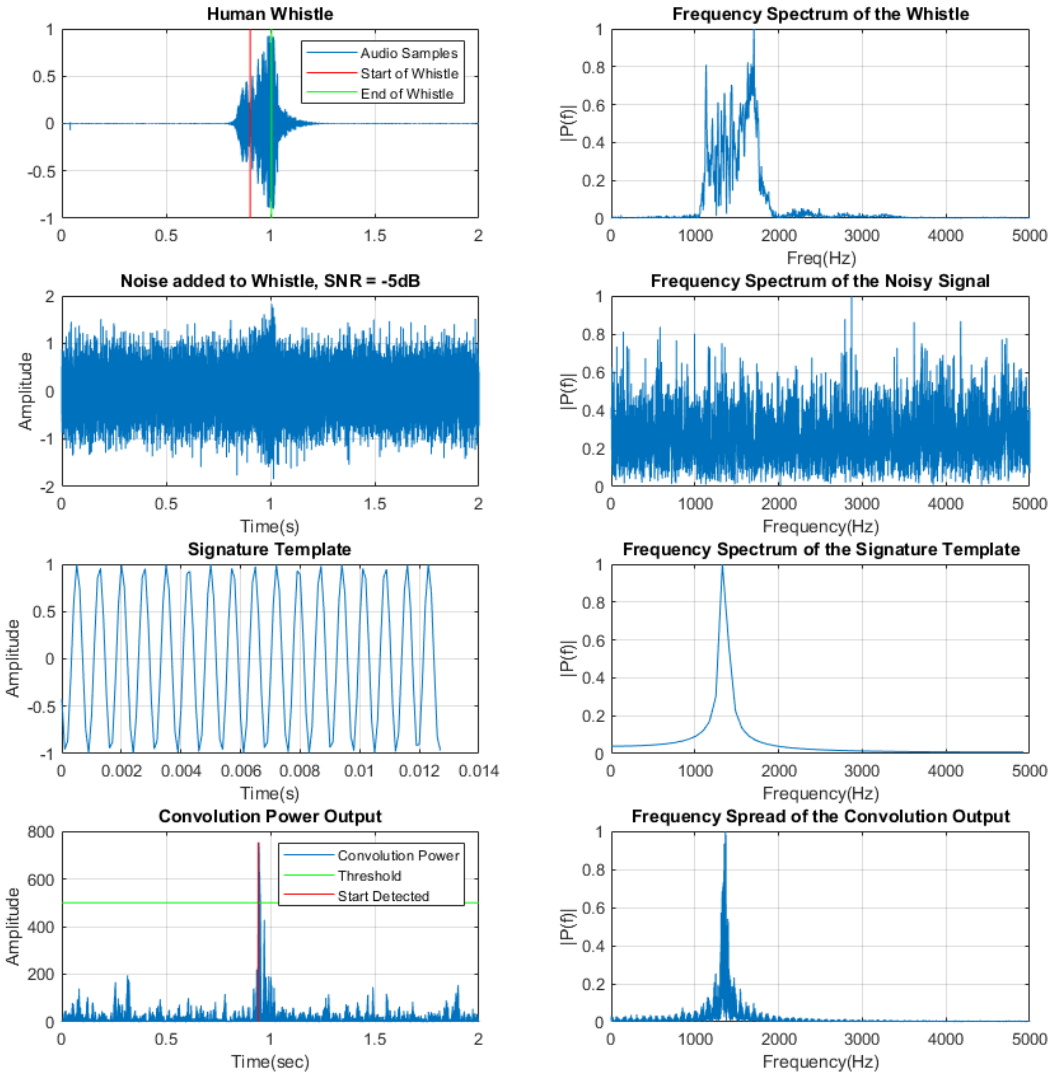


Figure 5.3: Figure shows a simulation study for the detection of a Human Whistle with an SNR of -5dB

Figure 5.4 shows the case when the SNR = -20dB and the noise completely overwhelms the system and causes an incorrect detection of the start.

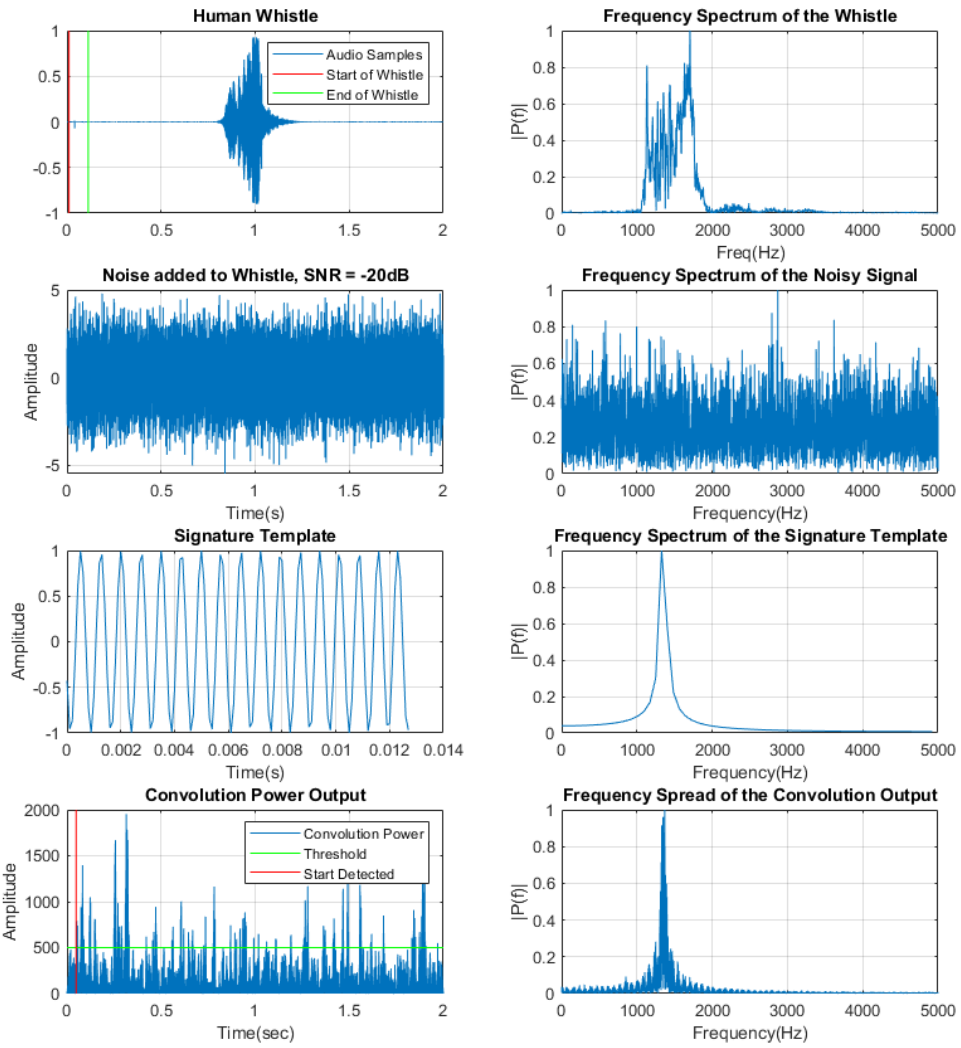


Figure 5.4: Figure shows a simulation study for the detection of a Human Whistle with an SNR of -20dB

These simulation studies helped impose a constraint on the methodology when detecting human whistles. It can be trusted to work accurately when the SNR is -5dB or better.

5.3 Designing the Experiments

The experiments were designed to test the system under a variety of different conditions. The following variables were considered for testing:

- **Room Conditions:** No Multi Path, Multi Path. Multi path refers to a situation where the sound reaching the microphone has gone through multiple reflections
- **Command Signals:** Computer Generated Chirp, Human Whistle
- **SNRs of Chirp with respect to Noise:** [$+\infty$, 5, 0, -2, -5] dB
- **Height between source and microphone array:** [0, 1.5]m
- **Angle of Arrival θ_A:** [0 15 22.5 30 45 67.5 90 112.5 135 157.5 180 270] degrees
- **Localization Methodology:** [Vector Sum (VS), Curve Fitting (CF)]

As mentioned in 2.3, θ_A represents the angle of arrival, θ_D represents the angle estimated by the proposed methods and the height represents the vertical distance between the sound source and the microphone array. The angle of arrival was known to a certainty of ± 10 degrees and hence an uncertainty in the estimated angle of up to $\epsilon = \pm 10^\circ$ is acceptable.

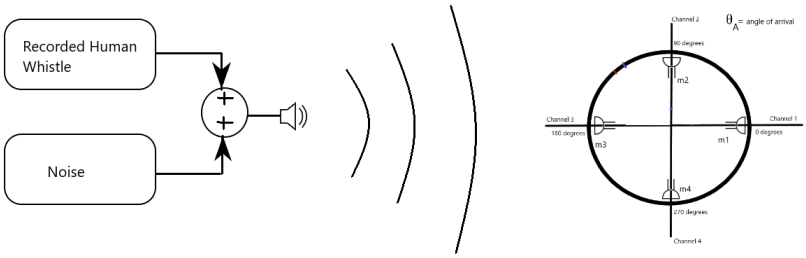


Figure 5.5: Figure shows the experimental setup used to test various SNRs

5.4 Antenna Pattern of the Microphone Array

An Ideal Chirp C was used as the source signal in a quiet room and the power output of each microphone due to the source signal was recorded as the angle of arrival θ_A was varied from 0° to 360° in increments of 30°.

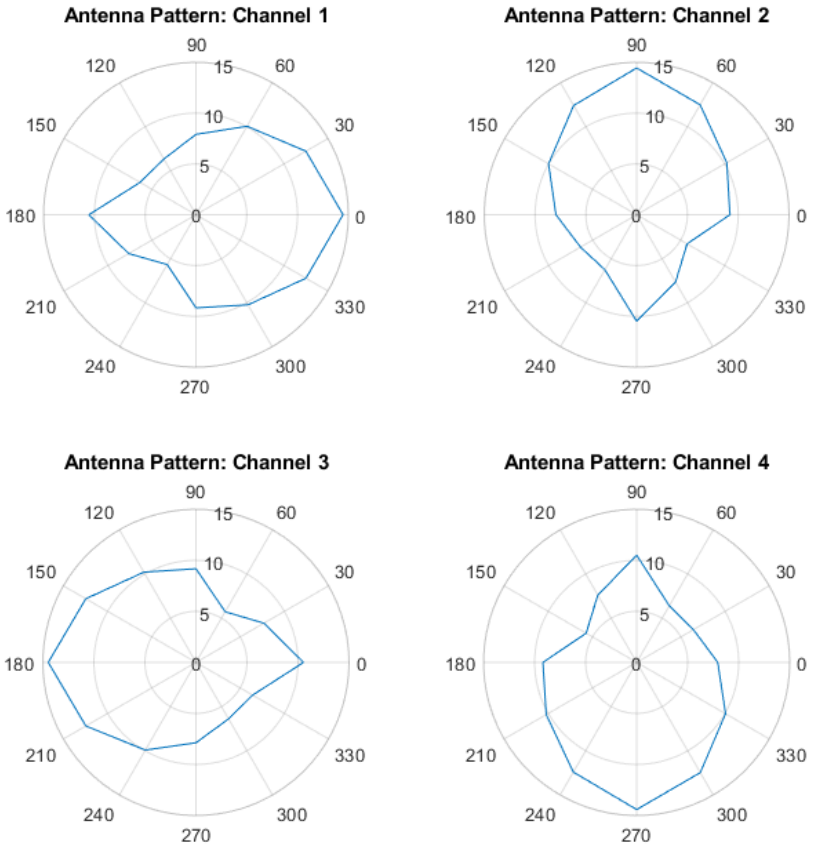


Figure 5.6: Figure shows the antenna pattern obtained for the 4 microphones

The maximum power of Channels 1, 2, 3 and 4 were 14.47, 14.44, 14.57 and 14.41 and they were obtained when the angle of arrival θ_A was $0°, 90°, 180°$ and 270°, which was as expected. The RMS power for Channels 1, 2, 3 and 4 at these angles when their power was maximum was calculated to be 0.4523, 0.4513, 0.4532 and 0.4505 respectively.

The minimum power was 5.66, 5.73, 5.62 and 5.79 and was obtained when the angle of arrival θ_A was $240°, 330°, 60°$ and $150°$ respectively. It was expected that the minimum power for Channels 1, 2, 3 and 4 would be obtained at $180°, 270°, 0°$ and $90°$ respectively, however the power at these angles were observed to be slightly higher than the minimum values. It can be speculated that this increase was due to unwanted reflections of the source signal off of the surrounding walls.

It was expected that the antenna pattern of the 4 microphones would be very similar to each other but rotated by 90° for each channel and the same can be observed in Figure 5.6.

Due to the physical design of the sensing platform, as discussed in Chapter 2, it was also expected that the antenna pattern of each channel would have a radial symmetry. This can also be observed in Figure 5.6, where Channels 1, 2, 3 and 4 are symmetric about the 0° to 180° line, 90° to 270° line, 0° to 180° line and the 90° to 270° line. At a few angles of arrival θ_A for each channel, namely 90° and 120° for Channel 1, 180° and 210° for Channel 2, 270° and 300° for Channel 3 and 0° and 30° for Channel 4 it can be observed that the symmetrical position to these angles of arrival are off in their power values by small amounts and once again it can be speculated to be so due to reflections from the surrounding walls.

5.5 Experimental Test Results

5.5.1 Varying the SNR when using a Human Whistle

Table 5.1: Detection and Localization of Human Whistle from height of 1.5m at Different SNRs from the same angle of arrival

SNR(dB)	θ_A°	$\theta_D^\circ(VS)$	$\lvert\theta_A^\circ - \theta_D^\circ\rvert < \epsilon$?	$\theta_D^\circ(CF)$	$\lvert\theta_A^\circ - \theta_D^\circ\rvert < \epsilon$?
∞	0	5.06	Yes	2.25	Yes
5	0	-1.71	Yes	-0.97	Yes
0	0	3.25	Yes	1.80	Yes
-2	0	7.11	Yes	4.05	Yes
-5	0	2.34	Yes	1.19	Yes

These results indicate that from a head-on position, both methods provide accurate answers up to -5dB. They also indicate that, from a head on position, the CF method is consistently better than the VS method.

When using an Ideal Chirp instead of a human whistle, at -2dB, the VS method gives a result of 6.48° and the CF method gives a result of 2.76° which is marginally better than when a human whistle is used.

5.5.2 Varying the Angle of Arrival when using a Human Whistle

Table 5.2: Detection and Localization of Human Whistle with an SNR of -5dB from height of 1.5m at Different Positions

Position	θ_A°	$\theta_D^\circ(VS)$	$\lvert\theta_A^\circ - \theta_D^\circ\rvert < \epsilon$?	$\theta_D^\circ(CF)$	$\lvert\theta_A^\circ - \theta_D^\circ\rvert < \epsilon$?
1	0	7.93	Yes	4.05	Yes
2A	15	14.53	Yes	8.20	Yes
2	22.5	27.20	Yes	16.18	Yes
2B	30	37.27	Yes	21.74	Yes
3	45	45.61	Yes	52.14	Yes
4	67.5	63.89	Yes	71.59	Yes
5	90	83.94	Yes	94.52	Yes
9	180	176.74	Yes	182.45	Yes
13	270	264.52	Yes	273.91	Yes

These results indicate that the Curve Fitting method works better than the Vector Sum method for head on position, while the Vector Sum method works better for other positions.

5.5.3 Varying the Angle of Arrival when Height is changed to 0m

Table 5.3: Detection and Localization of Human Whistle with an SNR of -5dB from a height of 0m at Different Positions

Position	θ_A°	$\theta_D^\circ(VS)$	$\lvert\theta_A^\circ - \theta_D^\circ\rvert < \epsilon$?	$\theta_D^\circ(CF)$	$\lvert\theta_A^\circ - \theta_D^\circ\rvert < \epsilon$?
1	0	1.37	Yes	0.87	Yes
2A	15	14.72	Yes	13.41	Yes
2	22.5	23.84	Yes	24.51	Yes
2B	30	31.96	Yes	32.39	Yes
3	45	50.67	Yes	47.32	Yes
4	67.5	66.47	Yes	68.65	Yes
5	90	91.40	Yes	90.50	Yes

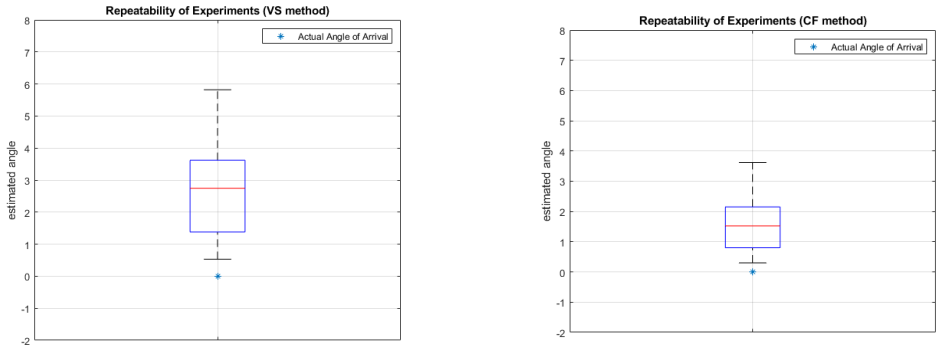
These results indicate that if the source were to be placed at the same height as the microphone array, both methods still provide accurate results.

5.5.4 Repeatability Test Results

Table 5.4: Repeatability of Results from a given position

Trial Number	θ_A°	$\theta_D^\circ(VS)$	$\lvert\theta_A^\circ - \theta_D^\circ\rvert < \epsilon$?	$\theta_D^\circ(CF)$	$\lvert\theta_A^\circ - \theta_D^\circ\rvert < \epsilon$?
1	0	5.82	Yes	3.62	Yes
2	0	1.04	Yes	0.61	Yes
3	0	1.42	Yes	0.80	Yes
4	0	2.78	Yes	1.51	Yes
5	0	3.99	Yes	2.28	Yes
6	0	2.71	Yes	1.53	Yes
7	0	3.49	Yes	1.99	Yes
8	0	0.53	Yes	0.30	Yes
9	0	3.62	Yes	2.15	Yes
10	0	1.38	Yes	0.81	Yes

For the curve fitting method, the mean is 2.67° with a Standard Deviation of 1.62° and for the Vector sum method, the mean is 1.56° with a Standard Deviation of 0.99°. This seems to indicate that, for a head on position, the Curve Fitting method is more accurate than the Vector Sum, as can be observed in Figure 5.7.



(a) VS method

(b) CF method

Figure 5.7: Figure shows the spread of the estimated angle of arrival for both methods.

Since the CF method performs better for head on positions and the VS method performs better at other angles, taking the average of the two estimates may yield a better estimate.

Table 5.5: Using the average of the two methods

Trial Number	θ_A°	$\theta_D^\circ(VS)$	$\theta_D^\circ(CF)$	$\frac{\lvert\theta_D^\circ(VS)+\theta_D^\circ(CF)\rvert}{2}$
1	0	5.82	3.26	4.72
2	0	1.04	0.61	0.82
3	0	1.42	0.80	1.11
4	0	2.78	1.51	2.14
5	0	3.99	2.28	3.13
6	0	2.71	1.53	2.12
7	0	3.49	1.99	2.74
8	0	0.53	0.30	0.41
9	0	3.62	2.15	2.88
10	0	1.38	0.81	1.09

The mean is now 2.11° and the Standard Derivation is 1.30°.

5.5.5 Multipath Case

Table 5.6: Multipath mode (multiple reflections)

Position	θ_A°	$\theta_D^\circ(VS)$	$\lvert\theta_A^\circ - \theta_D^\circ\rvert < \epsilon?$	$\theta_D^\circ(CF)$	$\lvert\theta_A^\circ - \theta_D^\circ\rvert < \epsilon?$
1	0	-65	No	-55	No

In this case, a wall was present very near to Channel D while the source was head on towards Channel A. Due to the wall, the derived angle is wrong. The net vector sum points in between Channels A and D, suggesting that the echo from the wall incorrectly increases the output from Channel D.

This shows that the method fails in the case where there are dominant reflections from close by walls and the resulting estimated angle does not point towards the source anymore but rather points towards the direction of the dominant echo. Since this system is meant to be the ears of an Unmanned Ground Vehicle as explained in Chapter 1, it may be worthwhile to have a distance sensor and not trust the readings when there is a wall close by.

Chapter 6

Conclusions

6.1 Summary and Conclusions

This chapter summarizes the system developed for the localization and detection of a human whistle, the goals achieved and the conclusions reached about the system.

A cost efficient sensing platform containing the microphone array was developed for audio sensing such that each microphone acted as a directional 'ear'. The development of this platform was of critical importance to the methods developed.

A method for the detection of a whistle was also developed with an emphasis on being memory and computationally efficient so that it may be used in real time on a microcontroller. This method relied on correlating the incoming audio signal with a signature template to look for a match.

After detecting the whistle, two methods were developed for localization of the sound source. Both methods were computationally efficient and could be implemented on a microcontroller without a large penalty. These methods were the Vector Sum method which relied on

estimating the angle of arrival using the vector sum of the power output of the 4 channels and the Curve Fitting Method which relied on fitting a quadratic curve through the dominant channel and its two neighboring channels and estimating the angle of arrival by the location of the peak of this quadratic curve.

The detection and localization methods were validated over a range of different scenarios. An ideal Chirp and a human whistle were considered. These were then corrupted with noise over an SNR range of $+\infty$ dB to -5dB. The height between the source and microphone array was also taken into consideration. The methods were also tested over a wide range of angle of arrivals varying from 0° to 360° in small increments.

During the testing of the localization methods, it was observed that the Curve Fitting method was more accurate than the Vector Sum method at head on positions whereas the Vector Sum method performed better at different angles. Thus, it was proposed to use the average of the two estimates as the final estimate of the angle of arrival.

Repeatability tests were also performed to check the correctness of both methods. The tolerance of the system was set to 10 degrees which was sufficient for the intended use case.

It was observed that the methods failed when there were multiple reflections from a nearby wall. Further research may be needed to develop methods more robust to such reflections.

In conclusion, a low cost hardware platform containing a four microphone array was developed for the detection and localization of a human whistle. Memory and computationally efficient methods were also developed so that they could be implemented in real time on a low cost microcontroller and used on a low cost Unmanned Ground Vehicle.

Bibliography

[1] Adafruit. *Electret Microphone Amplifier - MAX4466 with Adjustable Gain*, 2018. URL https://www.adafruit.com/product/1063.

[2] Jose A. Belloch, Maximo Cobos, Alberto Gonzalez, and Enrique S. Quintana-Ortí. Real-time sound source localization on an embedded gpu using a spherical microphone array. *Procedia Computer Science*, 51:201 – 210, 2015. ISSN 1877-0509. doi: https://doi.org/10.1016/j.procs.2015.05.226. URL http://www.sciencedirect.com/science/article/pii/S1877050915010340. International Conference On Computational Science, ICCS 2015.

[3] S. T. Birchfield. A unifying framework for acoustic localization. In *2004 12th European Signal Processing Conference*, pages 1127–1130, Sept 2004.

[4] S. T. Birchfield and R. Gangishetty. Acoustic localization by interaural level difference. In *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, volume 4, pages iv/1109–iv/1112 Vol. 4, March 2005. doi: 10.1109/ICASSP.2005.1416207.

[5] S. T. Birchfield and D. K. Gillmor. Fast bayesian acoustic localization. In *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages II–1793–II–1796, May 2002. doi: 10.1109/ICASSP.2002.5744971.

[6] Michael S. Brandstein and Harvey F. Silverman. A practical methodology for speech source localization with microphone arrays. *Computer Speech & Language*, 11(2):91 – 126, 1997. ISSN 0885-2308. doi: https://doi.org/10.1006/csla.1996.0024. URL http://www.sciencedirect.com/science/article/pii/S0885230896900248.

[7] Texas Instruments. *C2000 Delfino MCU F28379D LaunchPad*, 2018. URL http://www.ti.com/tool/launchxl-f28379d.

[8] Texas Instruments. *ARM® Cortex®-M4F Based MCU TM4C123G LaunchPad*, 2018. URL http://www.ti.com/tool/EK-TM4C123GXL.

[9] K. Kanagisawa, A. Ohya, and S. Yuta. An operator interface for an autonomous mobile robot using whistle sound and a source direction detection system. In *Proceedings of IECON '95 - 21st Annual Conference on IEEE Industrial Electronics*, volume 2, pages 1118–1123 vol.2, Nov 1995. doi: 10.1109/IECON.1995.483953.

[10] C. Knapp and G. Carter. The generalized correlation method for estimation of time delay. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 24(4):320–327, August 1976. ISSN 0096-3518. doi: 10.1109/TASSP.1976.1162830.

[11] K. F. Leung, F. H. F. Leung, H. K. Lam, and P. K. S. Tam. Recognition of speech commands using a modified neural fuzzy network and an improved ga. In *The 12th IEEE International Conference on Fuzzy Systems, 2003. FUZZ '03.*, volume 1, pages 190–195 vol.1, May 2003. doi: 10.1109/FUZZ.2003.1209360.

[12] Enzo Mumolo, Massimiliano Nolich, and Gianni Vercelli. Algorithms for acoustic localization based on microphone array in service robotics. *Robotics and Autonomous Systems*, 42(2):69 – 88, 2003. ISSN 0921-8890. doi: https://doi.org/10.1016/S0921-8890(02)00325-1. URL http://www.sciencedirect.com/science/article/pii/S0921889002003251.

[13] John C. Murray, Harry R. Erwin, and Stefan Wermter. Robotic sound-source localisation architecture using cross-correlation and recurrent neural networks. *Neural Networks*, 22(2):173 – 189, 2009. ISSN 0893-6080. doi: https://doi.org/10.1016/j.neunet.2009.01.013. URL http://www.sciencedirect.com/science/article/pii/S0893608009000136. What it Means to Communicate.

[14] M. Nilsson, J. S. Bartunek, J. Nordberg, and I. Claesson. Human whistle detection and frequency estimation. In *2008 Congress on Image and Signal Processing*, volume 5, pages 737–741, May 2008. doi: 10.1109/CISP.2008.415.

[15] Caleb Rascon and Ivan Meza. Localization of sound sources in robotics: A review. *Robotics and Autonomous Systems*, 96:184 – 210, 2017. ISSN 0921-8890. doi: https://doi.org/10.1016/j.robot.2017.07.011. URL http://www.sciencedirect.com/science/article/pii/S0921889016304742.

[16] Ping Song, Chuangbo Hao, Jiangpeng Wu, and Cheng Yang. Acoustic source localization using 10-microphone array based on wireless sensor network. *Sensors and Actuators A: Physical*, 267:376 – 384, 2017. ISSN 0924-4247. doi: https://doi.org/10.1016/j.sna.2017.10.019. URL http://www.sciencedirect.com/science/article/pii/S0924424717309494.

[17] Darren B. Ward and Robert C. Williamson. Particle filter beamforming for acoustic source localization in a reverberant environment. In *in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP-02*, 2002.