Audio-based Motion Detection

Filip Dujmic
fi5546du-s@student.lu.se
Gopal Gomatam
go3751go-s@student.lu.se

Department of Electrical and Information Technology
Lund University

Supervisor: Buon Kiong Lau (EIT) and Christoffer Cronström (Axis Communications)

Examiner: Fredrik Rusek

6th July 2021

Printed in Sweden
Tryckeriet i E-huset, Lund

Abstract

Motion detection is an essential technology and it has numerous use-cases, such as security tracking, automated door opening systems, and IP cameras. Commonly, passive infrared sensors or radio frequency sensors are used for motion detection. However, this thesis focuses on performing motion detection using existing hardware at audio frequencies. Specifically, it aims to implement a motion detection algorithm using existing speakers and microphones on each of two intercom devices designed by Axis Communications, a smaller one called Device 1 and a larger one called Device 2. The goal of implementing such an algorithm is to save power and cost by re-using existing hardware.

In this thesis, data was collected and studied for both stationary and moving objects to understand the systems' behaviour. The range detection capability for different signal parameters were compared and analysed using a stationary object, at different positions within the range of the systems. This was done to find the optimum signal parameters for the motion detection algorithm. Once these were determined, data was collected and analysed for a moving object. Based on this data, an optimisation and tracking algorithm was designed to obtain better results from the systems. The results show that for Device 1, the algorithm can detect any movement of a person within the range of the system. However, for Device 2, the algorithm can differentiate between different movements, allowing the detection of a person who is approaching the device, rather than just someone who is just passing by. The latter enables power saving for the device, since its camera will only turn on when someone approaches the device with the intention of using it.

Acknowledgements

We have received a great deal of support and assistance throughout the writing of this thesis.

We would first like to thank our supervisors, Professor Buon Kiong Lau and Christoffer Cronström. Their expertise has been invaluable in the formulation of the research questions and methodology. Their insightful and constant feedback pushed us to sharpen our thinking and brought our work to a higher level.

We would then like to thank our colleagues at Axis Communications - Steve Darmadi, Niklas Hjern, Marie Åkesson and others. Their constant support and energy has helped us complete the project successfully. We would also like to thank them for the opportunity and tools to carry out this project in Axis Communications' offices when it was possible.

Finally, we would like to thank Sriram Gomatam for helping us with editing and Naveen Kumar for helping with the graphic images on the report. We want to thank all our friends and family for always being there for us throughout our education and more so during our master's thesis. We could not have completed this without their invaluable support and encouragement.

Popular Science Summary

Over the past few decades, motion detection has played a vital role in automating various tasks. These tasks fall into the domains of security monitoring, automated doors and lighting control, among others. Common sensors for motion detection include those that emit and receive electromagnetic waves (RF) as well as sensors that just detect electromagnetic (infrared) radiation emitted from the human body. However, when these sensors are not readily available, other technologies can be modified to achieve the same purpose. In this thesis, one such example of reusing existing speaker and microphones to perform motion detection at audio frequencies was studied. The modification was implemented on two intercom devices built by Axis Communications, one small in size (Device 1) and one larger (Device 2). The main goal of implementing motion detection capability in the intercom devices is to enable power saving by turning on the camera in the device only when a person is detected to approach the device rather than just walking by. Moreover, the use of existing hardware for this purpose allows for significant cost saving.

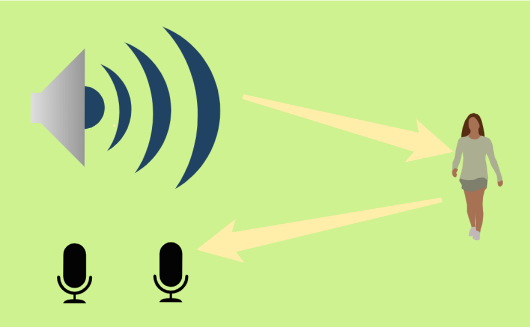


Figure 0.1: Motion detection at audio frequencies using speaker and microphone

The first step in the implementation was to study the behaviour of the systems for a stationary object. This was carried out by collecting data for different object locations with different signal parameters within the effective ranges of both devices. The collected data provided the mean and standard deviation of the error between the measured distance and the actual distance from the device to the object. The results were analyzed to determine the optimum duration, type, shape and frequency range of the signal for obtaining good distance estimation. The next step was to employ the chosen parameters to study the systems' behaviour under different scenarios for a moving object. Data was collected for five different scenarios with the aim to investigate whether both devices can differentiate between different types of movements in front of the device. Based on this data, an optimization and tracking algorithm was implemented to obtain more accurate results from both systems.

The results show that although both devices have a mean error in the range of 10 cm for a stationary object, they can still detect motion within their effective ranges. Device 1 cannot differentiate between the different scenarios considered and it will always turn on when there is any movement within its range. However, Device 2 can differentiate between different scenarios considered, hence allowing it to determine if a person is approach the device or just passing by. This means that it is more efficient in saving power and the associated cost. This is in contrast to using passive infrared sensors, as these recognise even small movements within their effective range, and they cannot differentiate between different kinds of movements, i.e. the device turns on even when it is not required.

Acronyms

AC	Alternating Current
DC	Direct Current
IEEE	Institute of Electrical and Electronics Engineers
IP	Internet Protocol
MP	Megapixel
PIR	Passive Infrared
PoE	Power over Ethernet
RF	Radio Frequency
RFID	Radio Frequency Identification
RTT	Round Trip Time
SONAR	Sound Navigation and Ranging
SPL	Sound Pressure Level
TT	Trip Time

Contents

List of Figures

List of Tables

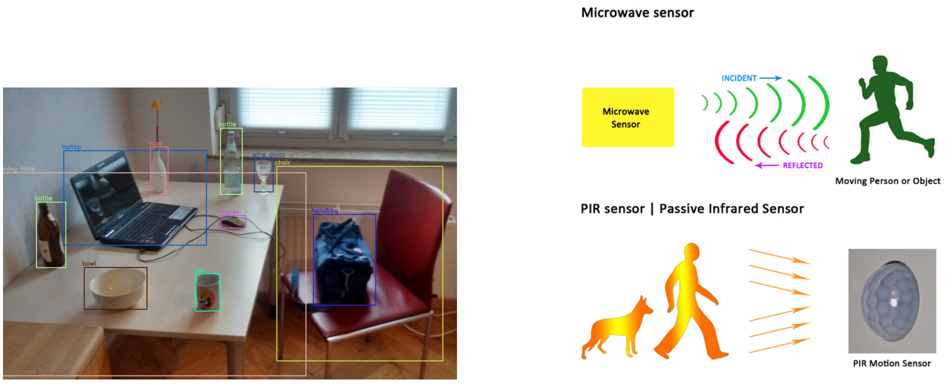
Chapter 1

Introduction

1.1 Background and Motivation

With the increasing interest in technologies for implementing automated actions over the past decade, object detection is being investigated for numerous applications. Object detection has various use-cases such as security, information collection, tracking and detecting of physical presence. Since object detection involves a wide range of scenarios, it can be classified into image-based and non-image based. Computer vision is usually employed for object detection and classification for technologies involving cameras (see Fig. 1.1(a)). Lately, Convolutional Neural Networks have been studied extensively for such applications, e.g., [3].

On the other hand, motion detectors using different technologies are implemented for detecting movements of objects, and commonly these include radio frequency (RF) or passive infrared sensors [4] (also see Fig 1.1(b)). Motion tracking is vital in various scenarios, such as security monitoring, automated lighting control, automatic door opening and IP camera control.



(a) Image based object detection

(b) Microwave and PIR based motion detection

Figure 1.1: Different methods used for object/motion detection

Reprinted from `https://commons.wikimedia.org/wiki/File:Detected-with-YOLO--Schreibtisch-mit-Objekten.jpg` and `https://www.newskysolarlight.com/blog/radar-sensor-vs-pir-motion-sensor-on-solar-light_b0018.html`

Using acoustic frequencies to realize motion detection is less common than using RF or infrared. It has been studied by implementing new hardware specific to the application, e.g., [5]. However, in cases where the hardware is already available and usable, similar acoustic-based methods could be used to implement motion detection, such as SOund NAvigation and Ranging (SONAR). SONAR is used primarily in underwater scenarios due to the lower path loss for acoustic frequencies than RF frequencies [6]. But SONAR can also be used in air to detect and track the presence of an object. Since there are already several intercom devices on the market having speakers and microphones, there is a possibility of implementing a motion detection algorithm on these devices using SONAR techniques, so as to not invest in additional resources or change the hardware layout of the devices to install RF or infrared sensors.

In this thesis, we will be studying the performance of SONAR for motion detection using a speaker and two microphones present on two intercom devices designed by Axis Communications.

1.2 Goal of Thesis and Related Literature

The main goal of this project is to design and implement a motion detection algorithm using one speaker and two microphones on each of two intercom devices made by Axis Communications. There are plenty of these devices already present in the market, and it would be beneficial for Axis Communications and the customers to update the device to have better functionality instead of investing in new hardware. Although Axis Communications has produces intercom devices that contain infrared sensors, there is no space available in some intercom devices to implement these additional sensors. Thus, it would be advantageous to use existing hardware to implement a motion detection algorithm on the device. Implementing this algorithm is not primarily for security, but for saving cost. The idea behind the algorithm is to automatically turn on the camera on the intercom device only in the presence of a person who intends to use the device. Thus, the camera does not need to be turned on perpetually, which facilitates power saving (and hence cost saving). Also, when a person is just passing by and has no intention of using the intercom device, the camera need not turn on. Hence, the algorithm also includes motion tracking. It is noted that one of the two intercom devices considered has infrared sensor for motion detection. It was added to the study since the more compact one (the first one) has the limitation of the microphones being close to the speaker. Nevertheless, the ability to differentiate different motions allows this device to be smarter than only when infrared sensors are available. This is because the SONAR implement will enable it to detect not just motion, but person(s) approaching the intercom (i.e., those intending to use it). This can lead to saving in power and cost even for this device.

An investigation similar to this thesis project, to track single or multiple persons in an indoor environment by making use of inaudible signals hidden in music, was carried out in [7]. However, the purpose of the work was to focus more on using smartphones to raise concerns for privacy threats. The results show that they were able to track walking subjects in line-of-sight, up to 6 m, and through barriers, up to 3 m. The mean tracking error for moving targets was 18 cm, and for stationary targets it was 8 cm. The audio frequency range used for this study was from 18 - 20 kHz.

Another study on implementing active sonar to track minute body movements and finger tracking using smartphones can be found in [1]. In this paper, three different scenarios involving various hand movements were successfully tested. The three scenarios were: a) scrolling a webpage, b) playing Tetris, and c) browsing pictures. These were tested in two different environments, one at a home environment (at a noise level of ~45 dB sound pressure level (SPL)) and one at a noisy cafeteria (noise level of ~72 dB SPL). The average percentages of correctly recognizing the hand movements, in both the environments, are given below in Table 1.1:

Location	Two Handed	Pull Back	Flick	Quick Taps	Slow Taps
Home	96.67	95.00	98.33	86.67	96.67
Cafe	100	96.67	93.33	88.33	93.33

Table 1.1: Average percentages of correctly recognized gestures [1]

However, one limitation of this study is that it was conducted at 18 - 19 kHz range, and these frequencies can be audible to children and pets. And unlike [7], the audio signal is not hidden in music.

This thesis focuses on detecting and tracking the presence of a human by an intercom device, potentially saving power and cost.

1.3 Thesis Outline

The structure of this thesis report is as follows: Chapter 2 introduces the concepts of SONAR and sound propagation to help readers understand the rest of the thesis. In Chapter 3, the technical specifications of the intercom devices are presented. The methods and parameters used for carrying out the motion detection algorithm under perfect conditions inside an acoustic chamber are also presented in Chapter 3. Furthermore, in Chapter 4, the results for stationary object range estimation and moving object scenario classification are presented and discussed. Finally, in Chapter 5, a conclusion to the thesis is given and the scope for future work is outlined.

1.4 Limitations

There are a few limitations or challenges that limit the performance of the motion detection algorithm utilized in this work:

- The highest sampling rate for the microphones is 32 kHz on Device 1, which allows us to only use frequencies of up to 16 kHz, and these frequencies are audible to the human ear.

- There is a delay in the turning on of the microphones to record the reflected signals, compared to the turning on of the speaker, which can cause a mismatch in the synchronization. Moreover, this turn-on delay is also not consistent over time.
- Since the duration of the signals being used is less than 10 ms, the speaker can in some cases cut off the beginning part of the signal and does not play the whole length of the signal. This causes difficulties to accomplish numerous measurements within a certain time period (since some trials are corrupted), and potentially affects the performance of the algorithm.
- The last limitation is that the speaker can also unexpectedly prolong the signal that is transmitted. Therefore, some settling time is required to account for this problem.

Chapter 2

Theory of SONAR

2.1 Introduction

SONAR is a technology that uses the propagation of sound waves to achieve specific applications. In the simplest form, a SONAR system transmits an audio signal from a transducer. The transducer (or another transducer, depending on the type used) then records the reflection of this audio signal from an object (commonly called the target). The time taken to reach the object and arrive at the receiving transducer, at the speed of sound in the environment, can be processed to give the distance to the object. SONAR was first proposed as a technique to detect icebergs. In order to detect the threat of submarines during World War I [8], the technical field of SONAR saw a spike in research and development. Underwater applications such as underwater object detection, navigation and communication make use of this technology. There are two types of SONAR: a) active SONAR and b) passive SONAR. Active SONAR involves emitting pulses of sound and listening to the reflections, and based on the time difference between the sound emitted and echoes received, calculates the distance and other parameters of the object. On the other hand, passive SONAR does not emit any sound, and it only listens to the sound waves emitted by other objects or vessels underwater and uses this to detect natural calamities and identify objects by detecting their acoustic signals. However, it cannot measure the distance to an object. SONAR usually operates in three different bands of frequencies: as shown in Fig. 2.1.

2.2 Active SONAR

2.2.1 Background

A transducer is used in an active SONAR system to convert the electrical energy from a transmitter into acoustical energy. If the transducer only operates for receiving sound waves, then the device is called a hydrophone, and if the transducer can perform both operations of transmitting and receiving, then the device is called a projector [9]. Since active SONAR performs both transmitting and receiving operations, it is usually just referred to as a transducer. In an active SONAR system, the transmitter and receiver can be placed close together (or colocated, if the same transducer is used), and this configuration is known as a mono-static configura-

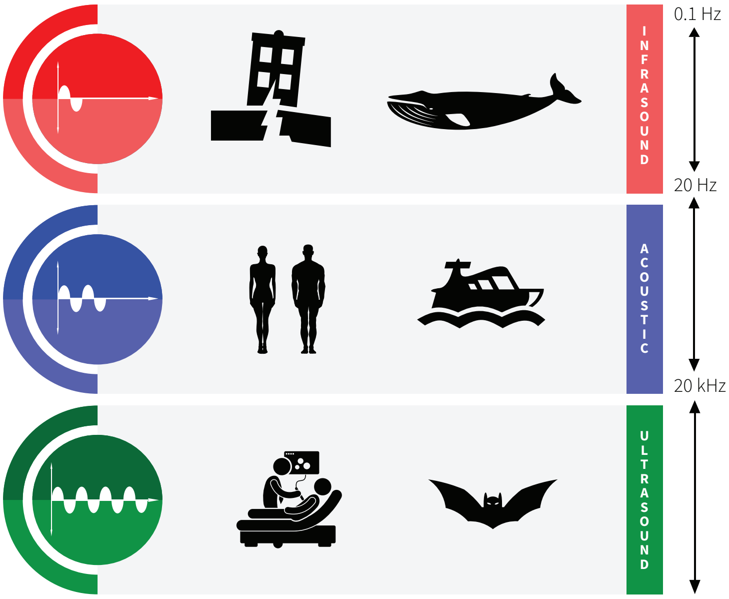


Figure 2.1: Approximate frequency ranges for SONAR

tion. When the transmitter and receiver are separated by a distance comparable to the target distance, it is known as a bi-static or multi-static configuration (see an example in Fig. 2.2). Most SONAR systems have a mono-static configuration [10].

2.2.2 Applications

Some of the most common applications involving SONAR are sea depth measurement (also known as echo sounding), measurement of the distance from one ship to other ships or submarines, and detection of shoals of fish on fishing vessels (see Fig. 2.3(a)). Sonography also uses active SONAR. It is a significant application in healthcare, and it uses ultrasound to create images of the body organs and tissues (known as a sonogram) in a safe, radiation-free and non-invasive manner. Sonograms are made by transmitting ultrasound pulses to the body area using a probe. These pulses echo/reflect off the tissues, and they are recorded and processed to form an image.

2.3 Passive SONAR

2.3.1 Background

In a passive SONAR system, the transducer only performs the operation of receiving a signal. In this case, the transducer is called a hydrophone. The source itself acts as the target because the transducer only 'listens' to the signals emitted by

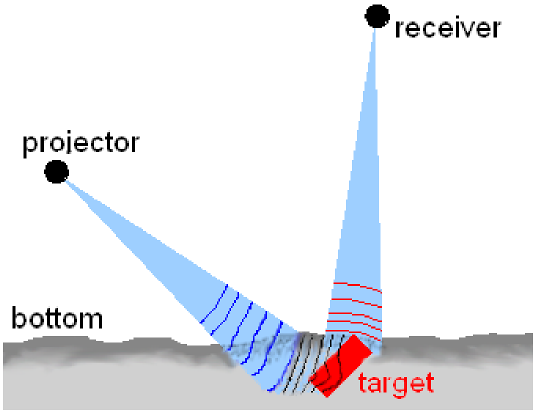
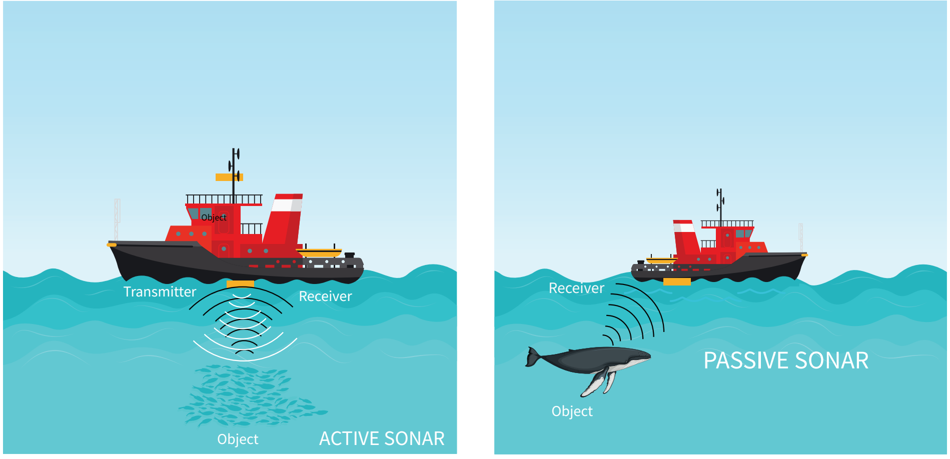


Figure 2.2: Bi-static SONAR configuration for detecting buried objects.
Reprinted from `https://commons.wikimedia.org/wiki/File:Buried_objects_dectection.png`

the source. These signals are recorded and used to identify and classify objects. However, passive SONAR cannot localise the target as precisely as active SONAR because it does not emit signals, i.e., there is no reference point to calculate the round trip time (RTT) of the signal from the source to the target and back. Hence, it cannot calculate the distance to the target [11]. Using the signals that it listens to, it can identify an object based on large databases consisting of reference acoustic signals [12]. Although passive SONAR cannot measure the range of an object, it is possible to use multiple passive SONAR systems for triangulation of a target [13].

2.3.2 Applications

Passive SONAR has been primarily used in military applications from a historical research and development perspective, and specifically to detect submarines [14]. An important reason for this is that passive SONAR does not transmit any signals, thereby not revealing its location and presence, and only detects the presence of the enemy vessels or submarines. Passive SONARs have also been used in identifying, tracking and classification of marine animals (see Fig. 2.3(b)), detecting earthquakes, as well as monitoring the testing of nuclear weapons and detonations [15].



(a) Active SONAR

(b) Passive SONAR

Figure 2.3: Active vs. Passive SONAR
Adapted from
`https://www.rfwireless-world.com/Terminology/Active-Sonar-vs-Passive-Sonar.html`

2.4 Acoustic Frequencies

2.4.1 Infrasound waves

Background

Infrasound refers to the frequencies below the audible range. It usually ranges from 0.1 Hz to 20 Hz (see also Fig. 2.1). The sources of infrasound may be artificial or natural. Some of the artificial or human-made sources of infrasound are nuclear tests, aircraft, and industrial machinery. Examples of natural sources of infrasound are earthquakes, avalanches, volcanic eruptions, and some animals such as humpback whales and elephants [16].

Applications

Infrasound is usually used for long-range detection because of the wavelength of signals at these frequencies. As the wavelength is inversely proportional to frequency, infrasound signals have comparatively longer wavelengths (17.2 m - 3430 m, assuming velocity of sound in air at 20 degree celsius, i.e., 343 m/s)). This property of wavelength is essential because objects interact with different wavelengths in different manners. Whereas the applications of industrial or human-made infrasound are very limited, the wavelength property is often exploited to detect natural calamities such as earthquakes [17], avalanches, or volcanic eruptions [18]. Infrasound is also often used or detected by animals including whales, elephants [19] and rhinoceroses for communication and mating.

2.4.2 Audible waves

Background

Audible wave signals fall in the range of 20 Hz to 20 kHz. Within this range, human ears can hear the sounds, even though the audible range may vary between individuals. For example, older people's hearing tends to be less sensitive to higher frequency sounds than younger people's hearing.

Applications

Audible range SONAR is commonly used for underwater applications such as echo sounding, to detect the distance from the source to the seafloor, or for military purposes to detect enemy's submarines and vessels. An example of this is a passive SONAR to detect an enemy's submarine without giving up its presence and location. Since European vessels usually operate at 50 Hz AC, and U.S. vessels at 60 Hz AC [20], they can detect the nationality of the other vessel or submarine if the vibration insulation is incorrectly installed. SONAR is also used on torpedoes, with the purpose of self-navigation to the target [21]. Another application of audible sonar is to detect shoals of fish; this is also known as a "fish-finder". However, these are just some of the applications of audible sonar. The complete list of applications is quite extensive, and the application areas are very diverse.

2.4.3 Ultrasound

Background

Ultrasound refers to sound waves higher than the human audible range, i.e., frequencies above 20 kHz. Signals of these frequencies have much shorter wavelengths (17.15 cm and below, assuming speed of sound in air at 20 degree celsius) compared to audible or infrasound signals, and therefore the reactions of various objects with these waves differ. This also means that ultrasound experiences higher levels of absorption in the environment, and can only be used for communication or applications with a shorter range [2]. Ultrasound has been studied for many years, since echolocation in bats was discovered in 1794 by Lazzaro Spallanzani. Since then, it has been studied extensively, and is now used for various purposes.

Applications

The applications of ultrasound can be divided into medical, industrial and domestic applications. Out of these, the most common application of ultrasound is in medicine. Ultrasound is used to create images, known as sonograms, of various parts of the human body as required. Sonography is used in medicine as it is a relatively inexpensive and portable method of medical imaging. It is also safe as it does not use ionizing radiation, and the heat generated by these waves are low compared to ionizing radiation, as the power levels of these waves are low [22]. Industrial and domestic applications of ultrasound are primarily for the purposes of cleaning. Ultrasonic jets are used to clean objects such as jewellery, lenses, surgical instruments and surgical parts.

2.5 Sound Propagation

2.5.1 Background

According to the Acoustical Society of America, sound is defined as: (a) oscillation in pressure, stress, particle displacement, particle velocity, etc., propagated in a medium with internal forces (e.g., elastic or viscous), or the superposition of such propagated oscillation; (b) auditory sensation evoked by the oscillation described in (a) [23]. Sound originates by the vibration of an object or source, and it propagates in an elastic medium such as air or water. Sound can be originated either naturally or artificially, and it behaves in the same way for both cases in the environment in which it can propagate. When sound is created, the pressure causes vibrations of the molecules in the medium. These vibrations get transferred to the neighbouring molecules, like the compression and retraction of two connected springs. Even though the molecule only vibrates around its equilibrium, the transfer of the vibrations to adjacent molecules causes sound to propagate in the medium [24].

2.5.2 Properties

The propagation properties of sound are different in different media. The speed of sound propagation of sound in a particular environment depends on the stiffness of chemical bonds between the molecules present in the medium and the molecules' mass density. Thus, the stiffer the chemical bonds among the molecules, the faster sound propagates. This is because the compression and retraction mechanism of the vibrations can travel faster between the molecules. On the other hand, the higher the mass density of the molecules in the medium, the more effort it requires for the other molecules to induce the vibrations, and thus the speed of sound propagation is lower. For example, the speed of sound in ice is over twice as fast as the speed of sound in liquid water. Additionally, the speed of sound in water is 1482 m/s, and in the air, it is 343 m/s, at an environmental temperature of 20°C. This is because the stiffness between the bonds in the molecules is higher in water than in air, even though water is denser compared to air, i.e., the stiffness overcomes the loss in speed due to density [25]. The speed of sound in the air can be calculated using the formula below [26]:

$$\begin{aligned} v_w &= \sqrt{\frac{\gamma RT}{M}} \\ &= \sqrt{\frac{\gamma RT}{M}(\frac{273K}{273K})} \\ &= \sqrt{\frac{(273K)\gamma R}{M}}\sqrt{\frac{T}{(273K)}} \approx 331\frac{m}{s}\sqrt{\frac{T}{(273K)}} \end{aligned} \tag{2.1}$$

Here, v_w is the speed of sound in the medium, γ is a constant that depends on the gas, and for air, $\gamma = 1.4$. M is the molar mass of air and $M = 0.02897$

kg/mol. R is the ideal gas constant and $R = 8.314$ J/(mol $\cdot$ K). Finally, T is the temperature.

The distance that the sound can propagate largely depends on the wavelength. That being said, high frequency sounds are absorbed to a higher degree by air than those with low frequencies, and this is because the wavelength of the high frequency waves is smaller and the wavelength of low frequency waves are larger. The relationship between wavelength and frequency is given in:

$$v_w = f\lambda \tag{2.2}$$

Here, v_w is the speed of sound, λ is the wavelength of the signal, and f is the frequency of the signal. Because of this, higher frequency waves have shorter wavelengths, and the distance that these waves can travel is smaller. However, the speed of sound is higher in water than in air (as discussed earlier), which also influences the wavelength. This explains why sound with the same frequency can travel further away in water compared to air, i.e., the signal of the same frequency has a larger wavelength in water as compared to air, as shown in Table 2.1.

Ultrasound			Infrasound		
Frequency (kHz)	Wavelength (mm)		Frequency (Hz)	Wavelength (m)	
	Air	Water		Air	Water
200	1.7	7.5	100	3.4	15
100	3.4	15	0	17	75
20	17	75	1	340	1500
10	34	150	0.1	3400	15000

Table 2.1: Illustrative wavelengths of Ultrasound and Infrasound signals [2]

Another essential property affecting sound is the sound pressure level. The sound pressure level is the property used to measure the strength of a sound wave. It can be measured easily using inexpensive instruments, and it correlates well with the human perception of loudness. To be able to measure the sound pressure level, a reference value is required. The reference sound pressure level is set to the threshold of human hearing at around 1000 Hz. The reference pressure level in the air is 20 μPa. The sound pressure level can be calculated using the formula below:

$$L_p = 10log\frac{p^2}{p_{ref}^2} = 20log\frac{p}{p_{ref}} \tag{2.3}$$

Here, L_p is the sound pressure level, p is the Root Mean Square sound pressure, and p_{ref} is the reference sound pressure, as mentioned above [27].

Chapter 3

Methods

This chapter discusses the methods used for detecting the motion of an object in front of an intercom device designed by Axis Communications. The motion detection is based on estimating the distance from the intercom to the object over time. In the previous chapter, it was mentioned that SONAR mainly has underwater applications. However, it is implemented in air for this project. It was seen in the previous chapter that different ranges of acoustic frequencies have different applications. However, in this project, we are limited to the audible frequency range (20 Hz - 20 kHz) since the devices in use contain a speaker and two microphones built to work in this range. The second device was added to the study because the first device's speaker and microphones were close to each other. Therefore, another device without this restriction was also considered.

3.1 Intercom Devices

Axis Communications designed the devices used for implementing the motion detection algorithm. Two devices were used for testing the algorithm, one with a smaller form factor and one with a larger size. Henceforth, they will be referred to as Device 1 and Device 2, respectively.

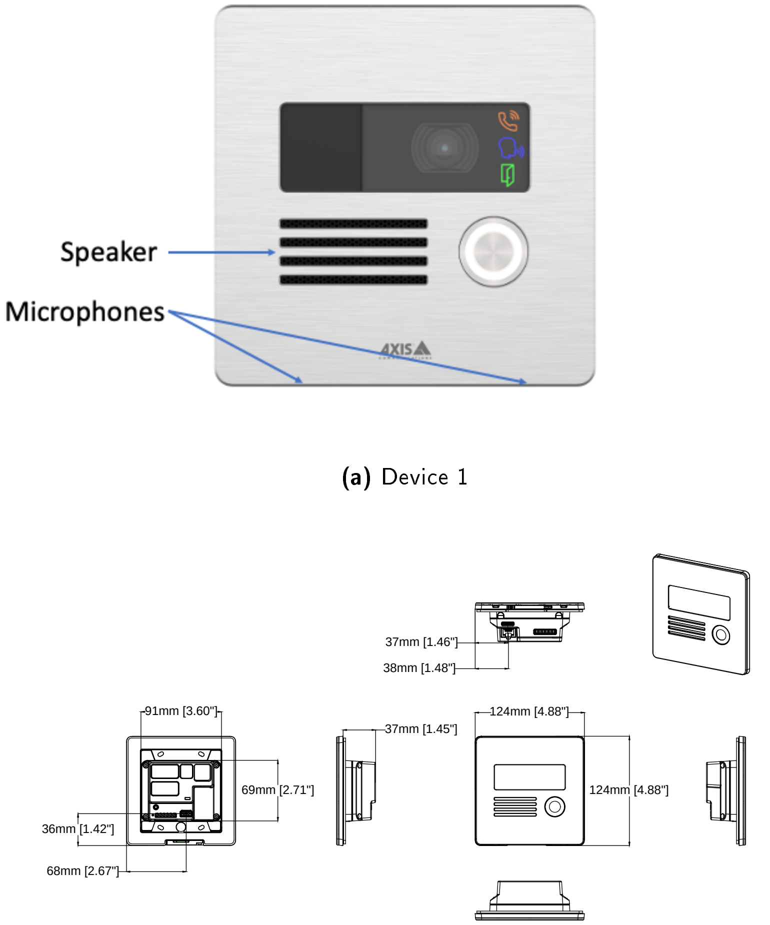
3.1.1 Device 1

Device 1 is an intercom device designed for two-way communication, identification and remote entry control. It provides high-quality audio and video interfaces, as it has a 5 megapixel (MP) camera with invisible infrared night vision capabilities and a 140° field of view. This device's audio capabilities also include echo-cancellation and noise reduction. However, echo-cancellation was turned off for this project's purpose to record reflected signals. The device is easy to install as just the Ethernet cable is required, i.e., it is Powered over Ethernet (PoE) based on the IEEE 802.3af/802.3at Type 1 Class 3 standard. It is usually installed at entry points, granting access to drivers at warehouses and loading docks.

Dimensions and Capabilities

The device is made of a stainless steel, zinc and plastic casing and it has a height and width of 124 mm, and it weighs 900 g (see a photo of the device in Fig. 3.1(a)).

A detailed picture of the dimensions of the device is provided in Fig. 3.1(b).



(a) Device 1

(b) Dimensions of Device 1

Figure 3.1: Device 1 and its dimensions
Reprinted from Axis Communications

The device has one speaker and two microphones. The speaker and the microphones operate at a default sampling rate of either 8 or 16 kHz. The audio streaming is two-way and full-duplex, and it has features of echo cancellation and noise reduction, as mentioned earlier. The speaker also has automatic gain control, and it has a sound pressure of 78 dB at a frequency of 1 kHz and a distance of 1 m (84 dB at 0.5 m at the same frequency).

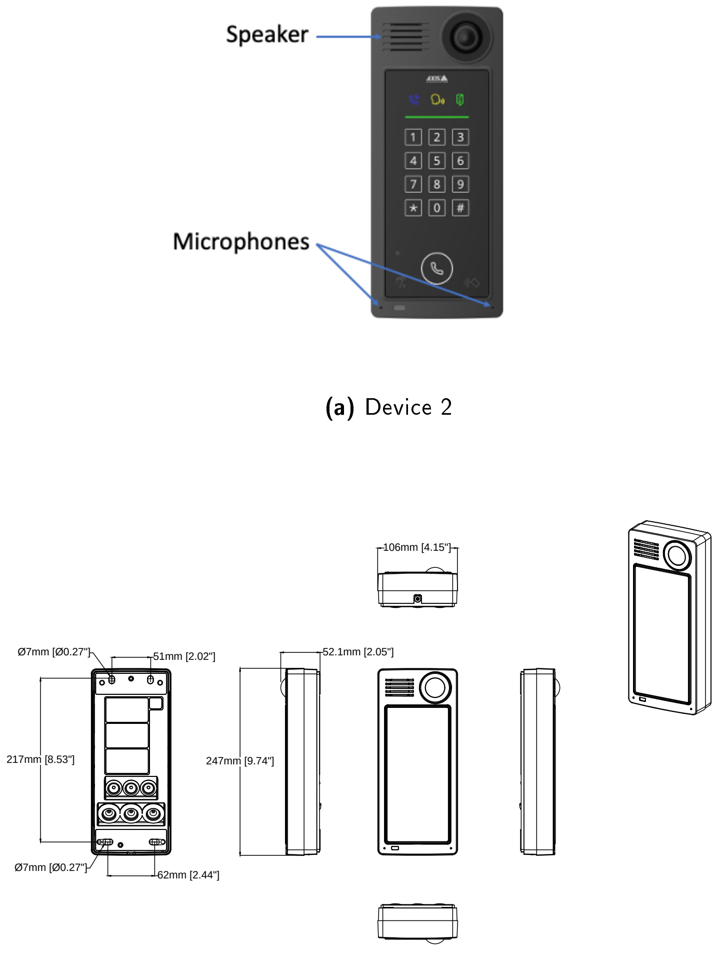
The default mode of operation for the microphones for this device is in Stereo mode, i.e., the data from both the microphones are recorded on two separate channels.

3.1.2 Device 2

Device 2 is also an intercom device designed for multiple functions such as video surveillance, two-way communication, and access control. This device provides a high-quality video and audio interface, and it has a 6 MP security camera. The device also has an RFID multi-frequency reader that supports most standard credential types. Additionally, the device contains a Passive Infrared (PIR) sensor, among many others. This PIR sensor is used for motion detection. Similar to Device 1, this device is also PoE, based on the same IEEE standard, along with PoE+ based on the IEEE 802.3at Type 2 Class 4 standard. However, this device can also be powered by 8 - 28 V DC.

Dimensions and Capabilities

This device is made of aluminium casing, with a polycarbonate hard-coated dome (see Fig. 3.2(a) for a photo of the device). The device has the following dimensions: $H \times W \times D$: 248 × 106 × 51 mm, and it weighs 1.3 kg. A detailed picture of the dimensions of the device is shown below in Fig. 3.2(b).



(a) Device 2

(b) Dimensions of Device 2

Figure 3.2: Device 2 and its dimensions
Reprinted from Axis Communications

In terms of the audio configuration for the device, it has one speaker and two microphones, similar to the previous device. The default sampling rate for the speaker and microphones is again either 8 or 16 kHz. Additionally, the audio is two-way full-duplex communication with features of echo cancellation and noise reduction. However, the sound pressure is 67 dB at a frequency of 1 kHz and a distance of 1 m (73 dB at 0.5 m at the same frequency).

The default mode of operation of the microphones here is double mono instead of stereo, i.e., the audio is recorded using only one channel and copied onto the other channel.

3.2 Signal Parameters

The goal of this thesis was to use the device as a SONAR to find the distance and the angle of approach of a person in front of the device, with the intention of turning on the camera on the device when a person is walking towards the device. The parameters of the SONAR signals used are described below.

3.2.1 Range

As mentioned in Section 1.4 (Limitations), initial experiments were carried out with the devices to understand how the speaker and microphones work. One such experiment was to play a signal of a particular duration while recording to understand if the signal was recorded accurately for the direct paths (from the speaker to the microphones). This was carried out by placing the device in an acoustic anechoic chamber without any object in front of the device that could reflect to find out if the signal played on the speaker matched the signal recorded by the two microphones. When this was analysed, it was observed in rows 2 and 3 of Fig. 3.3 that the duration of the signal that the speaker played was artificially prolonged relative to the original signal in row 1. Based on this, the measurement range was limited to beyond 0.8 m, since the prolonged part of the signal had much greater amplitude than the reflected signal. Otherwise, this "self-noise" will severely affect the accuracy of the distance estimation.

The movement in the distance between 0.8 m and 2 m was deemed sufficient to obtain satisfactory information of a person's intentions in the vicinity of the device. This helped to set the longest distance to be measured accurately, to determine if a person is walking towards the device or just passing by.

3.2.2 Duration

Based on the range mentioned above, and by considering the speed of sound in air (at 20°C), i.e., 343 m/s, experiments were carried out with signals of different duration. The calculations below show the Trip Time (TT) and Round Trip Time (RTT) for a distance (d) of 1 m. The duration of the signals was considered within this Round Trip Time not to exceed it and cause interference between the signal sent out and the signal reflected.

$$TT = \frac{d}{v_w} = \frac{1m}{343\frac{m}{s}} = 0.002915s = 2.915ms \tag{3.1}$$

$$\Rightarrow RTT = TT \times 2 = 2.915 \times 2 = 5.83ms \tag{3.2}$$

Thus, based on this, the signal duration should ideally be less than 5.83 ms and experiments were carried out with signals of duration between 2 - 5 ms. However, signals of duration 6 ms were also used to investigate the behaviour of reflected signals.

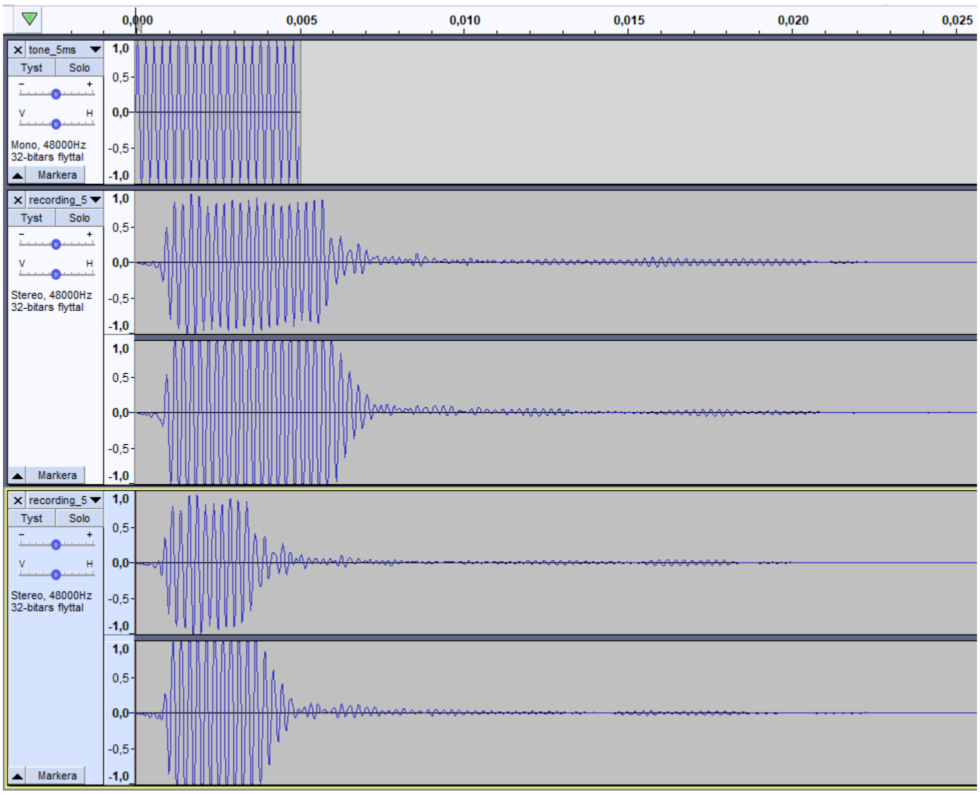


Figure 3.3: Signal prolongation and cutting-off of signal

3.2.3 Types of Signals

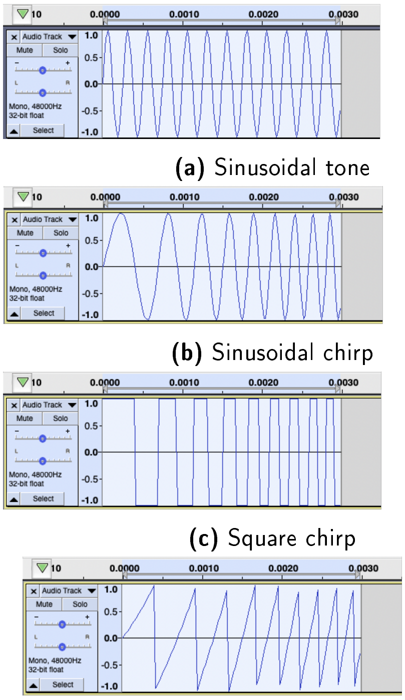
To begin with, two different kinds of signals were considered to be used for detection. These are:

1. Chirp - Chirp is a signal that changes the frequency throughout the duration of the signal. If the frequency is increasing, it's called an up-chirp and if the frequency is decreasing, it's called a down-chirp. A sinusoidal chirp signal generated at 48 kHz sampling rate, with a bandwidth of 1 - 5 kHz, and a duration of 3 ms, is shown in Fig. 3.4(b).

2. Tone - Tone is a regular single frequency sound wave with a certain duration. A sinusoidal tone signal generated at 48 kHz, with a constant frequency of 4 kHz and duration of 3 ms is shown in Fig. 3.4(a).

For the signal with better distance estimation accuracy (between sinusoidal tone and sinusoidal chirp), two other shapes than sinusoids were analysed in this project to determine which shape yields best performance. The shapes analysed other than sinusoidal were square (see Fig. 3.4(c)), and saw-tooth (see Fig. 3.4(d)). The sinusoidal shape is shown for both chirp and tone signal in Figs. 3.4(b) and

3.4(a), whereas the square and saw-tooth shapes are shown for the chirp signal only in Figs. 3.4(c) and 3.4(d), respectively.



(a) Sinusoidal tone

(b) Sinusoidal chirp

(c) Square chirp

(d) Saw-tooth chirp

Figure 3.4: Different types and shapes of the signals used (the unit for the time axis is second)

3.2.4 Frequency

As mentioned earlier, the default sampling rate for both devices is either 8 or 16 kHz. However, with the help of colleagues at Axis Communications, the speaker and microphones of both devices were re-configured to operate at higher sampling rates. Device 1 was re-configured to operate at 32 kHz sampling rate, whereas Device 2 was re-configured to operate at a sampling rate of 48 kHz. Additionally, since the default mode of operation of Device 2 was double mono, it was re-configured to operate in stereo mode to exploit the benefits of recording with two separate channels.

However, having obtained datasheets from the speaker manufacturers, it was observed that the amplitudes of the signals were much lower at certain frequencies compared to the amplitudes at certain other frequencies. This is seen in row 1 of Fig. 3.3, where a 4 kHz tone was played and the signal amplitude is high. Comparing this to Fig. 3.5, where signals of 8, 12, 16, and 20 kHz are shown in rows 1, 3, 5, 7 corresponding to Mic 1 and rows 2, 4, 6, 8 corresponding to Mic 2, respectively, it can be seen that the amplitudes are quite low. Thus, for

Device 2, frequencies between 1 - 5 kHz had the higher amplitudes, and hence, this bandwidth was chosen for that device. However, for Device 1, the higher amplitudes of the signals were observed to be between 1 - 9 kHz.

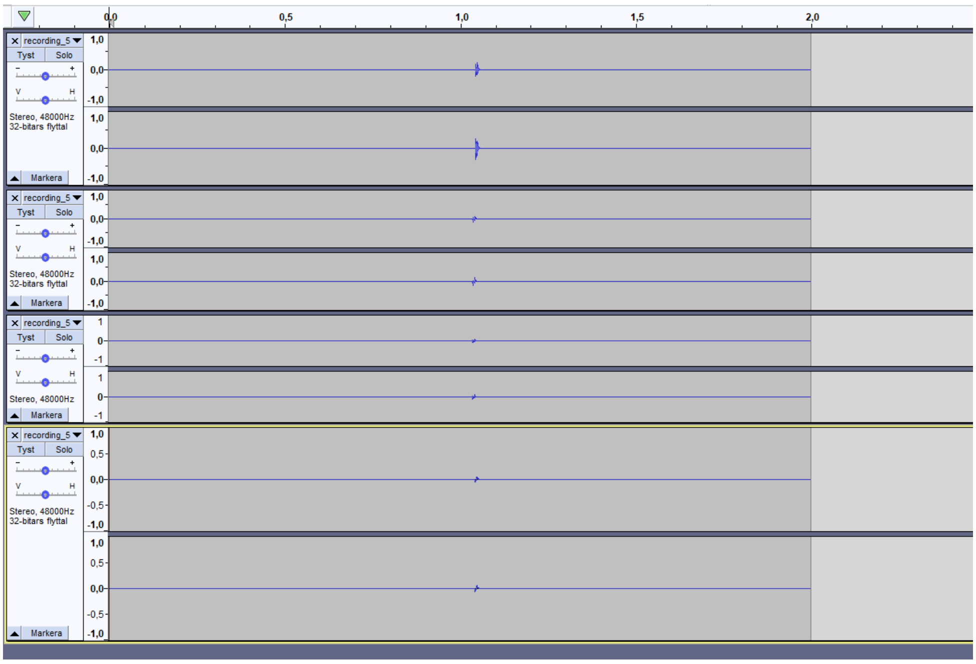


Figure 3.5: Amplitudes of signals at frequencies 8, 12, 16, and 20 kHz

3.3 Experimental Setup

As mentioned in Chapter 1, one of the devices' limitations is a significant delay in turning on the microphone compared to turning on the speaker, and this delay is not consistent over different instances of turning on the devices. Due to this limitation, the recording was started first and then a buffer period of 0.3 s was introduced in order to ensure that the microphone is turned on before the speaker and to be able to record the whole signal that was being transmitted as well as the reflected ones. Additionally, as mentioned in Section 1.4 (Limitations), it was observed that the speaker cuts off some parts of the signal in the beginning (as shown in rows 4 and 5 of Fig. 3.3), as there is a delay in turning on the speaker and this delay is not consistent. To deal with this problem, a silence period of 30 ms was introduced between consecutive measurements for Device 1 and 20 ms for Device 2. The silence period was calculated by allowing for the maximum RTT at the farthest distance in the range of the system, with plenty of buffer period to ensure that the speaker cuts off only the silence part of the signal and not the information part of the signal. As will be explained in more details in the next subsection, the difference of the chosen silence period for the two devices is mainly to account for the different distances from the speaker to the microphones in these devices.

3.3.1 Data Collection

The experiments were carried out inside an acoustic anechoic chamber with a person standing at distances of 1, 1.3, 1.5, 1.7, and 1.9 m from the device. Another person was behind the computer executing the commands to transmit and record the signals. The process was started with the idea to find the optimum duration of the signal. The first set of measurements were carried out with Device 1, followed by Device 2. At each distance, signals of duration 2 - 6 ms were used for the measurements. For each duration of the signal, 200 measurements were subsequently carried out 5 times, with around thirty seconds of gap between measurements. These measurements were carried out with a sinusoidal chirp signal, with a 1 - 5 kHz frequency range (1-9 kHz for Device 2). The optimal duration was found to be 3 ms.

Next, the above procedure was performed with a sinusoidal tone of 3 ms to investigate which signal provides more accurate data (chirp/tone). Further, measuring the distances was also carried out with different shapes of the signals, i.e., sinusoidal, square and saw-tooth.

The next step was to determine the system's behaviour when the object/person moves in front of the device. Considering the average walking speed of a person (1.4 m/s [28]), 40 measurements were carried out with these signals at a separation of 30 ms for Device 1 and 20 ms for Device 2. This difference in separation comes from the fact that the microphones and speaker are positioned closer in Device 1 than Device 2. This causes a longer prolongation in Device 1, and hence the need for a longer separation between measurements. Measurements were carried out for five different scenarios depicted in Fig. 3.6, i.e., walking towards the device along its broadside (Scenario 1), walking by the device (Scenario 2), walking towards the device at a 45°angle (Scenario 3), moving in parallel to the device (Scenario 4), as well as when no one was inside the 2 m range (Scenario 5). In Scenarios 1 and 3, the device is expected to turn on, since there is movement towards the device. The percentages of missed detection are calculated for these scenarios. In Scenarios 2 and 4, the device is not expected to turn on when there is movement in the directions shown in Fig. 3.6 for these scenarios. The percentages of false alarms are calculated for these scenarios and presented in Chapter 4. Finally, in Scenario 5, when there is no movement within the range of the system, the device is not expected to turn on, and the false alarm percentages for this scenario are presented.

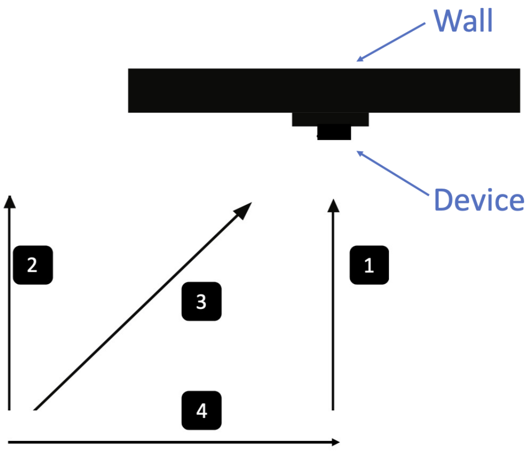


Figure 3.6: Different movement scenarios considered

3.3.2 Post Processing

Once the measurements were carried out and the data was collected, the next step was to perform cross-correlation between the signal that was played and the recorded signal. In order to achieve this, the information part of the reference signal (without prior silence) was compared and checked for its correlation with the entire recorded signal, i.e., the microphones recorded the signal that was played (direct signal) as well as the reflections.

The highest correlation was observed for the direct signal from the speaker to the microphone. Using this as the starting point, the time taken (in the number of samples) for the reflection to arrive at the microphones was determined by observing the second highest correlation peak. In order to achieve this, it had to be ensured that the algorithm would not pick peaks that were close to the highest peak (the starting point) since the signal played was prolonged, and that part would also give a relatively high correlation value. These were discarded by setting correlation values that corresponded to a distance of 80 cm or closer to 0. Thus, the algorithm picked only the second-highest correlation peak, and this would be from the reflection at the object. Additionally, the correlation values were checked for the highest peak only until a distance of 2 m, because it was enough to detect if a person was walking towards the device. Since the recording was carried out in stereo mode, the above post-processing was performed for the data sets from both microphones.

Once the two peaks were determined, the difference between them was calculated. This gave the RTT in the number of samples, which was later converted to the RTT in milliseconds, as shown below. For Device 1, the sampling rate used was 32 kHz, which meant that each millisecond contained 32 samples. Moreover, for Device 2, a sampling rate of 48 kHz was used, and this meant that there were 48 samples in 1 ms. With the above information, the target distance was calculated as shown below:

$$RTT = \frac{No.of Samples}{32\frac{Samples}{ms}} \tag{3.3}$$

or

$$RTT = \frac{No.of Samples}{48\frac{Samples}{ms}} \tag{3.4}$$

The TT was obtained by dividing RTT by 2, and this was then converted to time in seconds by dividing by 1000. Along with the speed of sound in air, this TT was used to determine the distance between the device and the object in metres, as shown in the formula below:

$$Distance = TT \times v_w \tag{3.5}$$

where v_w is the speed of sound in air at 20°C, which equals 343 m/s.

3.3.3 Optimization

For each set of 40 measurements, the correlation peaks of the direct and reflected signals were found using the post-processing method described above. An optimization algorithm (see Fig. 3.7) was designed to be able to track the movement of the object of interest. Distances closer than 1.95 m with a correlation value less than 0.4 and all distances higher than 1.95 m were set to zero to ensure that tracking was performed within the optimum range of the system. Once the first distance closer than 1.95 m with a correlation value greater than 0.4 was found, it was chosen as the starting point of tracking. The next point was calculated by finding the highest correlation value in the range of +/- 17.15 cm (the reason for this choice of 17.15 cm will be explained later) from the previous measurement, and this procedure was carried out until the 40^{th} measurement. Once the person reached a distance of 1.15 m, the algorithm would set the rest of the measurements to 0 to achieve a more precise slope. It is essential to notice that measurements set to 0 did not contribute to the slope value. The threshold was set to 1.15 m because it was learned from the initial results in Chapter 4 that the devices are inaccurate at closer distances. Assuming that the object is moving with a speed of 1.4 m/s (the average walking speed of humans [28]), the change in distance between two measurements is 3.5 cm. The margin of error (11.19 cm) was then calculated based on the standard deviations obtained in Chapter 4 and added to the expected change of distance of 3.5 cm. To allow for this change, at least 14.69 cm of change between samples had to be accommodated. The difference between the chosen 17.15 cm and 14.69 cm allows a person to walk faster and still turn on the device.

The motion detection algorithm is based on detecting the estimated slope and comparing it with a threshold to determine if the device (more precisely, camera) should be turned on. The threshold should then be set based on the statistics from multiple trials of each scenario, ensuring that the device will only turn on when required. 30 trials were performed for each scenario, and assuming normal distribution, mean and standard deviation were calculated. The threshold for turning on the device is set at two standard deviations from the average slope found for Scenario 3. This threshold was set because it provided the best results, i.e., the minimal percentages of missed and false detections. When the value of the slope is lesser than the boundary value for that microphone, the decision will be to turn on the device. The device will stay turned off only if both microphones provide slopes that fall outside the threshold value. More specifically, the threshold was set at two standard deviations from the average slope for Scenario 3 because it was calculated that the percentage of misses for both microphones would be 2.28%, assuming a normal distribution of slope values. Since the slope values for both microphones have to fall outside of the threshold for the device to stay off, the probability that the device will stay off for Scenario 3 becomes 0.08%, and 0% for Scenario 1. The decision to set the threshold value in this manner is also because it is more important to ensure that the device will turn on every time a person is approaching the device (i.e., very low probability of misses), even though it might increase the percentages of false alarms for Scenario 2 and 4. It is crucial to notice that if all 40 measurements are set to zero, the algorithm will not look for a slope, but it will just stay off. This happens when there is no movement inside the 2 m range from the device (Scenario 5).

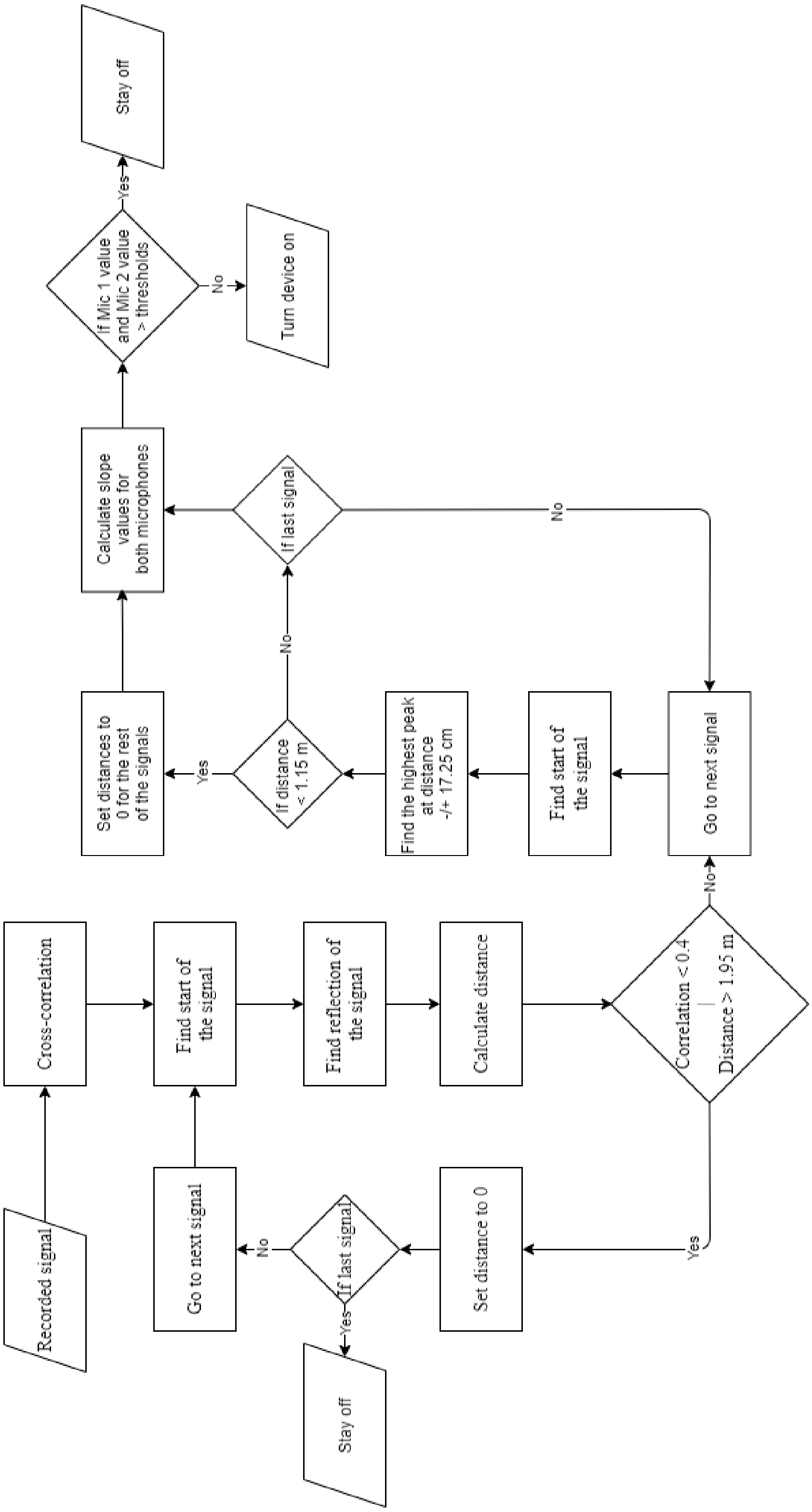


Figure 3.7: Optimization Flow Chart

Chapter 4
Results

This chapter presents and analyzes the results obtained from this project. It is divided into two sections for ease of comprehension: a) Stationary Object b) Motion Tracking. The stationary object study is intended to provide signal parameters that can support good distance measurement (from device to object), which is then used to determine motion (change of distance over time) in the second study. Investigations were carried out for both devices mentioned in the previous chapters. It is noted that all the measurements were carried out inside an acoustic anechoic chamber, i.e., acoustic reflections arrive only from the object of interest.

4.1 Stationary Object

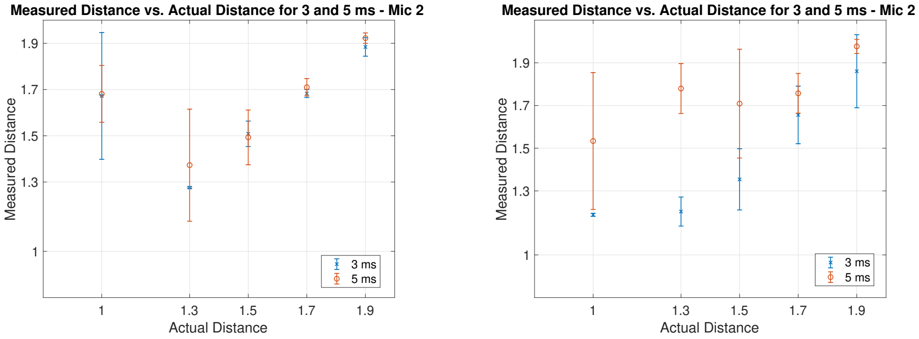
In this section, the precision of distance estimation for both devices is presented and compared when the object is stationary. The range mentioned in Chapter 3 was 0.8 - 2 m for the system. However, to be more precise, for Device 1, the effective range was determined to be 0.80 - 2.04 m, and for Device 2 it was 0.79 - 1.96 m. This part of the project consisted of performing 200 measurements at every position for five times yielding 1000 measurements in total for each position. The positions measured were at distances 1, 1.3, 1.5, 1.7 and 1.9 m from the device. The distance estimation accuracy, in terms of the mean error and standard deviation, is presented and discussed below for both devices.

4.1.1 Device 1

The most favourable parameter to determine the accuracy of these measurements is the difference between the actual distance and the mean of measured distances, i.e., mean error. The second parameter taken into account to compare the accuracy is the standard deviation. The most accurate measurement method would ideally give the smallest difference between the actual distance and measured distance while also having the least standard deviation. Assuming that the distribution of these measurements is normal, 68% of the measured values would then fall within one standard deviation from the mean [29].

Finding the optimum duration

These measurements were carried out using a sinusoidal chirp signal, with a bandwidth of 4 kHz, from 1 - 5 kHz. From Tables 4.1 and 4.2, it can be seen that the signals of duration 3 and 5 ms yield the most accurate data for all the distances measured, and not just one particular distance. The mean of the measurements is far from the actual distance for signals of duration 2, 4, and 6 ms. Additionally, they also have high standard deviations which make the distribution relatively flat. However, for 3 and 5 ms signal durations (also illustrated in Fig. 4.1(a)), the mean is not far away from the actual distance and, the standard deviation is lower than the others.



(a) Mic 1

(b) Mic 2

Figure 4.1: Comparison of 3 and 5 ms signals for Device 1

Mic 1										
	2 ms		3 ms		4 ms		5 ms		6 ms	
Range	Error	Std.	Error	Std.	Error	Std.	Error	Std.	Error	Std.
1.9	0.9858	0.0810	0.0161	0.0400	0.1326	0.0707	-0.0223	0.0226	0.1344	0.0336
1.7	0.5370	0.3599	0.0172	0.0172	0.0600	0.0725	-0.0098	0.0375	-0.0294	0.1082
1.5	0.4939	0.6035	-0.0086	0.0551	0.0944	0.0049	0.0069	0.1184	-0.0631	0.1186
1.3	0.1312	0.3894	0.0243	0.0031	-0.2882	0.2909	-0.0730	0.2424	-0.2488	0.2383
1.0	-0.7462	0.2118	-0.6723	0.2744	-0.3580	0.3730	-0.6811	0.1231	-0.6378	0.0322

Table 4.1: Accuracy of 2, 3, 4, 5 and 6 ms signals in meters - Device 1 (Mic 1)

Mic 2										
	2 ms		3 ms		4 ms		5 ms		6 ms	
Range	Error	Std.	Error	Std.	Error	Std.	Error	Std.	Error	Std.
1.9	0.2865	0.2782	0.0394	0.1712	0.0919	0.1145	-0.0771	0.0331	0.1139	0.0298
1.7	0.1890	0.2444	0.0442	0.1350	0.1165	0.0446	-0.0573	0.0931	-0.1459	0.1568
1.5	0.1703	0.2791	0.1461	0.1434	0.0147	0.1518	-0.2089	0.2548	-0.0921	0.1720
1.3	0.2011	0.6396	0.0968	0.0681	-0.3626	0.2599	-0.4796	0.1168	-0.4669	0.0048
1.0	-0.2850	0.4132	-0.1879	0.0068	-0.6589	0.2367	-0.5337	0.3209	-0.1343	0.4555

Table 4.2: Accuracy of 2, 3, 4, 5 and 6 ms signals in meters - Device 1 (Mic 2)

From Table 4.1 and Table 4.2, it can be seen that for distances of 1.7 and 1.9 m, mean error and standard deviation are very similar for both microphones for 3 ms and 5 ms signals. On the other hand, it is seen that for distances of 1.3 and 1.5 m, the mean error is comparable, but the standard deviation is much lower for the 3 ms signal than for the 5 ms signal. It can also be seen that distance estimation is quite inaccurate for both signal duration at a distance of 1 m. The reason for this could be that reflected signals reach the microphone when the prolonged part of the direct-path (with decaying amplitude) signal is still present, at $\sim$ 6 ms from the starting point. This can be observed in Fig. 3.3 (rows 2 and 3). Due to the interference from the prolonged part (of decaying amplitude), the cross-correlation values will be low since the signal being compared to, i.e., the signal played, has constant amplitude. The highest cross-correlation value is then observed at a random distance greater than 1 m, where the prolonged part of the signal has a relatively constant amplitude, even though the amplitude of the reflected signal is smaller at that distance than at a distance around 1 m where the amplitude is changing.

The above observations led to the conclusion that the 3 ms signal is well suited for this thesis. Even though 5 ms signal is as precise as 3 ms signal at longer distances, it is more critical for this algorithm to have accurate measurements at shorter distances, since it aims to turn on the device when a person reaches the distance of 1 m. Because of this, further investigations were carried out with 3 ms signal with purpose of finding out which type of signal is more precise, i.e., chirp or tone, as well as which shape of signal is more precise, i.e., sinusoidal, square or saw-tooth. The assumption is that 3 ms signal duration is the most accurate duration for all types and shapes of the signal, and the goal is to find the most accurate signal (duration, type, shape).

Comparing types of signals

Table 4.3 below show the mean error and standard deviation of 1000 measurements made with a sinusoidal chirp signal with frequency range 1 - 5 kHz and sinusoidal tone signal at 4 kHz frequency, both of 3 ms duration. This investigation was carried out to determine if a signal at a single frequency would yield better results than a signal with a range of frequencies.

	Mic 1				Mic 2			
	Chirp		Tone		Chirp		Tone	
Range	Error	Std.	Error	Std.	Error	Std.	Error	Std.
1.9	0.0161	0.0400	0.6042	0.1231	0.0394	0.1712	0.4702	0.4539
1.7	0.0172	0.0172	0.4367	0.0799	0.0442	0.1350	0.1466	0.2513
1.5	-0.0086	0.0551	0.4306	0.0731	0.1461	0.1434	-0.0270	0.0523
1.3	0.0243	0.0031	0.3017	0.0264	0.0968	0.0681	-0.4980	0.0291
1.0	-0.6723	0.2744	-0.0162	0.0699	-0.1879	0.0068	-0.6937	0.0049

Table 4.3: Accuracy of 3 ms sinusoidal chirp and sinusoidal tone signals in meters - Device 1

Comparing the results for a sinusoidal tone signal with sinusoidal chirp signal in Table 4.3, it is seen that a sinusoidal tone signal at 4 kHz bandwidth with the same duration, is relatively inaccurate. The mean distance is far away from the actual distance (i.e., relatively large mean error), with higher deviation for both microphones. Due to this inaccuracy, a sinusoidal chirp signal of the same duration was chosen for further investigations.

The decision to compare only a 3 ms tone and not tone signals of other duration was made based on the results from the previous sub-section. It was concluded there that a signal of duration 3 ms provides the most accurate results. Hence, it was sufficient to compare a 3 ms tone with a 3 ms chirp to find the optimum type of signal.

Comparing shapes of signals

Further, different shapes of signals were also compared to determine whether the shape has any relationship with the accuracy of the signals. Tables 4.4 and 4.5 show results for sinusoidal, square and saw-tooth shaped chirp signals, with the same frequency range and duration, i.e., 1 - 5 kHz and 3 ms, respectively.

Mic 1						
	Sin		Square		Sawtooth	
Range	Error	Std.	Error	Std.	Error	Std.
1.9	0.0161	0.0400	0.0637	0.0495	0.1047	0.1367
1.7	0.0172	0.0172	-0.0164	0.0816	0.1203	0.2667
1.5	-0.0086	0.0551	0.0482	0.0689	-0.0074	0.0963
1.3	0.0243	0.0031	-0.2258	0.2907	-0.3578	0.1986
1.0	-0.6723	0.2744	-0.6383	0.2820	-0.3817	0.4641

Table 4.4: Accuracy of different 3 ms chirp shape signals in meters - Device 1 (Mic 1)

Mic 2						
	Sin		Square		Sawtooth	
Range	Error	Std.	Error	Std.	Error	Std.
1.9	0.0394	0.1712	0.2416	0.2205	0.2125	0.3745
1.7	0.0442	0.1350	0.0067	0.0873	-0.0518	0.1659
1.5	0.1461	0.1434	-0.1208	0.1922	0.1430	0.3976
1.3	0.0968	0.0681	0.0393	0.1552	-0.4442	0.0797
1.0	-0.1879	0.0068	-0.4579	0.4436	-0.0191	0.2885

Table 4.5: Accuracy of different 3 ms chirp shape signals in meters - Device 1 (Mic 2)

It can be observed that a square-shaped signal yields better results than a sawtooth. However, when the square signal is compared with the sinusoidal signal, it is seen that the sinusoidal signal is quite precise for distances of 1.9 and 1.3 m, with a lower standard deviation. At 1.7 m for Mic 1, the mean measured distance is approximately 1 cm away for both square and sinusoidal signals, but the standard deviation is higher for the square signal. However, at the same distance, the mean is closer to the actual distance for Mic 2 for the square signal, and it also has a lower standard deviation. For 1.5 m, data from both the microphones are comparable for the two signals, but the sinusoidal signals turn out to be slightly better than the square signal.

A sinusoidal chirp signal seems to perform better than a tone signal because the pattern of amplitudes in a chirp signal is more varied than a tone, i.e., it is more 'unique'. The same can be said for the signal shapes as well. The changes in amplitudes for a sinusoidal chirp signal is more gradual than a square or saw-tooth signal, where the amplitudes rise or drop suddenly. This means the sinusoidal shape gives a more unique signature for the purpose of time correlation. Because of the reasons mentioned above, further investigations will be made with sinusoidal chirp signal of 3 ms duration, which in general yield the most accurate distance estimation out of all signal duration, types and shapes.

Finding the optimum frequency range

Finally, measurements at different frequency ranges were taken to determine the optimum range. As mentioned in the previous chapter, the optimal range of frequencies for Device 1 is 1 - 9 kHz. Presented below in Tables 4.7 and 4.2 are the measurement results for the frequency ranges of 1 - 5 kHz, 6 - 10 kHz and 11 - 15 kHz for a sinusoidal chirp signal of 3 ms duration. As before, 5 trials of 200 measurements each were conducted at each position.

Mic 1						
	1 - 5 kHz		6 - 10 kHz		11 - 15 kHz	
Range	Error	Std.	Error	Std.	Error	Std.
1.9	0.0161	0.0400	0.7670	0.1992	0.7552	0.2378
1.7	0.0172	0.0172	0.8480	0.2112	0.3343	0.3029
1.5	-0.0086	0.0551	0.0436	0.1762	-0.1016	0.3742
1.3	0.0243	0.0031	0.1821	0.2373	-0.1651	0.3580
1.0	-0.6723	0.2744	0.0525	0.2975	-0.4472	0.3678

Table 4.6: Accuracy of 3 ms sinusoidal chirp signal for frequency ranges 1 - 5 kHz, 6 - 10 kHz and 11 - 15 kHz in meters - Device 1 (Mic 1)

Mic 2						
	1 - 5 kHz		6 - 10 kHz		11 - 15 kHz	
Range	Error	Std.	Error	Std.	Error	Std.
1.9	0.0394	0.1712	0.7334	0.1121	1.0268	0.1428
1.7	0.0442	0.1350	0.7256	0.1058	0.7954	0.1796
1.5	0.1461	0.1434	0.3534	0.2435	0.7057	0.1039
1.3	0.0968	0.0681	0.2197	0.1558	0.4355	0.1400
1.0	-0.1879	0.0068	-0.3053	0.2168	0.1472	0.2311

Table 4.7: Accuracy of 3 ms sinusoidal chirp signal for frequency ranges 1 - 5 kHz, 6 - 10 kHz and 11 - 15 kHz in meters - Device 1 (Mic 2)

It can be seen from Tables 4.6 and 4.7 that the higher frequency range yield poor results. This is due to limitations in the speaker, as mentioned in Section 3.2. The speaker suppresses the amplitudes of signals at higher frequencies (above 10 kHz), resulting in relatively lower correlation levels between the transmitted and reflected signals. Thus, the system's post-processing function cannot effectively differentiate between noise and the reflected signal levels. However, lower frequency ranges work well with the system, especially for the 1-5 kHz range. Therefore, the motion tracking algorithm was implemented at the 1 - 5 kHz range. One drawback of using these frequencies is that they are audible to the human ear.

Results of 50 trials at one position

As mentioned in Chapter 1, one of the challenges to overcome in this project was the delay in turning on the microphone. This caused a deviation in the mean error between individual trials. It can be seen in Table 4.8, at a distance of 1.5 m, the standard deviation between individual trials is 2.29 and 6.18 cm for Mic 1 and Mic 2, respectively. The assumption is that bias of each individual trial is about the same and this investigation with 50 trials is done to study the statistics of the bias (deterministic error in distance introduced by the uncertain switched-on delay). The results suggest that, for the purpose of distance estimation, there is potential to compensate for this bias for Mic 1, since it is centered below the actual distance (1.3 cm) and has a relatively small standard deviation. However, for motion tracking, this bias is not an issue, since the focus is rather on the difference in the distance estimates (over time). For Mic 2, since the mean bias is close to the actual distance (and has a relatively large standard deviation), it is not as useful to compensate for the bias in the distance estimation.

	Mic 1	Mic 2
	Error	Error
Mean	0.1389	-0.034
Std.	0.0229	0.0618

Table 4.8: Mean and standard deviation of mean error at 1.5 m, 3 ms sinusoidal chirp signal in meters - Device 1

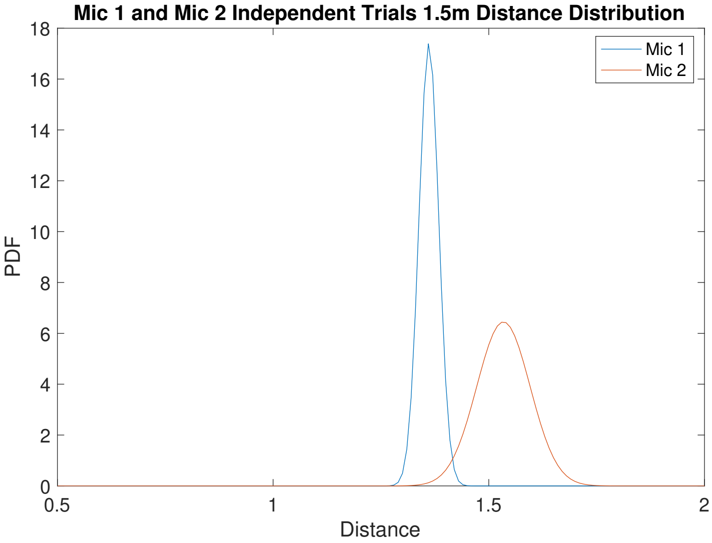


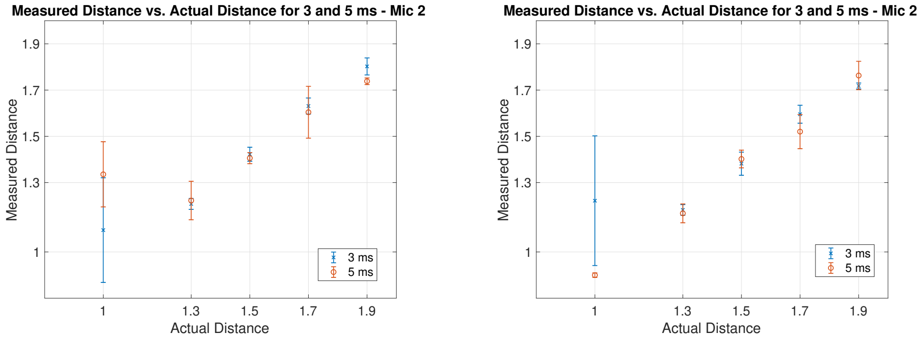
Figure 4.2: Distributions for 50 trials at 1.5 m - Device 1

4.1.2 Device 2

Similar to the Device 1 measurements, 200 measurements were made at each position for five trials. This investigation began with a sinusoidal chirp signal, generated at 48 kHz sampling rate. The procedure used to determine the optimum signal for this device to implement motion detection algorithm is similar to the procedure used for Device 1. The initial measurements were carried out to compare the performance of different duration signals, followed by comparisons of the results with different types and shapes of signals. Finally, the optimum frequency/range of frequencies for the system was determined.

Finding the optimum duration

From Tables 4.9 and 4.10, it can be observed that the sinusoidal chirp signals of duration 3 and 5 ms were more accurate compared to the others, and once again, for all distances and not only one particular distance. The 3 and 5 ms signals are comparable when it comes to mean error, but 3 ms signal has smaller standard deviations (see Fig. 4.3(a) and 4.3(b)). This device is also inaccurate at distance of 1 m for all signal durations because of the same reasons mentioned for Device 1. Thus, the 3 ms signal was chosen to carry out further investigations on this device as well, with purpose of finding out which type and shape of signal yields the most accurate results.



(a) Mic 1 **(b)** Mic 2

Figure 4.3: Comparison of 3 and 5 ms signals for Device 2

Mic 1										
	2 ms		3 ms		4 ms		5 ms		6 ms	
Range	Error	Std.	Error	Std.	Error	Std.	Error	Std.	Error	Std.
1.9	0.1948	0.0016	0.0975	0.0368	0.1279	0.0732	0.1616	0.0143	0.1492	0.0144
1.7	0.1610	0.0416	0.0689	0.0349	0.1873	0.0261	0.0957	0.1119	0.1853	0.0501
1.5	0.1439	0.1030	0.0774	0.0302	0.1392	0.0424	0.0942	0.0239	0.1627	0.0469
1.3	-0.0853	0.3349	0.0924	0.0236	0.0674	0.1150	0.0776	0.0830	0.1687	0.0407
1.0	-0.5674	0.3689	-0.0944	0.2263	-0.2790	0.1709	-0.3356	0.1410	-0.8137	0.1392

Table 4.9: Accuracy of 2, 3, 4, 5 and 6 ms signals in meters - Device 2 (Mic 1)

Mic 2										
	2 ms		3 ms		4 ms		5 ms		6 ms	
Range	Error	Std.	Error	Std.	Error	Std.	Error	Std.	Error	Std.
1.9	0.1662	0.0718	0.1820	0.0129	0.1249	0.0462	0.1369	0.0614	0.2111	0.0140
1.7	0.1171	0.1476	0.1044	0.0387	0.2178	0.0294	0.1795	0.0738	0.2295	0.0306
1.5	-0.1505	0.3049	0.1182	0.0501	0.2588	0.1176	0.0980	0.0385	0.2180	0.0315
1.3	0.1597	0.0253	0.1183	0.0231	0.2410	0.0650	0.1331	0.0406	0.2358	0.0201
1.0	-0.5571	0.3859	-0.2215	0.2805	-0.0950	0.2327	0.1004	0.0098	-0.7807	0.3217

Table 4.10: Accuracy of 2, 3, 4, 5 and 6 ms signals in meters - Device 2 (Mic 2)

Comparing types of signals

Comparing 3 ms sinusoidal chirp signal with 3 ms sinusoidal tone signal (4 kHz) in Table 4.11, it can be seen that the 3 ms sinusoidal chirp signal is more accurate and has lower standard deviations than the 3 ms sinusoidal tone signal. This shows that the 3 ms chirp sinusoidal signal is more suitable, and it was used to find out which signal shape yields the best results for the same reasons mentioned for Device 1.

	Mic 1				Mic 2			
	Chirp		Tone		Chirp		Tone	
Range	Error	Std.	Error	Std.	Error	Std.	Error	Std.
1.9	0.0975	0.0368	0.0851	0.0615	0.1820	0.0129	0.4186	0.1861
1.7	0.0689	0.0349	0.1056	0.2008	0.1044	0.0387	0.1387	0.1096
1.5	0.0774	0.0302	0.1376	0.1494	0.1182	0.0501	0.1332	0.1331
1.3	0.0924	0.0236	0.0086	0.1133	0.1183	0.0231	0.0266	0.1344
1.0	-0.0944	0.2263	-0.1988	0.4378	-0.2215	0.2805	0.0124	0.1615

Table 4.11: Accuracy of 3 ms sinusoidal chirp and sinusoidal tone signals in meters - Device 2

Comparing shapes of signals

Next, different signal shapes (sinusoidal, square, and saw-tooth) of the same duration, i.e., 3 ms, were compared in Tables 4.12 and 4.13. It was observed that the sinusoidal signal performed the best overall when both microphones were taken into account, and it was used further to determine if an inaudible frequency range (18 - 22 kHz) could be used for motion detection.

Mic 1						
	Sin		Square		Sawtooth	
Range	Error	Std.	Error	Std.	Error	Std.
1.9	0.0975	0.0368	0.5007	0.3350	0.3513	0.1373
1.7	0.0689	0.0349	0.2714	0.2347	0.2446	0.2859
1.5	0.0774	0.0302	0.4089	0.1397	0.3672	0.1322
1.3	0.0924	0.0236	0.0642	0.2699	0.2033	0.1565
1.0	-0.0944	0.2263	-0.5149	0.2971	-0.3556	0.2896

Table 4.12: Accuracy of different 3 ms chirp shape signals in meters - Device 2 (Mic 1)

Mic 2						
	Sin		Square		Sawtooth	
Range	Error	Std.	Error	Std.	Error	Std.
1.9	0.1820	0.0129	0.2780	0.1363	0.7368	0.3934
1.7	0.1044	0.0387	0.2480	0.1385	0.1384	0.1988
1.5	0.1182	0.0501	0.2056	0.0640	0.4045	0.4000
1.3	0.1183	0.0231	0.2227	0.2072	0.2822	0.1958
1.0	-0.2215	0.2805	0.0058	0.2073	-0.0474	0.1984

Table 4.13: Accuracy of different 3 ms chirp shape signals in meters - Device 2 (Mic 2)

Finding the optimum frequency

Frequencies at 1 - 5 kHz are audible for human beings, potentially making this system quite annoying to people nearby the device when implemented. In this section, distance estimation of frequencies at the higher end of the audible range and lower end of the inaudible range is investigated, i.e., 18 - 22 kHz.

In Table 4.14, accuracy of 3 ms sinusoidal chirp signals with frequency ranges of 1 - 5 kHz and 18 - 22 kHz are shown. The higher frequency range at 18 - 22 kHz yielded poor, inaccurate estimation of the distance with high standard deviations. This is due to the speaker's limitation, similar to Device 1. Amplitudes of higher frequency signals get suppressed, and because of this, signals at frequencies above 6 kHz were not suitable for the proposed system. Further investigations were carried out with signals in the 1 - 5 kHz frequency range.

	Mic 1				Mic 2			
	1 - 5 kHz		18 - 22 kHz		1 - 5 kHz		18 - 22 kHz	
Range	Error	Std.	Error	Std.	Error	Std.	Error	Std.
1.9	0.0975	0.0368	0.0917	0.1058	0.1820	0.0129	0.1068	0.0895
1.7	0.0689	0.0349	-0.0412	0.0201	0.1044	0.0387	0.0325	0.0546
1.5	0.0774	0.0302	-0.1958	0.2059	0.1182	0.0501	-0.1534	0.2216
1.3	0.0924	0.0236	-0.4884	0.1857	0.1183	0.0231	-0.3591	0.2660
1.0	-0.0944	0.2263	-0.3947	0.3266	-0.2215	0.2805	-0.7649	0.2592

Table 4.14: Accuracy of 3 ms sinusoidal chirp signal for frequency ranges 1 - 5 kHz and 18 - 22 kHz in meters - Device 2

Results of 50 trials at one position

Similar to Device 1, 50 trials were performed at one position even for this device. Due to the delay in turning on of the microphone, the average measured distance has some deviations. From Table 4.15, it is seen that the standard deviation between individual trials is 5.79 cm for Mic 1 and 19.39 cm for Mic 2. The purpose of this investigation is the same as for Device 1. As opposed to Device 1, both microphones in Device 2 show a clear bias in the estimation results (as opposed to only Mic 1 in Device 1), potentially allowing the bias to be compensated. However, further investigations involving other distances should be performed to ensure reliable compensation.

	Mic 1	Mic 2
	Error	Error
Mean	0.1432	0.2373
Std.	0.0579	0.1939

Table 4.15: Mean of the mean distances and standard deviations at 1.5 m, 3 ms sinusoidal chirp signal in meters - Device 2

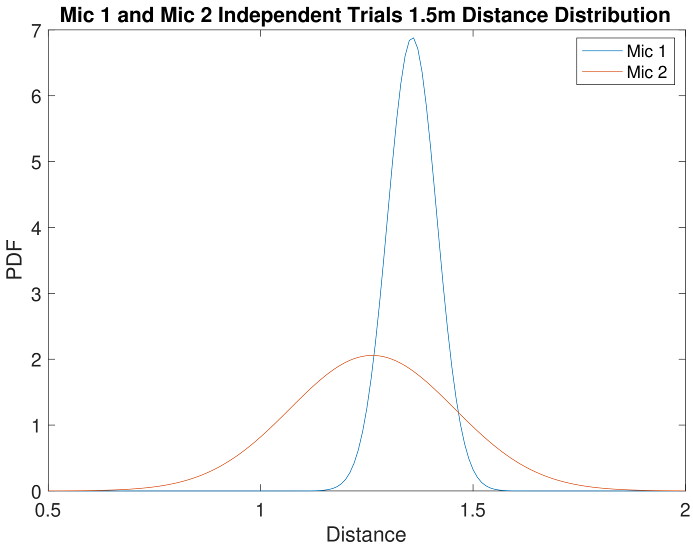


Figure 4.4: Distributions for 50 trials at 1.5 m - Device 2

4.2 Motion Tracking

The main goal of this thesis project was to implement a motion detection algorithm. In this section, results from the optimization and tracking algorithm described in Section 3.3 are presented and discussed, for both devices. Thus, the results in this section are also divided into two sub-sections, one for each device. The duration and type of signals used for tracking the object of interest in this range are the sinusoidal chirp signals of duration 3 and 5 ms. It is to be noted that motion detection is performed within the respective range for each device. Additionally, different types of movements were also measured and the performances for these movements for both devices are discussed.

4.2.1 Device 1 - Motion Tracking

The range for tracking an object for this device is 0.79 - 2.04 m. The algorithm decides whether to turn on the device based on the average slope of the distances measured within one trial, i.e., 40 measurements. From a set of 40 measurements conducted with a 3 ms sinusoidal chirp signal for a period of 1.32 s, the algorithm measures the average change in distance between two measurements. This average change is checked whether it lies within the decision threshold specific to this device. It decides whether the device has to be turned on if the slope lies within this threshold. Table 4.16 shows the percentages of misses and false alarms for each scenario considered. Figure 4.5 shows the average slope distribution for each scenario (obtained from fitting 30 trials of average slope measurements to normal distributions). As mentioned in Section 3.3.3, the threshold for turning on the device was set to two standard deviations above the average of Scenario 3 (as a compromise between percentages of misses and false alarms).

	Scenario 1	Scenario 3	Scenario 4	Scenario 2	Scenario 5
	Miss %	Miss %	False alarm %	False alarm %	False alarm %
Mic 1	0.05	2.28	99.99	99.98	0
Mic 2	3.88	2.28	100	100	0
Device	0	0.05	100	100	0

Table 4.16: Miss and false alarm percentage for different scenarios with 3 ms sinusoidal chirp signal - Device 1

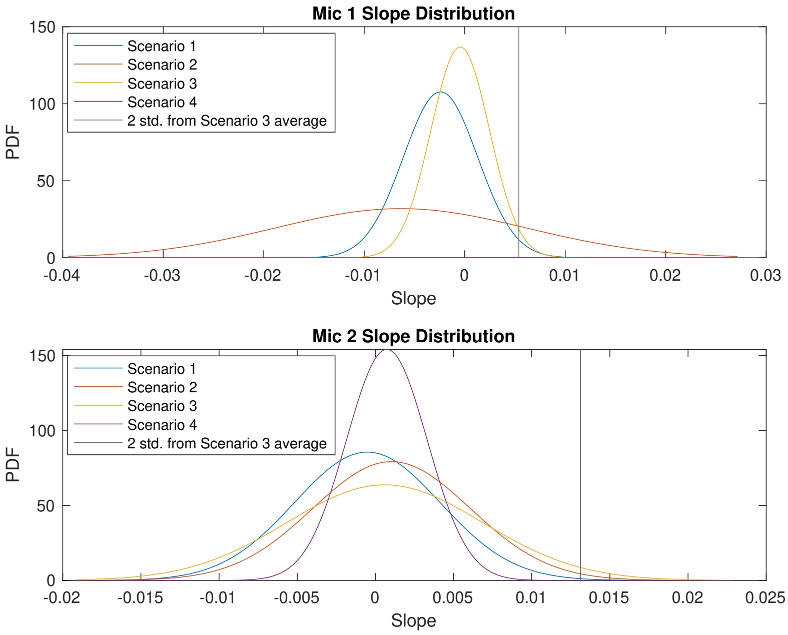


Figure 4.5: Average slope distribution for each scenario for 3 ms signal - Device 1

This device performs poorly because 100% of the time in Scenario 4, the device will turn on when it is not supposed to. The same happens in Scenario 2; the device will turn on 100% of the time when it is not supposed to. Even though the device gives poor results when there is any motion inside the 2 m range, and there is no difference between scenarios, the algorithm can still save some power and costs because it will never turn on when there is no motion inside the 2 m range.

Due to the poor performance of Device 1 with the 3 ms signal, it was decided to also study the 5 ms signal, since it showed similar distance estimation accuracy. Table 4.17 shows the percentages of misses and false alarms for different scenarios for the 5 ms signal duration. It can be seen that the results are similar to the 3 ms signal shown above in Table 4.16. The only difference is seen for Scenario 2, where the percentages of false alarms decreases by 1.21%. This still means that the device cannot differentiate between the scenarios considered. Figure 4.6 shows the average slope distribution for each scenario for the 5 ms signal (obtained from fitting 30 trials of average slope measurements to normal distributions).

	Scenario 1	Scenario 3	Scenario 4	Scenario 2	Scenario 5
	Miss %	Miss %	False alarm %	False alarm %	False alarm %
Mic 1	1.64	2.28	91.86	86.08	0
Mic 2	0.43	2.28	99.99	91.31	0
Device	0	0.05	100	98.79	0

Table 4.17: Miss and false alarm percentage for different scenarios with 5 ms sinusoidal chirp signal - Device 1

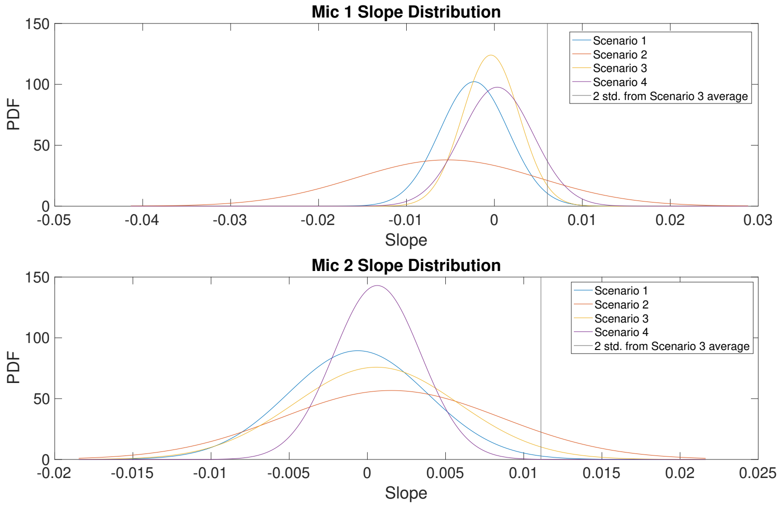
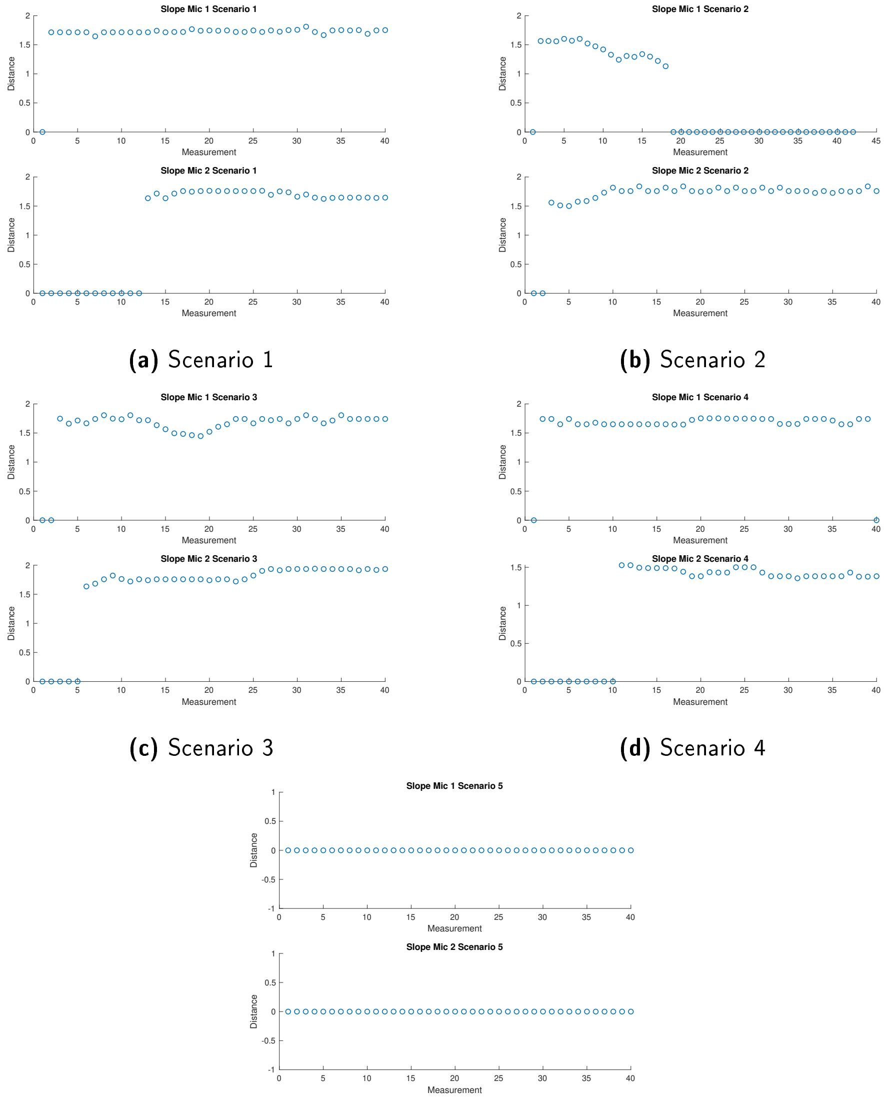


Figure 4.6: Average slope distribution for each scenario for 5 ms signal - Device 1

Figs. 4.7(a) to 4.7(e) show the typical motion tracking process within one trial for all the scenarios considered (using the 3 ms signal). It can be seen that there is no difference in the pattern of the slopes for the different scenarios that can be used to distinguish between them.



(a) Scenario 1

(b) Scenario 2

(c) Scenario 3

(d) Scenario 4

(e) Scenario 5

Figure 4.7: Motion Tracking for Device 1

4.2.2 Device 2 - Motion Tracking

As mentioned in the previous subsection, the optimization algorithm tracks moving objects within a certain range. The decision to turn on the device is decided by the algorithm based on the average slope. Tables 4.18 and 4.19 below show the percentages of misses and false alarms for each scenario for the 3 and 5 ms signals, respectively. Additionally, Figures 4.8 and 4.9 show the average slope distribution for each scenario for the 3 and 5 ms signals, respectively.

	Scenario 1	Scenario 3	Scenario 4	Scenario 2	Scenario 5
	Miss%	Miss%	False alarm%	False alarm%	False alarm%
Mic 1	1.77	2.28	50	82.62	0
Mic 2	0.17	2.28	100	99.18	0
Device	0	0.05	100	85.75	0

Table 4.18: Miss and false alarm percentage for different scenarios with 3 ms sinusoidal chirp signal - Device 2

	Scenario 1	Scenario 3	Scenario 4	Scenario 2	Scenario 5
	Miss%	Miss%	False alarm%	False alarm%	False alarm%
Mic 1	0	2.28	8.26	97.35	0
Mic 2	0.33	2.28	17.72	92.40	0
Device	0	0.05	24.52	99.86	0

Table 4.19: Miss and false alarm percentage for different scenarios with 5 ms sinusoidal chirp signal - Device 2

Implementing the 3 ms sinusoidal chirp signal with this algorithm will cause the device to turn on 100% of the time when it is not supposed to turn on in Scenario 4. Similarly, for Scenario 2, it will turn on 85.75% of the time when it is not supposed to.

On the other hand, using the 5 ms sinusoidal chirp signal with this algorithm will turn on the device only 24.52% of the time for Scenario 4, and 99.86% of the time for Scenario 2, when it is not supposed to.

Both devices will stay off 100% of the time in Scenario 5. It is concluded that even though the 3 ms sinusoidal chirp signal has a high percentage of false alarm for Scenario 2 and 4, it can still save a lot of power and reduce costs, since it will never turn on in Scenario 5. Further, the 5 ms sinusoidal chirp signal saves power and reduces costs even more, since for Scenario 4, it has only 24.52% of false alarm, which makes it the most suitable for purposes of this project.

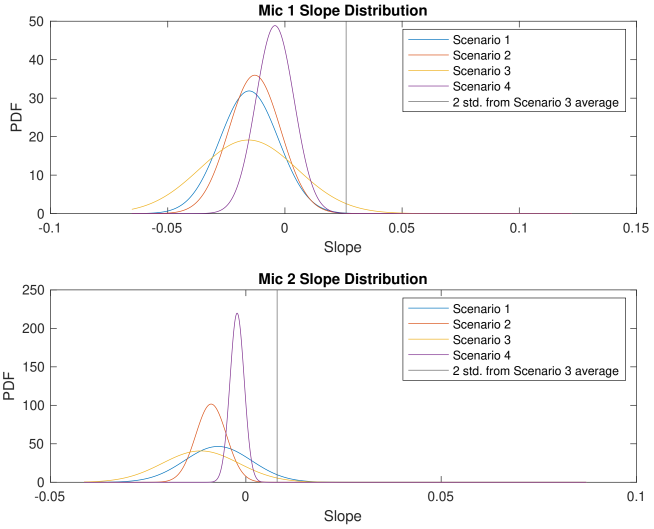


Figure 4.8: Average slope distribution for each scenario for 3 ms signal - Device 2

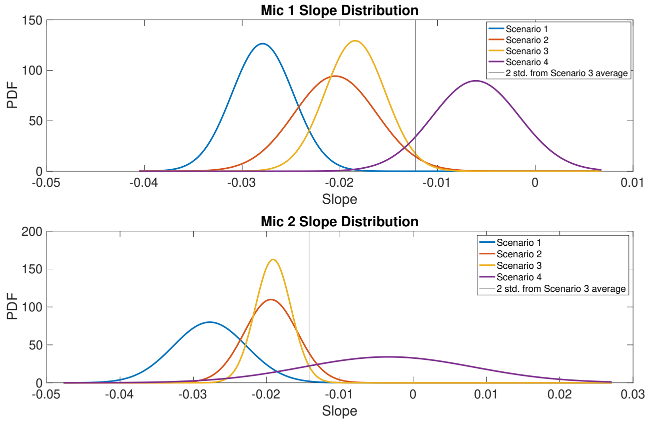
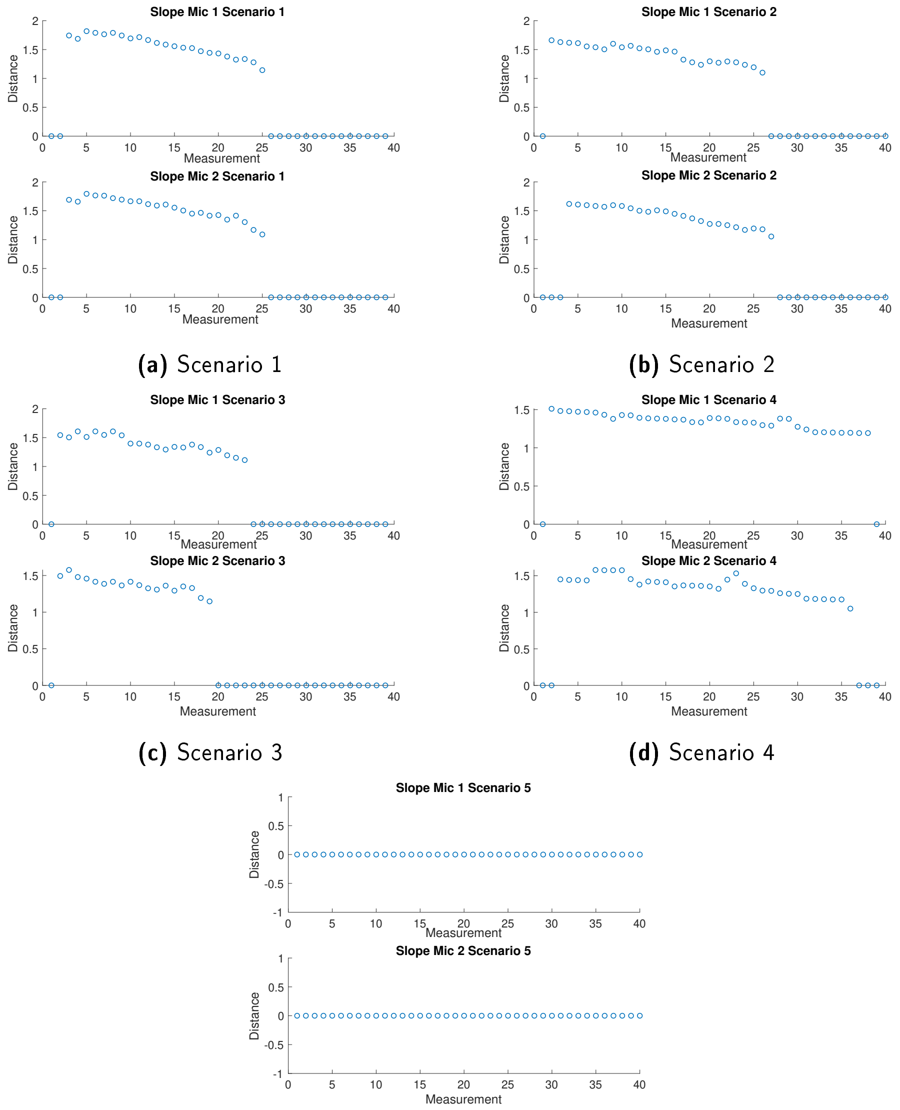


Figure 4.9: Average slope distribution for each scenario for 5 ms signal - Device 2



(a) Scenario 1

(b) Scenario 2

(c) Scenario 3

(d) Scenario 4

(e) Scenario 5

Figure 4.10: Motion Tracking for Device 2 with 5 ms sinusoidal chirp signal

Figures 4.10(a) to 4.10(e) show the typical motion tracking process within one trial for all the scenarios considered. In contrast to motion tracking in Device 1, for Device 2, the difference between the pattern of slopes for different scenarios can be observed. As mentioned before, this could be used in saving power, and reducing cost.

Chapter 5

Conclusion

All investigations and results presented in this thesis were made to determine if Device 1, built by Axis Communications, can use its speaker and microphones for motion detection. The goal of using its speaker and microphones in this way is to turn on the device only (i.e., the camera) when someone approaches the device with the intention to use it. If this can be done successfully, customers can save power and reduce cost.

It was found out that Device 1 could not distinguish between different types of motions, i.e., different scenarios, within its operating range. Thus, the device will always turn on, even when a person does not intend to use the device but is moving within the 2 m operating range of the device. Although this is the case, it was shown that the device would not turn on when there is no movement within 2 m from the device. This way, the algorithm created for this purpose will save power and cost, by not being kept on when there is no one within this range from the device.

Because the speaker and microphones are positioned close to each other in Device 1, more investigations were carried out using Device 2, where the speaker and microphones are positioned relatively far away from each other. This was to find out if the distance between speaker and microphones affects the accuracy of distance estimation and different scenarios.

It is shown that Device 2, like Device 1, will never turn on when there is no movement within the 2 m operating range from the device. On the other hand, it is also shown that Device 2 can differentiate Scenario 4 from other scenarios. This device will turn on only 24.5% of time when someone is walking in parallel with the device within the 2 m range, i.e., Scenario 4. This shows that the algorithm will save more power with this device and reduce cost even more than Device 1. It should also be noted that the scenario detection capability also goes beyond the simple motion detection functionality provided by its in-built PIR sensor. Also, it shows that a longer distance between speaker and microphones provides better accuracy for distance estimation and better results for motion tracking.

However, the proposed algorithm has some drawbacks as well. First, the system operates at the lower end of the audible range (1 - 5 kHz). This can make the algorithm quite annoying to those who can hear the sound signal, especially if it is running perpetually to detect motion. Second, all the measurements carried out in this project were made inside an acoustic anechoic chamber, i.e., under ideal conditions. There was no multi-path propagation of sound waves, and there were no other sound sources causing interference. Finally, when a person walks along the wall, the system cannot detect his motion. This is due to a limitation in the speaker directivity, i.e., the sound waves are weak in the endfire in that direction.

The three points above should be investigated further before the algorithm can be considered reliable enough for real usage. Working at higher sampling rates can aid in operating at frequencies close to the inaudible range. Thus, the algorithm can be implemented without being audible to the human ear. If the system is implemented under more realistic conditions, operating at higher frequencies can also act as an advantage since other sound sources usually have much lower frequencies. Additionally, bandpass filtering can also be performed to retrieve the required frequency range. Finally, if the system is implemented on different hardware, with speakers having a higher directivity (around 180°), the algorithm can also detect motion when a person walks towards the device along the wall.

Bibliography

[1] S. Gupta, D. Morris, S. Patel, and D. Tan. Soundwave: Using the Doppler Effect to Sense Gestures. *In Proc. SIGCHI Conference on Human Factors in Computing Systems*, page 1911–1914, 2012.

[2] J. D. Pye and W. R. Langbauer. Ultrasound and Infrasound. In Steven L. Hopp, Michael J. Owren, and Christopher S. Evans, editors, *Animal Acoustic Communication: Sound Analysis and Research Methods*, pages 221–250. Springer, Berlin, Heidelberg, 1998.

[3] K. Kang, W. Ouyang, H. Li, and X. Wang. Object Detection from Video Tubelets with Convolutional Neural Networks. In *Proc. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 817–825, Los Alamitos, CA, USA, jun 2016. IEEE Computer Society.

[4] X. Jin, S. Sarkar, A. Ray, S. Gupta, and T. Damarla. Target Detection and Classification using Seismic and PIR sensors. *IEEE Sensors Journal*, 12(6):1709–1718, 2012.

[5] A. Ekimov and J. M. Sabatier. Human Motion Analyses Using Footstep Ultrasound and Doppler Ultrasound. *The Journal of the Acoustical Society of America*, 123(6):EL149–EL154, May 2008.

[6] A. Elfes. Sonar-based Real-world Mapping and Navigation. *IEEE Journal on Robotics and Automation*, 3(3):249–265, 1987.

[7] T. Kohno R. Nandakumar, A. Takakuwa and S. Gollakota. Covertband: Activity Information Leakage Using Music. *In Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(3), September 2017.

[8] SONAR | Definition, Acronym, Uses, & Facts
`https://www.britannica.com/technology/sonar`.

[9] Introduction to SONAR,
`https://fas.org/man/dod-101/navy/docs/es310/uw_acous/uw_acous.htm`.

[10] E. J. Sullivan. Active Sonar. In David Havelock, Sonoko Kuwano, and Michael Vorländer, editors, *Handbook of Signal Processing in Acoustics*, pages 1737–1755. Springer New York, New York, NY, 2008.

[11] N. Kolev. *Sonar Systems.* BoD – Books on Demand, September 2011. Google-Books-ID: oAWaDwAAQBAJ.

[12] H. Peyvandi, B. Fazaeefar, and H. Amindavar. Determining Class of Underwater Vehicles in Passive Sonar using Hidden Markov Model with Hausdorff Similarity Measure. In *Proc. 1998 International Symposium on Underwater Technology*, pages 258–261, 1998.

[13] National Oceanic and Atmospheric Administration, US Department of Commerce. What is sonar?
`https://oceanservice.noaa.gov/facts/sonar.html`.

[14] B. H. Maranda. Passive Sonar. In David Havelock, Sonoko Kuwano, and Michael Vorländer, editors, *Handbook of Signal Processing in Acoustics*, pages 1757–1781. Springer New York, New York, NY, 2008.

[15] A.B. Baggeroer and B.M. Lucca. Sonar Systems. In J. Kirk Cochran, Henry J. Bokuniewicz, and Patricia L. Yager, editors, *Encyclopedia of Ocean Sciences (Third Edition)*, pages 319–327. Academic Press, Oxford, third edition edition, 2019.

[16] The Independent Advisory Group on Non-Ionising Radiation. Health Effects of Exposure to Ultrasound and Infrasound
`https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/335014/rce-14_for_web_with_security.pdf`, February 2010.

[17] W. Wei and S. Xinjian. Dynamic Fuzzy Clustering for Infrasound as a Precursor of Earthquakes. In *Proc. 2010 3rd International Congress on Image and Signal Processing*, volume 8, pages 3582–3586, 2010.

[18] J. Shimatani, H. Takahashi, M. Ichihara, T. Takahata, and I. Shimoyama. Monitoring Volcanic Activity with High Sensitive Infrasound Sensor using a Piezoresistive Cantilever. In *Proc. 2019 IEEE 32nd International Conference on Micro Electro Mechanical Systems (MEMS)*, pages 783–786, 2019.

[19] J. V. Wijayakulasooriya. Automatic Recognition of Elephant Infrasound Calls Using Formant Analysis and Hidden Markov Model. In *Proc. 2011 6th International Conference on Industrial and Information Systems*, pages 244–248, 2011.

[20] E.L. Owen. The origins of 60-Hz as a Power Frequency. *IEEE Industry Applications Magazine*, 3(6):8–14, 1997.

[21] ZOKA - Acoustic Torpedo Countermeasure Jammers and Decoys | ASELSAN
`https://www.aselsan.com.tr/en/capabilities/naval-systems/torpedo-and-torpedo-countermeasure-systems/zoka-acoustic-torpedo-countermeasure-jammers-and-decoys`.

[22] G. ter Haar. Ultrasonic imaging: Safety Considerations. *Interface Focus*, 1(4):686–697, August 2011.

[23] Acoustical Society of America Standards. Terminology Database - Sound,
`https://asastandards.org/{Terms}/sound-2/`.

[24] R. Mora, S. Penco, and L. Guastini. *The Effect of Sonar on Human Hearing*. IntechOpen, September 2011.

[25] How does sound going slower in water make it hard to talk to someone underwater?
https://wtamu.edu/{~}cbaird/sq/2013/11/12/how-does-sound-going-slower-in-water-make-it-hard-to-talk-to-someone-underwater/.

[26] 17.2 Speed of Sound | University Physics Volume 1 |
https://courses.lumenlearning.com/suny-osuniversityphysics/chapter/17-2-speed-of-sound/. Publisher: Lumen Candela.

[27] M. Long. *Architectural Acoustics*. Academic Press, 2nd edition, May 2014.

[28] M. Olfat and A. Heather. Clinical Assessment of Gait. In Kevin K. Chui, Milagros "Millee" Jorge, Sheng-Che Yen, and Michelle M. Lusardi, editors, *Orthotics and Prosthetics in Rehabilitation (Fourth Edition)*, pages 102–143. Elsevier, St. Louis (MO), Fourth edition, 2020.

[29] M Brodie. The Empirical Rule for Normal Distributions - Wolfram Demonstrations Project
http://demonstrations.wolfram.com/TheEmpiricalRuleForNormalDistributions/, January 2014.