Implementation of an Autonomous Impulse Response Measurement System

Abraham Martinez Ornelas

School of Electrical Engineering

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 15.7.2019

Thesis supervisor:

Prof. Ville Pulkki

Thesis advisor:

M.Sc. Ricardo Falcón Pérez

Aalto University
School of Electrical
Engineering

AALTO UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING

ABSTRACT OF THE
MASTER'S THESIS

Author: Abraham Martinez Ornelas

Title: Implementation of an Autonomous Impulse Response Measurement System

Date: 15.7.2019 Language: English Number of pages: 7+63

Department of Signal Processing and Acoustics

Professorship: Audio and Acoustics Technology (S-89)

Supervisor: Prof. Ville Pulkki

Advisor: M.Sc. Ricardo Falcón Pérez

Data collection is crucial for researchers, as it can provide important insights for describing phenomena. In acoustics, acoustic phenomena are characterized by Room Impulse Responses (RIRs) occurring when sound propagates in a room. Room impulse responses are needed in vast quantities for various reasons, including the prediction of acoustical parameters and the rendering of virtual acoustical spaces. Recently, mobile robots navigating within indoor spaces have become increasingly used to acquire information about its environment. However, little research has attempted to utilize robots for the collection of room acoustic data.

This thesis presents an adaptable automated system to measure room impulse responses in multi-room environments, using mobile and stationary measurement platforms. The system, known as Autonomous Impulse Response Measurement System (AIRMS), is divided into two stages: data collection and post-processing. To automate data collection, a mobile robotic platform was developed to perform acoustic measurements within a room. The robot was equipped with spatial microphones, multiple loudspeakers and an indoor localization system, which reported real time location of the robot. Additionally, stationary platforms were installed in specific locations inside and outside the room. The mobile and stationary platforms wirelessly communicated with one another to perform the acoustical tests systematically. Since a major requirement of the system is adaptability, researchers can define the elements of the system according to their needs, including the mounted equipment and the number of platforms. Post-processing included extraction of sine sweeps and the calculation of impulse responses. Extraction of the sine sweeps refers to the process of framing every acoustical test signal from the raw recordings. These signals are then processed to calculate the room impulse responses. The automatically collected information was complemented with manually produced data, which included rendering of a 3D model of the room, a panoramic picture. The performance of the system was tested under two conditions: a single-room and a multiroom setting. Room impulse responses were calculated for each of the test conditions, representing typical characteristics of the signals and showing the effects of proximity from sources and receivers, as well as the presence of boundaries. This prototype produces RIR measurements in a fast and reliable manner.

Although some shortcomings were noted in the compact loudspeakers used to produce the sine sweeps and the accuracy of the indoor localization system, the proposed autonomous measurement system yielded reasonable results. Future work could expand the amount of impulse response measurements in order to further refine the artificial intelligence algorithms.

Keywords: data collection, mobile robot navigation, room impulse response, room acoustics

Preface

Olorún oba tobi to

This thesis was developed in the Department of Signal Processing and Acoustics at Aalto University, founded by the Nordic Sound and Music Computing Network (NordicSMC). This work is also requirement for obtaining the master degree in Acoustics and Audio Technology.

I would like to express my appreciation to professor Ville Pulkki. I am thankful for giving me the opportunity to work in this unique project, for its guidance, observations, and patience. My gratitude also to Ricardo Falcon Perez for his ability to share any kind of knowledge, his time, attentive feedback and advice. Additionally, to Craig Rollo for enhancing the English language and contents of this work. Extended gratitude to the Aku-Lab staff for facilitating the material and conditions to perform the experiments.

In loving memory of Gilberto Martinez Chavez, the best Mexican hardware engineer. To my parents Araceli Ornelas Rosas and Jose Manuel Martinez Castillo for its infinite love, immense generosity, and strenuous support. To my family for being my main motivation: Hugo, Dali, Jovita, Graciela, Elvia, Fer, Teucro, Mari, and Panchito–π.

Also, I appreciate the support from my friends in Finland and beyond, especially the Acoustic Bants and Kisis Treeni Group.

Helsinki, 15.7.2019

Abraham Martinez Ornelas

Contents

Symbols and abbreviations

Symbols and Operators

$h_S(t)$	Response of the source
$h_{sys}(t)$	Response of the system
$h_R(t)$	Response of the receiver
A	Amplitude of the signal
ω_1	Start frequency from the sine sweep
ω_2	Stop frequency from the sine sweep
T_R	Exponential rate of frequency
N	Length of the sine sweep
τ_f	Time of flight
d	Distance from the target and reference node
c	Speed of electromagnetic waves
N_{item}	Number of items in the system
N_r	Number of mobile robot
N_{st}	Number of remote stations
N_s	Number of speakers
N_{cam}	Number of cameras
N_i	Number of microphones
N_{ch}	Number of channels
N_m	Number of measurements
N_l	Number of room impulse responses

Abbreviations

AIRMS	Autonomous Impulse Response Measurement System
API	Application Programming Interface
BLD	Back Left Down
BRU	Back Right Up
ESS	Exponential Sine Sweep
FCC	Federal Communication Commission
FLU	Front Left Up
FRD	Front Right Down
FTP	File Transfer Protocol
IMU	Initial Measurement Unit
IP	Internet Protocol
LTI	Linear Time-Invariant system
MDF	Medium-Density Fibreboard
MLS	Maximum Length Sequence
PLA	Polylactic Acid
PVC	Polyvinyl Chloride
TOA	Time of Arrival
TOF	Time of Flight
TWR	Two-Way Ranging
RIR	Room Impulse Response
RTLS	Real-Time Locating System
RTT	Round-Trip Time
TCP	Transmission Control Protocol
UWB	Ultra-Wideband

1 Introduction

Data acquisition from the physical world requires a considerable amount of effort, time, and economic resources [1]. Recently, indoor navigation using wheeled robots has attracted much attention as a method for gathering data in indoor environments, since these mobile robot platforms allow the collection of consistent and reliable data. Acquiring and accumulating information using this method is crucial in some fields of research, and the collected data could be used for various applications, including the prediction of parameters for artificial intelligence algorithms [2, 3, 4]. In this thesis, acoustic data is collected for predicting acoustic parameters, which would assist in improving acoustic conditions and satisfying auditory needs. This collected data is used to measure room impulse responses (RIRs). An impulse response defines the aspects of a Linear Time-Invariant (LTI) system [5]. Rooms are commonly modeled as an LTI system [6]. Therefore, a room impulse response describes the acoustic phenomena in a room at a specific point. Room impulse responses can be created manually or synthetically. For instance, manual methods use sine sweeps [7, 8], while synthetic methods employ geometrical [9, 10], wave-based [11, 12, 13], and hybrid models [14].

Currently acoustic parameters are predicted based on manually or synthetically generated datasets. These datasets compile room impulse responses, which define acoustic phenomena from a room, such as reverberation time and early decay time [15, 16, 17, 18]. The majority of these datasets are manually gathered under varying and different conditions. These datasets are typically restricted in terms of measurement locations and also include insufficient information related to the experimentation room (i.e., wall materials and arrangement). Therefore, using the information from these datasets would not produce fully accurate results. Although various databases include room impulse responses, none of these have attempted to automate the gathering of these signals from several locations in a room. One solution to this problem would be to use a wheeled robot in order to measure room impulse responses.

Numerous indoor mobile navigation robots have recently been developed for various purposes, including surrounding recognition [19], obstacle avoidance [20, 21], sensor implementation [22, 23], algorithm development [24, 25], and educational prototyping [26, 27]. Although a great number of these robots have implemented devices that incorporated built-in camera microphone arrays [28, 29, 30, 31], few studies have attempted to collect and process audio signals [30, 31]. Moreover, both of these studies failed to provide sufficient details on documentation to enable further development.

This thesis aims to design, implement, and test an adaptable automated system for the playback and recording of sine sweeps (acoustic test signals) in various locations of multiroom environments, which is called as the Autonomous Impulse Response Measurement System (AIRMS). In order to achieve this objective, this thesis develops a system based on a mobile robot platform and remote stations for automatic collection. This collected information is composed by manual and automatic collected data. From this data structure, the sine sweeps are extracted and processed to provide in automated room impulse response measurements.

This thesis is organized as follows. Chapter 2 reviews relevant concepts that AIRMS is based on. Chapter 3 describes the structure of the collected data, the automatic data collection, specially the development of the mobile robot platform, and the data post-processing. Chapter 4 explains the methodology and setups of the experiment tests. Chapter 5 presents and discusses the results of the measurement tests. Finally, Chapter 6 concludes the thesis.

2 Background Research

This chapter details important concepts that are relevant to this thesis. Firstly, the impulse response is introduced, with methods for measuring the room impulse response. Following this, mobile robots are discussed and how they are relevant to data collection. The positioning and localization of the robotic system implemented is detailed after. This section concludes with a brief overview on how the measurement platforms communicated wirelessly.

2.1 Impulse Response

Sound propagation in an enclosed space is a complex phenomenon [32]. Sound interacts with boundaries of a room and furniture, occurring several acoustic events including absorption, reflection, scattering, refraction and diffraction [33]. These phenomena can be described, illustrated, and quantified physically and mathematically at an observation point [34]. The room impulse response (RIR) characterizes acoustic events at an observation point captured by a microphone [35]. The next section explains the impulse response and relevant factors that involve its measurement: the excitation signal, the source, and the receiver.

2.1.1 Measurement of the Impulse Response

Ideally, a room can be modelled as a Linear Time-Invariant (LTI) system [6]. A LTI system can be described by feeding the system with an input signal $x(t)$ and comparing this signal with the output signal $y(t)$, where t is time in seconds [5]. Figure 1 shows the block diagram of an impulse response measurement. A source (loudspeaker), with response $h_S(t)$, feeds the system with the input signal. This signal is modified in the system, which is then measured at an observation point, $h_{sys}(t)$. At the observation point, a receiver (microphone), $h_R(t)$, captures the resultant signal. The representations of the excitation signal from the source and the receiver are denoted by $x'(t)$ and $y'(t)$ respectively [36].

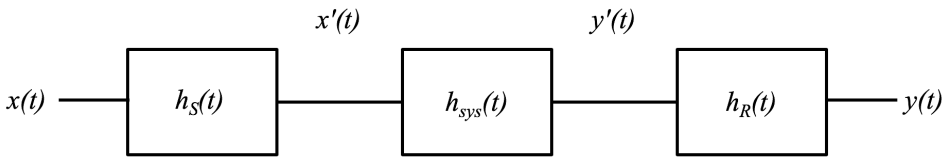

Figure 1: Impulse response measurement block diagram. Adapted from [36]

As a result, the output signal $y(t)$ results in the convolution of the input signal $x(t)$, the response of the source, the acoustic system and the response of the receiver [36].

$$y(t) = x(t) \otimes h_S(t) \otimes h_{sys}(t) \otimes h_R(t). \tag{1}$$

When the excitation signal is an impulsive sound, this signal decays in the enclosure as a function of time as shown in Figure 2 [34]. It can be seen that three parts can be identified form the impulse response: the direct sound, early reflections and the reverberation [37]. In theory, the direct sound fully resembles the input signal with time-delayed inferences within the first milliseconds [35]. The early reflections correspond to the sound interactions occurring during 50 [38] to 80 [39] milliseconds right after the direct sound. The reverberation describes the signal decay after a significant number of reflections.

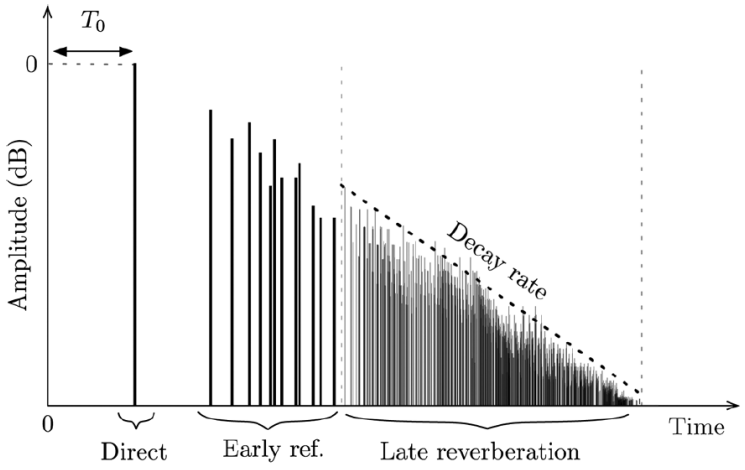

Figure 2: Parts of an impulse response. Adopted from [37]

2.1.2 Exponential Sine Sweep

Several sources have emulated an ideal impulsive sound, such as pistols, balloons, firecrackers, clap machines, interrupted noise or music. Since an ideal impulsive sound should contain sufficient energy at all frequencies of interest, electroacoustic devices are implemented, commonly known as continuous signal methods. The Exponential Sine Sweep (ESS) emerges from the time delay spectrometry method [40, 41, 42] to produce impulse responses, implemented for measuring room acoustics [34, 7, 8]. The ESS method utilizes sinusoidal signals varying at different frequencies over time. The exponential sine sweep is defined as:

$$x\left[n\right] = A \sin\left(\omega_1 T_R \left[\exp\left\{\frac{n}{T_R}\right\} - 1\right]\right), \tag{2}$$

where n are the samples, A is the gain of the sine sweep, ω_1 the start frequency of the sine sweep in rad/sec, and T_R is the exponential rate of frequency defined as:

$$T_R = \frac{N}{\ln\left(\frac{\omega_2}{\omega_1}\right)} \tag{3}$$

where N is the length of the sweep in samples, and ω_2 the stop frequency of the sine sweep in rad/sec.

The estimation of the impulse response uses the inverse ESS excitation signal convolved with the recorded ESS at an observation point.

Among the other continuous signal methods, such as the Maximum Length Sequence (MLS) and Linear Sine Sweep, the ESS has proven the most optimal for the estimation of room impulse responses. Comparing the ESS to MLS, the swept-sine mitigates harmonic distortion from the loudspeaker and time variances [43, 44, 6]. The ESS allows to isolate and remove nonlinear harmonic distortion, finding these distortion components separately [6, 8]. Because the sweeps vary at different frequencies over time, the signals contain no repetitions. Therefore, group delays increment that signifies a high tolerance for time variances [45]. Moreover, comparing the ESS to LSS, exponential sine sweeps carry higher energy at low frequencies [44]. Furthermore, the ESS signal decreases 3 dB per octave, whereas linear sweeps are flat, in other words, ESS preserve more energy at low frequencies compared to LSS.

2.1.3 Loudspeaker

A loudspeaker is a device that transforms electrical energy into acoustic energy [46]. Generally, the term loudspeaker refers to the two parts, driver and enclosure, which this device is composed of [47]. Both parts contribute to radiate acoustical energy to the air in certain direction and with a specific response in frequency, directivity and frequency response, respectively [47].

Ideally, the measurement of RIR requires a loudspeaker which radiates its energy with enough power, a flat frequency response, and an omnidirectional pattern [34]. However, this conditions are difficult to accomplish due to inherit properties of the loudspeakers.

Moreover, the usage of loudspeakers allows the reproduction of impulsive input signals as the ESS [48]. Additionally, deconvolution can be applied to avoid the loudspeaker distortions, by implementing the inverse response of the loudspeaker. However, this inversion is limited to a certain frequency range, if the signal-to-noise ratio is sufficient [44]. Ultimately, it is aimed to use a robust source with repeatability and reproducibility, in other words, constantly perform experiments under same conditions with minimal variations.

2.1.4 First-Order Microphone Array

Microphones transform acoustical to electrical energy [46]. A microphone captures sound signals within a range of frequencies, called frequency response. It is desired that a good microphone counts with a flat response at all the audible frequency range, from 20 Hz to 20 kHz, and produce the minimum of noises or distortions [48]. Another property on microphones is the polar pattern. The polar or directional pattern of a microphone describes the frequency response and sensitivity of a device; examples of polar patterns include cardioid, omnidirectional and figure-of-eight [49]. A microphone array consist of the positioning of multiple microphones that can capture spatial acoustic information [50].

First-order microphone array, commonly known as Ambisonics B-format, is a tetrahedral transducer array that counts with four cardioid diaphragm capsules [51]. The four capsule arrangement represents, at the centre of the array, an omnidirectional and figure-of-eight patterns [52]. Consequently, the microphone records four channels W, X, Y, and Z as shown in Figure 3 b). B-Format consists in a virtual configuration: channel W as omnidirectional, denominated as the pressure, while X, Y and Z as figure-of-eight, denominated as pressure-gradient. Therefore, the B-format represents the acoustic pressure and particle velocity in the three-dimensional space at an observation point [53]. A-format refers to the output signals of the microphone capsules without any further processing [51].

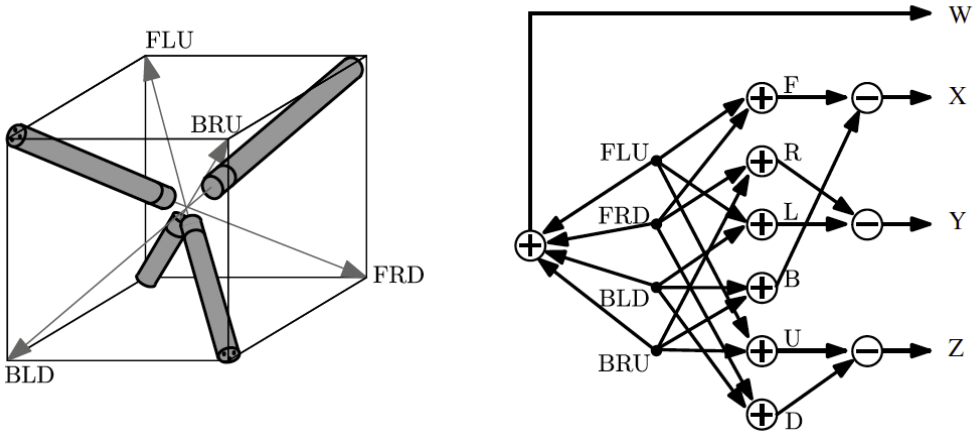

Figure 3: a) Four capsule array placement. b) Encoding from channels W,X,Y,and Z from the First order microphone. Adopted from [53]

Each of the capsules point in different directions and pairs on the horizontal plane [51]. The capsules are identified as Figure 3 a) shows, Front Left Up (FLU), Front Right Down (FRD), Back Left Down (BLD) and Back Right Up (BRU). The four capsules of the tetrahedron must be identical; therefore, an equal frequency and phase response compensation can be implemented. In theory, first order microphone arrays aim an omnidirectional pattern to capture a diffuse field response. However, as illustrated in Figure 4, Channel W is only omnidirectional at low frequencies. The

polar pattern becomes more directional at higher frequencies, due to spatial aliasing [54].

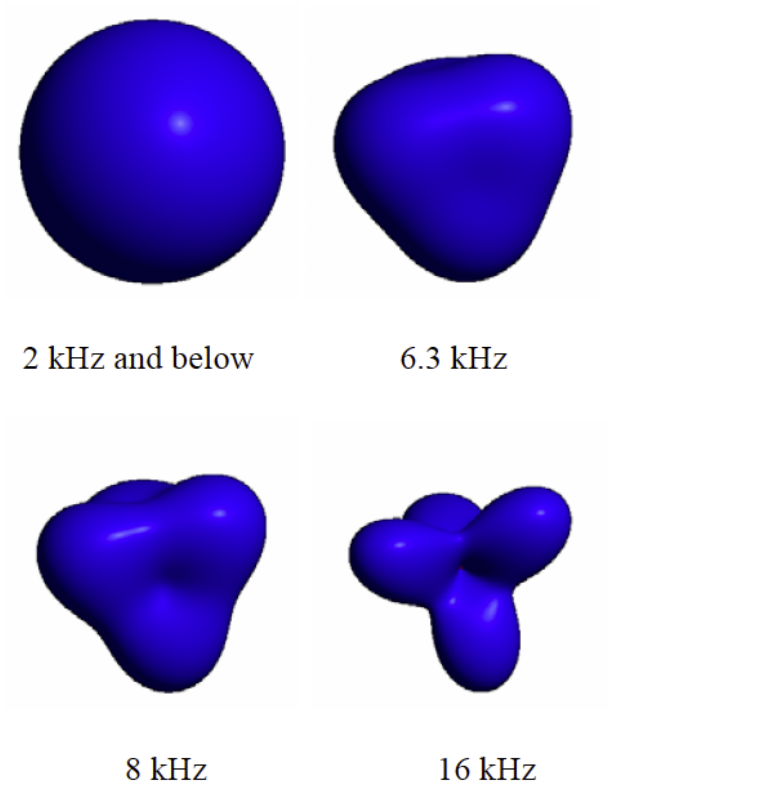

Figure 4: First order microphone polar patterns on Channel W at 2 kHz, 6.3 kHz, 8 kHz, and 16 kHz. Adopted from [52]

Furthermore, outputs from X, Y, and Z vary their frequency response on low and high frequencies. Therefore, to reduce these issues, it is recommended that the manufacturer properly aligns the capsules [52].

Microphone arrays have been used for measuring room impulse responses and sound field analysis for spatial acoustic parameters [55, 56, 57], reverberance [55], early room reflections [58], directionality [59, 60], and ancient theatre acoustical parameters [61]. Recently, the implementation of microphone arrays has expanded through several applications, including virtual reality and 360-degree audio internet broadcast [53].

2.2 Mobile Robotic Navigation

The term robot originates from the Czech word 'robota', meaning forced work [62]. Robotics apply knowledge of areas, such as mechanics, electronics and computer science. A robot performs specific tasks for reducing human effort or risks with repeatability and precision [62]. In some cases, these actions are executed under complicated situations or environments [63]. In these days, the implementation of

robots has expanded from the food service industry to home appliances. Autonomous robots follow preprogrammed actions and their decision-making is based on the situation and environment [64]. Furthermore, the degree of automation varies from simple to complex tasks, for instance, from stopping the robot's motion to avoid obstacles to map a building layout on 3D. Autonomous mobile robots applications include self-driving cars, warehouse robots and delivery aerial drones as Figure 5 illustrates on a), b) and c) respectively. Mainly, mobile robot navigation utilizes a mechanic platform that allows the device to move across areas without human intervention, orderly executing programmed or guided actions [63].

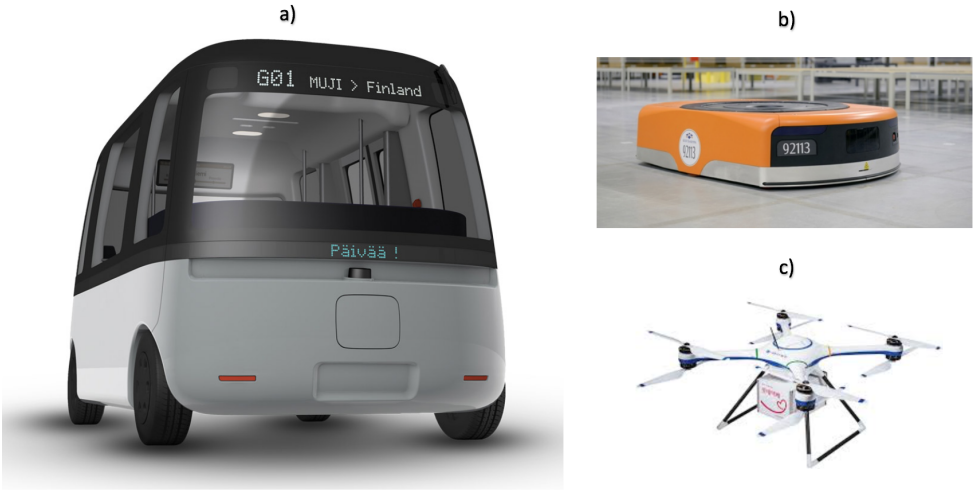

Figure 5: Examples of autonomous mobile robots a) Self-driving vehicle (Sensible 4). Adapted from [65], b) Warehouse vehicle (Amazon Robotics). Adopted from [66], c) Delivery aerial drone (Uconsystem). Adapted from [67]

On robotic navigation, the primary interrogatives formulated during the design phase cover three main questions: Where am I?, Where should I go? and How to get there? [68]. These questions relate to the location, path-planning, and trajectory of the robot [69]. In specific for indoor environments, the development of robot wheeled navigation originated during the 1950s, facilitated due to the controlled conditions on interiors. However, technology rapidly advanced two decades later when NASA explored the outer space using wheeled navigation rovers: Spirit and Opportunity [63].

2.2.1 Indoor Mobile Robots

Advances in technology have permitted to develop indoor navigation systems through history. In the 1950s, G. Walter built a mechanical 'tortoise' capable of avoiding obstacles helped with a photoelectric sensor [70]. Introduction of digital computers in 1969, allowed researchers from the Stanford University Artificial Intelligence Centre to develop 'Shakey the robot' [71]. A mobile platform for data collection with enhanced cognition and path planning [64]. In 1984, the Hilare-I employed laser, ultrasonic

sensors and cameras for navigation. The Hilare-I reported its position and orientation, enhancing control motion and path planning from this device [72]. Two decades later, in 2002, indoor robot navigation found a niche on home appliances. Wheeled mobile robots were developed as vacuum cleaners, such as the iRobot. Figure 6 presents the previously mentioned indoor robot platforms a)Mechanical tortoise, b)Shakey the robot, c)The Hilare-I, d)Roomba S series.

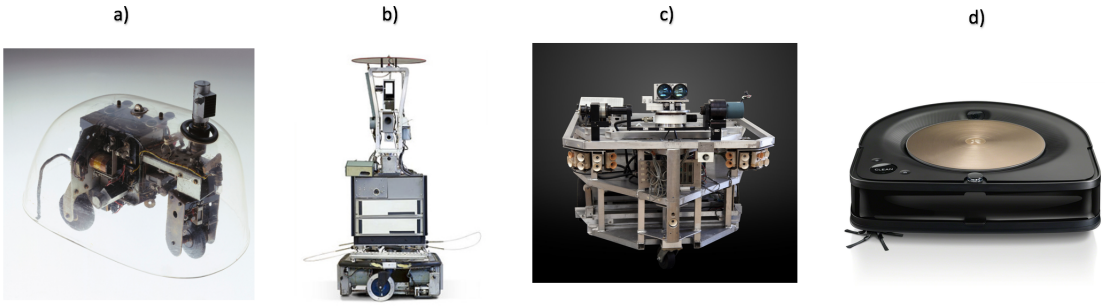

Figure 6: Examples of indoor mobile robots a) Mechanical tortoise. Adopted from [73], b) Shakey the robot. Adopted from [71], c) The Hilare-I. Adopted from [74], d) Roomba S series. Adopted from [75]

Current indoor mobile navigation has focused on activities, such as perception, exploration, mapping, localization, path planning and path execution [76]. These activities, on autonomous systems, are carried without or with minimum human-intervention and are adaptable to different scenarios [76]. Mobile robots consist of four elements: mechanical platform, control, cognition and sensors [77]. The mechanical platform relates to the machinery that allows the movement from one place to another. The control relates to the management of the system to stay on task and the execution of the commanded actions. The cognition is the perception of the sensed information and decision-making of further actions. Sensors employ apparatuses that acquire the environment features, perception and reply to the cognition [77]. Although these elements have reached a high sophistication level, such as cognition and sensors are constantly upgrading their capabilities [78]. In the case of sensors, several types have been implemented like vision, laser and odometry; however, the abstraction of the environments arises non-trivial problems, for instance, the position and orientation of robot platforms and object identification. In recent years, enhanced recognition and obstacle identification and avoidance [79] have been accomplished, utilizing vision [75] and sensor fusion, for example, the visual Simultaneous Localization and Mapping (vSLAM) [80].

2.2.2 Related Works

MICBOTS MICBOTS are robot platforms targeted for automatic speech recognition and acquisition of room impulse responses [30]. A MICBOT platform is composed of a wheeled mobile robot, a single-board computer, a WIFI dongle, two cameras with built-in four-microphone linear array, and a speaker [81], as Figure 7

illustrates. The robot navigates through an area playing and recording utterances from existing speech databases. Additionally, the impulse responses are collected by delimiting a room into small cells. For example, it is projected that MICBOTS would be able to capture 1.3×10^6 room impulse responses in a $12\,\mathrm{m}^2$ room. Although MICBOTS have been tested in cocktail-party scenarios reproducing speech utterances, the documentation about the acquisition of room impulse responses lacks from details. Furthermore, questions and challenges concerned to the measurement procedure remain pending. The main aspects to be implemented include the localization of the robot, Simultaneous Localization and Mapping (SLAM) complemented with laser sensors, and noise reduction of the robot. Some pending questions relate to the measurement setups, position between the robot platforms, indoor robot location, and defining robot movements during the utterances.

Figure 7: MICBOTS, robot platforms targeted for speech recognition and collection of RIR. Adopted from [81]

The CREATE Mobile Robot CREATE multimodal database employs a mobile robot for acquiring visual, kinesthetic, and auditory data for environment representations. The usage of this data is meant for unsupervised learning, multimodal prediction, self-robot and scenery movement detection [82, 31]. The robot platform consisted of a wheeled mobile robot, an embedded computer, an Initial Measurement Unit (IMU), and adapter, and a webcam with built-in two-stereo microphone array. The robot, seen in Figure 8, collected pictures, optical flow fields, IMU measurements, odometry, atmospheric pressure, air pressure, and audio recordings from indoor spaces of a school facility.

During the experimentation, the audio tracks contained the motor noises and chassis vibrations of the robot, the wheel frictions with the floor, and collision noises of the robot with obstacles. Additionally, in some experiments, speech from people was present in the recordings. The analysis of the audio signals included the time

domain plotting and spectrograms of audio excerpts. Furthermore, a user commanded the movements of the robot with a joystick, and the localization of the robot was tracked based on the IMU complemented with odometry of the wheeled robot.

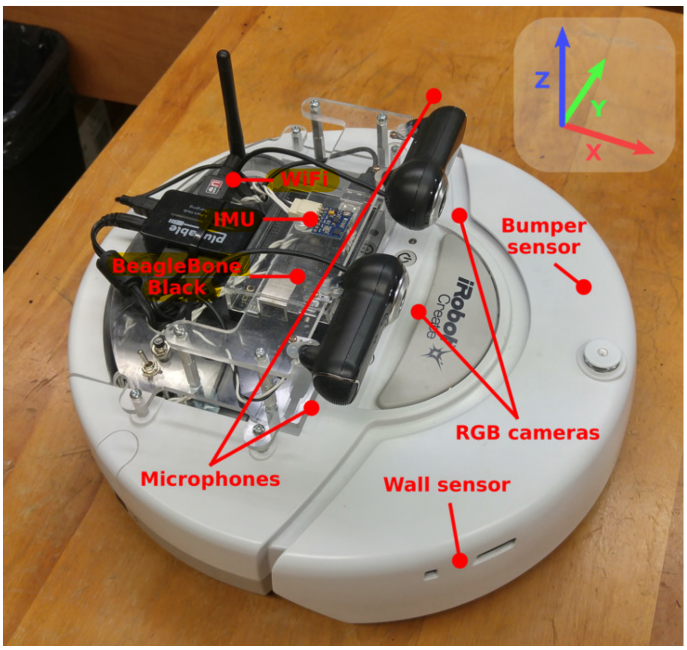

Figure 8: CREATE mobile robot targeted for multimodal collection, including stereo audio recording. Adopted from [31]

2.3 Indoor Positioning Systems

Indoor positioning systems permits to track an object within a delimited area at an interval of time. Recently, these systems are commonly employed in diverse areas, for instance, to track expensive objects in warehouses or locate museum visitors. It is expected that these systems demonstrate more accuracy compared to outdoor locating systems. Thus, locating indoor targets remain within a known area, which movements and speeds could be restricted. Advantages of indoor environments include the acknowledgment of factors inside the buildings, such as the layout, temperature, humidity and air circulation; however, signal interactions with building elements and objects might translate into issues for the system. The interactions include reflections, multipath, and signal delays. Therefore, the stability of the system varies in terms of power and signal scattering. The objects in the room affect the signal strength, according to the material composition of the objects, insulators that allow passing the signals and conductors that scatter the signals. A classification of indoor positioning systems suggests to organize them according to the medium used for determining the position, mediums such as infrared, ultra-sound, radio-frequencies, magnetic, vision, and audible [83]. Ultra-wideband is comprised into the radio-frequency group.

2.3.1 Ultra-Wideband Localization

Ultra-Wide Band is a radio technology based on extremely short pulses implementing techniques to broaden its radio energy. Compared to other technologies, UWB supports non-line-of-sight, and the communication is established based on large bandwidth and signal modulation. Due to the large UWB bandwidth, there is no signal interference at broadband or narrowband systems [84]. Compared to these systems, the UWB relative bandwidths are 20% larger, and their absolute bandwidth exceeds 50 MHz; thus, high-speed communication is established [85]. The time modulation on the signal, or known as pulse-position, permits to encode the properties of the signal, such as amplitude and polarity. Because of their short time-resolution and wavelength, the signal presents optimal power levels; therefore, multipath interference is avoided. The bandwidth of operation of the signals is restricted by recommendation of The Federal Communication Commission (FCC)[86]. For the European Union, the frequency mask must operate within 3.1 – 10.6 GHz range, to avoid interference with other radio devices.

2.3.2 Positioning Algorithm

Time of Arrival Positioning is estimated based on the time of travelling between radio signals. These radio signals are produced by two devices of the system, target nodes and reference nodes, where the reference node location is known in advance [87]. Furthermore, the position estimation between these elements relies on ranged-based, angle-based, and proximity-based estimation [88]. Time of Arrival (TOA) is a technique utilized in UWB that offer several advantages under complex environments that is shown in Figure 9. TOA overcomes problems such as obstacle avoidance and multipath fixing. Since TOA considers the first signal to arrive for its estimation, problems on the first signal, including noise, time-delayed signals, multipath components, and clock drift can exist. However, its advantages include its cost, accuracy, and hardware simplicity.

Additionally, the position location are based on the delay of the propagated signal, known as the time of flight (TOF) [88]. $\tau_f = d/c$, where d is the actual distance between the target and reference node, while c stands for the electromagnetic wave speed $c \approx 3 \times 10^8$ m/s. It is important to note that TOA synchronize the reference node signal's while the target can remain asynchronized. Since the system records the departure and arrival time of the signal, any delays on the signal using TOA produce errors. For instance, an error of 3 ns on the TOF approximately causes a distance deviation of 1000 mm [88]. Conversely, asynchronous systems use their own clock to measure the TOF, which translates into less usage of memory. Therefore, the distance calculation between two points follows as, the reference node times the signal when it completes a round trip, in order to estimate the propagation time, and then this quantity is multiplied by the TOF. In other words, the time of a round-trip by the speed of light. Therefore, TOF is a crucial factor in measuring precisely [89].

Two-Way Ranging Two-Way Ranging (TWR) is a position algorithm that implements an asynchronous system. The distance estimation supports on the Round-Trip

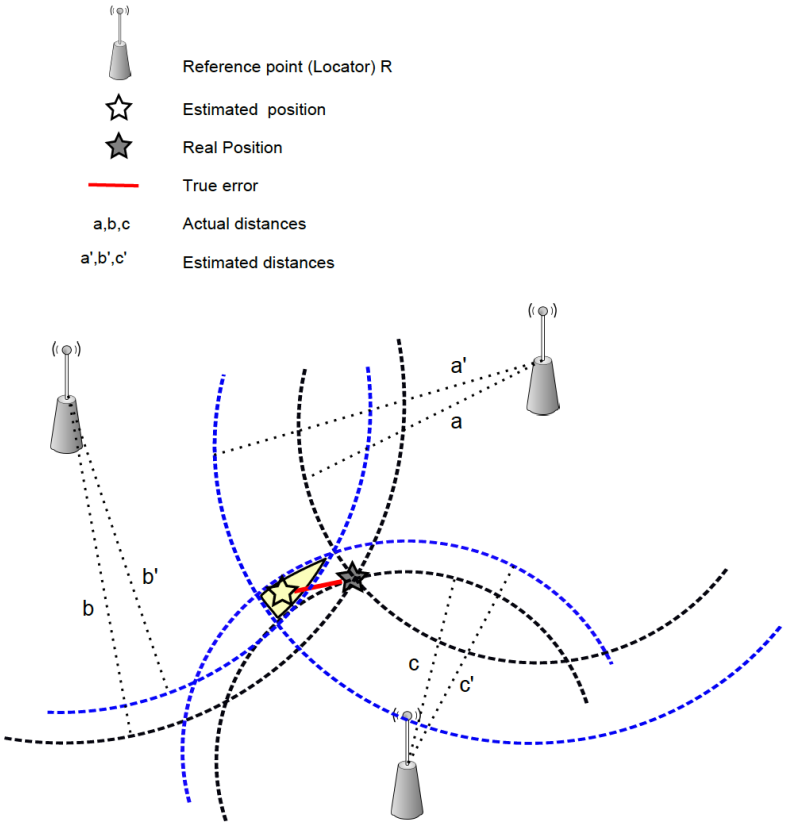

Figure 9: Time of Arrival diagram. Adopted from [87]

Time (RTT)[88]. Device A sends information to device B, once acknowledge device B sends back messages to device A at a certain propagation time, in other words, TWR calculates the distance between two elements by using the information from two packages, as Figure 10 illustrates. The RTT at device A is expressed as:

$$\mathrm{RTT} = 2\tau_f + \tau_d, \tag{4}$$

where τ_d is the known distance between elements [88].

Although TWR needs no synchronization, ranging errors are produced due to the clock drift. These errors might be caused for the cheap WSNs built-in the reference nodes. Low-cost oscillators can produce frequency drifts, where errors from the order of 1 ns translate into centimetres of inaccuracy. In general, for UWB its tiny wavelengths allow penetrating objects and partitions. However, some materials permit the pass of the signals better than others. Therefore, the materials are classified into two groups, insulators and conductors. Insulators are materials that, by their composition, can be easily trespassed, such as cardboard, plastic, fabrics, wood, glass, bricks, whereas conductors would cause diminution or scattering or time-delays to the signals.

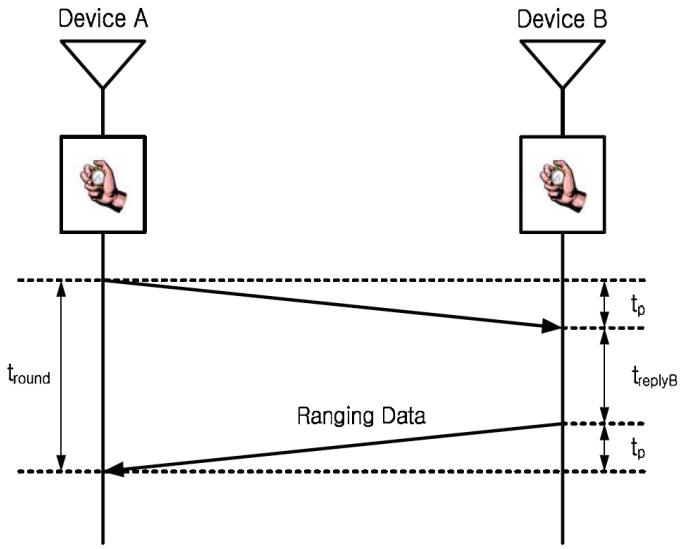

Figure 10: Two-way ranging communication. Adopted from [89]

2.4 Wireless Communications

2.4.1 Internet Protocol Suite

The Transmission Control Protocol (TCP) and the Internet Protocol (IP) are conceptual models that convey communication protocols. These protocols consolidate specifications for end-to-end data communication, including the package, address, transmission, route, and receive [90]. The Internet Protocol Suite model is divided into four categories that rank the protocols according to their level of networking. Figure 11 illustrates these categories, divided in the Network Access Layer, Internetwork Layer, Transport Layer and Application Layer [91].

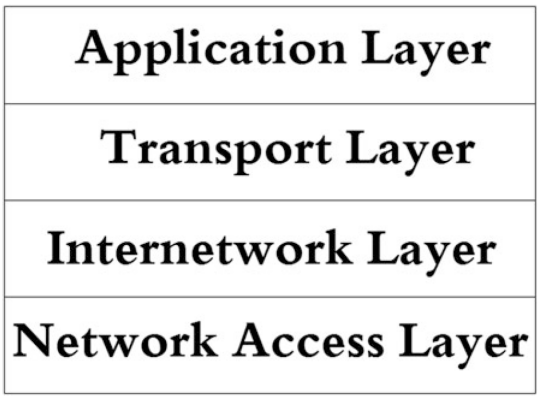

Figure 11: Transmission Control Protocol model layers. Adopted from [91].

2.4.2 Network Sockets

Network sockets are implemented to transfer data inside a computer network. Nodes, internal or external endpoints, are established to send and receive data [92]. For example, TCP/IP utilizes internal sockets as a one-to-one connection, such as server-client sockets [93], in this circumstance, the sockets are created and logged into a network, where IP addresses and port numbers are required to complete a local connection. This distributed structure divides the actions in two parts, servers and clients. Servers are the suppliers that contain a directory of actions, whereas clients demand the performance of these actions [94]. Network sockets based on client-server architecture follow an specific structure, including bind, listen, accept, connect, send and receive, and close, shown in Figure 12. Furthermore, during the data exchanging, additional functions can be performed such as read, write, open and close. In principle, the server socket waits to perform requests that the client socket demands. Thus, the server establishes an address so that the client can find the server; this process is known as bind. Once the bind is set, the server waits for requests from the client. The socket is created when the client demands an action from the server, data exchange. Finally, the server performs the demanded action and sends back a reply to the client [95].

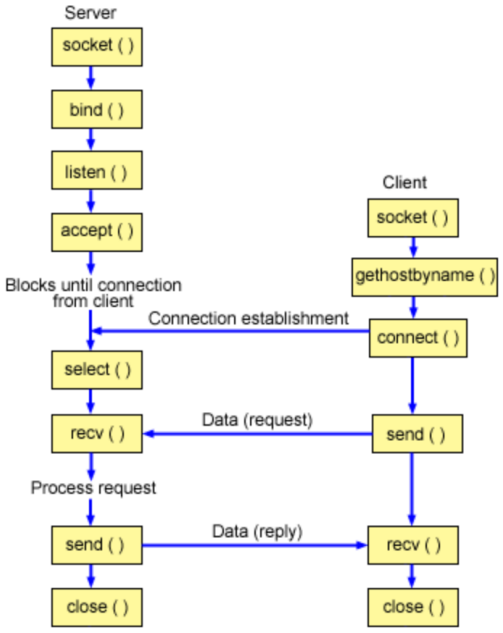

Figure 12: Description of the process on client-server network sockets. Adopted from [95].

The steps included in the socket API includes:

1. Socket. The one-to-one connection is established for communication. Additionally, a description of the socket is sent back, which specifies the endpoint.
2. Bind. The client receives the description of the socket. The server assigns a

specific name and direction to the socket. Moreover, these addresses can be accessible to the rest of the network.

3. Listen. The port accepts client requests. First, the confirmation must be received after accepting any requests from the client.
4. Connect. The client and the server establish a connection. Therefore, the socket is created.
5. Accept. The server allows the data exchange from a client. Then, the client can request the server for actions.
6. Functions. The client and server can manage the data transferred, including send, receive, read, write.
7. Close. The connection between client and server is stopped.

Sockets can be implemented with protocol stacks in an Application Programming Interface (API) called socket API.

3 Case of Study

Once understood the principles in which the equipment employed in AIRMS is based in, it is important to describe the development of AIRMS and its relevant features. Following to the introduction about the explanation of how a system can be conformed, the data structure is explained to present fields of data, mode, amount, and format. Later, the automatic data collection is described, which includes the development of a mobile robot platform. The mobile robot development is presented as its components are described: mechanical platform, cognition and control, sensors, as well as the equipment. Finally, the data post-processing is detailed, which includes the framing of the signals and the measurement of the room impulse responses.

The main design consideration of AIRMS is to constitute a highly adaptable system for researcher needs. Based on these needs, the investigator can select a desired number of items to conform the system. These items are classified into three groups, platform, equipment and results. Platform is the vehicle or structure capable to perform and accumulate acoustical information from the spaces under study. Platforms can be mobile in case of the robot platform or stationary in case of a remote station. Equipment are the apparatus that rig the platform. These devices acquire relevant information for the system, such as location, reproduction, and recording of acoustical signals. Results are the outcomes of the experimentation process and the amount produced is directly proportional to the selected platform and equipment. N_{item}, where item is the respective element of the system, such as robots N_r, remote stations N_{st}, speakers N_s, cameras N_{cam}, and microphones N_i. Furthermore, number of channels N_{ch} are an outcome inherent from the microphone. For example, first-order spherical microphone array has four channels. The outcomes of the system, number of measurements N_m, where a measurement consists of the exponential sine sweep recorded played by a speaker at a location, room impulse responses N_l are the calculated signals based on the recorded sweeps. Table 1 lists the items, their denomination and type that compose AIRMS.

Denomination	Item	Type
N_r	Mobile robots	Platform
N_{st}	Remote stations	Platform
N_s	Speakers	Equipment
N_i	Microphones	Equipment
N_{cam}	Cameras	Equipment
N_m	Number of measurements	Result
N_{ch}	Number of channels	Result
N_l	Room impulse responses	Result

Table 1: Main elements of the Autonomous Impulse Response Measurement System

The principal products of AIRMS are the recorded sine sweeps and the room impulse responses at a point in a room or place. To maintain light computational costs, AIRMS may be divided into two stages: data collection and post-processing. Data collection implements automated mobile robots and/or remote stations to

gather acoustical signals. While the data collection is real-time process, the post-processing is a non-real-time. The post-processing includes the sweep framing and the calculation of impulse responses. The framing of the sweeps is the extraction of the Exponential Sine-Sweep (ESS) signals, see Section 2.1.2, from the experiment recordings. Based on the ESS signals, the calculation of impulse responses are performed. Furthermore, the obtained data is arranged into an structure.

3.1 Data Structure

The information collected in the experiment was compiled using two methods, manual and automatic. In the manual method, the user generates the data, whereas, in the automatic method, the mobile robot gathers the information. The manual information includes a panoramic picture, a 3D model of the room under study, the raw recordings, and the metadata concerned to the room, for instance, building materials and layout of the room. The automatic data includes the position, orientation and a log of movements performed by the robot, the temperature of the room, pictures from the webcams, the timestamps of the sweeps, and metadata from the microphones and speakers, position and orientation, for example, speaker and microphone type, and the framed sine sweeps of the measurements.

The information compiled from the rooms under study is presented in Table 2, describing the name of the field, the mode of compilation —manual (M) or automatic (A), the quantity of the field, the format of the field, and additional comments.

Field	Mode	Quantity	Format	Comments
Panoramic picture	M	1	2D matrix	
3D-model	M	1	3D matrix	
Room metadata	M	1	TXT	ID/Wall mat./Layout
Raw recordings	M	N_i	WAV	4 channel recordings
Measurements	A	N_m	Object	
Robot	A	N_r	-	
Position	A	1	Vector	X, Y, Z
Orientation	A	1	Vector	Roll, yaw, pitch
Movements	A	1	Vector	Robot sensors
Temperature	A	1	Scalar	Celsius
Cameras	A	N_{cam}	2D matrix	images
Speakers	A	N_s	-	
Mic and speaker metadata	A	1	TXT	Type/Pos./Orien.
Framed sine-sweeps	A	N_i	array	
Room impulse responses	A	N_l	Vector	RIRs each channel

Table 2: Structure from the manual and automatic data of a room

*matrix, timesteps x nl

3.2 Automatic Data Collection

Data collection consists in the gathering of the measurement data using the mobile robot platform and/or the remote stations. For this thesis, the mobile platform integrates a wheeled vacuum cleaner, the tag of an indoor localization system, two portable loudspeakers, a microphone recorder, a webcam and a single-board computer, as it is shown in Figure 13. Remote stations are composed by a single-board computer, a portable loudspeaker, and a microphone recorder.

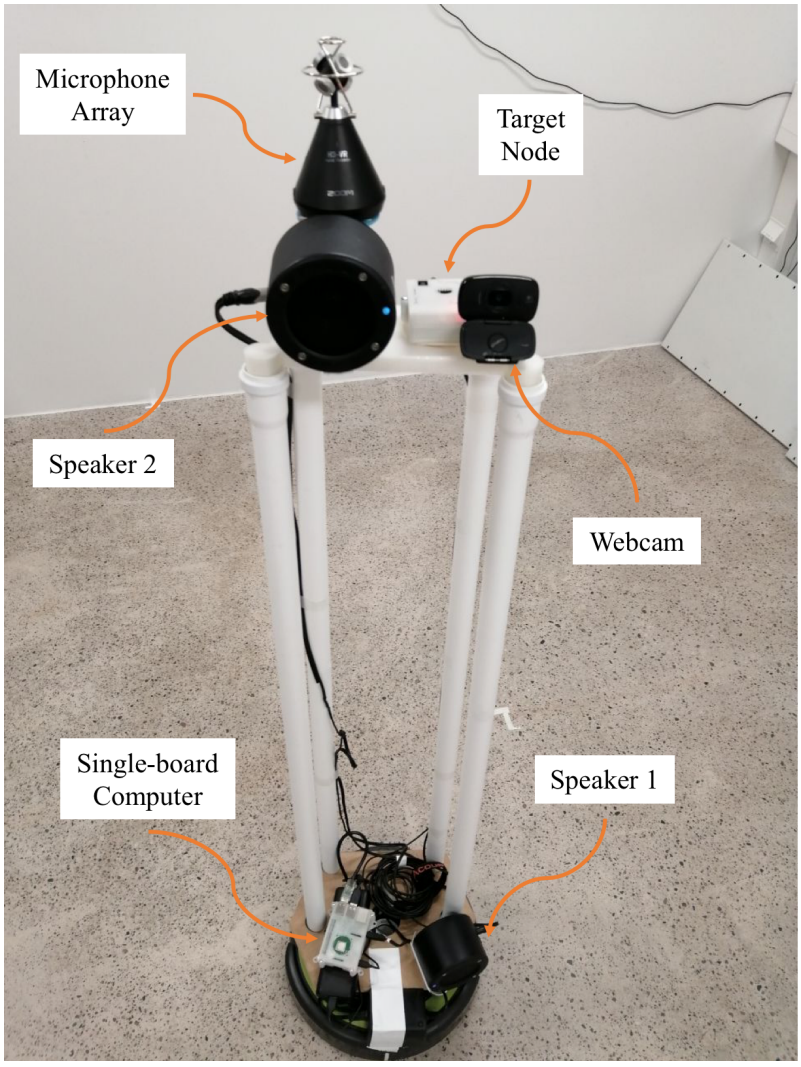

Figure 13: Mobile robot platform, equipped with a Raspberry Pi, a positioning tag, two speakers, a microphone, and a webcam

The following section, the platforms and their items concerned to the data collection stage are explained.

3.2.1 Wheeled Robot

The mobile robot chosen was the commercially available Roomba Create®2 as the manufacturer released a version that can be customized [96].

This robot offers several advantages including chassis customization, open interface, actuators control, rechargeable battery, and low cost. The dimensions of the robot are 93 mm in height, 349 mm external circumference and 3.5 kilograms in weight. The chassis of the robot be modified and is capable to carry a maximum 9 kilograms weight. This consideration is important because the robot has a payload when moving across the room.

Additionally, this robot includes a serial port for receiving commands and interfacing with microprocessors or computers. The robot's actuators bases on one-byte opcodes, called open interface [97]. This open interface has been adapted to several programming languages, including Python [98]. Due to the numerous libraries, Python was chosen for command actions in this prototype. Since the velocity can be controlled, it is chosen that the velocity should be slow, therefore, the robot brings two advantages, generate less noise when it is moving and keep a better balance when transporting its rigged equipment.

Built-in Robot Sensors Using the robot built-in sensors [97], this robot enhances its navigation to recognize the surroundings. Activating these sensors permit to raise flags, that once acknowledged, following actions can be performed [99]. The sensors used during these experiments were bumper, infrared, and thermometer.

The bumper consists of a moving frontal part that when pushed activates. Once the frontal part hits an obstacle, the moving circumference is pushed, and the sensor raises a flag. The bumper structure is armed with an infrared sensor array across half of the robot's circumference. This array recognizes walls or objects in proximity. Additionally, a couple of infrared sensors are located at the bottom body to identify cliffs. The thermometer locates in the battery pack.

Mobility It is important to note that due to the limitation of the built-in sensors in the robot, the area in which the robot navigates must be delimited, as the navigated area must be confined during the experiments, as Section 4.1.1 explains. The robot navigates in arbitrary directions sorting obstacles and walls across a room, commonly referred as pseudo-random walk. The main parameters used for mobility are velocity, actuators and duration. To meet the pseudo-random walk, the velocity of the movements is randomized and obstacles and walls are avoided.

Velocity is the speed that actuators perform. The range of velocity of Roomba is between -500 to 500 mm/s. However, the type of ground varies the speed of the robot, depending on the surface it should be easier or difficult to navigate. By observation during the experiments, it was found that the optimal low velocity of displacement is 150 mm/s for bare floors, whereas for carpet flooring the velocity should be increased due to the effort that it takes the Roomba to move. For example, on carpet flooring the speed increases to 250 mm/s.

Actuators are the type of movements that the robot is capable to execute. The experiment routine in this prototype only uses five actuators: drive forward, drive backwards, spin right, spin left, and stop. The pseudo-random walk routine is implemented, utilizing a combination of these actuators repeated in every experiment. Moreover, when an obstacle or a wall intrudes or is detected into the robot's path, the robot sorts the obstacles executing two additional movements.

To randomize the routines, the duration of the movements performed in every experiment varies. Duration is related to the time lapse in seconds of movements. The duration of the movements relies on the velocity of the robot. For example, driving forwards or backwards at the reference velocity of 150 mm/s for 9 seconds, the Roomba moves 1 meter. Similarly, at the same velocity, spinning left or right for 3 seconds, the Roomba turns 360 degrees corresponding to the indicated direction. For the purpose of this prototype, this parameter is randomized from 1 to 8 seconds.

Construction of the Mounted Structure on the Robot A robust and light structure is required to support and transport the equipment. The materials employed for the construction of the structure were wood, Polylactic acid (PLA) filament, and Polyvinyl chloride (PVC) tubes. In the next section, the structure mounted on the robot is explained from the bottom part to the top. The mounted structure is composed by two parts, a bottom level and an top level. The bottom level surface is a Medium-density fibreboard (MDF) table and the top level is a PLA 3D printed surface held by PVC tubes, as Figure 14 shows.

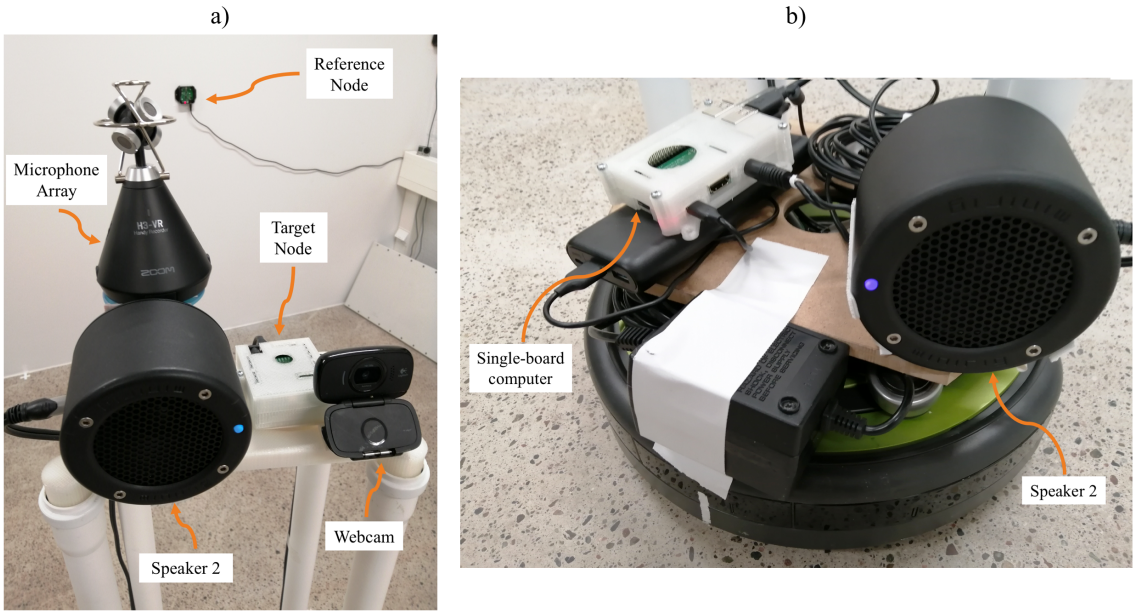

Figure 14: a)Robot mobile top platform, b)Robot mobile bottom platform

The bottom level couples with the robot attaching the external structure to the external plastic layer, as Figure 15 depicts. Moreover, the top surface adapts a

pedestal to raise and support the microphone recorder, to reduce the interactions with the speaker.

Figure 15: Bottom platform ensembled to the external plastic layer of the Roomba

Equipment Placement The placement of the equipment mounted on the robot platform is shown in Figure 16. The dimensions are expressed in millimetres, considering the mid-body of the wheeled robot projected to the ground as the reference. The distances from this reference to a mounted device refer to the central part of the body of each device.

On the top level, the microphone located 90 mm left from the center, at 1550 mm height of the ground. The portable loudspeaker, Speaker 1, located 90 mm left from the center, 1454 mm height of the ground. The tag placed at 40 mm right from the center, 1416 mm height of the ground. On the bottom level, the portable loudspeaker, Speaker 2, placed 100 mm right from the center, tiled 5 degrees from the normal plane, 190 mm height of the ground.

Since the robot navigates in a pseudo-random walk, it should report its location constantly. The position reported in AIRMS' robot platform refers to the place where the tag is mounted on the robot, therefore, from this position, the location of the rigged equipment can be calculated, for example, the location of the Speaker 1, Speaker 2.

Mounted Equipment Considerations Considerations should be applied when selecting the equipment mounted on the robot. Therefore, the devices should be light to optimize the robot mobility.

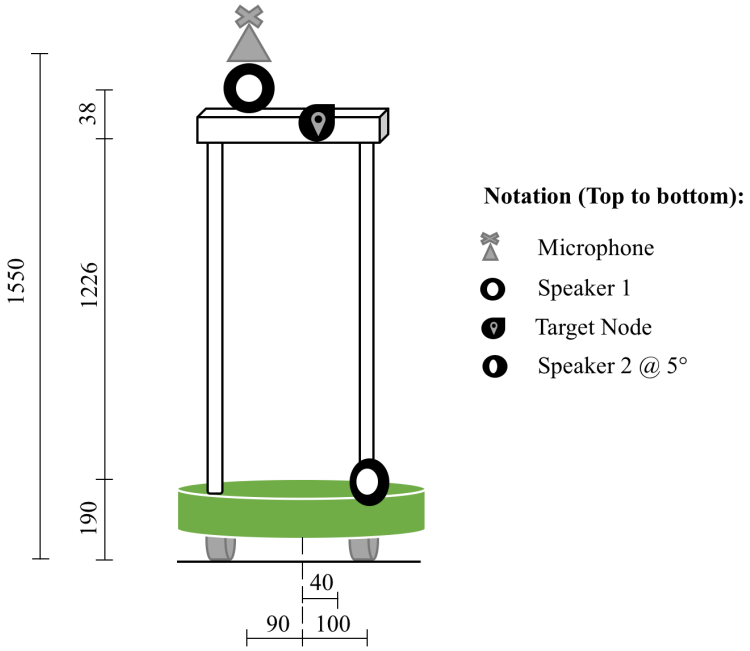

Figure 16: Schematics of the robot platform and equipment with dimensions in millimeters, considering the mid-body of the robot projected on the ground as the reference. The distances refer from this reference to the mid-body of each of the devices.

Light loads on the robot result on enhanced performance such as battery endurance, effortless displacement, wheels and machinery noise reduction and mounted structure stability. Weight loads that Roomba can carry are unspecified. However, 9 kilograms is considered as the maximum load.

Additionally the speed at which the robot operates increases the noise. Increasing the speed of the robot generate more wheel traction on the floor and the machinery increments. Therefore, light payloads would reduce noise. It is also important the distribute the payloads according to the wheel axels. Since the robot has two parallel wheels and one omni wheel, the center of gravity remains on the two-wheel axels. Hence, maintaining the balance between these three wheels is recommended.

3.2.2 Single-Board Computer

The Raspberry Pi 3 Model B+ is a single-board computer based on Linux that has 1.4 GHz processor, Ethernet, WIFI and USB units, as well as video and audio outputs [100]. This device was chosen as the computer to interface, command, control devices of the AIRMS and storage the collected information. The Raspberry Pi interfaced, controlled and commanded actions including the robot mobility, the indoor localization system and graphical processing of the measurement locations, storage of the information collected automatically. The stored information and

system accessing was performed remotely via FTP, including the execution of the measurement scripts and file transferring. Additionally, the Raspberry Pi audio output facilitated the connection to the loudspeakers mounted on the robot and the playback on the remote platforms via network sockets.

Playback in the Mobile Robot Platform The playback on the robot platform speakers was performed using wires. A stereo 3.5 mm audio cable connected the output of the Raspberry Pi to each of the two portable loudspeakers, (i.e., the stereo cord connected the left channel to the Speaker 1, and right channel to the Speaker 2). A mono test signal was sent to a single channel at sequential times; therefore, the signal reproduced on different speakers at different times.

Playback in the Remote Stations The playback on the remote platform speakers was performed wirelessly by using network sockets. The network sockets implemented the server-client configuration, as Section 2.4.2 describes. The computer mounted on the robot worked as the client while the remote stations functioned as the servers. A local list, hosted by the servers, included commands of actions to be performed. When a command was called, the server executed the action, and it sent back a confirmation message to the client. The number of remote stations defines the number of sockets linked. The client communicated at a single time with each of the servers, when multiple servers were communicating, this communication was established only with one server at a single time, pausing the rest of servers.

3.2.3 Indoor Localization System

An indoor localization system should be installed in the room under study to track the position and orientation of the robot. This prototype implemented an off-the-shelf solution for indoor localization. An Ultra-wideband (UWB) indoor localization system called Pozyx was chosen. Despite the fact that Pozyx is a ready-made-solution, its manufacturer released a version of this system for developers, which allows modifications of the system, through an application programming interface (API). Pozyx is a Real-Time Locating System (RTLS) that tracks the position and orientation of an object inside a room based on

Pozyx employs two hardware elements to estimate indoor localization: tags and anchors, target nodes and reference nodes respectively, as depicted in Figure 17.

As seen in Section 3.2.3, the location of the tag is recognized by the anchors, the tag establishes double way communication with each of the anchors placed on the walls. Moreover, the minimum number of anchors is three, the hardware could be scaled from this minimum requirements. Anchors can be added to increase the area of coverage until 70 m^2, and tags until five devices [101]. Pozyx software provides an API that consists of functions as autocalibration, real-time tracking, and location algorithm.

Accuracy The margin of error that Pozyx documents is 100 mm [102]. For this thesis, localization tests were performed, as A details, in the same environment as

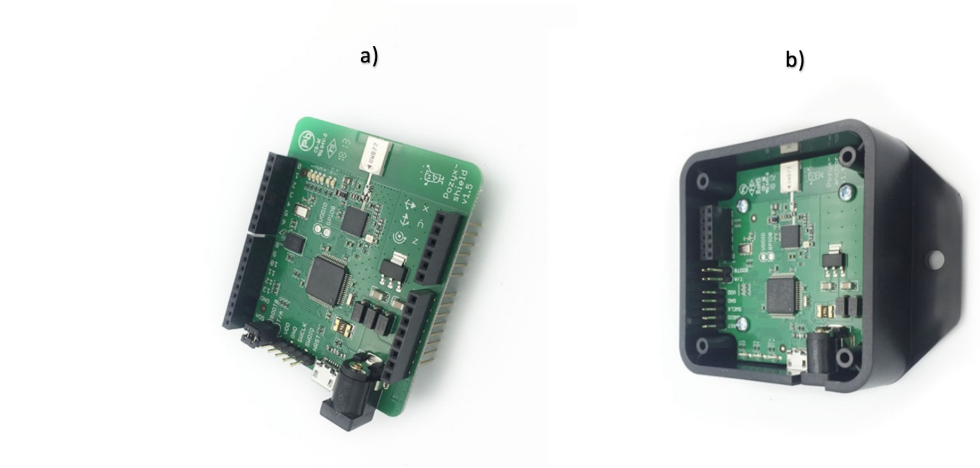

Figure 17: Positioning system hardware a) Tag (Target node), b) Anchor (Reference node). Adopted from [101]

the acoustical tests. Preliminary results point that one of the axis is estimated more accurately than the other. The margin of error tended to range from 80 mm to 400 mm compared to the reference position. It is possible that this deviation would be caused due to metal-based structures or furniture presented in the room, for instance, a white board, a sheet metal cable cover, air conditioning pipes.

Other localization systems were considered, such as ultrasound and trackers from virtual reality headsets. Ultrasound was not chosen as the shorter wavelength would not able to penetrate the partitions in the room. Trackers were dismissed due to issues with the creation of reference planes [103].

3.2.4 Loudspeakers

Minirig is a neodymium based portable loudspeaker, as Figure 18 shows. Minirig was chosen due to its portability, which allows the speaker to be placed and pointed on any level and positioned at any direction. Furthermore, portable features resulted convenient for the tests, such as power amplification (15 Watts RMS), weight (0.5 Kilograms), and lasting self-power battery.

The frequency response of the Minirig ranges from 70 to 16000 Hertz [104].

3.2.5 Microphone

This prototype aims to record acoustical signals at all directions. Therefore, a first order microphone, detailed information is found in Section 2.1.4. The Zoom H3-VR 360 was chosen for its compact size, recording resolution, battery and storage, this microphone is shown in Figure 19.

The recordings were done with a sampling rate of 96 kHz with a bit depth of 24 bits. For convenience, the recordings were captured in A-format, conversions can be applied if needed. A-format records the incoming sounds on the capsules, see Section 2.1.4. During the testings, another approach was tested for the recording, a

Figure 18: Minirig, portable neodymium speaker used during the experimentation. Adopted from [104]

Figure 19: Zoom H3-VR 360, microphone used during the measurement experiments. Adopted from [105]

fourth-order spherical microphone array was employed. However, this device was unsuitable for mobility. This microphone array must be wired to an interface or a field recorder to record the acoustical tests, which is unpractical for mobility.

3.2.6 Webcam

Two webcams were tested on the mobile robot to capture pictures at the measurement locations, a standard webcam and a GoPro. The aim was to reinforce the localization by capturing these pictures.

The webcam pictures depicted a narrow field of view, thus, some of the pictures resulted undescriptive as the measurement position was hardly identified.

Another approach was to continuously record video with the use of a GoPro. Post-processing was applied to extract snapshots at the locations based on the measurement timestamps. This approach was dismissed as it resulted in a computationally heavy process.

3.3 Data Post-Processing

The post-processing refers to the procedure for framing the sine sweep signals, calculating the room impulse responses of every measurement positions, and the compilation of the gathered data.

3.3.1 Framing the Sine Sweeps

The raw recordings contain the sine sweeps and background noises manifested during the test, for example, the robot machinery noises. Therefore, it is crucial to extract the sine sweeps from the noises. The identification and extraction of the sine sweeps is referred as the framing process. The framing process employs the manual setting when the first sine sweep occurs. This reference is called calibration point and it is selected over the X-axis on the raw recording track.

Setting a Calibration Point By visualizing the entire recording track, a calibration point is selected at the start point of the first sine sweep onto the X-axis (time). Since the duration of the sweep is known, it sums to the calibration point for framing the first sweep. Therefore, this point is marked as the end point of the first sweep.

For the second sine sweep, the measurement timestamp converted into a date vector. The date vector represented the start point on the X-axis of the second sweep. Thus, this point is turned into the new start point.

This process is iterated summing the start points to the length of the sweeps in all four channels, and the subsequent timestamps are referenced as new start points. Moreover, to ensure that the sweeps are appropriately framed, a pad of 0.5 seconds added to the start and end points of the framed sweeps.

Figure 20 illustrates the definition of the calibration points, the length of the sweep, the ending point and pads of a framed sine sweep, and the timestamp of the subsequent sine sweep.

Finally, a list compiles the start and end points of the sine sweeps and saves the audio signals between these limits. This information is presented along with the acquired by the robot, (i.e., position, orientation, movements, temperature).

During the experiments, it was noticed that the calibration point could be selected automatically by using the timestamps. However, drawbacks initiating the playback

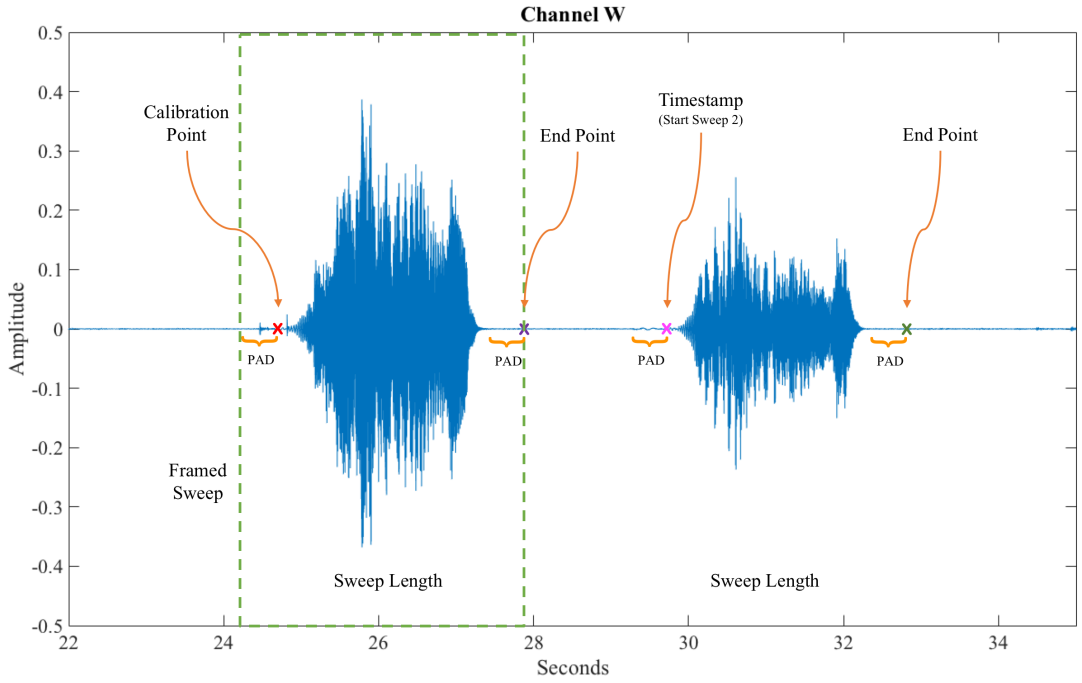

Figure 20: Framing of a sine sweep signal, defining start and end points.

of first the sine sweep were noted. These drawbacks were interruptions or clippings on the sine sweep playback. Therefore, it was opted for the manual setting of the calibration point.

3.3.2 Calculating the Room Impulse Responses

Following to the framing process, the impulse responses are calculated. The calculation of the room impulse responses are based on the sine sweep signals, as suggested [7, 44, 8]. The original unprocessed reference sweep is inverted and then convolved with a recorded sine sweep. The result is a room impulse response at that location in the room. This process is repeated for all recordings in the measurement.

4 Implementation

AIRMS aims to acquire a better understanding about acoustical interactions. In this manner, it is desired to not only study the acoustical phenomena inside a enclosure, also between contiguous spaces. Therefore, multiroom environments are aimed to be study by using the AIRMS. Once the components of AIRMS have been explained, this chapter describes the setups in which the prototype was implemented. The first setup concern to the study of the acoustics inside a room, whereas the second setup includes the study of the room and its contiguous space.

4.1 Environment

AIRMS aims to measure interactions inside and between unoccupied rooms in multiroom environments. Therefore, AIRMS implements mobile and stationary platforms able to reproduce and record acoustical measurements in diverse areas. Mobile robot platforms navigate in various parts inside the room, meanwhile, remote stations, placed in fixed locations, capture phenomenon in specific areas, allowing interactions between platforms. An example of a multiroom building layout is shown in Figure 21. To strengthen the understanding of acoustical behaviour, such as sound propagation, amount of absorption, complementary data is gathered.

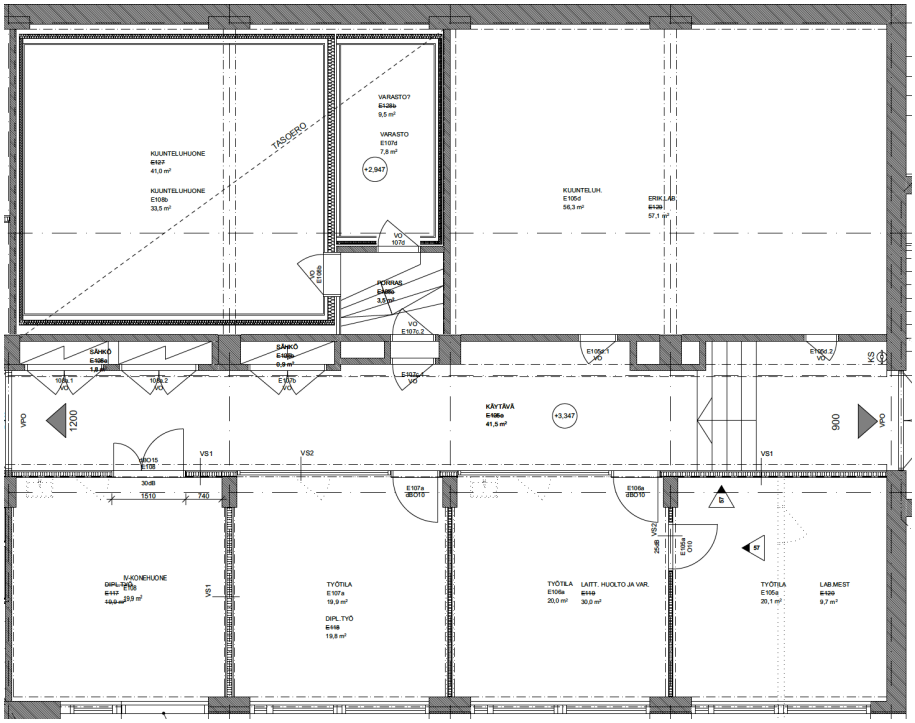

Figure 21: Typical multiroom environment, rooms studied during the testing process.

The measurements are performed in a multiroom arrangement in unoccupied and confined surroundings due to the robot built-in sensor limitations. These sensors were unable to perceive boundaries above the body of the robot. The

rooms were unoccupied reducing the robot's physical interactions only with objects. Therefore, the furniture and appliances were stocked in the room and then confined in restricted areas. The restricted areas were limited with wooden boards placed vertically on the floor, which permit the robot sensors perceive these boundaries. The rooms were comprised of common furniture items, such as tables, chairs, televisions. Due to the limitations on the indoor localization system, the area of the rooms for experimentation should size below 70 m^2. The indoor localization system was installed in the room, as Section A.2.1 details.

4.1.1 Room for Experimentation

The room utilized for the test was a 23.5 m^2 meeting room in Aalto University. The room was located in Otakaari 5 wing D.

Figure 22 presents the room with a panoramic view. It can be seen the materials of the wall, the arrangement of its elements, such as windows and doors and the placement of the furniture before the tests.

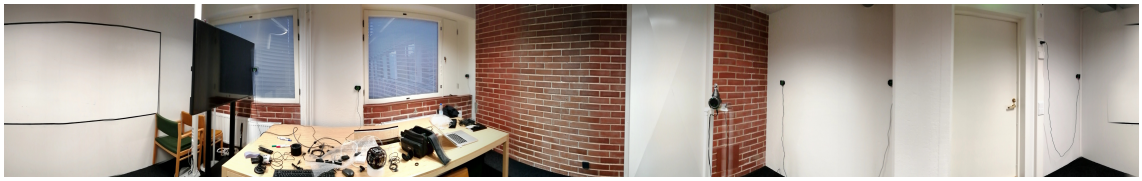

Figure 22: A panoramic picture of the room under study.

As seen from Figure 22, the walls of the room were arranged as it follows. The far-right wall possessed two windows separated by a concrete foundation. The front wall was brick and a white metal sheet which stretched from the floor to the ceiling. The far-left wall held a door, and a concrete foundation in the middle. The back wall possessed a whiteboard, the hard floor was covered with a carpet, and air conditioning pipes and perforated gypsum panels hung from the ceiling. The total dimension of the room was 4935 x 4808 x 3000 mm.

A 3D model of the room was also developed. This representation modelled the geometry of the room, as well as the openings in the room. The restricted area of the room was highlighted in the model. Figure 23 displays a three-axis perspective of the 3D model. The model is included in the manually collected data for visualization and testing purposes.

Figure 24 presents a top-down view perspective of the 3D modelled room and the restricted area is indicated.

4.1.2 Metadata of the Room

The name of the field referred as room metadata includes relevant information and description from the room. This field describes the name of the room and location, the restricted area, the area occupied by furniture, the volume of the room, and additional comments about the room arrangement.

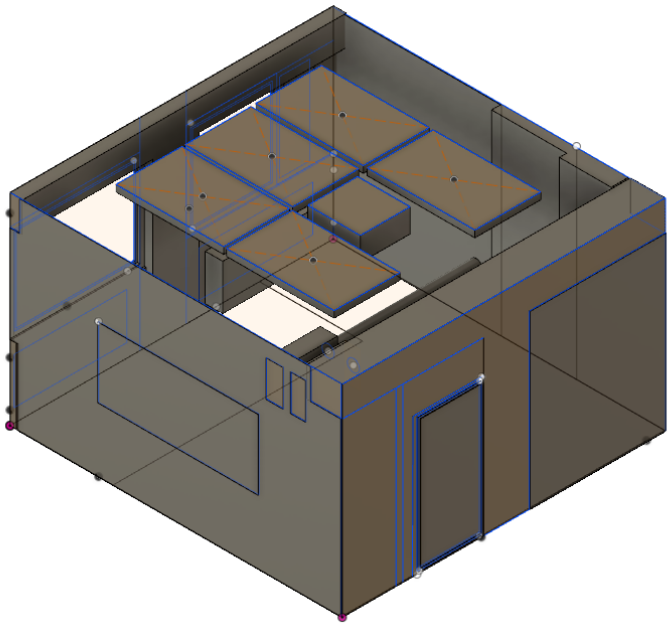

Figure 23: 3D room model, three axis perspective

4.2 Test Setups

A setup is defined based on the researcher needs, the environment, as well as the arrangement and number of items composing AIRMS. As previously described in Table 1, the items that compose AIRMS must be defined. Items including the number of mobile robots (N_r), remote stations (N_{st}), speakers (N_s), microphones (N_i), cameras (N_{cam}), measurements (N_m), channels (N_{ch}), and room impulse responses (N_l). For this thesis, two setups were implemented during the testing of the prototype. In the next sections, the location of the elements are given with Cartesian coordinates, where points (X, Y) are expressed in millimetres.

4.2.1 Single Closed Room (Setup 1)

The single closed room test, known as Setup 1, involved the acoustical characterization of previously described room. The platforms implemented during the setup were one mobile robot and one remote station, as Figure 25 shows. The aim was to understand acoustical interactions inside the room, particularly around the door. Table 3 contains the listing of elements in Setup 1. Note that the item dimensions of the robot, remote platforms, speakers and microphones are not to scale.

The mobile robot was equipped with two loudspeakers (Speaker 1 and Speaker 2), one microphone, that recorded four tracks, and one webcam. Additionally, Setup 1 employed a remote station that was 600 mm in front of the door, at the coordinates (3238, 4318) mm. The remote station used a Raspberry Pi 3, Speaker 3, and a microphone at 1500 mm high, the distance between Speaker 3 and the microphone was 300 mm. The stand supporting the remote station was covered with acoustical

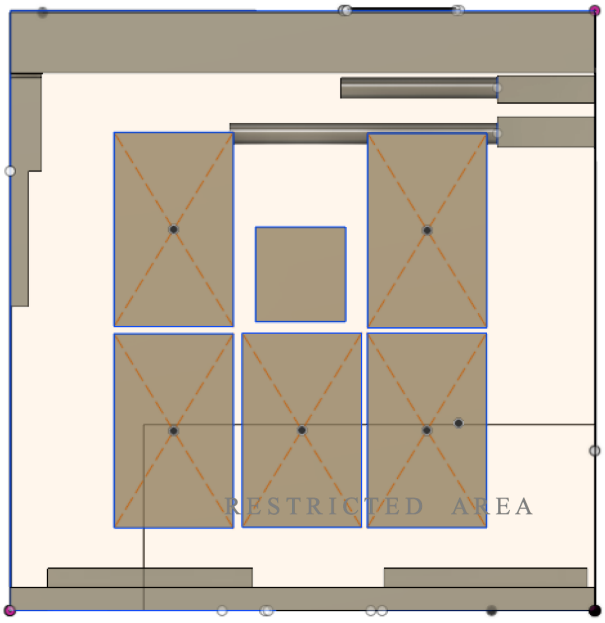

Figure 24: 3D room model, top-down perspective

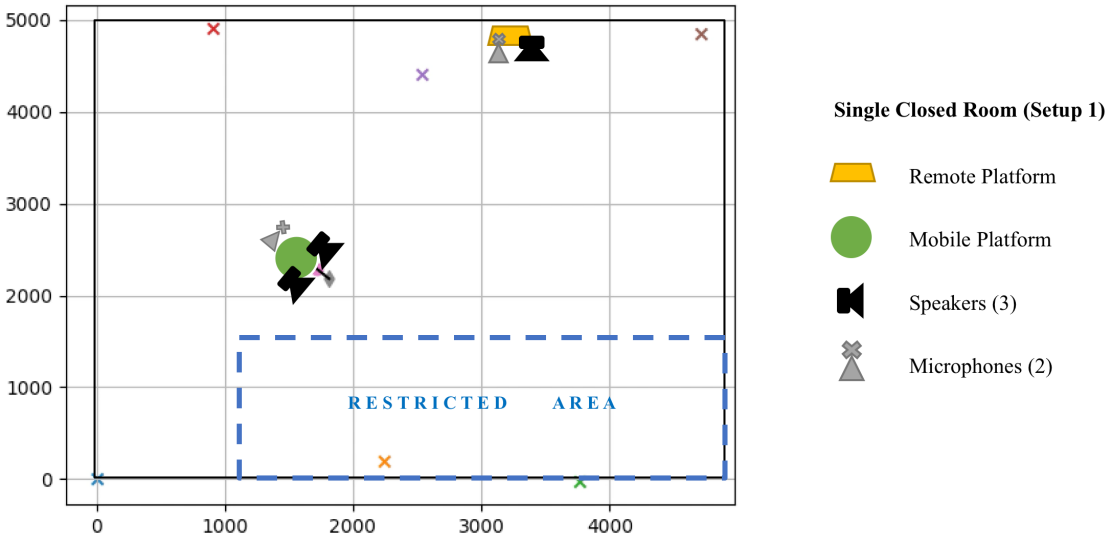

Figure 25: Single closed room setup scenario.

foam, as Figure 26 illustrates.

An acoustical foam screen covered completely the door, to mitigate acoustical interactions between the equipment and the surfaces.

Item	Robot (N_r)	Remote Station (N_{st})
Speaker (N_s)	2	1
Microphone (N_i)	1	1
Channels (N_{ch})	4	4
Cameras (N_{cam})	1	0
Measurements (N_m)	40	20

Table 3: Items in Setup 1. Mobile platforms, remote platforms, speakers, microphones and anchor locations are depicted.

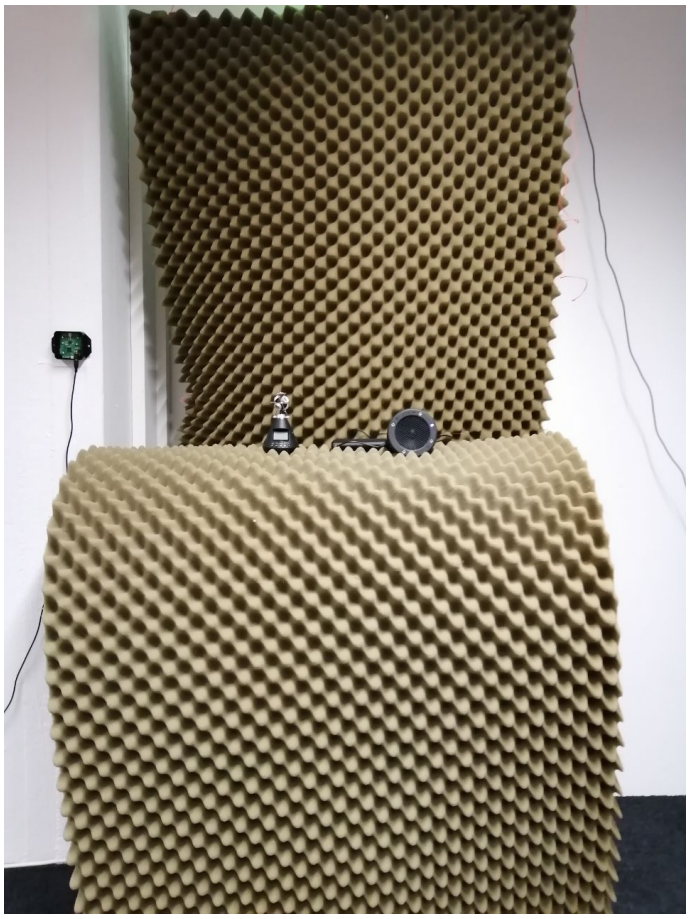

Figure 26: Remote platform in Setup 1

Methodology In Setup 1, the mobile robot platform navigated in a pseudo-random walk, selecting twenty different measurement locations across the room. Once the mobile robot stopped at a place, the Pozyx estimated its position and orientation, reproduced two sine sweeps, and waited to the remote station to reproduced one sine sweep, this process is known as a measurement. Since every measurement consisted of three sine sweeps at every location, the total number of measurements was sixty.

4.2.2 Single Open Room with External Source (Setup 2)

Setup 2 involved the acoustical characterization of two areas, the room and its adjacent hallway. The platforms implemented in this setup were one mobile robot and two remote stations, as Figure 27 illustrates. The aim was to understand acoustical inside the room and interactions between the two separated areas and their building elements, especially at the boundary between the room and the hallway, and the complete hallway. Table 4 contains the list of items in Setup 2. Note that the item dimensions of the robot, remote platforms, speakers and microphones are not to scale.

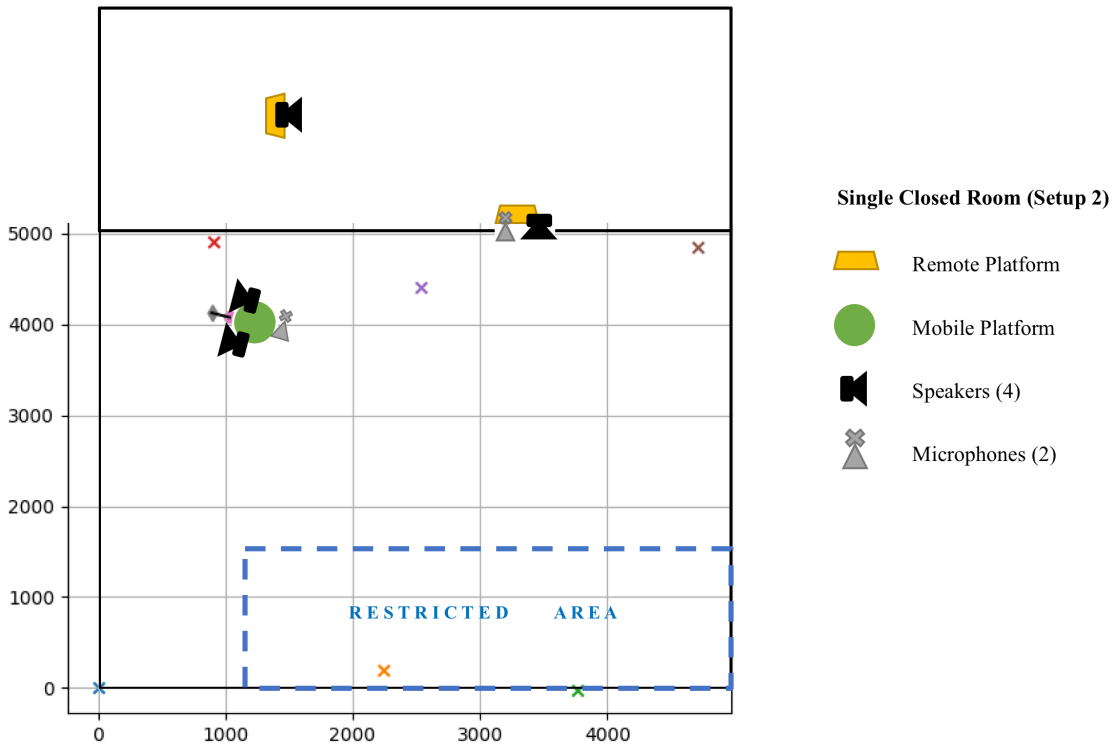

Figure 27: Single open room with external source setup scenario. Mobile platforms, remote platforms, speakers, microphones and anchor locations are depicted.

As Table 4 refers, Setup 2 was equipped with two loudspeakers (Speaker 1 and 2), one microphone, that records four tracks, and one webcam. Setup 2 implemented two remote stations. The first remote station was positioned in the middle of the open door, while the second remote station was placed in the hallway. The first remote station, at the coordinates (3238, 4918) mm, had a Raspberry Pi, Speaker 3, the microphone at 1500 mm high, the distance between Speaker 3 and the microphone was 300 mm. The remote stations were covered in acoustical foam as in Setup 1. The first remote station is depicted in Figure 28.

It can be seen that the microphone was capable to register the incoming sounds from inside the room and the hallway.

Item	Robot (N_r)	Remote Station 1 (N_{st1}) [Open door]	Remote Station 2 (N_{st2}) [External]
Speaker (N_s)	2	1	1
Microphone (N_i)	1	1	0
Channels (N_{ch})	4	4	0
Cameras (N_{cam})	1	0	0
Measurements (N_m)	40	20	20

Table 4: Items in Setup 2

Figure 28: Remote station 1 in Setup 2

On the other hand, the second remote station was positioned in the hallway, at the coordinates (1238, 6418) mm. This remote station only had Speaker 4 connected to a Raspberry Pi, as Figure 29 shows.

Methodology The mobile robot platform navigated in a pseudo-random walk, selecting twenty different measurement locations across the room. Once the mobile robot stopped at a place, the Pozyx estimated its position and orientation, performed

Figure 29: Remote station 2 in Setup 2

two sine sweeps, and waited to the two remote stations to reproduce their test signals, this process is known as a measurement. Because every measurement consisted of four sine sweeps at every location, the total number of measurements was eighty.

5 Results and Discussion

This chapter presents and discusses the measurement recordings from the setups described in Chapter 4. From these recordings, exponential sine sweeps are framed to calculate the room impulse response. The room impulse responses are represented and their features are discussed.

5.1 Single Closed Room (Setup 1)

This setup implemented a robot mobile platform navigating across the room and a remote station placed in front of the door. This experiment consisted of twenty different measurement positions. Figure 30 shows the twenty locations that the robot platform obtained, the dimensions of the floor are displayed in millimetres along the axes.

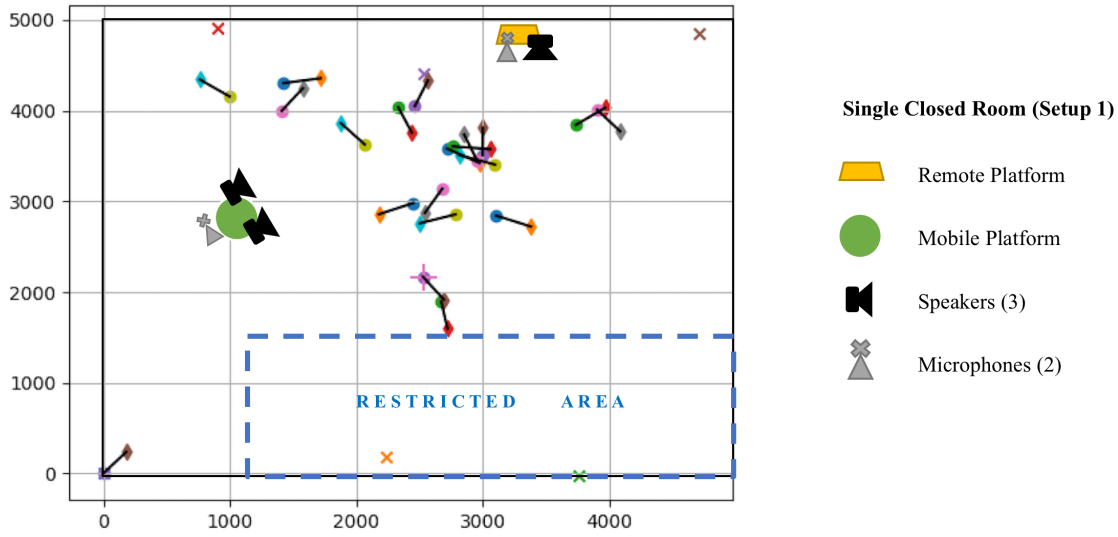

Figure 30: Top-down view of the floor with measurement locations of the robot platform during Setup 1.

On Figure 30, the circles denote the robot position, the diamonds denote the robot orientation, and a line links these symbols, this line does not show the movement of the robot. Additionally, the anchor placement is indicated by an **X**. During this test, the robot mostly navigated on the central part of the room. Even though the pseudo-random walk aims to cover most of the available areas, the navigation concentrated in similar zones. Therefore, the range values of the pseudo-random walk should be extended. In contrast, the values on the orientation demonstrated an optimal performance, the orientation of the robot varies at every location.

As shown in Figure 30, a measurement position sets at the point of origin (0,0) mm. In this case, this position represents a target node loss. This occurred when the Pozyx lost track of the tag and plots the last position at 0, 0.

5.1.1 Recording from the Robot

Once the exponential sine sweeps have been recorded, the full recordings should be processed to frame the test signals. The framing process included the setting of a calibration point, to mark the start point of the first sweep. Subsequently, the timestamps process the start and end points, delimiting each of the sine sweeps. Considering the first measurement location, where the robot platform located at (1411, 3999) mm facing 213.75 degrees.

Figure 31 shows the time domain signals of a recording segment to visualize the start point. In this case, the calibration point was set at 24.7 seconds, and the following starting points were calculated based on the data vectors of the timestamps. It is important to note that setting the calibration point on a single channel, also defines the calibration point for the rest of channels.

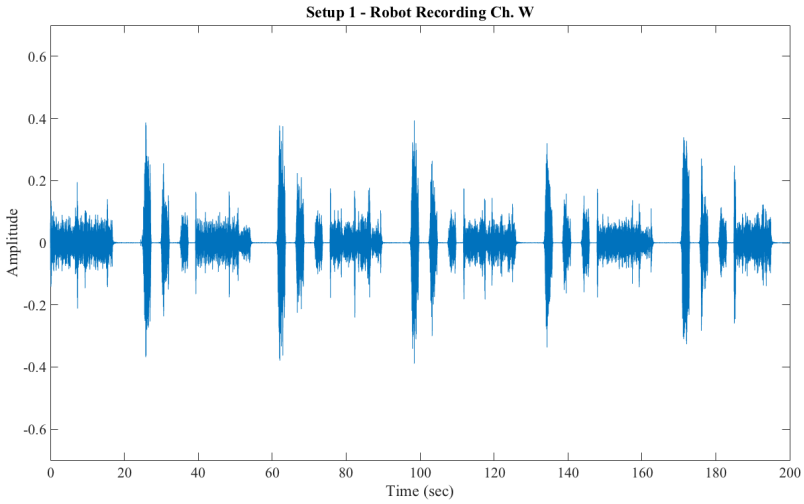

Figure 31: Time domain signal from Channel W of the robot platform recording in the test Setup 1.

During Setup 1, twenty locations were sampled. At each measurement location three sweeps were recorded. Therefore, sixty sine sweeps were recorded in total. The first sweep had the highest peak of the three sine sweeps, due to the positioning between the speaker and the recorder.

It can be seen from Figure 31 that every three sine sweeps are followed by background noise. This noise was the sound of the robot moving to the next position.

Figure 32 presents time domain plots of the three framed exponential sine sweeps measured at the first location. The framed signals correspond to the three sine sweeps from the Speaker 1, Speaker 2, and Speaker 3 on the Channels W (Blue), X (Orange), Y (Yellow) and Z (Purple).

It is seen that the framed sweeps occurred at different times, therefore, the framing agrees with the sequential playback of the signals from the speakers.

Considering the first sweep recorded on Channel W, played by Speaker 1, its impulse response was calculated. The representation of this impulse response is shown in Figure 33.

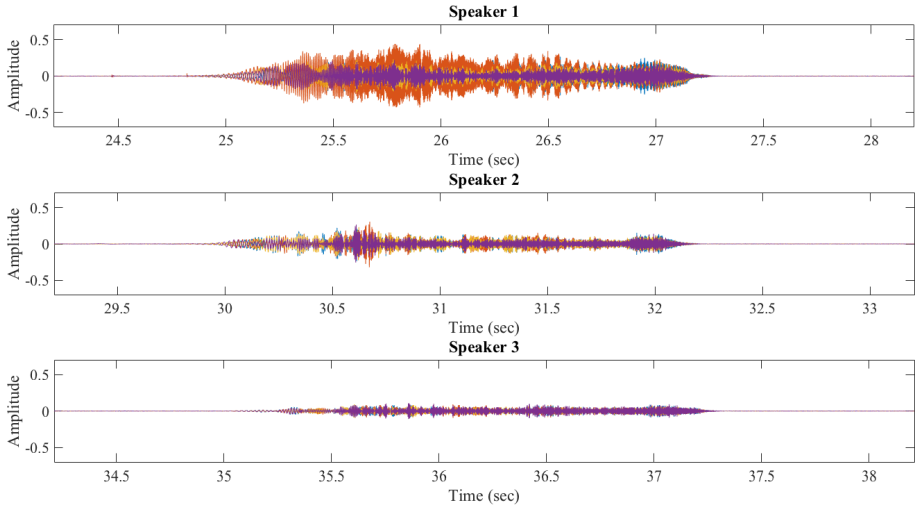

Figure 32: Time domain framed sine sweep plots from Speaker 1 (Top), Speaker 2 (Middle) and Speaker 3 (Bottom) in Setup 1, in Channels W (Blue), X (Orange), Y (Yellow) and Z (Purple).

From Figure 33 a) The signal presents an adequate signal to noise ratio, capable to be employed for calculating acoustical parameters. b) The power of the signal presented energy mostly on the range of 0 to 15 kHz.

5.1.2 Recording from the Remote Station

In Setup 1, the remote station recorded the sine sweeps at a fixed position in front of the door. The process described in 5.1.1 for obtaining the impulse responses was repeated for the recordings from the remote station. The calibration point of this recording was set at 29.5 seconds. Figure 34 depicts the time domain signals of an audio excerpt from Channel W from the recording on the remote station, where the X-axis shows the amplitude and Y-axis the duration.

The background noises significantly reduced in comparison to Figure 32 due to the distance from the robot. Furthermore, the last sine sweep was more prominent because of the proximity of the Speaker 3 and the recorder.

The three sine sweeps measured at the first location are depicted in Figure 35 on the channels W, X, Y and Z.

In Figure 35 the framed sine sweep recordings from the four channels overlapped at same times.

Once the framing of the sweeps is completed, a sweep was employed to calculate its impulse response. The impulse response of Speaker 1, W Channel is depicted in Figure 36. X-axis depicts the duration and Y-axis depicts the amplitude.

The signal of Speaker 3 was prominent in comparison with the other recorded sine sweeps, due to the proximity between the microphone and Speaker 3.

Figure 36 presents a) No room modes because the loudspeaker and microphone were near to each other. b) The magnitude of this signal prevails longer compared to 33. This effect could be produced because the recorder was situated in front a wall.

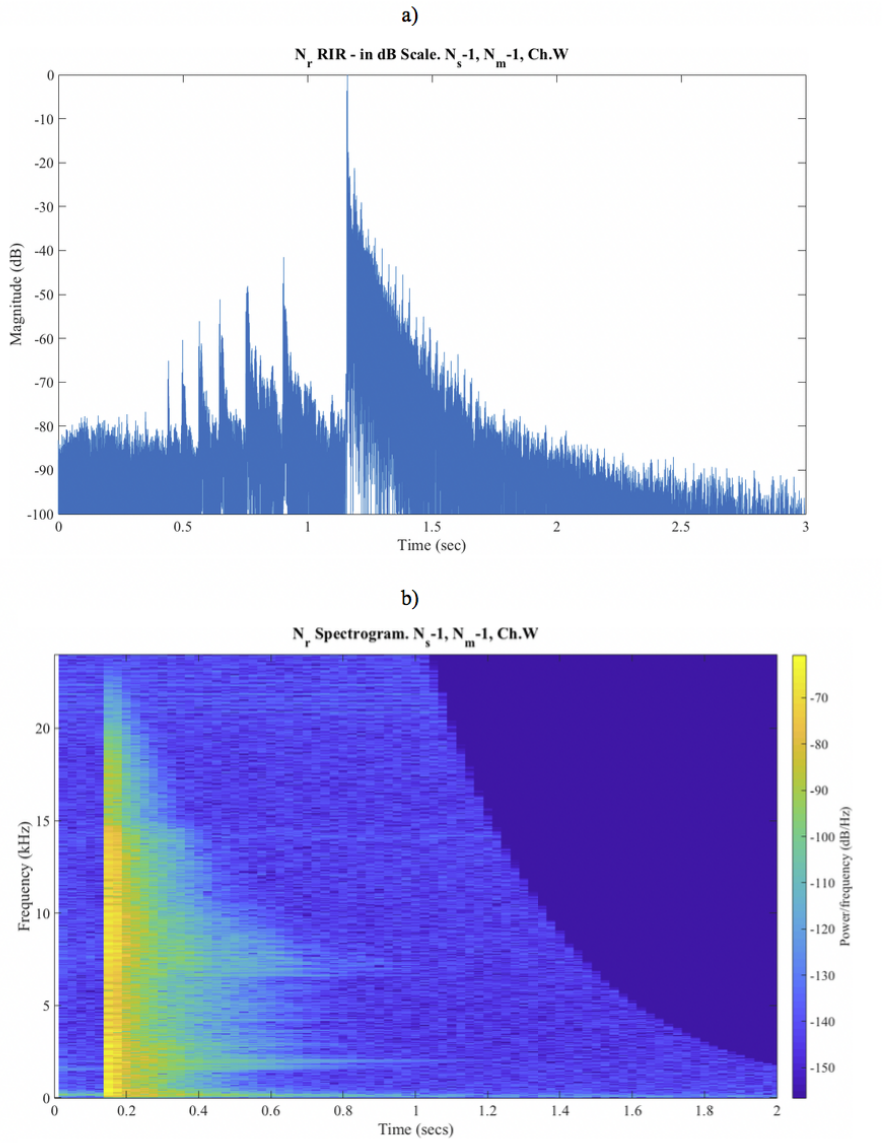

Figure 33: Room impulse response from the Speaker 1 during the first measurement location by the robot platform recorder, represented in: a) dB scale, b) Spectrogram

5.2 Single Open Room with External Source (Setup 2)

Setup 2 implemented a robot mobile platform and two remote platforms, one located in the middle of the open door, and one was set on the adjacent hallway. During this test, the robot platform performed measurements at twenty different experiment locations. Figure 37 presents the circles as the robot position, the diamonds as the robot orientation, and a line to link these symbols,note that this line does not show the movement of the robot. Additionally, the anchor placement is indicated by an **X**

In this setup, the robot platform covered similar areas in the previous test.

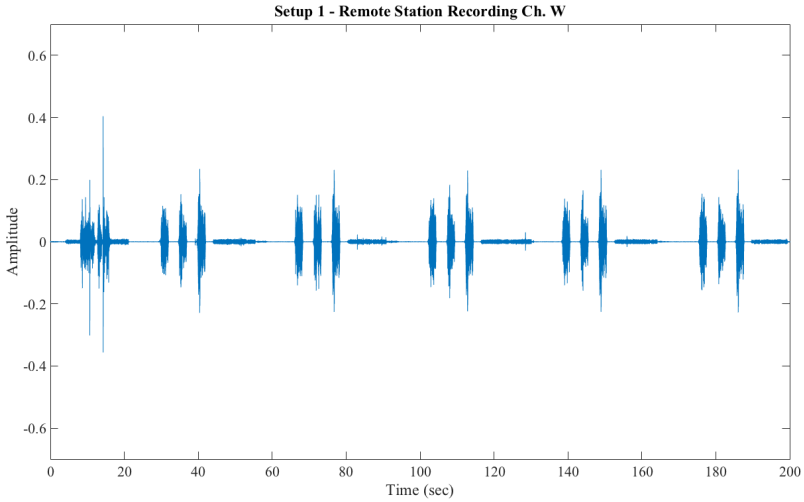

Figure 34: Time domain signal from Channel W of the remote recording in Setup 1.

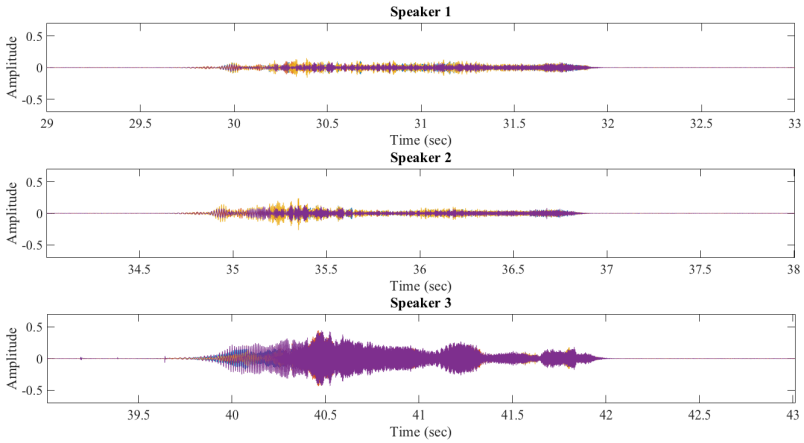

Figure 35: Time domain framed sine sweep plots from Speaker 1 (Top), Speaker 2 (Middle) and Speaker 3 (Bottom) in Setup 1, on Channels W (Blue), X (Orange), Y (Yellow) and Z (Purple).

5.2.1 Recording from the Robot

The measurement in Setup 2 recorded test signals from four speakers. Figure 38 presents the time domain plot of an audio measurement segment recorded on Channel W, displaying the amplitude on the X-axis and the duration Y-axis.

At the twenty measurement locations, four sine sweeps were recorded. Therefore, a total of eighty sweeps were recorded. For this example, the twelfth location is examined, where the location of the robot platform occurred at (2765, 1883) mm orientated 248.87 degrees. As seen in Figure 38 four consecutive spiked signals were presented along the recording, that corresponded to the sine sweeps played at each loudspeaker. These four signals were the sine sweeps of each of the speakers at a location. The signals also presented background noises, mainly the robot noise. Furthermore, Speaker 1 shows predominacy in amplitude compared to the rest of

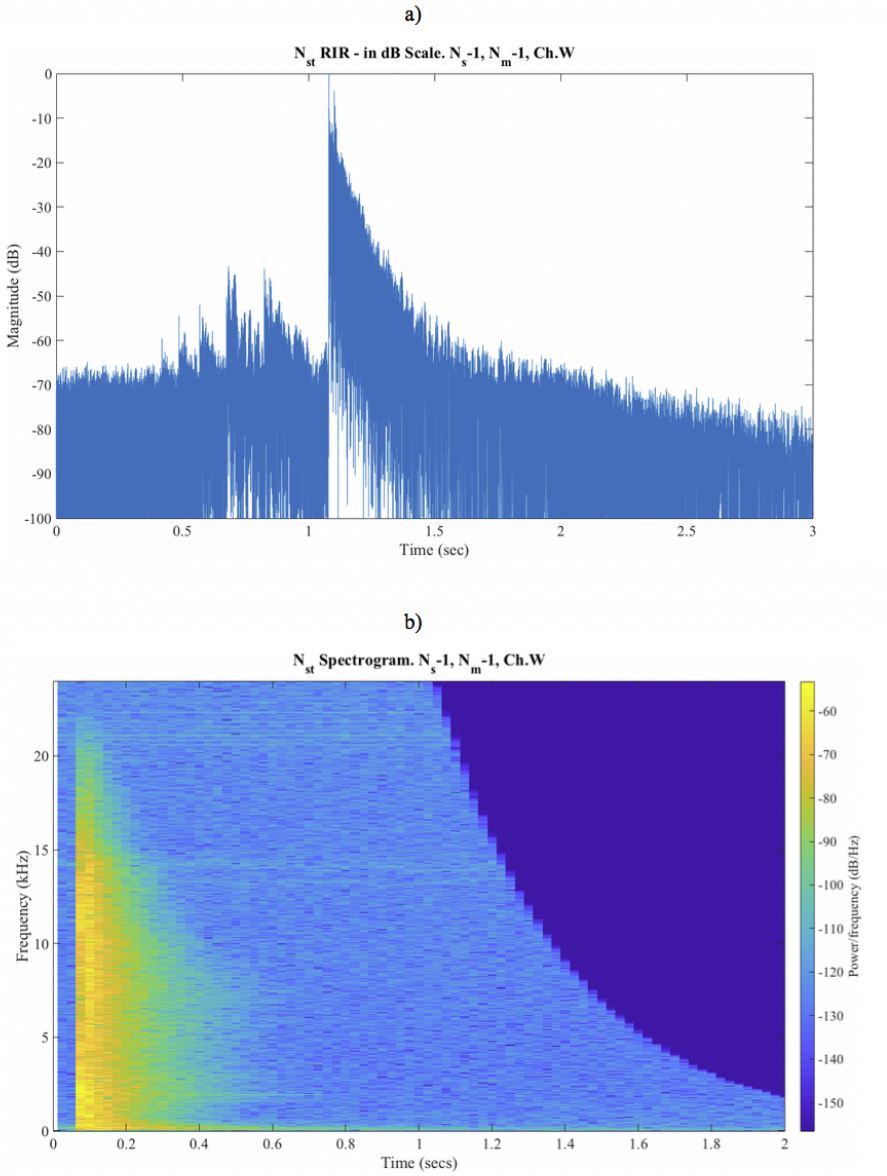

Figure 36: Room impulse response from Speaker 1 at the first measurement location by the remote station recorder, represented in: a) dB scale, b) Spectogram.

the speakers, because of the positioning between the microphone and speakers. The third and fourth sweeps have lower amplitude because of their location. Speaker 3 faced inside of the room while, Speaker 4 was located on the hallway.

The calibration point was set at 25 seconds. The framed signals on the twelfth measurement are depicted on Figure 39, showing the time domain signals of the four sweeps on the channels W, X, Y, and Z. The X-axis presents the amplitude and Y-axis presents the duration.

From these framed sweeps, the impulse response was calculated. The sweep on Speaker 3 was considered to calculate its impulse response. Figure 40 presents the

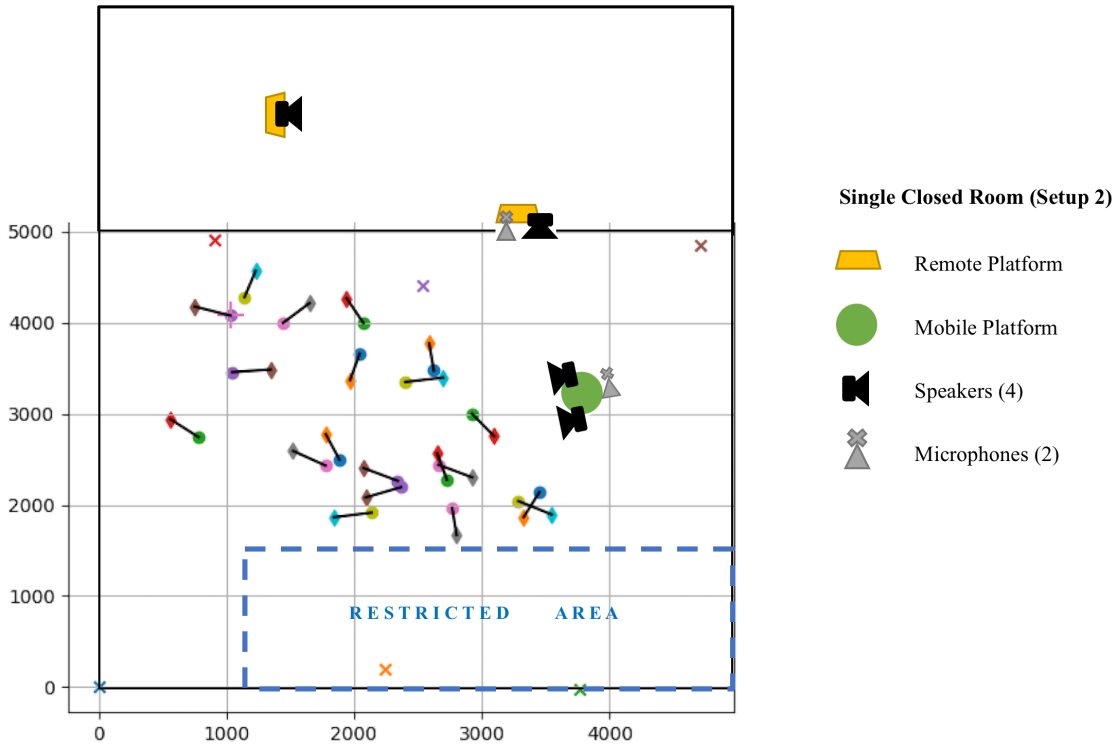

Figure 37: Measurement locations of the robot platform during Setup 2, anchor placement, and room dimensions.

impulse response played by Speaker 3, a) On a dB scale and b) Spectrogram.

It can be seen from Figure 40 that a) There is an absence of room modes, due to the proximity between the microphone and loudspeaker. b) The energy decays at 0.2 seconds.

5.2.2 Recording from the Remote Station

In Setup 2, the remote platform recording was placed in the middle of the open door. Similarly to the previous process, the calibration point was selected at 18.6 seconds. Figure 41 shows a segment of the measurement captured by the recorder, where the X-axis shows the amplitude and Y-axis the duration on Channels W.

Subsequently, the framing process was performed in the recording. Figure 42 presents the framed time domain sweeps of the twelfth measurement location on the channels W, X, Y and Z. These signals depict in X-axis the amplitude and the Y-axis the duration.

The signal from Speaker 3 presented a higher amplitude compared to the rest, due to the microphone-speaker position. Furthermore, Speaker 4 sweeps, device placed on the hallway, presented a lower amplitude.

Calculating the impulse response from the sweep played on Speaker 3, in the twelfth measurement position. Figure 43 illustrates this impulse response, a) On a dB scale, and b) Spectrogram.

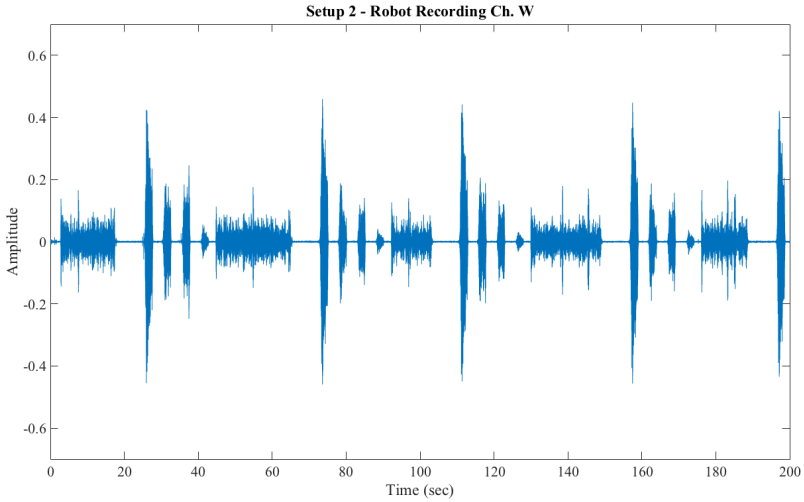

Figure 38: Time domain signals from Channel W of the robot platform recording in Setup 2.

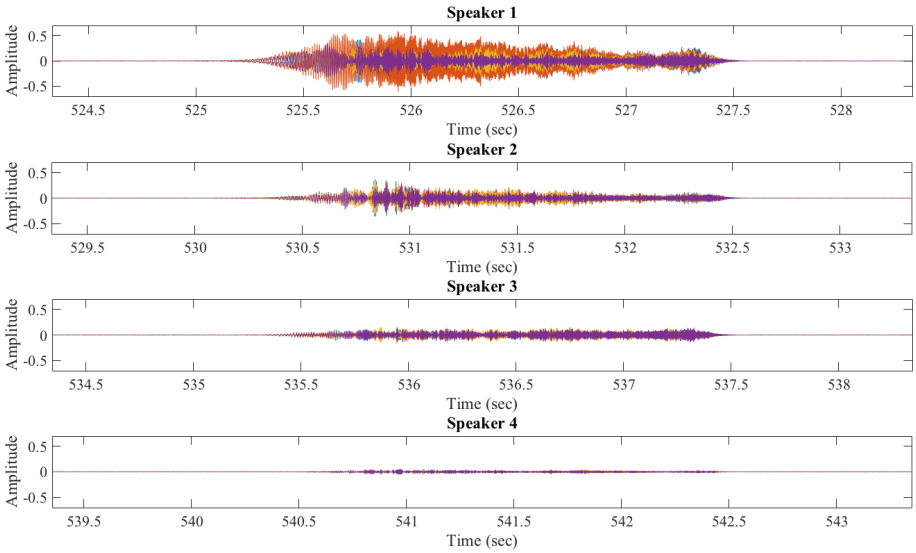

Figure 39: Time domain framed sine sweeps plots from Speaker 1 (Top), Speaker 2 (Mid-top), Speaker 3 (Mid-bottom), Speaker 4 (Bottom), in Channels W (Blue), X (Orange), Y (Yellow) and Z (Purple).

It can be seen from Figure 43 b) The power decays in 1.5 seconds, the propagation of the sound occurrs rapidly compared to the recordings where the microphone was set next to a partition.

Although the measured impulse responses in this prototype does not follow the recommendations of the ISO 3382 standard, these measurements shown consistent results in both of the experiment setups. Similar equipment was implemented on the platforms; thus, it is clear that the compact speakers presented shortcomings. For instance, the low power signal levels coming from speakers at a significant distance or behind a partition. Optimal levels should be set in order to capture properly the

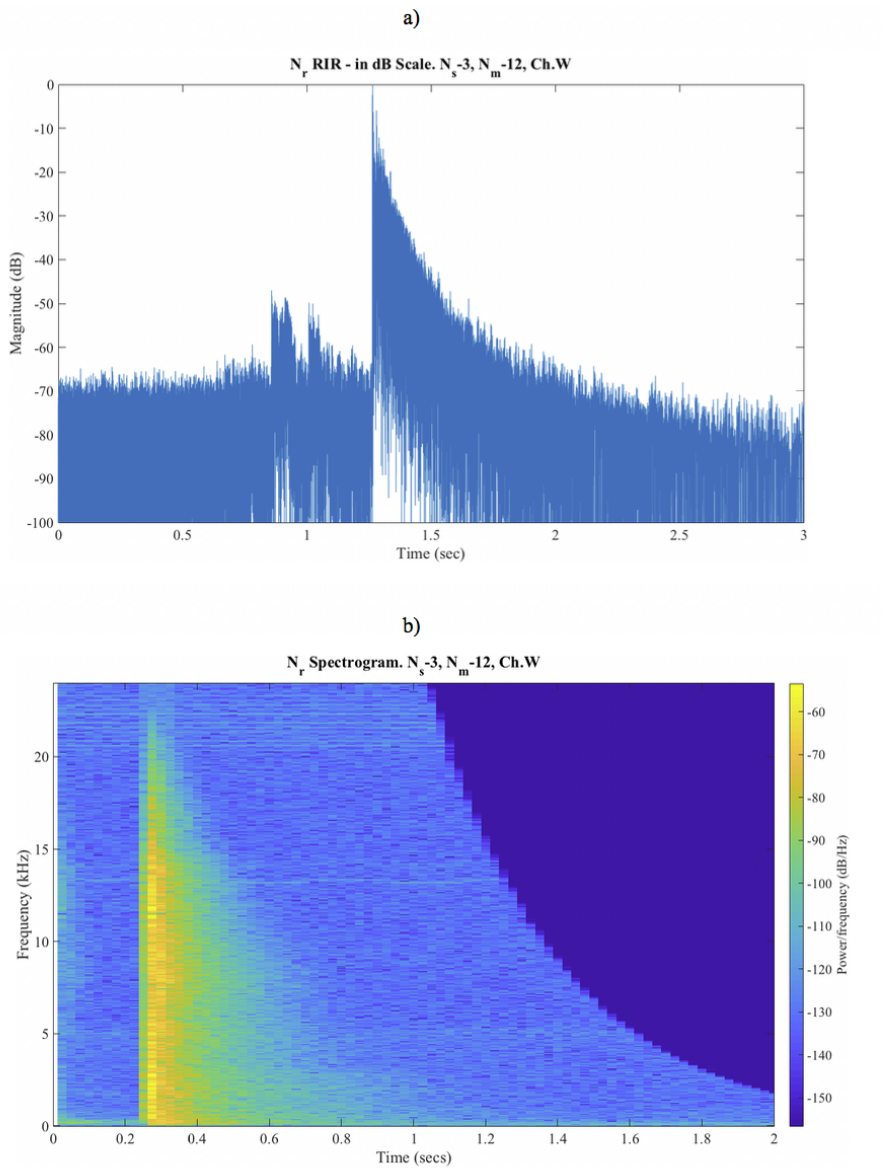

Figure 40: Room Impulse Response from the Speaker 3 during the twelfth measurement location by the robot platform recorder, represented in: a) dB scale, b) Spectrogram.

signals. Additionally, it has to be considered that the loudspeakers do not generate distortion or clipping. At this point, the results need further testing in artificial intelligence algorithms to analyze their outcome. For this reason, a vast amount of information should be collected.

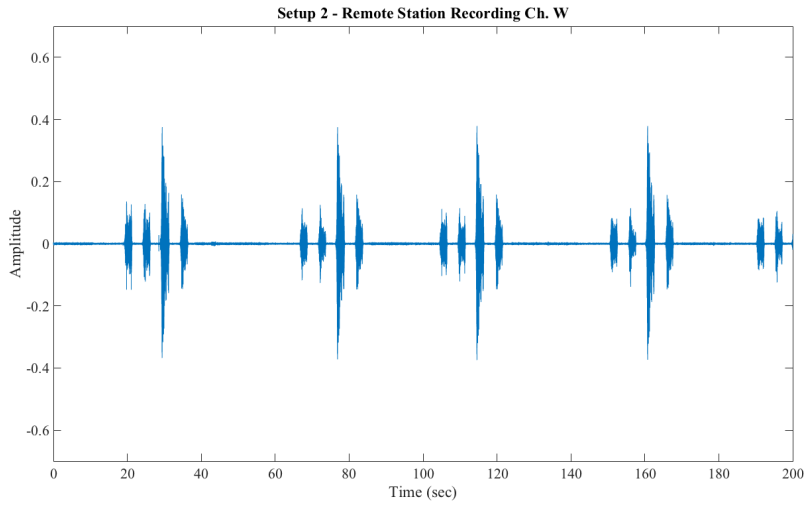

Figure 41: Time domain signal from the Channel W of the remote platform in Setup 2.

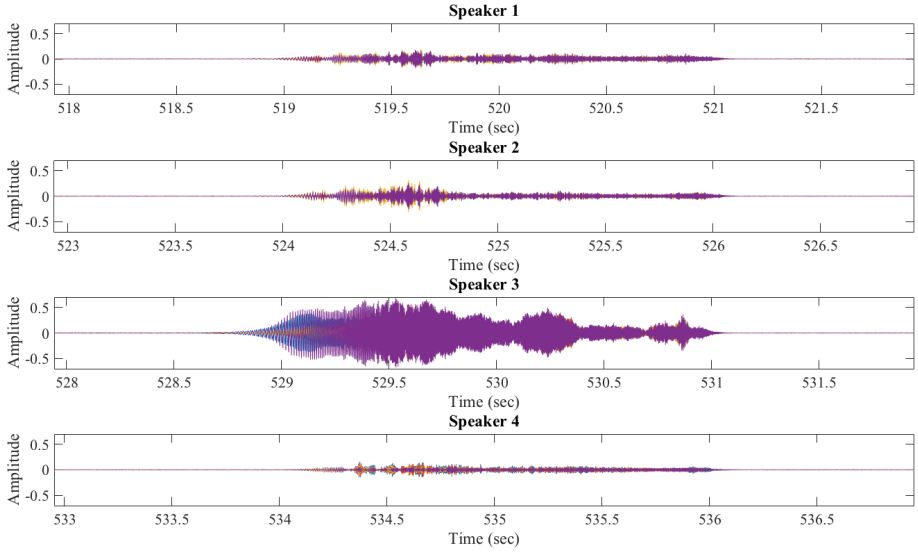

Figure 42: Time domain framed sine sweeps plots from the Speaker 1 (Top), Speaker 2 (Mid-top), Speaker 3 (Mid-bottom), and Speaker 4 (Bottom) in Setup 2, in Channels W (Blue), X (Orange), Y (Yellow) and Z (Purple).

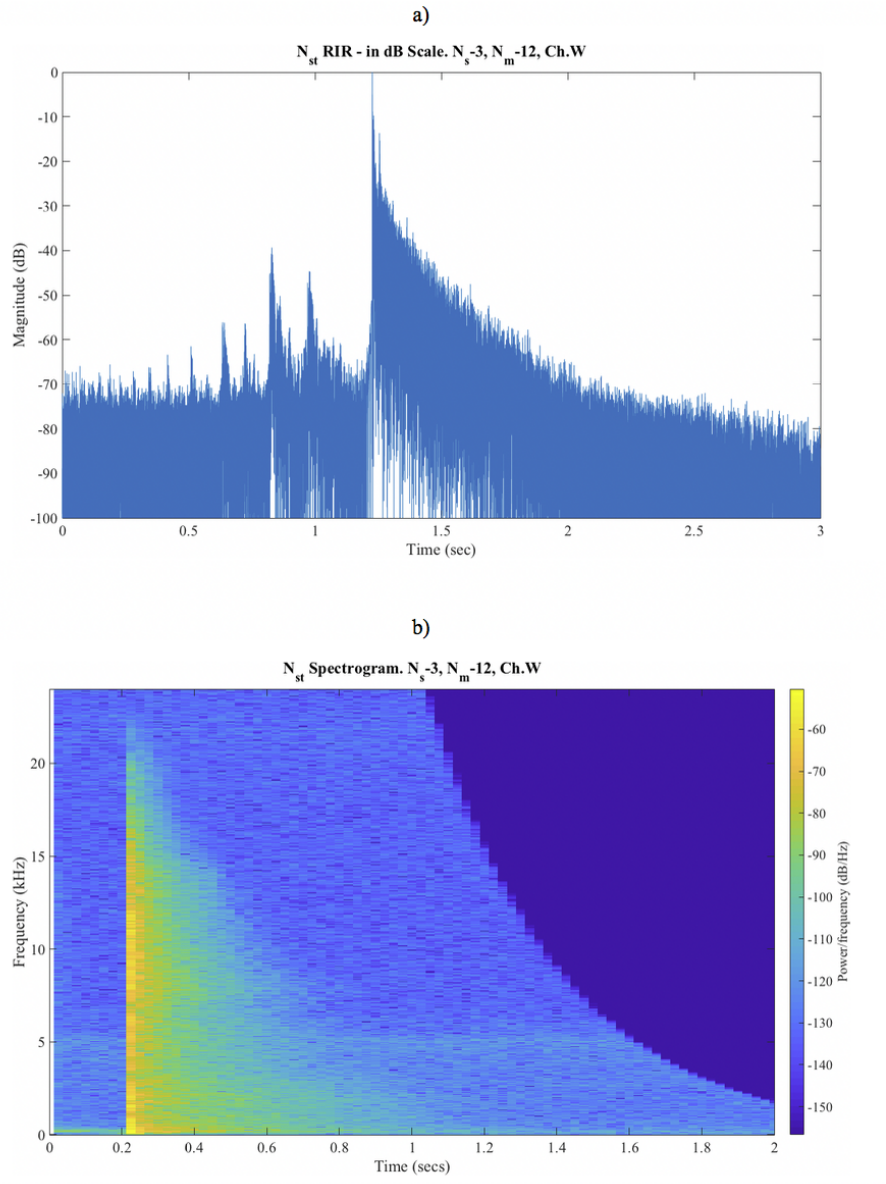

Figure 43: Room Impulse Response from Speaker 3 during the twelfth measurement location by the remote station recorder, represented in: a) dB scale, b) Spectrogram

6 Conclusion

This thesis designed, constructed and tested an adaptable automated system (AIRMS) to collect exponential sine sweeps in confined and unoccupied spaces of multiroom environments. AIRMS allows researchers, according to their purpose, to define the number of items to be implemented in the system, such as rooms to be tested, measurements, platforms and mounted equipment. This novel approach employs exponential sine sweeps to perform acoustical tests, an indoor localization system to track its real-time location, as well as multiple speakers and spatial microphones to play and record the test signals.

AIRMS measures RIRs in two stages: data collection and data post-processing. For data collection, the system uses mobile and stationary measurement platforms to record test signals in the measurement environment. To perform acoustical measurements, the mobile platform navigates around a room equipped with electroacoustic devices. Meanwhile, the remote stations are set in fixed locations to capture phenomena in specific areas. During measurements, mobile and remote platforms wirelessly communicate to systematically perform the tests. Furthermore, the interactions between the mobile and remote platforms provides vast amounts of information concerning the measurement space.

Data post-processing involves framing the exponential sine sweeps and measuring the impulse response. During framing, every exponential sine sweep is extracted from the raw audio recordings. All the sine sweeps are stored in a data structure. The automatically stored data is supplemented by manually gathered data. Thus, the data structure allocates all the compiled information from the measured environment. To test the prototype, AIRMS was assembled and implemented in two measurement environments: a single room and a multiroom environment.

Furthermore, no previous solutions have implemented a complete set of features to gather acoustical test signals. Although AIRMS is designed to collect a vast amount of data from a theoretically unlimited number of locations, tests for the proof of concept measured only twenty locations. Since this prototype does not follow the ISO 3382 standard for performing acoustic measurements, feeding a vast amount of the measurements into artificial intelligence algorithms would determine the accuracy of acoustical parameter predictions. The current system has two limitations inherit to the equipment implemented: the indoor localization and mobility of the robot. Preliminary measurements for the localization accuracy of the robot showed that it had a margin of error of ± 400 mm. The area of the rooms was restricted to $23.5\,\mathrm{m}^2$, due to limitations of the built-in robot sensors. Nevertheless, the robot requires no human intervention, as it can automatically move pseudo-randomly to capture the measurements. Furthermore, the robot produced little background noise audio measurements owing to the methodology of the tests and speed of the robot. The automatically collected sine sweeps were used to calculate room impulse responses. This enabled reliable data to be implemented into the artificial intelligence algorithms, which reproduces manual data collection.

6.1 Future Work

Future work could focus on enhancing the functionalities of AIRMS in its two measurement stages: data collection and post-processing. For data collection, the wheeled robot should be optimized by randomizing the mobility parameters. This could be accomplished by incrementing the range of the pseudo-random walk movements, as this would enable the robot to reach more areas. Additionally, the wheeled robot could integrate more sensors, such as LIDAR, to enhance avoidance or recognition of obstacles in the room.

For indoor localization, autocalibration should be revised in order to recognize the reference anchors. Furthermore, to reinforce the location, it would be important to implement wide field of view cameras, this would enable capture of panoramic snapshots.

For post-processing, the sweep framing process could be automated, as the calibration point is currently set manually by the user. using two different approaches: by setting the calibration point automatically or by tracking the experiments. Such automatic calibration could be accomplished by using timestamps. However, this method would also recognize inherit noises from the loudspeaker, such as clippings. Another approach would be to track the measurement recordings. Communication via the USB port might be established between the single-board computer and the microphone, thus allowing the computer to send signals for initiating the start and end of a measurement.

Furthermore, the measurement process could also be remotely monitored by building a web interface. This web interface could display real-time information concerning the measurement process, such as images from the installed cameras and the files stored on the computer.

References

[1] C. Sun, A. Shrivastava, S. Singh, and A. Gupta, "Revisiting unreasonable effectiveness of data in deep learning era," in *Proceedings of the IEEE international conference on computer vision*, pp. 843–852, 2017.

[2] J. M. Carmena, M. A. Lebedev, R. E. Crist, J. E. O'Doherty, D. M. Santucci, D. F. Dimitrov, P. G. Patil, C. S. Henriquez, and M. A. Nicolelis, "Learning to control a brain–machine interface for reaching and grasping by primates," *PLoS biology*, vol. 1, no. 2, p. e42, 2003.

[3] J. H. Min and Y.-C. Lee, "Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters," *Expert systems with applications*, vol. 28, no. 4, pp. 603–614, 2005.

[4] R. Falcon Perez, "Machine-learning-based estimation of room acoustic parameters," Master's thesis, Aalto University, 11 2018.

[5] A. V. Oppenheim *et al.*, *Signals and Systems*. Prentice Hall International, 2 ed., 1997.

[6] G.-B. Stan, J.-J. Embrechts, and D. Archambeau, "Comparison of different impulse response measurement techniques," *Journal of the Audio Engineering Society*, vol. 50, no. 4, pp. 249–262, 2002.

[7] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio Engineering Society Convention 108*, Audio Engineering Society, 2000.

[8] A. Farina, "Advancements in impulse response measurements by sine sweeps," in *Audio Engineering Society Convention 122*, Audio Engineering Society, 2007.

[9] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.

[10] J. H. Rindel, "The use of computer modeling in room acoustics," *Journal of vibroengineering*, vol. 3, no. 4, pp. 219–224, 2000.

[11] L. L. Thompson, "A review of finite-element methods for time-harmonic acoustics," *The Journal of the Acoustical Society of America*, vol. 119, no. 3, pp. 1315–1330, 2006.

[12] S. M. Kirkup, *The boundary element method in acoustics*. Integrated sound software, 2007.

[13] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Computer simulation of hrtfs for personalization of 3d audio," in *2008 Second International Symposium on Universal Communication*, pp. 435–440, IEEE, 2008.

[14] A. Southern, S. Siltanen, and L. Savioja, "Spatial room impulse responses with a hybrid modeling method," in *Audio Engineering Society Convention 130*, Audio Engineering Society, 2011.

[15] J. Y. Wen, N. D. Gaubitch, E. A. Habets, T. Myatt, and P. A. Naylor, "Evaluation of speech dereverberation algorithms using the mardy database," in *Proceedings of the International Workshop on Acoustic Echo and Noise Control (IWAENC*, Citeseer, 2006.

[16] R. Stewart and M. Sandler, "Database of omnidirectional and b-format room impulse responses," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 165–168, IEEE, 2010.

[17] D. T. Murphy and S. Shelley, "Openair: An interactive auralization web resource and database," in *Audio Engineering Society Convention 129*, Audio Engineering Society, 2010.

[18] E. Hadad, F. Heese, P. Vary, and S. Gannot, "Multichannel audio database in various acoustic environments," in *2014 14th International Workshop on Acoustic Signal Enhancement (IWAENC)*, pp. 313–317, IEEE, 2014.

[19] P. Kamol, S. Nikolaidis, R. Ueda, and T. Arai, "Rfid based object localization system using ceiling cameras with particle filter," in *Future generation communication and networking (FGCN 2007)*, vol. 2, pp. 37–42, IEEE, 2007.

[20] H. Hisahara, Y. Ishii, M. Ota, T. Ogitsu, H. Takemura, and H. Mizoguchi, "Human avoidance function for robotic vacuum cleaner through use of environmental sensors: Roomba® making way for humans," in *2014 5th International Conference on Intelligent Systems, Modelling and Simulation*, pp. 64–67, IEEE, 2014.

[21] W. Poomarin, R. Chancharoen, and V. Sangveraphunsiri, "Automatic docking with obstacle avoidance of a differential wheel mobile robot," *International Journal of Mechanical Engineering and Robotics Research*, vol. 5, no. 1, p. 11, 2016.

[22] A. Gallardo, J. Taylor, C. Paolini, H.-K. Lee, and G. K. Lee, "An anfis-based multi-sensor structure for a mobile robotic system," in *Computational Intelligence in Control and Automation (CICA)*, pp. 143–148, IEEE, 2011.

[23] D. A. Lazewatsky and W. D. Smart, "An inexpensive robot platform for teleoperation and experimentation," in *2011 IEEE International Conference on Robotics and Automation*, pp. 1211–1216, IEEE, 2011.

[24] P. Nattharith, "Fuzzy logic based control of mobile robot navigation: A case study on irobot roomba platform," *Scientific Research and Essays*, vol. 8, no. 2, pp. 82–94, 2013.

[25] J. Xu, S. Zhu, H. Guo, and S. Wu, "Automated labeling for robotic autonomous navigation through multi-sensory semi-supervised learning on big data," *IEEE Transactions on Big Data*, 2019.

[26] M. J. Mataric, N. P. Koenig, and D. Feil-Seifer, "Materials for enabling hands-on robotics and stem education.," in *AAAI spring symposium: Semantic scientific knowledge integration*, pp. 99–102, 2007.

[27] A. Araújo, D. Portugal, M. S. Couceiro, and R. P. Rocha, "Integrating arduino-based educational mobile robots in ros," *Journal of Intelligent & Robotic Systems*, vol. 77, no. 2, pp. 281–298, 2015.

[28] S. Schmitt, H. Will, B. Aschenbrenner, T. Hillebrandt, and M. Kyas, "A reference system for indoor localization testbeds," in *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–8, IEEE, 2012.

[29] E. Ruiz, R. Acuña, N. Certad, A. Terrones, and M. E. Cabrera, "Development of a control platform for the mobile robot roomba using ros and a kinect sensor," in *2013 Latin American Robotics Symposium and Competition*, pp. 55–60, IEEE, 2013.

[30] J. Le Roux, E. Vincent, J. R. Hershey, and D. P. Ellis, "Micbots: collecting large realistic datasets for speech and audio research using mobile robots," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5635–5639, IEEE, 2015.

[31] S. Brodeur, S. Carrier, and J. Rouat, "Create: Multimodal dataset for unsupervised learning, generative modeling and prediction of sensory data from a mobile robot in indoor environments," *arXiv preprint arXiv:1801.10214*, 2018.

[32] H. Kuttruff, *Room Acoustics*. CRC Press, 2016.

[33] F. A. Everest, K. C. Pohlmann, *et al.*, *Master Handbook of Acoustics*. McGraw Hill Professional, 5 ed., 2009.

[34] A. J. Berkhout, D. de Vries, and M. M. Boone, "A new method to acquire impulse responses in concert halls," *The Journal of the Acoustical Society of America*, vol. 68, no. 1, pp. 179–183, 1980.

[35] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, "Spatial decomposition method for room impulse responses," *Journal of the Audio Engineering Society*, vol. 61, no. 1/2, pp. 17–28, 2013.

[36] J. G. Bolaños, *Features and Applications of the Laser-induced Spark as a Monopole Source for Acoustic Impulse Response Measurements*. PhD thesis, Aalto University, 2017.

[37] V. Valimaki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel, "Fifty years of artificial reverberation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 5, pp. 1421–1448, 2012.

[38] D. R. Begault, "Perceptual effects of synthetic reverberation on three-dimensional audio systems," *Journal of the Audio Engineering Society*, vol. 40, no. 11, pp. 895–904, 1992.

[39] J. S. Bradley and G. A. Soulodre, "The influence of late arriving energy on spatial impression," *The Journal of the Acoustical Society of America*, vol. 97, no. 4, pp. 2263–2271, 1995.

[40] R. C. Heyser, "Acoustical measurements by time delay spectrometry," *Journal of the Audio Engineering Society*, vol. 15, no. 4, pp. 370–382, 1967.

[41] R. C. Heyser, "Loudspeaker phase characteristics and time delay distortion: Part 1," *Journal of the Audio Engineering Society*, vol. 17, no. 1, pp. 30–41, 1969.

[42] M. A. Poletti, "Linearly swept frequency measurements, time-delay spectrometry, and the wigner distribution," *Journal of the Audio Engineering Society*, vol. 36, no. 6, pp. 457–468, 1988.

[43] P. Svensson and J. L. Nielsen, "Errors in mls measurements caused by time variance in acoustic systems," *Journal of the Audio Engineering Society*, vol. 47, no. 11, pp. 907–927, 1999.

[44] S. Müller and P. Massarani, "Transfer-function measurement with sweeps," *Journal of the Audio Engineering Society*, vol. 49, no. 6, pp. 443–471, 2001.

[45] X. Wang, *Model based signal enhancement for impulse response measurement*, vol. 18. Logos Verlag Berlin GmbH, 2014.

[46] L. L. Beranek, L. L. Beranek, L. L. Beranek, and L. L. Beranek, *Acoustical measurements*. Acoustical Society of America Melville, NY, 1988.

[47] M. Kleiner, *Acoustics and audio technology*. J. Ross Publishing, 2011.

[48] V. Pulkki and M. Karjalainen, *Communication acoustics: an introduction to speech, audio and psychoacoustics*. John Wiley & Sons, 2015.

[49] J. Eargle, *The Microphone Book: From mono to stereo to surround-a guide to microphone design and application*. CRC Press, 2012.

[50] J. Benesty, J. Chen, and Y. Huang, *Microphone array signal processing*, vol. 1. Springer-Verlag Berlin Heidelberg, 2008.

[51] K. Farrar, "Soundfield microphone," *Wireless World*, vol. 85, no. 1526, pp. 48–50, 1979.

[52] E. Benjamin and T. Chen, "The native b-format microphone: Part 1," in *Audio Engineering Society Convention 119*, Audio Engineering Society, 2005.

[53] F. Zotter and M. Frank, *Ambisonics: A practical 3D audio theory for recording, studio production, sound reinforcement, and virtual reality.* Springer Nature, 2019.

[54] B. Rafaely, B. Weiss, and E. Bachmat, "Spatial aliasing in spherical microphone arrays," *IEEE Transactions on Signal Processing*, vol. 55, no. 3, pp. 1003–1010, 2007.

[55] B. N. Gover, J. G. Ryan, and M. R. Stinson, "Measurements of directional properties of reverberant sound fields in rooms using a spherical microphone array," *The Journal of the Acoustical Society of America*, vol. 116, no. 4, pp. 2138–2148, 2004.

[56] D. A. Dick and M. C. Vigeant, "A comparison of measured room acoustics metrics using a spherical microphone array and conventional methods," *Applied Acoustics*, vol. 107, pp. 34–45, 2016.

[57] L. McCormack, S. Delikaris-Manias, and V. Pulkki, "Parametric acoustic camera for real-time sound capture, analysis and tracking," in *Proceedings of the 20th International Conference on Digital Audio Effects (DAFx-17)*, pp. 412–419, 2017.

[58] B. Rafaely, I. Balmages, and L. Eger, "High-resolution plane-wave decomposition in an auditorium using a dual-radius scanning spherical microphone array," *The Journal of the Acoustical Society of America*, vol. 122, no. 5, pp. 2661–2668, 2007.

[59] B. Rafaely, "Spherical microphone array with multiple nulls for analysis of directional room impulse responses," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 281–284, IEEE, 2008.

[60] D. Pavlidi, S. Delikaris-Manias, V. Pulkki, and A. Mouchtaris, "3d doa estimation of multiple sound sources based on spatially constrained beamforming driven by intensity vectors," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 96–100, IEEE, 2016.

[61] A. Farina and L. Tronchin, "3d sound characterisation in theatres employing microphone arrays," *Acta acustica united with Acustica*, vol. 99, no. 1, pp. 118–125, 2013.

[62] T. R. Kurfess, *Robotics and automation handbook.* Taylor & Francis, 2005.

[63] S. G. Tzafestas, *Introduction to mobile robot control.* Elsevier, 2013.

[64] G. A. Bekey, *Autonomous robots: from biological inspiration to implementation and control.* MIT press, 2005.

[65] Sensible 4/Self-Driving Technology, "Gacha – self-driving shuttle bus for all weather." https://www.sensible4.fi/gacha/. Last accessed on August 2019.

[66] Amazon Robotics, "We reimagine now." https://www.amazonrobotics.com/. Last accessed on August 2019.

[67] Uconsystem, "Parcel delivery drone." http://www.uconsystem.com/eng/products/industrial/remocopter-010.asp. Last accessed on August 2019.

[68] A. Koubâa, H. Bennaceur, I. Chaari, S. Trigui, A. Ammar, M.-F. Sriti, M. Alajlan, O. Cheikhrouhou, and Y. Javed, *Robot Path Planning and Cooperation*, vol. 772. Springer, 2018.

[69] T. S. Levitt, "Qualitative navigation for mobile robots," *Int. J. Artificial Intelligence*, vol. 44, pp. 305–360, 1990.

[70] O. Holland, "The first biologically inspired robots," *Robotica*, vol. 21, no. 4, pp. 351–363, 2003.

[71] SRI International, "Timeline of innovation. shakey the robot." https://www.sri.com/sites/default/timeline/timeline.php?tag=1960s#!&innovation=shakey-the-robot. Last accessed on August 2019.

[72] F. Lamiraux, S. Sekhavat, and J.-P. Laumond, "Motion planning and control for hilare pulling a trailer," *IEEE Transactions on robotics and automation*, vol. 15, no. 4, pp. 640–652, 1999.

[73] Science Museum Blog, "What to think about machines that think." https://blog.sciencemuseum.org.uk/what-to-think-about-machines-that-think/. Last accessed on August 2019.

[74] Wikidata, "Hilare-1 robot." https://www.wikidata.org/wiki/Q55139668. Last accessed on August 2019.

[75] iRobot Corp., "irobot roomba s series." https://media.irobot.com/media-kits?item=29. Last accessed on August 2019.

[76] A. Barrera, "Robot topological mapping and goal-oriented navigation based on rat spatial cognition," in *Mobile Robots Navigation*, IntechOpen, 2010.

[77] J. C. Andersen, *Mobile robot navigation*. PhD thesis, Technical University of Denmark, 2006.

[78] E. Consortium *et al.*, "Building the european robotics platform europ." https://ec.europa.eu/research/industrial_technologies/pdf/robotics-ppp-roadmap_en.pdf. Last accessed on August 2019.

[79] E. Ackerman, "irobot completely redesigns its floor care robots with new m6 and s9." https://spectrum.ieee.org/automaton/robotics/home-robots/irobot-completely-redesigns-its-floor-care-robots-with-new-m6-and-s9. Last accessed on August 2019.

[80] N. Karlsson, E. Di Bernardo, J. Ostrowski, L. Goncalves, P. Pirjanian, and M. E. Munich, "The vslam algorithm for robust localization and mapping," in *ICRA*, pp. 24–29, 2005.

[81] Le Roux, Jonathan, "Micbots: collecting large realistic datasets for speech and audio research using mobile robots." http://www.jonathanleroux.org/research/micbots.html. Last accessed on August 2019.

[82] S. Brodeur, "Create: Multimodal dataset for unsupervised learning and generative modeling of sensory data from a mobile robot." https://ieee-dataport.org/open-access/create-multimodal-dataset-unsupervised-learning-and-generativemodeling-senso Last accessed on August 2019.

[83] Y. Gu, A. Lo, and I. Niemegeers, "A survey of indoor positioning systems for wireless personal networks," *IEEE Communications surveys & tutorials*, vol. 11, no. 1, pp. 13–32, 2009.

[84] J. Reed, *An introduction to Ultra wideband communication systems*. Prentice Hall Press, 2005.

[85] M.-G. Di Benedetto, *UWB communication systems: a comprehensive overview*, vol. 5. Hindawi Publishing Corporation, 2006.

[86] R. ITU, "Characteristics of ultra wideband technology," 1755.

[87] A. Alarifi, A. Al-Salman, M. Alsaleh, A. Alnafessah, S. Al-Hadhrami, M. Al-Ammar, and H. Al-Khalifa, "Ultra wideband indoor positioning technologies: Analysis and recent advances," *Sensors*, vol. 16, no. 5, p. 707, 2016.

[88] D. Dardari, A. Conti, U. Ferner, A. Giorgetti, and M. Z. Win, "Ranging with ultrawide bandwidth signals in multipath environments," *Proceedings of the IEEE*, vol. 97, no. 2, pp. 404–426, 2009.

[89] M. Kwak and J. Chong, "A new double two-way ranging algorithm for ranging system," in *2010 2nd IEEE InternationalConference on Network Infrastructure and Digital Content*, pp. 470–473, IEEE, 2010.

[90] Braden, R., "Requirements for internet hosts - communication layers." https://tools.ietf.org/html/rfc1122. Last accessed on August 2019.

[91] M. M. Alani, *Guide to OSI and TCP/IP models*. Springer, 2014.

[92] W. R. Stevens, B. Fenner, and A. M. Rudoff, *UNIX Network Programming: The Sockets Networking API*, vol. 1. Addison-Wesley Professional, 2004.

[93] A. Watters, J. C. Ahlstrom, and G. V. Rossum, *Internet programming with Python*. Henry Holt and Co., Inc., 1996.

[94] A. G. Blank, *TCP/IP Foundations*. John Wiley & Sons, 2006.

[95] IBM Knowledge Center, "How sockets work." https://www.ibm.com/support/knowledgecenter/ssw_ibm_i_71/rzab6/howdosockets.htm. Last accessed on August 2019.

[96] iRobot Corp., "irobot create® 2 programmable robot." https://www.irobot.com/about-irobot/stem/create-2. Last accessed on August 2019.

[97] iRobot, "irobot® create® 2 open interface (oi) specification based on the irobot® roomba® 600," tech. rep., April 2015.

[98] Matthew Witherwax, "irobot. a python implementation of the irobot open interface.." https://github.com/julianpistorius/irobot. Last accessed on August 2019.

[99] T. E. Kurt, *Hacking Roomba: ExtremeTech*, vol. 48. John Wiley & Sons, 2006.

[100] Raspberry Pi Foundation. https://www.raspberrypi.org/. Last accessed on August 2019.

[101] Pozyx, "Store." https://www.pozyx.io/shop. Last accessed on August 2019.

[102] Pozyx, "Positioning accuracy." https://www.pozyx.io/product-info/creator/system-performance. Last accessed on August 2019.

[103] D. C. Niehorster, L. Li, and M. Lappe, "The accuracy and precision of position and orientation tracking in the htc vive virtual reality system for scientific research," *i-Perception*, vol. 8, no. 3, 2017.

[104] Minirig, "Minirig 2: Portable bluetooth speaker." https://minirigs.co.uk/speakers/bluetooth-minirig-2. Last accessed on August 2019.

[105] Zoom, "Zoom h3-vr handy recorder." https://www.zoom-na.com/products/field-video-recording/field-recording/zoom-h3-vr-handy-recorder. Last accessed on August 2019.

[106] M. Yavari, "Indoor real-time positioning using ultra-wideband technology," *Master's thesis, University of New Brunswick*, 2015.

A Localization Tests

This appendix describes the test performed with the indoor localization system in the experimentation room, see Section 4. The purpose of testing this system was to inspect the location accuracy. The test methodology recreated the behavior of a mobile robot, moving the tag across the room and setting onto a specific reference position. Five reference positions were selected around the room. A measurement consisted of moving the tag for 15 seconds and place it at the reference location, then the system reported its position and orientation every second for 45 seconds more. The total of experiments performed at a reference position were 30.

A.1 Environment and Equipment

The tests were implemented at the Aalto University meeting room, with dimensions of the room are 4935 x 4808 x 3000 mm. The equipment employed for the tests consisted on the indoor localization system (Pozyx), 6 anchors and 1 tag, the Raspberry Pi, a station, and a laser meter. The station was composed by a 1000 mm stand with a 400 mm cardboard box on top, the height of this structure was 1400 mm. The stands employed insulated materials for avoiding any type of interferences. The laser meter measured the reference points where the tag position was tested.

A.2 Methodology

The test consisted in the setup of 6 anchors locating 1 tag, during one minute, registering every second the position and orientation of the tag. The tag was in movement for the first 15 seconds and after that, it was set in a reference position. The anchors of the Pozyx were installed on the walls of the room as it is described below.

A.2.1 Anchor Placement

For the installation of the anchors, recommendations based on device manufacturer manuals and best practices during the tests were followed. The anchors were installed on the walls of the room maintaining a clear line of sight, pointing their antenna at the ceiling, placing them on insulator surfaces, such as wood, brick, glass, and setting the anchors at different distances, depths, and heights.

From the six anchors, groups of three anchors were installed on two parallel walls. A height 1400 mm functioned as reference for setting above and below the anchors. The distance between the three-grouped anchors was randomly assigned. The consideration was to form a triangle between the two parallel walls, therefore, these devices were intercalated. The depths on the anchors varied according to the surface that were mounted.

A point of origin, I, was set in the lower left corner of the room (0, 0), represented at the bottom right side as Figure A1 depicts. Based on this position, the rest of the anchors were installed as Table A1 contains.

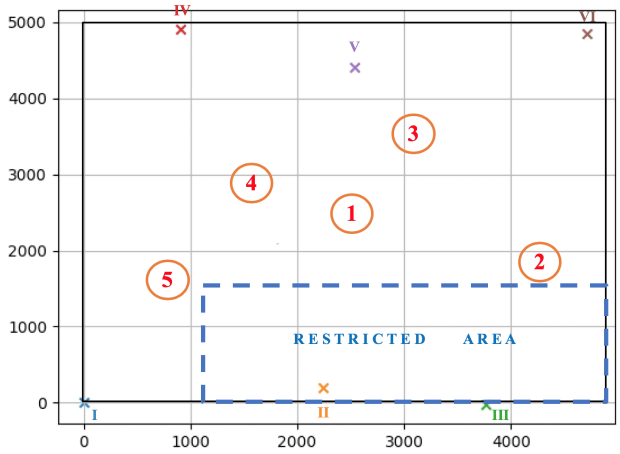

Figure A1: Map of the meeting room depicting six anchors, and five reference positions and dimensions of the room expressed in millimeters

Anchor	Distance (X)	Depth (Y)	Height (Z)
I	0	0	0
II	1393	0	1367
III	3424	-82	1494
IV	768	4550	175
V	2702	4550	1582
VI	4596	4525	1557

Table A1: Anchor position location on the tested room, distances expressed in millimeters

Pozyx was adjusted following the autocalibration function provided by the manufacturer, which consider as references four of the anchors of these system.

A.2.2 Reference Positions

Five random locations in the room were assigned as reference positions, as A1. During this test, the tag was rotated in increments of 90 degrees on each of the reference positions. Table A2 contains the five reference positions on the X and Y coordinates, the distance are expressed in millimetres.

The test methodology imitated the actions as the robot platform, when performing an acoustical measurement. A measurement consisted on randomly moving the tag across the room in for 15 seconds, then, the tag was set on the reference position for 45 seconds. This process was repeated 30 occasions at single reference position. In every experiment, the position of the tag was tracked constantly for one minute. Furthermore, moving the tag permits to update its position from previous experiments, and allows to observe its stabilization time.

Reference Position	X (Reference X)	Y (Reference Y)	Orientation
1	2486	2412	0
2	4254	1804	90
3	3023	3513	180
4	1667	2923	270
5	712	1455	0

Table A2: Reference positions implemented in the localization tests, distances expressed in millimiters.

A.3 Preliminary Results and Discussion

The preliminary results showed a better accurary in one of the axis compared to the other axis. It was noted that typically, the localization of Pozyx deviates from 80 mm to 400 mm from the reference value in X and Y axis, as Figure A2 shows. Furthermore, this preliminary results agreed with the test performed in other works evaluating Ultra-Wideband localization systems [106].

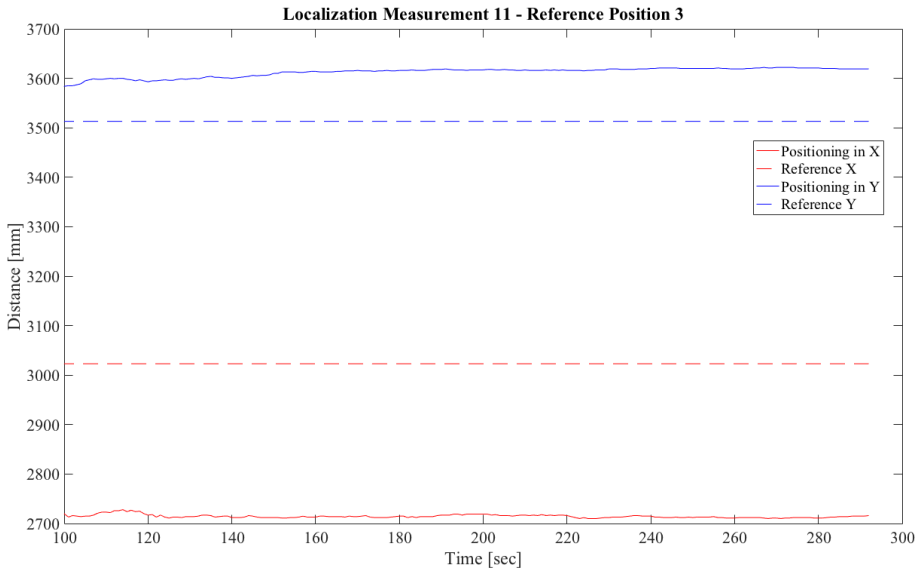

Figure A2: Time domain plot from a sample measurement of a reference position, displaying real time X and Y positions, and reference X and Y positions.

However, during the preliminary tests, the best cases agreed with the margin of accuracy reported by the manufacturer, 100 mm [102]. Figure A3 illustrates a histogram of a measurement point from a reference position. From Figure A3 that estimated position from the Pozyx tends to be constant, thus, the errors occur in a consistent value or around it.

The possible reason of major deviations might be caused from the materials present in the environment. The meeting room is partially built with metal materials as well as it has elements which are built upon metal. For instance, the gypsum

boards functioning as partitions are supported by metal studs, the air conditioning pipes hanging from the ceiling, materials placed on the wall, such as a white board and a metal sheet cover.

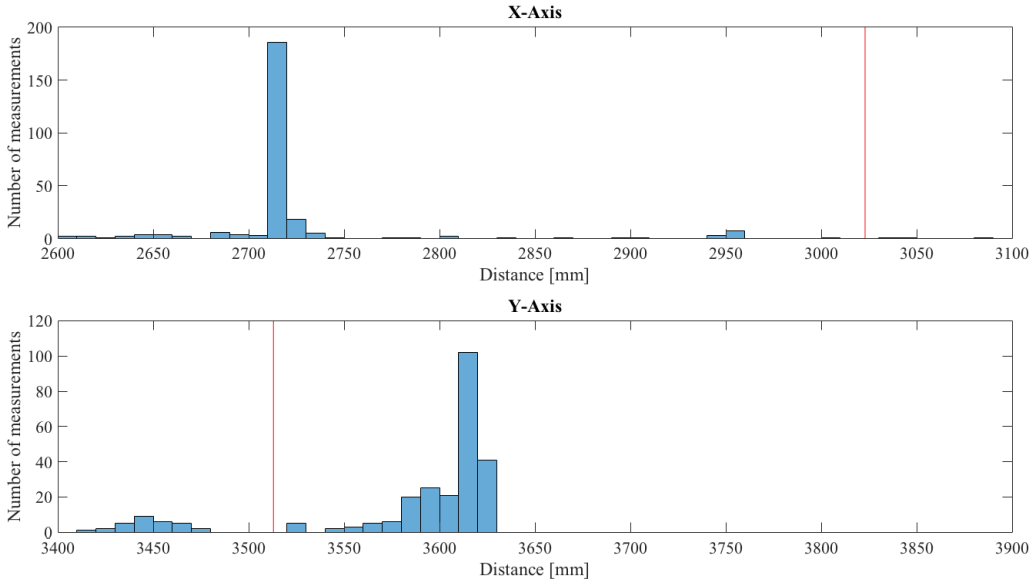

Figure A3: Histogram of the localization measurement 11, from the reference position 3

Additional test including, the stabilization of the measured location, the orientation of the target node, the implementation of rooms with different areas were performed. However, the process of these results remaind pending.

The localization system gives an approximate position of the robot. However, more accuracy is required to locate the robot. It is needed to finalize the processing of the data and found a representation to show the results. In the meantime, it is recommended to consider the spaces in which this system will be implemented.

B User Manual

1. Install the indoor localization system.
    - Take and note the dimensions of the room, arrangement and materials.
    - Decide the anchor placement arrangement.
    - Annotate heights and spacing of the anchors (X, Y, Z).
    - Supply electrical power to all anchors.
2. Prepare the robot mobile platform.
    - Ensure that the self-powered devices utilized are fully charged.
    - Mount devices equipped on the robot, such as the target node, loudspeakers, microphones.
    - Annotate distances and angles from the equipment implemented.
    - Connect the devices and ensure solid connections and tight all wires.
3. Establish wireless communication between the single-board computers.
    - Identify the IP addresses of each of the single-board computers.
    - Connect and establish communication with the single-board computer set as client. In this case, the computer equipped to the robot mobile platform. This communication must be established via remote desktop protocol, so that the graphical user interface properly displays the figures from the measurement script.
    - Connect and establish communication with the single-board computer set as servers. In this case, the computers supplied to the remote stations. Terminal can be used to login into the systems.
    - Establish sockets between the servers and client. It is recommended to initiate the programs from the client first, so that the socket would be ready to be established by the client.
4. Prepare and define AIRMS on the measurement script.
    - Type in the anchor positions and observations concerning to the room under study.
    - Capture panoramic picture from the room under study.
    - Define system parameters, such as number of measurements, speakers and remote stations, robot speed, speaker volume, IP addresses and ports of the remote stations.
    - Ensure the peripheral devices connected via USB are recognized by the single-board computer.
    - Transfer the measurement script to the single-board computer via FTP.

5. Initiate the measurement process.
    - Turn on the wheeled platform.
    - Initiate the recording on the microphones.
    - Ensure that the devices were recognized by the system.
6. Perform the post-processing.
    - Convert the measurement files, in this case, from pickle to mat.
    - Transfer the files from the single-board computer to other workstation via FTP.
    - Transfer the audio files from the microphones to the workstation via USB. Additionally, allocate and order all the transferred information into a local folder.
    - Run the post-processing script. Identify the calibration point from the recordings.
    - Input the calibration point and ensure the proper framing of sweeps signals.
    - Generate the file which contains all the information from the measurement process.
    - Once framed every sweep, and identified on the listing, the sweeps can be recalled to calculate their impulse response.
7. Integrate all the information. The automatically collected and the manually collected data should be allocated and order into a folder, following the data structure.