Circuits and Systems
Mekelweg 4,
2628 CD Delft
The Netherlands
http://ens.ewi.tudelft.nl/

M.Sc. Thesis

Towards the Integration of Acoustics and LiDAR

Ellen Riemens B.Sc.

Abstract

Loudspeakers are placed in an environment unknown to the loudspeaker designers. The room influences the acoustic experience for the user. Having information about the room makes it possible to better reproduce the sound field as intended. Using microphone measurements, the location of acoustic reflectors can be inferred. Current state-of-the-art methods for room boundary detection focus on a two-dimensional setting. Detection of arbitrary reflectors in three dimensions increase complexity due to practical limitations, i.e. the need for a spherical array and the increase of computational complexity. The presence of horizontal reflectors cause inaccuracy for wall detection due to model mismatch. Loudspeakers may not present an omni-directional directivity pattern, as usually assumed in the literature, thus making the detection of acoustic reflectors in some directions more challenging.

In this thesis, a LiDAR sensor is added to a smart loudspeaker to improve wall detection accuracy and robustness. This is done in two ways. First, the horizontal reflectors that are not present in the acoustic model are sought detected with the LiDAR sensor to enable elimination of their detrimental influence. Second, a method is proposed to compensate for the challenging regions for wall detection in highly directive loudspeakers, using the LiDAR sensor. Experimental results, evaluated in different simulated scenarios are shown for comparison of the proposed method and the state-of-the-art method, that exclusively uses acoustic information.

Towards the Integration of Acoustics and LiDAR

A LiDAR-aided approach for detection of acoustically reflective surfaces from microphone measurements

---

Thesis

submitted in partial fulfillment of the
requirements for the degree of

Master of Science

in

Electrical Engineering

by

Ellen Riemens B.Sc.
born in 's-Gravenhage, The Netherlands

This work was performed in:

Circuits and Systems Group
Department of Microelectronics
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology

Delft University of Technology

Delft University of Technology
Department of
Microelectronics

The undersigned hereby certify that they have read and recommend to the Faculty of Electrical Engineering, Mathematics and Computer Science for acceptance a thesis entitled **"Towards the Integration of Acoustics and LiDAR"** by **Ellen Riemens B.Sc.** in partial fulfillment of the requirements for the degree of **Master of Science**.

Dated: 06-09-2021

Chairman: _____
dr.ir. R.C. Hendriks

Daily Supervisor: _____
dr. J. Martínez-Castaneda

Committee Members: _____
dr. M. Mastrangeli

_____
dr. P. Martínez-Nuevo

_____
M. Bo Møller, PhD

Abstract

Loudspeakers are placed in an environment unknown to the loudspeaker designers. The room influences the acoustic experience for the user. Having information about the room makes it possible to better reproduce the sound field as intended. Using microphone measurements, the location of acoustic reflectors can be inferred. Current state-of-the-art methods for room boundary detection focus on a two-dimensional setting. Detection of arbitrary reflectors in three dimensions increase complexity due to practical limitations, i.e. the need for a spherical array and the increase of computational complexity. The presence of horizontal reflectors cause inaccuracy for wall detection due to model mismatch. Loudspeakers may not present an omnidirectional directivity pattern, as usually assumed in the literature, thus making the detection of acoustic reflectors in some directions more challenging.

In this thesis, a LiDAR sensor is added to a smart loudspeaker to improve wall detection accuracy and robustness. This is done in two ways. First, the horizontal reflectors that are not present in the acoustic model are sought detected with the LiDAR sensor to enable elimination of their detrimental influence. Second, a method is proposed to compensate for the challenging regions for wall detection in highly directive loudspeakers, using the LiDAR sensor. Experimental results, evaluated in different simulated scenarios are shown for comparison of the proposed method and the state-of-the-art method, that exclusively uses acoustic information.

Contents

Abstract **v**

1 Introduction **1**
1.1 Research Question and Outline 2

2 Problem Description **5**
2.1 Background Theory 5
2.1.1 Room Impulse Response 5
2.1.2 Image Source Method 6
2.1.3 Sensing modalities 7
2.1.4 Point clouds 9
2.1.5 Hough Transform 10
2.2 Prior Art 10
2.2.1 Room Boundary Estimation 10
2.2.2 Plane detection from point clouds 11

3 Method **13**
3.1 Problem Scenario 13
3.2 State-of-the-art approach for reflecting surface detection 14
3.2.1 Discrete measurement model 14
3.2.2 Inverse problem 16
3.3 State-of-the-art approach for planar surface detection from a point cloud 17
3.3.1 Clustering 18
3.3.2 Computing Gaussian Kernels 18
3.3.3 Spherical Accumulator 19
3.3.4 Kernel-based Voting 19
3.3.5 Peak detection 19
3.3.6 Planar surface equation 20
3.4 Proposed method: Combined approach 20
3.4.1 Compensation for detected horizontal reflectors from the point cloud 20
3.4.2 Loudspeaker Directivity Compensation 21
3.4.3 Complete Algorithm for detection of acoustically reflecting surfaces 23

4 Results **25**
4.1 Experiment 1 — Single wall scenario 26
4.2 Experiment 2 — Scenario with windows 28
4.3 Experiment 3 — Detection of a wall in presence of a floor 31

5 Conclusion & Discussion **33**
5.1 Discussion 33
5.2 Future Work 34

List of Figures

2.1 Simplified example of a room impulse response with the direct sound in red, the early reflections in yellow and the late reverberation in green. Taken from [1]. 5
2.2 Sketch of mirrored source in a reflective surface. 6
2.3 Scanning LiDAR, taken from [2] 9
2.4 Flash LiDAR, taken from [2] 10

3.1 Problem Scenario. The microphones are placed in a circular array colocated with the loudspeaker. The LiDAR sensor is oriented towards the back of the loudspeaker. 14
3.2 Grid of candidate source locations from [3]. The co-located system is placed in a room. The corresponding image sources are visualized on the radial grid of candidate locations. 15
3.3 Loudspeaker directivity visualization [3] 15
3.4 Microphone array response [3] 16
3.5 Workflow for the D-KHT. First the depth image is divided into approximately coplanar clusters. Next, the accumulator is incremented in the voting procedure based on the clusters. Finally, peaks are detected in the accumulator resulting in the detection of planes. 18

4.1 Top view of system setup 25
4.2 Side view of system setup 26
4.3 The setup used for experiment 1. The thick black line illustrates a wall that is placed at different angles α around the system. 26
4.4 Directivity characteristic of the loudspeaker. The magnitude scaling used for the directivity. In the back of the loudspeaker the energy is lower than in the front. 27
4.5 Detection of a rotating wall around the co-located system. The mean hitrate is shown against SNR. The performance at $0°, 60°, 120°$ and $180°$ is shown in blue, orange, yellow and purple respectively. The challenge of detection of the walls in the low-energy regions of the speaker is resolved with the proposed method. 27
4.6 Detection of a rotating wall around the co-located system. The mean hitrate is shown against SNR. The performance at $0°, 60°, 120°$ and $180°$ is shown in blue, orange, yellow and purple respectively. The challenge of detection of the walls in the low-energy regions of the speaker is not completely resolved as it was using method 1, but the hitrate has increased compared to Zaccá's method. 28
4.7 experiment 2 29

4.8 Left: Detection of a wall at different angles using Zaccá's method, while simultaneously attempting to detect a window. Right: Detection of a window at different angles using Zaccá's method, while simultaneously attempting to detect a wall. 30
4.9 Left: Detection of a wall at different angles using Zaccá's method, while simultaneously attempting to detect a window. Right: Detection of a window at different angles using Method 1, while simultaneously attempting to detect a wall.The detection of the wall improves with the proposed method, while the detection of the window goes to zero. . . . 30
4.10 Left: Detection of a wall at different angles using method 2, while simultaneously attempting to detect a window. Right: Detection of a window at different angles using method 2, while simultaneously attempting to detect a wall. The detection of the wall improves with the proposed method, while the detection of the window remains similar. 31
4.11 Side view of the set-up for Experiment 3. The co-located system is placed at height d_{f} at a distance $R = 1$ m from a wall at angle $\alpha = 0°$ 31
4.12 Left: The detected distance R for the method of Zaccá [3] and the proposed method at different floor heights. The ground truth is $R = 1$. Right: The mean square error with the ground truth normal vector for Zaccá's method and the proposed method at different floor heights. While the floor is in the LiDAR FOV, it is possible to detect the wall reliably. 32

1 Introduction

Recently, more time has been spent at home than ever [4][5]. This has led to an increased interest luxury home experiences, including high quality music reproduction[6][7]. The room has a large influence on the sound field reproduction, meaning the highest quality can only be obtained when the room influence is known. When a loudspeaker is placed closely to a wall or corner, the lower frequency range in the room is amplified compared to the mid and high frequency, resulting in an unbalanced sound experience [8].

Working from home also led to a large increase of teleconference meetings. In order to use teleconference meetings as a feasible alternative to in-person meetings, speech intelligibility is a crucial factor. Speech intelligibility can be degraded by echoes introduced by the room, making it crucial to be aware of the nearby walls that introduce echoes [9].

The introduction of smart loudspeakers gives rise to opportunities to obtain extra information to improve the sound experience of the user. The important piece of information is the proximity of walls. If the wall locations are known, their effects can be compensated for using digital filters [10][11].

How is it then possible to detect the walls in close proximity to the speaker? Modern smart loudspeakers have microphone arrays and other sensors built-in that can be exploited for this purpose. The microphone array can together with the loudspeaker detect the reflective surfaces. The general principle of this process works as follows. The loudspeaker emits a known sounds though its drivers. The sound travels through the room and reflects on the large surfaces. The microphones on the loudspeaker measure the emitted signal and its reflections. Since the configuration is known and constant over time, the direct path contribution can be eliminated. Then, from the reflections the reflective surfaces can be found. The distance to the wall can be determined by estimating the time-delay via the wall to the microphone array. The angle of the wall is determined by the time difference of arrival of the echoes between the microphones in the array. In practice, loudspeakers do not emit an equivalent amount of energy in all directions; they are directive [12]. This makes it more challenging to detect surfaces in low-energy regions, i.e. the back of the loudspeaker [3]. In addition to that, in practice the detection is limited to the horizontal plane due to the limits on computational complexity.

Is it then possible to take advantage of the presence of multiple sensors in a smart loudspeaker? Other sensing modalities have been equipped to detect walls in numerous applications.

An example of this is sound in the non-audible range; the ultrasonic range. This includes ultrasonic parking detectors and Sonar. Sound sources become more directive as frequency increases. In order to measure reflectors with an ultrasonic transducer, reflectors must be approximately perpendicular to the transducer. Many transducers

are needed to then successfully detect acoustic reflectors [13]. Other modalities rely on electromagnetic waves, either in the radio-frequency (RF) range or in the light range. In the RF range, Radar or ultra-wideband (UWB) sensing is equipped. The RF signals partially reflect on walls and partially penetrate. This makes them suitable for behind-the-wall sensing as well as mapping applications [14], but not the obvious candidate for wall detection applications. In the light range, passive sensors and active sensors are employed. In the visible range the camera can be used to reconstruct walls using image recognition algorithms where object with known geometry can be used as reference to obtain dimensions or machine learning algorithms can be used for depth estimation [15]. The stereo camera construction avoids this by exploiting the known geometry of the system and casting depth estimation into a stereo matching problem, where pixels in two images need to be matched [16]. Active sensors such as Light Detection and Ranging (LiDAR) illuminate the scene with light in the non-visible range. The time-of-flight for this light to return to the sensor in combination with the angle of arrival give the coordinates of the points of refraction [17]. Since the infrared light refracts on most surfaces, it is not required for object to be perpendicular to the sensor to be detected. LiDAR sensors do not perform well under natural lighting conditions due to noise from the sun [18]. It was decided that LiDAR sensor is suitable in this work, due to the indoor application. The processing cost to obtain depth information is low compared to other camera solutions.

In this thesis, an approach is presented where a co-located system of a loudspeaker, a microphone array and a LiDAR sensor is used to robustly detect acoustic reflective surfaces. This system is constructed in such a manner that the region where acoustic detection is challenging is aided with the LiDAR sensor. The proposed algorithm also uses the information from the LiDAR sensor to eliminate the effect of the reflections from non-vertical reflectors, causing a more robust detection of walls in presence of such reflectors.

1.1 Research Question and Outline

In this thesis, the following general research question is addressed:

Can the robustness of acoustically reflective surface detection using a microphone array be increased by adding information from a LiDAR Sensor?

- How can the information from a LiDAR sensor be equipped to compensate for the reduced energy in the backside of a loudspeaker?
- How can the accuracy of acoustic detection be improved by including information from the LiDAR sensor about the 3D scenario?

The rest of the thesis continues as follows. Chapter 2 provides the necessary background information, describes the prior art and presents the problem statement. In Chapter 3 the state-of-the-art methods for acoustic reflective surface detection and a method for plane extraction from point clouds are presented in depth. A combined approach that improves the robustness is proposed. Chapter 4 presents simulation results that evaluate the performance of the proposed method with the state-of-the-art

method and investigates the performance in a real-world scenario. Finally, Chapter 5 presents the discussion and conclusion as well as future research directions.

2 Problem Description

A smart loudspeaker consists of at least one microphone and loudspeaker driver in one co-located system, where a supplementary LiDAR sensor is added to this system. The challenge is to estimate the nearby walls. The emitted sound is assumed to be known, as well as the co-located system configuration, i.e. the position of the loudspeaker, microphones and LiDAR with respect to each other.

This chapter presents the background theory needed for the work in this thesis and a brief overview of prior art for acoustic estimation of reflective surfaces and plane detection from point clouds.

2.1 Background Theory

2.1.1 Room Impulse Response

The room impulse response is the transfer function from the room and depends on the room dimensions, the properties of the room boundaries and the location of transmitters and receivers. It consists of the direct path from the transmitter to the receiver and from wall reflections. These wall reflections can be first, second or higher order. The first part of the Room Impulse response is called the early reflections, this part contributes most strongly and is the part that can be used to recover the location of the reflectors. The late (higher-order) reflections are typically perceived as reverberation and can mask the important components of the signal. A simplified example of such a room impulse response is shown in Figure 2.1. The room impulse response can be modeled in several ways [19][20]. Commonly, the Image Source Method [20] is used, as described in subsection 2.1.2.

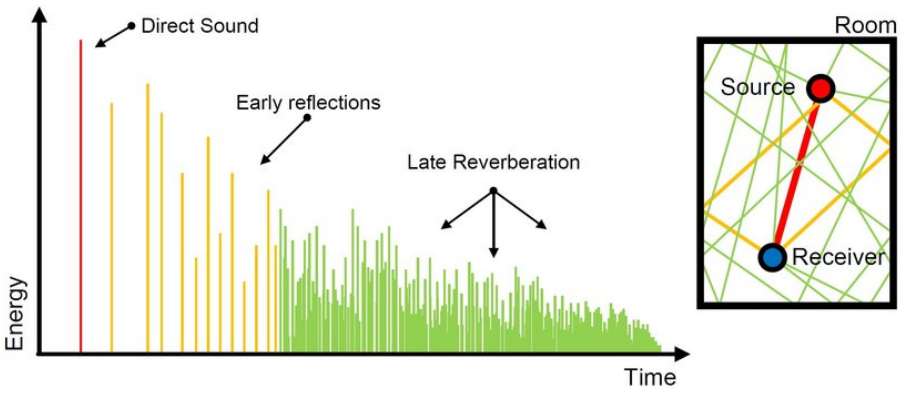


Figure 2.1: Simplified example of a room impulse response with the direct sound in red, the early reflections in yellow and the late reverberation in green. Taken from [1].

2.1.2 Image Source Method

In order to model the acoustic room transfer function, the image source method (ISM) is used. This method assumes specular reflections of sound in the walls to find the transfer function between an omnidirectional point source to an omnidirectional point receiver. It is based on the construction of a virtual point source that is across the wall, substituting the sound source itself. This can be applied for the first-order reflection that are of interest here, but extends to higher order reflection. The Room Impulse Response can be modeled with the image source model for arbitrary polyhedra [21].

The image sources are found by mirroring the source in the reflective surface, as visualized in Figure 2.2. The plane can be described by its normal vector $\boldsymbol{\nu}$ and the distance from the origin s. The distance from point $\mathbf{s}$ to the plane is given by $d = s - < \mathbf{s}, \boldsymbol{\nu} >$. An expression for the position vector of the image point $\mathbf{s}_I$ is then $\mathbf{s}_I = \mathbf{Q} + 2d\boldsymbol{\nu}$. Generating virtual walls allows the computation of higher order reflections by mirroring image sources.

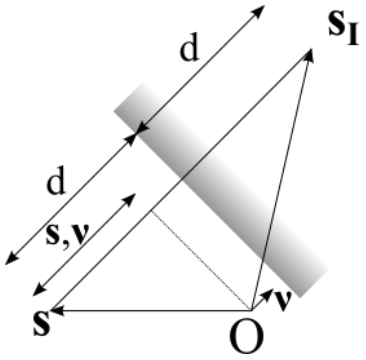


Figure 2.2: Sketch of mirrored source in a reflective surface.

The generated image sources for a shoebox shaped room are visualized in figure for reflections up to the second order. The distance from the image source to the receiver is equivalent to the distance from the original source to the receiver via the reflector.

The time-delay of arrival (TOA) at receiver $\mathbf{r}$ of a sound ray from (image) source position $\mathbf{s} \in \mathbb{R}^3$ is given by

$$\tau = \frac{d}{v_c} = \frac{||\mathbf{s} - \mathbf{r}||_2}{v_c},$$

where v_c is the speed of sound.

The Room Impulse Response can then be constructed using Equation 2.1, where γ_i is the attenuation coefficient corresponding to the i-th wall, s_i is the position of the i-th (image) source and τ_i is its associated TOA. $\mathcal{S}$ is the set of considered (image) sources.

$$g(\mathbf{r}, \mathbf{s}, t) = \sum_{i=0}^{\mathcal{S}} \gamma_i \frac{\delta(t - \tau_i)}{4\pi||\mathbf{r} - \mathbf{s}_i||_2} \tag{2.1}$$

2.1.3 Sensing modalities

Different sensing modalities were explored with the purpose of finding a sensing modality that complements the acoustic method best. The main considerations are the necessity of finding planes that have the acoustic properties of a wall. The main limitations of acoustic methods are that detection is limited to one plane due to computational complexity and practical constraints and that reflections from floor are then disturbances, the directivity that obstructs detections of wall in low-energy regions and the presence of higher-order reflections causing 'ghost' walls. For the explored sensing modalities, a distinction is made between modalities that rely on sound waves in subsubsection 2.1.3.1, and modalities relying on electromagnetic waves in subsubsection 2.1.3.2.

2.1.3.1 Ultrasound-based

There are several wall-sensing methods that are, just like the audible sound measurement methods, based on the reflection of sound. The first is ultrasonic sensors, such as used in parking sensors. These sensors transmit narrow-band ultrasound pulses and measure the time-of-flight from the reflection. Since ultrasound is highly directive, reflections can only be measured if the reflecting surface is approximately perpendicular to the sensor (8 degrees) [22]. Another problem arises with the size of the objects that reflect. If a small object is obstructing the field of view, it is impossible to recover the large objects that affect the sound field reproduction. For example, if a vase is placed in front of the ultrasonic transducer, an object is detected that does not have a large influence on the sound field in a room, whereas the large reflector, i.e. the wall, is missed.

Sonar uses a wideband ultrasonic beam and is historically inspired by bats. In Bat-G net [23], an exponential sine sweep ultrasound signal over a frequency range of 20-120 kHz, similar to that of bats, is emitted. Four receivers are used to measure the reflections and the image is reconstructed with a neural network.

A benefit of the ultrasonic approaches is that in most indoor environments the ultrasonic interference is low. Also, the acoustic reflective properties of the room are measured, which is the relevant information in the application. However, the acoustic properties depend on frequency, and since sensing is done with a different frequency than in the application, this might not be accurate. Error can be introduced since the speed of sound is dependent on the temperature and pressure in the environment, however this error is assumed to be negligible, since in this project, the focus is on indoor environments. In indoor environments, the range is short and the temperature is fairly constant. Also, the issue of interfering higher order reflections is still present, like in the audible range acoustic method.

2.1.3.2 Electromagnetic waves

The use of electromagnetic (EM) waves as opposed to sound, has several inherent benefits. EM waves travel with the speed of light, which is much higher than the speed of sound, creating a potential for faster imaging. However, in the stationary home environment this is not a concern. The speed of EM waves is not dependent on

environmental circumstances which can be the case for acoustic waves. EM waves also experience less attenuation when traveling through air compared to acoustic waves, creating stronger echos[24][25]. The distinction is made between radio-frequency (RF) and light spectrum waves.

Radio-frequency wave-based The same principles as used in the audio-based approaches, can also be applied with RF waves. A sine sweep radio signal can be transmitted, from which the reflections are measured with an antenna array. The signal partially travels through the wall and partially reflects. This means that if there is another reflector behind the wall, a reflection from this can be mistakenly seen as a wall. Also, this modality also suffers from interference from higher order reflections [26].

Ultrawide band (UWB) radar is used for autonomous reconstruction of permanent structures [27], for example in SLAM (Simultaneous Localization and Mapping) applications, where a robot used the UWB sensors together with inertial measurement units to find a map and localize itself in it.

There is also potential in existing communication technology that is already present in the environment, such as Wi-Fi. The advantage here is that a module is present in every home, as well as on modern speakers. One of the challenges here is that with the Wi-Fi protocols, it is difficult to get access to the raw signal [28].

Light-based Cameras are widely available through smartphones. With the help of image recognition, it is possible to detect objects. This could be exploited by detecting a speaker and the surrounding walls. The benefit of the approach is the availability, however there are several obstacles. In this approach, depth is not measured, meaning the distance from a speaker to a wall cannot be quantified. Another downside is that performance relies heavily on lighting conditions [29].

When two cameras are used in a stereo camera setup, it is possible to determine depth. Objects are recognized in the images from both cameras, and by matching pixels, the displacement of the pixels can be found. When the cameras are calibrated, this displacement can be used to find the depth in the form of a point cloud [30]. Again, the performance of this method decreases when the light exposure decreases. Both the single and the stereo camera system require extra processing steps because of the image recognition algorithms that are needed.

LiDAR sensors emit light in the infrared spectrum. The Field-of-view (FOV) is illuminated by a beam, and the reflections are measured. Based on the time-of-flight (TOF), the distance to surfaces is determined. It is clear that this approach overcomes the limitation of poor lighting conditions that are present in other camera approaches [17]. However, in the presence of sunlight, which contains infrared light, the performance of this method degrades. Also, windows and mirrors cannot be sensed using LiDAR because these surfaces do not scatter the incident light [22].

With all camera-based solutions, privacy is a larger issue than other solutions.

2.1.3.3 Comparison

The above modalities can be evaluated in terms of how well they complement acoustics in the audible range. As discussed, a difficulty in acoustics is that higher order reflections cannot easily be distinguished from the first order reflections that the models need, causing 'ghost' walls. Both RF-wave-based sensors also suffer from this. Ultrasonic sensors need surfaces to be perpendicular to the sensor to be able to sense them. Camera-based methods do not have these limitations. In terms of computational complexity, RF-wave based methods are very similar to acoustics as well. Ultrasound sensors are very low power and low-complexity, whereas in a stereo camera setup, first image recognition needs to be done to find the depth information, and then a surface extraction algorithm is equipped. A LiDAR sensor directly returns a point cloud, on which a surface extraction algorithm is applied.

It is concluded that a point cloud representation of the environment would complement the acoustic information. This gives the possibility to use the stereo camera or a LiDAR sensor. Since the LiDAR sensor has less processing steps to obtain the point cloud and this is an indoor application, the LiDAR sensor is used in this application.

2.1.4 Point clouds

A point cloud is a set of data points $\mathbf{p} = (x, y, z)$ in space. The point cloud is often used to express depth measurements. These measurements are typically obtained from LiDAR sensing. There are two types of LiDAR sensors: the scanning LiDAR and the flash LiDAR.

2.1.4.1 Scanning LiDAR

A Scanning LiDAR sensor emits a narrow laser beam in a certain direction. The laser refracts on a surface it coincides with, this refraction is measured at the sensor. The time the light is travelling is combined with the information of the angle of transmission, to retrieve the refraction-point. The Scanning LiDAR scans a grid of different angles, which results in the point cloud. The process is visualized in Figure 2.3. This point cloud is unorganized, meaning it points that are in close in the point cloud, are not necessarily close in space.

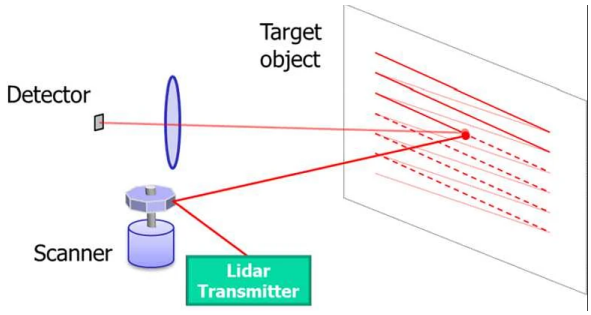


Figure 2.3: Scanning LiDAR, taken from [2]

2.1.4.2 Flash LiDAR

The Flash LiDAR illuminates the scene with a wide diverging laser. Again, the light refracts on the coinciding surfaces. Now, a detector grid is used with a lens, focusing the refractions on the detector grid. This procedure is visualized in Figure 2.4. This point cloud is naturally organized due to the detector grid, similar to an RGB camera, but can also be used in an unorganized way.

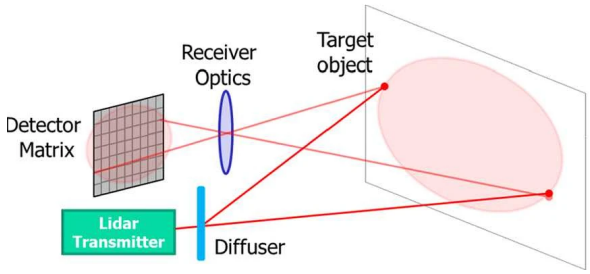


Figure 2.4: Flash LiDAR, taken from [2]

2.1.5 Hough Transform

The Hough transform is used in digital image processing to extract parametric curves [31]. The Hough transform to extract planes is given by Equation 2.2.

$$x \cdot \cos\theta \cdot \sin\varphi + y \cdot \sin\varphi \cdot \sin\theta + z \cdot \cos\varphi = \rho \tag{2.2}$$

For each point (x, y, z), the transform is done to the Hough space (ρ, φ, θ). The plane that is spanned by three points (x, y, z), can be found by finding the intersection of the curves in Hough space. For all points that lie in the same plane, the curves in Hough space intersect in one point (ρ, φ, θ). The value of ρ at this point, gives the distance of this plane to the origin in Euclidean space. The values of ρ and θ describe the direction of the plane normal in the following way:

$$\boldsymbol{\nu} = \begin{pmatrix} \nu_x \\ \nu_y \\ \nu_z \end{pmatrix} = \begin{pmatrix} \sin\varphi\cos\theta \\ \sin\varphi\sin\theta \\ \cos\varphi \end{pmatrix} \tag{2.3}$$

2.2 Prior Art

The focus of this thesis is on combining a loudspeaker with built-in microphones with a LiDAR camera in order to estimate acoustic reflectors. The literature on acoustic reflector location estimation using microphones is described in subsection 2.2.1, while an overview of approaches for plane detection from point clouds is given in subsection 2.2.2.

2.2.1 Room Boundary Estimation

Estimating the locations of acoustic reflectors from an acoustic measurement is typically done in the following steps. First, the channel is estimated using a channel estimation

algorithm, where the excitation signal is known. Here, the channel is both the room impulse response and the response of the system. Then, the (known) loudspeaker influence needs to be removed using deconvolution, to obtain the Room Impulse Response for each microphone. The peaks in the RIR can be detected and consecutively the acoustic echoes should be labeled. This is done according to which wall produced them. Most methods available in the literature for obtaining the acoustic reflectors assume that the RIR is known and focus on the echo labeling problem. Dokmanić et al. in [32] aims to reconstruct a convex polyhedral room from impulse responses exploiting the properties of Euclidean distance matrices (EDMs) in the general 3D case. Since this combinatorial problem is NP-hard, the computational complexity is high. De Jager et al. proposed in [33] a method that solves this problem at the same accuracy but much lower complexity by posing it as a graph. Coutino et al. proposed a greedy method to further reduce this computational complexity in [34]. Even though the recent algorithms are much faster, this is still a limitation. In addition to that, these methods rely on perfect peak detection from the RIR. More recently, Zaccà et al. [3] posed this problem as a linear system. This system directly maps the image source locations to the received impulse response, including the loudspeaker directivity. This method does not rely on peak detection from the impulse response, however the computational complexity of solving the inverse problem limits the method in practical applications to 2D. In addition to that, the results from Zaccà et al. show that the performance of the wall detection algorithm is improved compared to other methods, due to a one-step approach that is used that does not rely on peak detection. Also, assuming a directive loudspeaker model as opposed to an omnidirectional one improves the accuracy of the model. However, the performance in the low directivity regions of the loudspeaker i.e. the back is considerably lower than would be desirable.

2.2.2 Plane detection from point clouds

Plane detection from point clouds can be separated into model fitting approaches and region growing approaches. The model fitting approaches assume the plane to be modeled using the plane equation in Equation 2.4.

$$\nu_x x + \nu_y y + \nu_z z = \rho \tag{2.4}$$

The first category of popular model fitting approaches are Random Sample Consensus (RANSAC) [35][36] approaches. In RANSAC, randomly a sample of the minimum number of points to fit the model is drawn from the point cloud. For detecting a plane, three points from the point cloud are selected and for all other points it is determined whether they fit this model or if they are outliers. This procedure is repeated for several iterations and afterwards the model for which the least points were considered outliers is chosen to be the plane. A disadvantage of RANSAC is that when the number of iterations is limited, the solution may not be optimal. In addition to that, using this approach the solution cannot be obtained analytically, which could be exploited when combining information of multiple modalities. In order to find multiple surfaces with RANSAC, the algorithm requires multiple runs and merging of detected planes. Considering these fundamental limitations of RANSAC approaches, this category is not further considered.

The second popular category of model fitting approaches is based on the Hough Transform. The procedure consists of four steps:

1. **Transform to Hough space** The Hough transform that is described in subsection 2.1.5 is used to convert all points in the point cloud to the 3D Hough space, for the Standard Hough Transform (SHT) [31].

2. **Increment an accumulator** The accumulator is a discretization of the Hough space. Each voxel in this discrete space is called a cell. The cells in the accumulator are empty at the start of the process. The value of a cell is incremented by one if a surface intersects this cell. For each surface, all cells in the accumulator are incremented by one, this process is often called voting.

3. **Find the local maxima** After the voting is done, the winning cells are selected by finding the local maxima.

4. **Transform back to Euclidean space** The winning cells are converted back to Euclidean space to find the plane equations.

This procedure is computationally very demanding, resulting in many adaptations of this algorithm.

The Probabilistic Hough Transform (PHT)[37] and the Adaptive Probabilistic Hough Transform (APHT)[38] use a random selection of points from the point cloud instead of using all points, whereas the Randomized Hough Transform (RHT)[39] uses an approach where randomly three points from the point cloud are selected. The accumulator cell corresponding to the plane spanned by the three points is incremented. If there is a large plane present in the point cloud, it is likely that the randomly selected points lie in that plane. Once a cell reaches a certain threshold value, it is classified as a plane and the corresponding points are removed from the point cloud. These methods depend heavily on the choices of parameters, such as number of iterations or number of selected points. Another approach is based on clustering, such as the Depth Kernel-based Hough Transform (D-KHT). First, an attempt is made to cluster the point cloud into approximately co-planar clusters. The accumulator is incremented based on the mean and variance of each cluster. Then the local maxima are found by employing a hill-climbing strategy starting from the means of the clusters. This approach leads to an analytical solution with a reduced computational demand.

3 Method

So far the importance of identifying reflecting surfaces have been discussed, and the challenges of the identifying such surfaces have been highlighted. Consecutively, the opportunity in using a complementary sensing modality in the form of a LiDAR sensor has been presented. The problem that is addressed in this thesis is that of combining the sensor information in a meaningful way, considering individual limitations and strengths.

Consider a system with one loudspeaker, a microphone array with N microphones and a LiDAR system, in a known geometry. The system is located in an unknown room and can actively sense the room, both by emitting a sound signal and scanning with LiDAR. The acoustic transmitter and receiver are assumed to be synchronized and coincide geometrically. The room response is modeled using the image source method. The point cloud from the LiDAR sensor has a limited and known field of view.

Building on the work from [3], the information from the point cloud is included to obtain a more robust solution. This chapter will first define the setup in the room, after which the state-of-the-art methods in both acoustic reflecting surface detection and plane detection from point clouds are presented. Following, the proposed method combining information from both sensing modalities for overcoming limitations regarding directivity is given in section 3.4, as well as how to use the information about the horizontal surfaces.

3.1 Problem Scenario

Consider a smart loudspeaker system with a uniform circular array (UCA) of radius r with M microphones and a LiDAR sensor with a predefined FOV of $\beta^{\circ}_{\text{hor}} \times \beta^{\circ}_{\text{ver}}$. The coordinate system is defined such that the center of the microphone array, the loudspeaker point source and the LiDAR sensor are at the origin, $\theta = 0^{\circ}$ corresponds with the direction at which the loudspeaker transmits on axis. The LiDAR is oriented such that the center of its FOV is at 180°. The method is generalized for any number of walls, as an example the shoe-boxed shaped room is taken.

In Figure 3.1 the corresponding problem scenario is given. The loudspeaker and the LiDAR sensor are located in the center of the microphone array where the number of microphones $M = 6$. The positive horizontal axis is $\theta = 0$. The red dashed lines indicate the horizontal field of view of the LiDAR sensor $\beta^{\circ}_{\text{hor}}$ and is centered around $\theta = 180^{\circ}$. The black lines describe the shoe-boxed shaped room, and are the boundaries that need to be detected. Now, given the sensor inputs and the directivity response of the loudspeaker, the problem is to detect the walls.

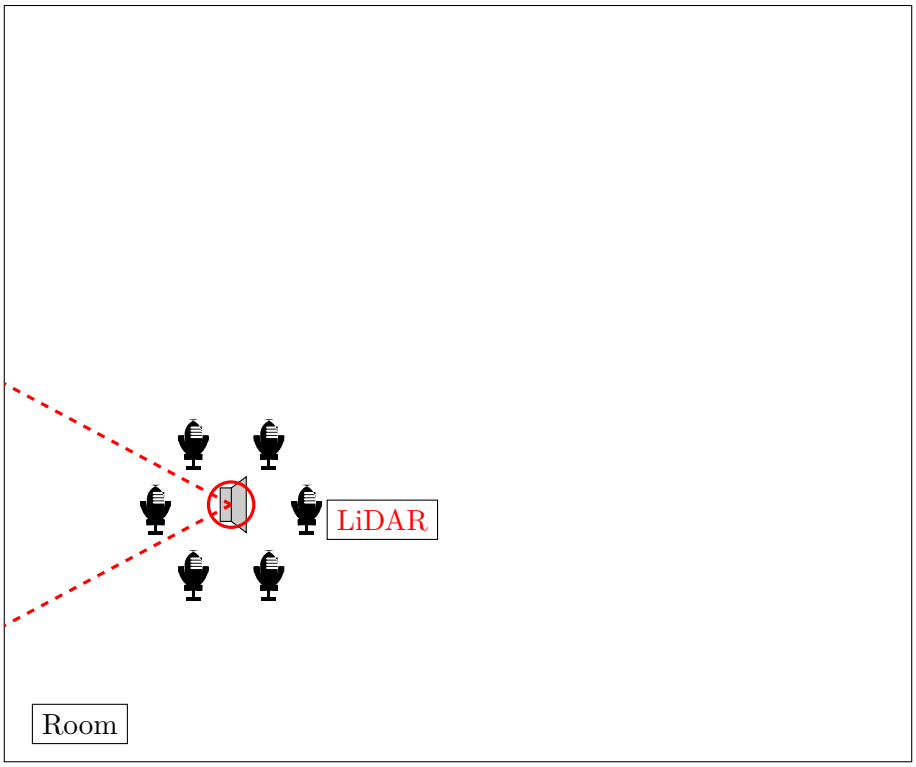


Figure 3.1: Problem Scenario. The microphones are placed in a circular array colocated with the loudspeaker. The LiDAR sensor is oriented towards the back of the loudspeaker.

3.2 State-of-the-art approach for reflecting surface detection

The method from [3] is given in Chapter 2 as the state-of-the-art for acoustic surface detection. In this method, a microphone signal model was presented that maps the location of image sources to microphone measurements. The signal model can be described in the continuous domain and can be formulated in matrix-vector form after discretization. Estimating the location of reflecting surfaces is solved as an inverse problem. In the following sections, this model is described and analysed in-depth, following the publication closely.

3.2.1 Discrete measurement model

The system can be described in three main components, the first being the image source locations, the next the loudspeaker directivity and finally the microphone array response. It is assumed that a discrete version of the received impulse response $h(t, \theta)$, $h[n, m]$ is accurately known, as well as the loudspeaker directivity and the microphone array geometry. The measurements are sampled in time at f_s Hz. The direct path of is assumed to be removed in a pre-processing stage. The time steps are given as $n = 0, ..., N_h - 1$ and the microphone indices are $m = 0, ..., M - 1$.

Image source locations The image source locations are determined on a grid in polar coordinates, as shown in Figure 3.2.

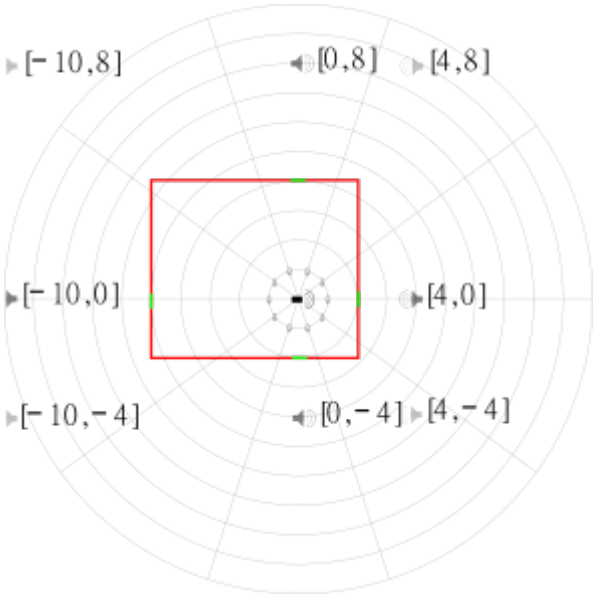


Figure 3.2: Grid of candidate source locations from [3]. The co-located system is placed in a room. The corresponding image sources are visualized on the radial grid of candidate locations.

The radial step size is given by $\Delta R = \frac{v_c}{f_s}$ and the angular step size by $\Delta\tau = \frac{2\pi}{M}$. The candidate locations are between $R_{\min} = R_a$ and $R_{max} = \frac{Tv_c}{f_s} + R_a$ for an integer T. If an image source location is not present in the discrete set, it is assigned to the closest grid point in the set.

Loudspeaker directivity The loudspeaker directivity $\gamma_0(t, \theta)$ is characterized for the discrete transmit angles, as visualized in Figure 3.3.

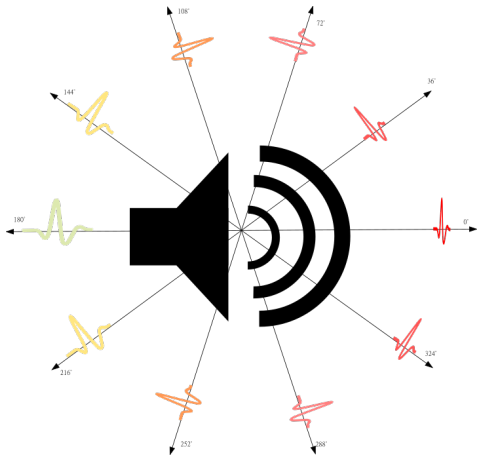


Figure 3.3: Loudspeaker directivity visualization [3]

This loudspeaker model is discretized according to Equation 3.1.

$$v[n,m] = \gamma_0(\frac{n}{f_s} - \frac{R_0 f_s}{v_c}, \frac{2\pi m}{M}) \tag{3.1}$$

for $n = 0, ..., N_v - 1$, $m = 0, ..., M - 1$, where R_0 is the far field distance and γ_0 is the function describing the continuous directivity characteristic.

Array Geometry The microphone array structure gives a spatial sampling of the signal along the continuous aperture. As depicted in Figure 3.4, the signal is modeled as a plane wave incident on a circular array, from a certain angle θ_r.

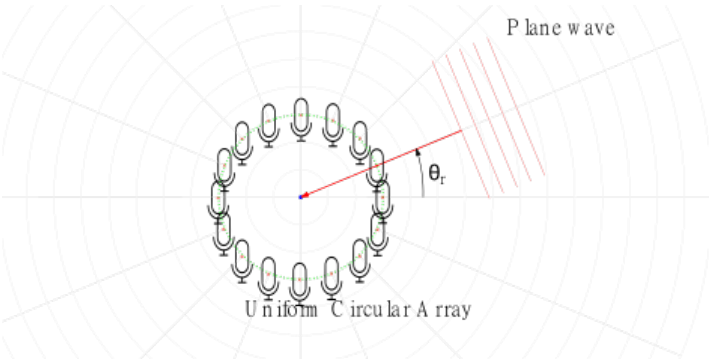


Figure 3.4: Microphone array response [3]

This is modeled with the Kronecker delta as given in Equation 3.2, for time instance $n = 0, ..., N_\delta - 1$, where N_δ is the number of samples corresponding to the array diameter.

$$\delta[n,m] = \begin{cases} 1, & \text{if } n = \left\lceil f_s \frac{R_a}{v_c}(1 - \cos(\frac{2\pi m}{M})) \right\rceil \\ 0, & \text{otherwise} \end{cases} \tag{3.2}$$

Now, the room impulse response can be expressed as three discrete convolutions, as given in Equation 3.3.

$$\begin{aligned} h[n,m] &= \sum_{m'=0}^{M-1} \sum_{n_h=0}^{N_h-1} \delta[n - n_h, (m - m')] \sum_{n_v=0}^{N_v-1} v[n_h - n_v, m'] s[n_v, m'] \\ &= \delta[n,m] *_{n,m} (v[n,m] *_n s[n,m]) \end{aligned} \tag{3.3}$$

3.2.2 Inverse problem

The problem can be posed as a linear system of equations. $\mathbf{s}^{(p)}$ is a vector of size T with the elements $\mathbf{s}_q^{(p)} = s[q,p]$. $\mathbf{s}$ of size TM is then given in Equation 3.4.

$$\mathbf{s} = [[\mathbf{s}^{(0)}]^T, ..., [\mathbf{s}^{(M-1)}]^T]^T \tag{3.4}$$

The channel responses are arranged similarly, where $\mathbf{h}_n^{(m)} = h[n,m]$ and has size N_h:

$$\mathbf{h} = [[\mathbf{h}^{(0)}]^T, ..., [\mathbf{h}^{(M-1)}]^T]^T \tag{3.5}$$

The model is then posed as:

$$\mathbf{h} = \mathbf{\Phi}\mathbf{s} + \mathbf{n} \tag{3.6}$$

The matrix $\mathbf{\Phi}$ is constructed to represent the expression with the three convolutions. For this $\mathbf{I}_N$ is the identity matrix of size $N \times N$ and the zero-padding matrix $\mathbf{W}_{a\times b}$ is given by: $\begin{bmatrix} \mathbf{I}_b \\ \mathbf{0}_{a-b\times b} \end{bmatrix}$.

$\mathbf{F}_M$ is the normalized DFT matrix of size $M \times M$. Then $\mathbf{\Phi} = \mathbf{AD}(\mathbf{I}_{MP} \otimes \mathbf{W}_{N_h\times T})$, where $\otimes$ is the Kronecker product.

The loudspeaker directivity matrix $\mathbf{D}$ is formed by constructing a vector v by concatenating the angle-dependent directivity responses: $\mathbf{v} = [[\mathbf{v}^{(0)}]^T, ..., [\mathbf{v}^{(M-1)}]^T]^T$. Then $\mathbf{D}$ is defined in Equation 3.7.

$$\mathbf{D} = (\mathbf{I_M} \otimes \mathbf{F}_{N_h})^{-1}\mathbf{\Lambda_v}(\mathbf{I_M} \otimes \mathbf{F}_{N_h}) \tag{3.7}$$

where $\mathbf{\Lambda_v} = \text{diag}\{(\mathbf{I}_M \otimes \mathbf{F}_{N_h}\mathbf{W}_{(N_h\times N_v)})\mathbf{v})\}$.

The array geometry matrix is defined in a similar fashion. $\mathbf{m} = [[\mathbf{m}^{(0)}]^T, ..., [\mathbf{m}^{(M-1)}]^T]^T$, where the elements are $\mathbf{m}_n^m = \delta[n, m]$. Then, A is given in Equation 3.8.

$$\mathbf{A} = (\mathbf{F}_M \otimes \mathbf{F}_{N_h})^{-1}\mathbf{\Lambda_m}(\mathbf{F}_M \otimes \mathbf{F}_{N_h}) \tag{3.8}$$

where $\mathbf{\Lambda_m} = \text{diag}\{(\mathbf{F}_M \otimes \mathbf{F}_{N_h})(\mathbf{I}_M \otimes \mathbf{W}_{(N_h\times N_h)})\mathbf{m}\}$.

Having defined this linear system from Equation 3.6, it is possible to solve for $\mathbf{s}$, since $\mathbf{\Phi}$ is defined by the known loudspeaker directivity and array configuration and $\mathbf{h}$ is measured. This is done by solving the minimization problem from Equation 3.9.

$$\min_{\mathbf{s}} \quad ||\mathbf{\Phi}\mathbf{s} - \mathbf{h}||_2^2 + \lambda||\mathbf{s}||_1 \tag{3.9}$$

From Image Source Locations to Planar Surface Equation

Once $\mathbf{s}$ is found, the next challenge lies in getting the planar surface equation parameters. $\mathbf{s}$ is separated to M vectors $\mathbf{s}^{(m)} \in \mathbb{R}^T$ for $m = 0, ..., M-1$. From $\mathbf{s}^{(m)}$ the peaks $m_{\text{peak}}, q_{\text{peak}}$ are estimated. The distance of R the image source is found by $R_{\text{imgsrc}} = q_{\text{peak}}\frac{v_c}{f_s}$, while the angle of the image source location follows from $\alpha_{\text{imgsrc}} = 360° \frac{m_{\text{peak}}}{M}$.

How to then get the planar surface equation parameters from the image source distance and angle? The plane equation is described with the plane distance ρ and the plane normal $\boldsymbol{\nu} = [\nu_x, \nu_y, \nu_z]$. Since the plane is located exactly halfway between the source and the image source, $\rho = \frac{R_{\text{imgsrc}}}{2}$. The plane normal values are $\boldsymbol{\nu} = [\cos\alpha_{\text{imgsrc}}, \sin\alpha_{\text{imgsrc}}, 0]$.

3.3 State-of-the-art approach for planar surface detection from a point cloud

In Chapter 2, the D-KHT [40] is presented as a suitable approach for planar surface detection in this application. This method consists of three main stages, for which the

workflow is given in Figure 3.5. Stage 1 is clustering, where the point cloud is divided into clusters of co-planar points. In Stage 2, the accumulator is populated based on the Gaussian kernels of each cluster. In Stage 3, the peaks in the accumulator are detected. When the peaks are found, the Hough-space coordinates are transformed to the plane equation in Euclidean space. In the following sections, this method is described and analysed in-depth, following the publication closely.

Figure 3.5: Workflow for the D-KHT. First the depth image is divided into approximately coplanar clusters. Next, the accumulator is incremented in the voting procedure based on the clusters. Finally, peaks are detected in the accumulator resulting in the detection of planes.

3.3.1 Clustering

The points in the structured point cloud are divided into clusters of approximately co-planar points. This is done by using Principal Component Analysis (PCA) [41], to decide whether the current cluster are associated with the same plane. If they are not, the cluster must be subdivided into four clusters, on which the procedure can be repeated. When the number of points in the cluster α is smaller than a minimum $\alpha_{\min}$ and the points are not co-planar, the cluster is regarded as a non-planar cluster. To determine whether the cluster is co-planar, the sample mean $\mu_{(x,y,z)}$ and the covariance matrix $\Sigma_{(x,y,z)}$ are defined using Equation 3.10 and Equation 3.11.

$$\boldsymbol{\mu}_{(x,y,z)} = \begin{pmatrix} \mu_x \\ \mu_y \\ \mu_z \end{pmatrix} \tag{3.10}$$

$$\boldsymbol{\Sigma}_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})} = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{xy} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_{zz} \end{pmatrix} \tag{3.11}$$

Then, it is checked whether $2\sqrt{\lambda_1} < \chi_{\text{cluster}}$, where λ_1 is the smallest eigenvalue of $\Sigma_{(x,y,z)}$ and χ_{cluster} is the maximum allowable spread. χ_{cluster} is determined empirically.

3.3.2 Computing Gaussian Kernels

Now that the point cloud is divided into clusters of co-planar points, a Gaussian kernel in Hough space can be computed. The plane that best fits the samples in the cluster passes through the mean $\boldsymbol{\mu}_{(x,y,z)}$ and has a normal vector $\boldsymbol{\nu} = (\nu_x, \nu_y, \nu_z)$. $\boldsymbol{\nu}$ is the eigenvector associated with the smallest eigenvalue of $\Sigma_{(x,y,z)}$. The Gaussian kernel in Hough space is centered at $\mu_{(\rho,\varphi,\theta)}$ as given in Equation 3.12. The covariance matrix in Hough space can be computed using Equation 3.13, where J is the Jacobian of $\mu_{(\rho,\varphi,\theta)}$.

$$\boldsymbol{\mu}_{(\rho,\varphi,\theta)} = \begin{pmatrix} \mu_\rho \\ \mu_\varphi \\ \mu_\theta \end{pmatrix} = \begin{pmatrix} \nu_x\mu_x + \nu_y\mu_y + \nu_z\mu_z \\ \cos^{-1}(\nu_z) \\ \tan^{-1}(\frac{\nu_y}{\nu_x}) \end{pmatrix} \tag{3.12}$$

$$\mathbf{\Sigma}_{(\rho,\varphi,\theta)} = \mathbf{J}\mathbf{\Sigma}_{(x,y,z)}\mathbf{J}^T \tag{3.13}$$

3.3.3 Spherical Accumulator

In [42], different accumulator designs were compared. The cubical accumulator is a linear discretization of the Hough space in all dimensions. This leads to a non-uniform cell size, since the length of the circle that θ makes depends on φ. In the spherical accumulator design, the cell size in the θ direction depends on φ, attempting to make the cell size uniform. It was shown that the spherical accumulator is favorable over the cubical accumulator design, since this gives a more even distribution of cell size in the full accumulator. The axis of this accumulator are $\theta \in [-\pi, \pi)$, $\varphi \in [0, \pi)]$ and $\rho \in [0, \rho_{\max}]$ where $\rho_{\max}$ is chosen to be the distance between the origin and the furthest point in the point cloud. The discrete values in ρ and φ directions are defined by linear interpolation where the number of samples N_ρ and N_φ are defined by the user. The discretization of θ is given dependent on the current latitude circle φ_i. The largest possible circle is at $\varphi = 0$, the equator. For the unit sphere the length of this circle is $l_{\max} = 2\pi$. The length of the latitude circle at the segment located at φ_i is given by $l_i = 2\pi(\varphi_i + \varphi/N_\varphi)$. Then the step size of θ at φ_i, $d\theta_{\varphi_i}$ is computed as:

$$d\theta_{\varphi_i} = \frac{360^\circ l_{\max}}{l_i N_\theta}. \tag{3.14}$$

3.3.4 Kernel-based Voting

The bins in the accumulator are increased per cluster. For each cluster, the Gaussian kernels were computed. For a given kernel defined by a mean $\boldsymbol{\mu}_{(\rho,\varphi,\theta)}$ and a covariance matrix $\mathbf{\Sigma}_{(\rho,\varphi,\theta)}$, the accumulator bins that fall within two standard deviations of the mean are incremented. The contribution of a kernel is considered significant at the bins whose parameter vector $\mathbf{q} = [\rho, \varphi, \theta]^T$ satisfies $f(\mathbf{q}) \geq f(\boldsymbol{\mu}_{(\rho,\varphi,\theta)} + 2\sqrt{\lambda}\mathbf{u})$, where $\mathbf{u}, \lambda$ are an eigenpair of $\mathbf{\Sigma}_{(\rho,\varphi,\theta)}$ and f gives the probability density function of the Gaussian distribution as described in Equation 3.15.

$$f(\mathbf{q}) = \frac{1}{\sqrt{2\pi^3|\mathbf{\Sigma}_{(\rho,\varphi,\theta)}|}} \exp(-\frac{1}{2}(\mathbf{q} - \boldsymbol{\mu}_{(\rho,\varphi,\theta)})^T \mathbf{\Sigma}^{-1}_{(\rho,\varphi,\theta)}(\mathbf{q} - \boldsymbol{\mu}_{(\rho,\varphi,\theta)})) \tag{3.15}$$

3.3.5 Peak detection

The peaks in the accumulator after the voting step correspond to planes. A hill climbing strategy for peak detection is employed [43]. The mean of each kernel that is used in the voting procedure, is used as a starting point. It is checked whether its neighbors are smaller, after which it can be concluded that the point is a local maximum. If this is not the case, the neighbor bin having more votes is selected and the procedure is repeated.

3.3.6 Planar surface equation

Once all local maxima have been detected, the plane equation can be found. The distance from the origin is given by ρ, and the plane normal is found as follows:

$$\boldsymbol{\nu} = \begin{pmatrix} \nu_x \\ \nu_y \\ \nu_z \end{pmatrix} = \begin{pmatrix} \sin\varphi\cos\theta \\ \sin\varphi\sin\theta \\ \cos\varphi \end{pmatrix} \tag{3.16}$$

.

3.4 Proposed method: Combined approach

The detection of acoustic reflectors using a microphone array has several shortcomings. The first one is computational complexity. Solving the inverse problem from section 3.2 is computationally demanding. Due to this, it would be not feasible to extend this method to a 3D representation. Then, the reflections from horizontal reflectors (i.e. floors) are a disturbance in the measurements since they are not included in the model.

The loudspeaker directivity is modeled and included in the presented method, however it was found that surfaces that are positioned in low-energy regions (generally behind a speaker) are more challenging to detect due to the low signal power. This is most strongly the case in the presence of other reflectors or noise.

Point clouds stemming from LiDAR measurements have limitations as discussed in subsection 2.1.3. The presence of natural light degrades the performance of the sensor, as well as having other LiDAR sensors present. The windows and mirrors are challenging to detect since the detection depends on refraction on objects, while windows and mirrors reflect and/or let light through. This causes mostly issues when a large window is present.

The flash LiDAR technology limits the FOV of the LiDAR camera. A typical FOV is $50° \times 70°$, meaning that the configuration of the co-located system is crucial. Considering the limitations in acoustics, a decision is made to position the LiDAR in such a way that the low-energy region is within the LiDAR FOV. Also, a position is chosen such that the floor is likely to be within the accessible range. The floor is considered to be a more likely disturbance than a ceiling, seeing that in a home environment loudspeakers are typically positioned closer to the floor. The combined approach focuses on overcoming the limitations in the acoustic method, while taking into consideration the limitations in the LiDAR method. The first limitation is the ambiguity introduced by horizontal reflectors. In subsection 3.4.1, the procedure for overcoming this limitation is described. The other challenge is the loudspeaker directivity, for which the procedure is described in subsection 3.4.2. The full procedure in pseudo-code is described in subsection 3.4.3.

3.4.1 Compensation for detected horizontal reflectors from the point cloud

The unmodelled reflections from non-vertical structures are complicating the recovery of walls. Given the location of a horizontal reflector, i.e. a floor, it can be accounted

for in pre-processing. Using the LiDAR camera and the D-KHT algorithm, the surfaces within the LiDAR FOV are found. If a non-vertical plane, a plane where the z-component of the plane normal is non-zero, is detected, its contribution is eliminated from the acoustic response.

There are two elements of information known about these horizontal reflectors. The first one is that the response is equal on all microphones in the array, secondly the distance of the reflector is known from the LiDAR measurement. The distance cannot be used directly, since delays from the loudspeaker response need to be considered. Here, this is done by taking the distance from the peak of the direct path, which is at a known distance. Once the expected time delay with reference to the direct path peak is found, the expected sample of the floor reflection is known. The goal is to exploit this information to eliminate this reflection before employing the acoustic wall detection algorithm (UCA). This is done by minimizing the following optimization problem.

$$\begin{aligned} \min_{\mathbf{h}_{\text{floor}}} \quad & ||\mathbf{h} - \mathbf{I} \otimes \mathbf{h}_{\text{floor}}||_2^2 \\ \text{s.t.} \quad & ||\mathbf{L}\mathbf{h}_{\text{floor}}||_2^2 \leq b \end{aligned} \tag{3.17}$$

This optimization problem is constructed in such a way that both pieces of information are incorporated. $\mathbf{h}$ is the stacked response from all microphones $[[\mathbf{h}^{(0)}]^T, ..., [\mathbf{h}^{(M-1)}]^T]^T$, where $\mathbf{h}_n^{(m)} = h[n, m]$ and has size N_h. $\mathbf{h}_{\text{floor}}$ is the response from the floor that is aimed to estimate. By using the Kronecker product with the identity matrix, it is ensured that the estimated floor response is the same for all microphone channels. The matrix $\mathbf{L}$ is a weighting matrix. This weighting matrix is constructed such that the samples that lie far from the expected floor location sample p, have a high weight and the expected floor location sample has a weight of 0. The purpose of this condition is to limit the response of $\mathbf{h}_{\text{floor}}$ outside the area where the floor is expected to be. The weighting factor scales with the squared distance to the expected location. First a vector $\mathbf{l} \in \mathbb{R}^{N_h}$ is defined, where the entries of l are $l[n] = |n - p|^2$. The vector $\mathbf{l}$ is put in the diagonal matrix $\mathbf{L}$. After minimizing the problem from Equation 3.17, the floor response that is recovered is subtracted from each response $\mathbf{h}^{(m)}$ as given in Equation 3.18, where the remainder is the wall response that is required.

$$\mathbf{h}_{\text{walls}}^{(m)} = \mathbf{h}^{(m)} - \mathbf{h}_{\text{floor}} \tag{3.18}$$

3.4.2 Loudspeaker Directivity Compensation

The directivity of the speaker makes it more challenging to detect walls in low-energy regions, due to a lack of energy in the echoes. The proposed method relies on a system where a LiDAR camera is positioned such that the low-energy region is within the FOV of the LiDAR. Given that the large planar surfaces in this region can be detected from the point cloud, there are several ways to combine this with the acoustic information.

In this thesis, two methods are used. The first one is a straightforward approach that relies on the LiDAR sensor in the region where the accuracy of the acoustic detection is low. This highly increases the accuracy, however this does not ensure the detection of a surface that has the acoustic properties of the wall, only the visual properties of

a wall. This method is described in subsubsection 3.4.2.1. For the other approach, first the plane detection from the point cloud is done. The angle and the distance of this detected plane are used as prior when solving the acoustic problem, making it less challenging to detect this wall applying acoustics. This method is described in subsubsection 3.4.2.2.

3.4.2.1 Method 1 - LiDAR is always right

The LiDAR sensor is positioned such that its FOV covers the low-energy region of the loudspeaker. The plane detection from the point cloud is done separately using the D-KHT algorithm. Likewise, the acoustic reflectors are estimated from the acoustic measurements. These two estimations are then compared. If there is a plane detected from the point cloud, this plane is assumed to be a wall. For the regions that do not lie in the LiDAR FOV, the acoustic detection is used. The pseudocode for the decision-making process is presented below.

Algorithm 1: Decision between LiDAR and acoustic detection

```
Result: Detected Planes
LidarPlanes = D-KHT(pointcloud);
AcousticPlane = UCA(microphoneresponse);
foreach AcousticPlane do
    if not in LiDARRegion then
        DetectedPlane = AcousticPlane;
    end
end
DetectedPlane = [DetectedPlane; LiDARPlanes];
```

3.4.2.2 Method 2 — LiDAR used as prior

The LiDAR is again positioned such that the low-energy region of the loudspeaker is covered by its FOV. Now, first the plane detection from the point cloud using the D-KHT is performed. The output of this algorithm is a list of plane equations, consisting of the distance ρ and the normal vector $\boldsymbol{\nu}$. From the normal $\boldsymbol{\nu}$, the angle in the xy-plane at which the wall is located compared to the system is easily recovered using Equation 3.19.

$$\alpha = \tan^{-1}(\frac{\nu_y}{\nu_x}) \tag{3.19}$$

The acoustic algorithm (UCA) solves for the image sources rather than the plane equation directly. The angle α_{pc} at which the plane from the point cloud is detected corresponds to the angle of the image source. The distance of an image source ρ_{imgsrc} corresponds to two times the distance of the planar surface that corresponds to that image source:

$$R_{\text{imgsrc}} = 2\rho_{\text{pc}} \tag{3.20}$$

where ρ_{pc} is the distance from the system to the planar surface. Now the angle and the distance at which an image source is expected, if the planar surface is indeed a wall, are known. This is used as a prior in solving the optimization problem from section 3.2

in Equation 3.9, by including it as a constraint. The optimization problem is then presented in Equation 3.21.

$$\begin{aligned} \min_{\mathbf{s}} \quad & ||\mathbf{\Phi s} - \mathbf{h}_{\text{walls}}||_2^2 + \lambda ||\mathbf{s}||_1 \\ \text{s.t.} \quad & ||\mathbf{Ls}||_2^2 \leq b \end{aligned} \tag{3.21}$$

Now, $\mathbf{L}$ is a diagonal matrix of size $MT \times MT$. This matrix is constructed by forming the vectors $\mathbf{l}^{(m)}$, that is filled in a similar way as in subsection 3.4.1. If the candidate location is the location that is found using the plane detection from the point cloud, its entry is zero. The entries are larger if the candidate location in the grid is further away from the expected location from the point cloud: $l[m,n] = |n - n_{\text{pc}}|(1 + |m - m_{\alpha_{\text{pc}}}|)$. The vector $\mathbf{l}$ is the combination of all vectors $\mathbf{l}_{(m)}$, $[[\mathbf{l}^{(0)}]^T, ..., [\mathbf{l}^{(M-1)}]^T]^T$. The procedure is summarized in the pseudocode below. Now, the `UCA` algorithm has as input the detected LiDAR planes. These are included in the detection following the procedure as described above.

Algorithm 2: Using the LiDAR detection as a prior in the acoustic algorithm

```
Result: Detected Planes
LidarPlanes = D-KHT(pointcloud);
AcousticPlane = UCA(MicrophoneOutput,LidarPlanes);
```

3.4.3 Complete Algorithm for detection of acoustically reflecting surfaces

The above procedures are combined in one complete algorithm. Beforehand, a decision needs to be made if Method 1 or Method 2 is employed. Then, the plane detection from the point cloud is done. If a plane with a z-normal is detected, the `eliminateZnormal` function is employed before eliminating the direct path, otherwise only the direct path elimination step is performed. If the first method is used, the acoustic detection is done after which the decision-making step is done, while in the second method the acoustic detection with the `UCA` algorithm is done with the consideration of the detected planes

in the point cloud. This process is summarized below.

Algorithm 3: Complete algorithm for detection of acoustically reflecting surfaces

```
Result: Plane equations
initialization;
LidarPlanes = D-KHT(pointcloud);
if non-zero z-normal detected then
    MicrophoneOutput = eliminateZnormal(plane, MicrophoneOutput);
    MicrophoneOutput = eliminateDirectPath(AnachoicOutput,
     MicrophoneOutput);
else
    MicrophoneOutput = eliminateDirectPath(AnachoicOutput,
     MicrophoneOutput);
end
if Method 1 then
    AcousticPlane = UCA(MicrophoneOutput);
    Decision(LidarPlanes,AcousticPlane);
else
    if xy-plane detected then
        UCA(MicrophoneOutput, LidarPlanes);
    else
        UCA(MicrophoneOutput);
    end
end
```

4 Results

The proposed method is evaluated in different simulation scenarios to isolate the effect of certain situations. The proposed method is evaluated against the state-of-the-art method of acoustic reflector detection by Zaccá [3]. First, the performance in a scenario with a directive speaker is evaluated. The noise levels are varied over different SNRs and it is attempted to detect a wall that is located at different angles from the loudspeaker. This experiment is described in 4.1. Next, it is attempted to estimate two acoustic reflectors, one of which is a wall and can be detected with LiDAR, the other one is a window and is invisible for the LiDAR sensor, for which the results are presented in 4.2. Finally, a simulation that includes a horizontal reflector, i.e. a floor, is done. It is attempted to detect a wall when a disturbing floor reflection is present at different heights. The results are shown in 4.3.

The loudspeaker is modelled as a directive point source at the origin, where the front at $\alpha = 0°$ is in the positive horizontal direction. The LiDAR sensor is also located in the origin, but its front is directed in the negative horizontal direction $\alpha = 180°$. Its FOV is $\beta_{\text{hor}} \times \beta_{\text{ver}}$. A circular microphone array containing M microphones surrounds the loudspeaker and the LiDAR sensor. A top view of the set-up of the co-located system is shown in Figure 4.1 and a side view is displayed in Figure 4.2.

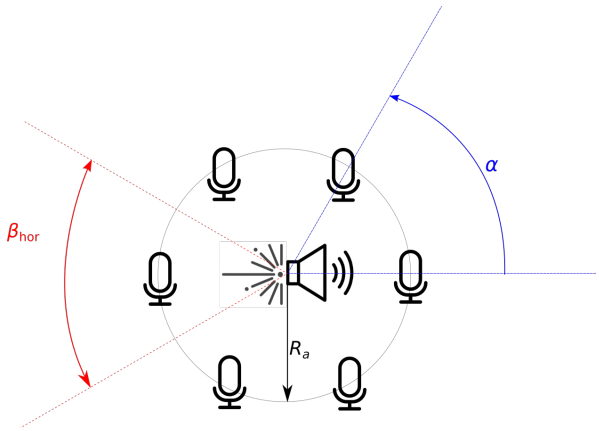


Figure 4.1: Top view of system setup

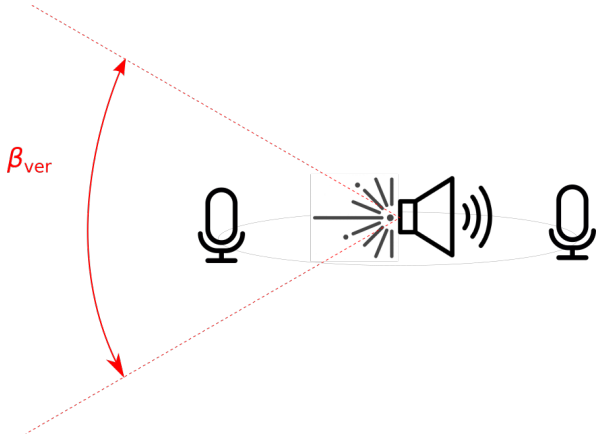


Figure 4.2: Side view of system setup

4.1 Experiment 1 — Single wall scenario

The performance of wall detection decreases when a loudspeaker is not omnidirectional. The low emitted energy on the backside of the loudspeaker makes it challenging to detect reflecting surfaces. The purpose of this experiment is to evaluate the performance of the proposed method compared to the state-of-the-art method in the low-energy regions of the loudspeaker. This is demonstrated with a single wall scenario. This wall placed at 0.5 m and is rotated around the co-located system, as shown in Figure 4.3.

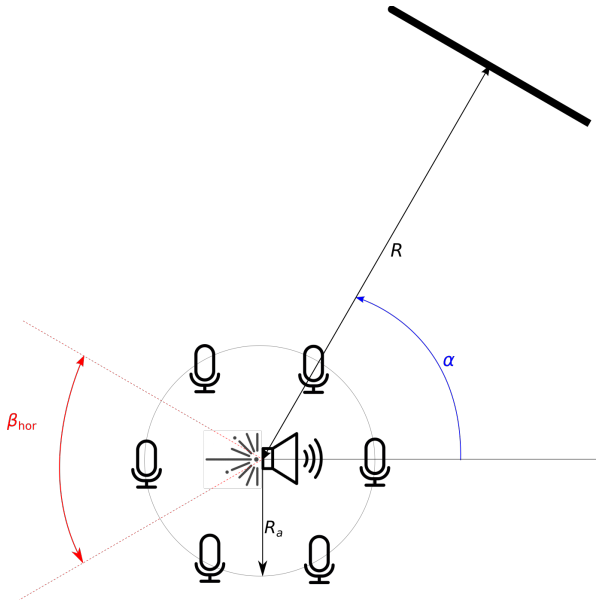


Figure 4.3: The setup used for experiment 1. The thick black line illustrates a wall that is placed at different angles α around the system.

The field of view of the LiDAR is $70° \times 50°$. The loudspeaker directivity characteristic is given in Figure 4.4.

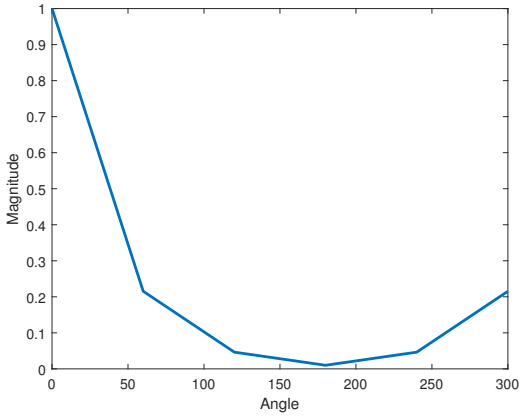
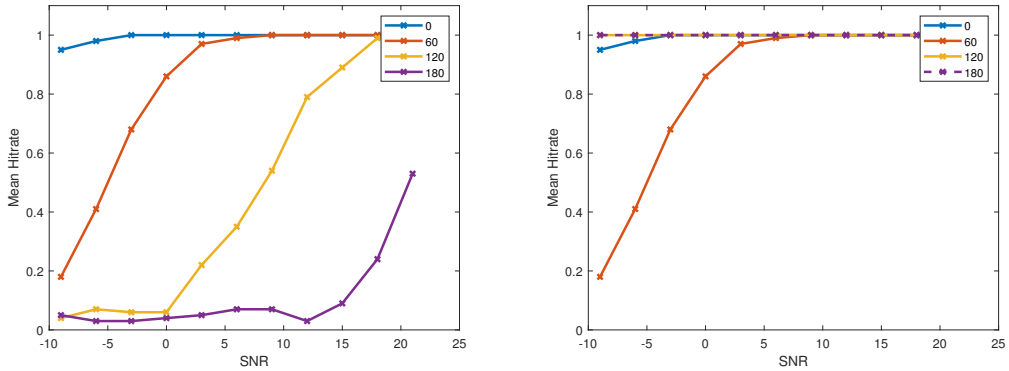


Figure 4.4: Directivity characteristic of the loudspeaker. The magnitude scaling used for the directivity. In the back of the loudspeaker the energy is lower than in the front.

The LiDAR measurements are noiseless and the acoustic SNR is varied from -9 dB to 21 dB. A Monte-Carlo simulation of 100 runs is performed. In Figure 4.1a, the mean hitrate of Zaccá's method is given for each angle and for different SNR values. The hitrate is 1 if a wall is detected correctly and averaged over the runs.



Zaccá's method [3] (a) Proposed method 1 (b)

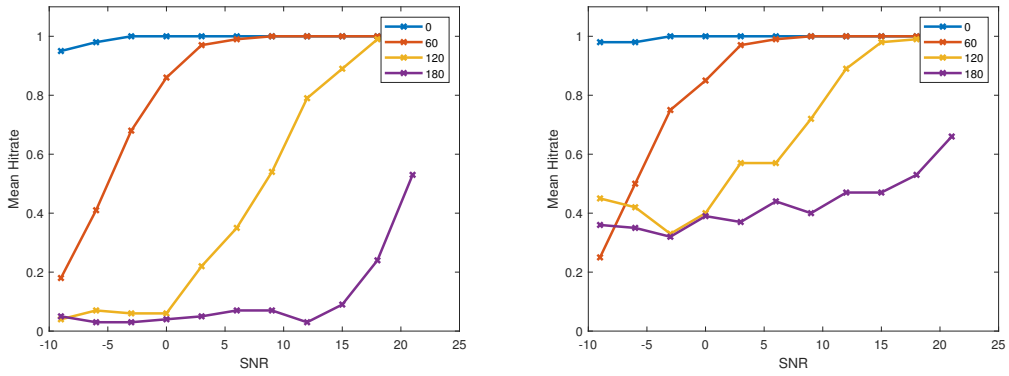
Figure 4.5: Detection of a rotating wall around the co-located system. The mean hitrate is shown against SNR. The performance at $0°, 60°, 120°$ and $180°$ is shown in blue, orange, yellow and purple respectively. The challenge of detection of the walls in the low-energy regions of the speaker is resolved with the proposed method.

What can be seen from this figure, is that it is challenging to detect a wall at a low SNR, and that it becomes difficult as the angle increases to $180°$.

In Figure 4.1b, the results of the same experiment, where the proposed method is used to detect the wall. Method 1 is used here. From this result, it can be concluded that the challenge of detection of the walls in the low-energy regions of the speaker is resolved as they enter the LiDAR FOV.

In Figure 4.1b, the mean hitrate, when the second proposed method is used, is given for each angle and for different SNR values. For easier comparison the results

with Zaccá's method are repeated in Figure 4.1a.



Zaccá's method [3] (a) Proposed method 2 (b)

Figure 4.6: Detection of a rotating wall around the co-located system. The mean hitrate is shown against SNR. The performance at $0°, 60°, 120°$ and $180°$ is shown in blue, orange, yellow and purple respectively. The challenge of detection of the walls in the low-energy regions of the speaker is not completely resolved as it was using method 1, but the hitrate has increased compared to Zaccá's method.

Here, a significant improvement is achieved for the angles in the LiDAR sensor FOV, i.e. $120°$ and $180°$. The problem is not completely resolved as it was using method 1, since here a combination of the noisy acoustic measurements is used, aided with the detection from the LiDAR sensor.

4.2 Experiment 2 — Scenario with windows

Using the LiDAR sensor, it is challenging to detect windows. In this experiment, the consequence is shown in a dual wall scenario. One of the walls is a window, i.e. it behaves like a wall acoustically, but cannot be detected with a LiDAR sensor. Again, the configuration is rotated around the co-located system as demonstrated in Figure 4.7.

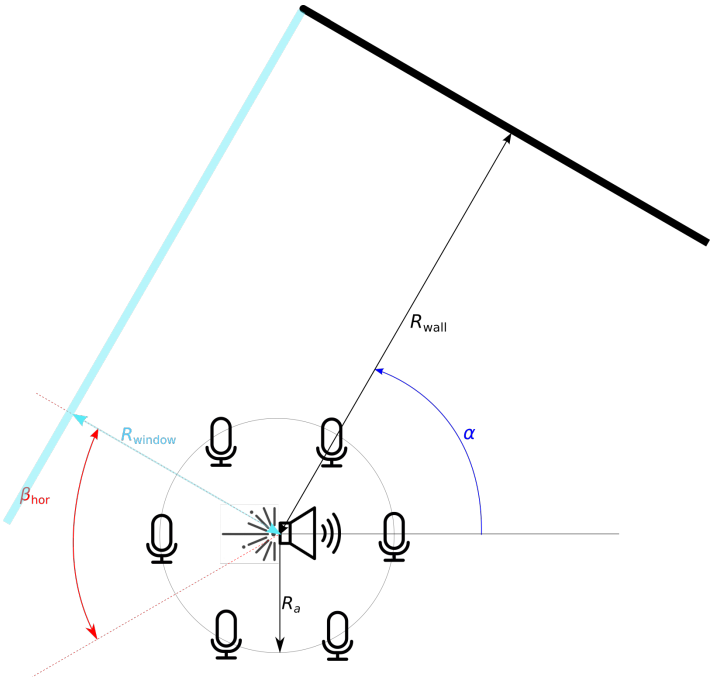


Figure 4.7: experiment 2

The distance $R_{\text{window}} = 0.6$ m and the distance $R_{\text{wall}} = 0.4$ m. The FOV of the LiDAR is $70° \times 50°$ and is positioned such that it captures the low-energy region of the loudspeaker. The loudspeaker directivity from Figure 4.4 is used.

The LiDAR measurements are noiseless and the acoustic SNR is varied from -9 dB to 20 dB. A Monte-Carlo simulation of 100 runs is performed. In Figure 4.8, the mean hitrate of Zaccá's method is given for each angle and for different SNR values. The detection of the wall is given on the left and the detection of the window is displayed on the right. For a hitrate of 1, the wall or window is detected correctly. The detection of the wall seems similar to the detection of the window. This is mainly due to the fact that the wall is placed closer to the system. In addition to that, the wall angles are perfectly aligned with the discrete detectable angles of the acoustic system, whereas the window is on an intersection of two discrete grid points. However, since it is placed at an intersection, both angles are considered correct, e.g. a window at 150° is classified as correctly detected if the found angle is either 120° and 180°, or anywhere between. If either the window or the wall is detectable, it can be more challenging to detect the other surface, since the detection relies on finding the maxima in $\mathbf{s}$. Often, two such maxima are detected next to each other for one surface.

The experiment is repeated with the first proposed method. The results are given in Figure 4.9. Like in experiment 1, the detection of the wall in low-energy regions is perfect in all noise conditions. However, now in the case that the wall is in the LiDAR sensor FOV, the system fully relies on this information, and it becomes impossible to detect the window, giving a mean hitrate of zero at all SNR levels.

Now, the experiment is repeated with the second proposed method. The results are presented in Figure 4.10. Here, the mean hitrate of the wall detection improves in

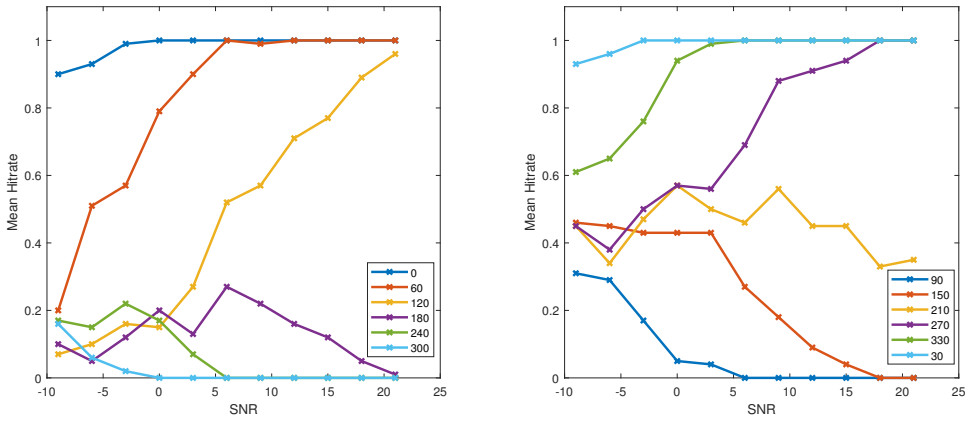


Figure 4.8: Left: Detection of a wall at different angles using Zaccá's method, while simultaneously attempting to detect a window. Right: Detection of a window at different angles using Zaccá's method, while simultaneously attempting to detect a wall.

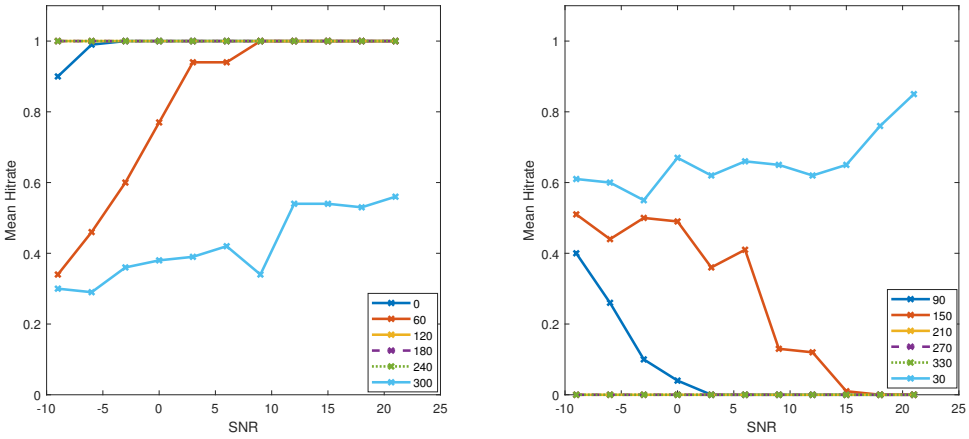


Figure 4.9: Left: Detection of a wall at different angles using Zaccá's method, while simultaneously attempting to detect a window. Right: Detection of a window at different angles using Method 1, while simultaneously attempting to detect a wall.The detection of the wall improves with the proposed method, while the detection of the window goes to zero.

low-energy regions, while keeping approximately the same performance for the window detection. For in-home scenarios, this behaviour is considered preferable than the obtained results from method 1, since windows are a main component of the house.

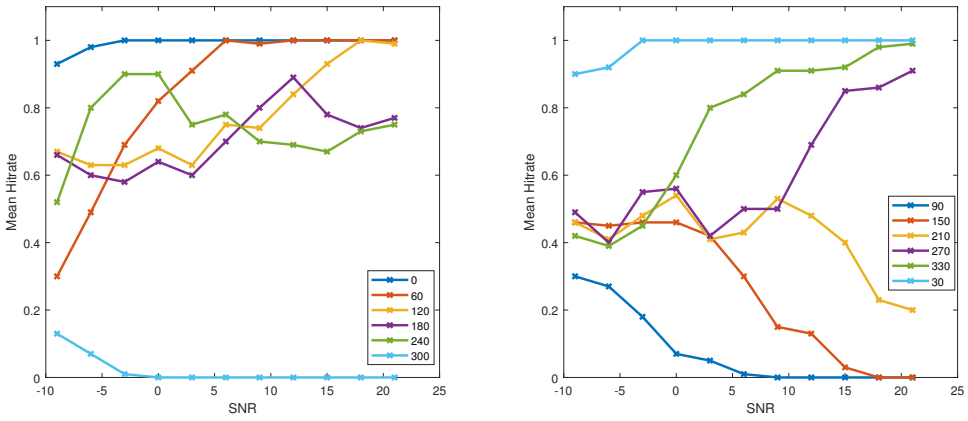


Figure 4.10: Left: Detection of a wall at different angles using method 2, while simultaneously attempting to detect a window. Right: Detection of a window at different angles using method 2, while simultaneously attempting to detect a wall. The detection of the wall improves with the proposed method, while the detection of the window remains similar.

4.3 Experiment 3 — Detection of a wall in presence of a floor

A significant ambiguity is introduced when reflections that were not included in the model are present. The purpose of this experiment is to show that the proposed method reduces this ambiguity. The co-located system is placed in front of a wall at angle $\alpha = 0°$ and distance $R = 1$ m. A floor is introduced at varying distance d_f, as shown in Figure 4.11. The loudspeaker is assumed to be omnidirectional and the LiDAR FOV is now $50° \times 70°$. This scenario is noiseless.

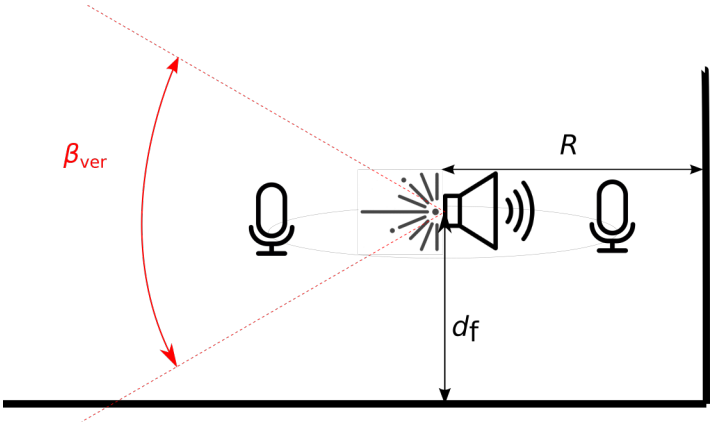


Figure 4.11: Side view of the set-up for Experiment 3. The co-located system is placed at height d_f at a distance $R = 1$ m from a wall at angle $\alpha = 0°$

In Figure 4.12, the results of this experiment are shown. In the first figure, the distance that is detected using either Zaccá's algorithm or the proposed method is

given. In the second figure, the mean square error (MSE) of the plane normal is given.

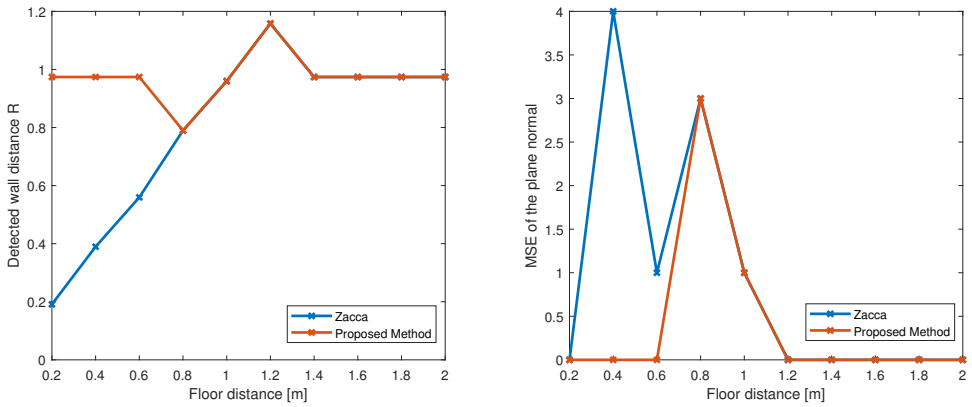


Figure 4.12: Left: The detected distance R for the method of Zaccá [3] and the proposed method at different floor heights. The ground truth is $R = 1$. Right: The mean square error with the ground truth normal vector for Zaccá's method and the proposed method at different floor heights. While the floor is in the LiDAR FOV, it is possible to detect the wall reliably.

From this simulation, what is seen is that using Zaccá's algorithm, the first arriving reflection is detected. When a floor is present closer than the wall, a wall is detected at this distance, where the angle is unpredictable. When the floor is further away than the wall, the wall can be found reliably. Using the proposed method, it is possible to eliminate this effect. What is seen is that when the floor is within the LiDAR sensor FOV, the detection can be done reliably. Above a distance of 0.6 m, the floor gets below the LiDAR sensor's FOV and the same results as with Zaccá's method are achieved.

5 Conclusion & Discussion

Detection of reflective surfaces is crucial in obtaining the optimal listening experience for the user. Knowing the location of the reflectors, the effects of the echos can be compensated for in smart loudspeakers. Estimating the location of these surfaces is challenging using the smart loudspeaker, i.e. a loudspeaker and a co-located microphone array, due to computational constraints and loudspeaker directivity. The potential for using alternative sensors on a smart loudspeaker was shown, where the choice was made for a LiDAR sensor. A method was presented where the information from the microphone array and the LiDAR sensor was combined to obtain a more robust system.

The method presented in Chapter 3 exploits the information of the additional LiDAR sensor in regions where the loudspeaker is emitting less energy. It was shown that combining the two sensing modalities leads to a better performance of surface detection in low-energy regions. Two methods were implemented and compared. The naive approach of using the detected planar surfaces from the LiDAR data makes detection in the low-energy region near-perfect. However, when a window is introduced, the performance is degraded significantly. When instead the detected surfaces from the LiDAR data are included as a prior while solving the acoustic problem, there is again an improvement in detection. This method has as an advantage that the focus is on the acoustic properties, which are the properties that are most important. In the case of the wall/window scenario, an improvement of wall detection while the window detection is not influenced. In addition to that, it was shown that using this additional sensing modality, the issue of having interfering horizontal reflectors, e.g. floors, could be addressed. Using an omnidirectional model, it is demonstrated how such a reflector makes the wall detection challenging. In addition, it is shown how detecting this horizontal reflector from the point cloud gives rise to the opportunity to eliminate this negative effect, as long as the horizontal reflector is in the LiDAR FOV. The additional steps in the processing increases the computational complexity. However, the steps are less demanding than if the linear system would be extended to 3D and the practical implications of a LiDAR sensor are smaller than those of a spherical array.

5.1 Discussion

In this thesis a method to improve acoustic reflector detection with the aid of LiDAR was considered. Different assumptions were made to isolate the problem to make it relevant in practice. In this section, the decision for several assumptions is discussed briefly. In Zaccá's method, it was assumed that only the relevant first order reflections will be present in the measured impulse response. This assumption is followed in this work, since only walls in close proximity to the co-located system are considered. In the case that the full room geometry is estimated, this assumption is not valid.

Furthermore, in this work it is assumed that the channel, i.e. the room impulse response including the loudspeaker model, can be estimated perfectly. In practice, this is not as straightforward. An appropriate wideband excitation signal must be selected and a suitable channel estimation algorithm is required.

The experiments in Chapter 4 are conducted with noisy acoustic data. In this case, the LiDAR data, i.e. the point cloud, is assumed to be noiseless in order to isolate the problem scenario. In addition to that, the circumstances in which plane detection from point clouds is more challenging are only considered briefly in the form of a scenario including a window.

The choice of the D-KHT algorithm was made with the potential for a joint optimization problem in mind. In the current implementation, this algorithm could be replaced with any planar surface detection algorithm. A more simple RANSAC algorithm could be more appropriate for the current method.

5.2 Future Work

- **Extending to an external sensor** With the introduction of LiDAR sensors on smartphones and tablets, it is possible to extend the algorithm in such a way that the LiDAR sensor is not co-located. With image recognition, the loudspeaker location can be inferred from the point cloud, as well as the nearby surfaces. Using transformation of coordinate systems, the proposed method can be used. This is an advantage for users who have a smart loudspeaker and a device with the LiDAR sensor and can have the advantages of the extended system with this algorithm. This also reduces privacy invasion since a LiDAR sensor is not permanently located on the smart loudspeaker system.

- **Solve the problem jointly** Using the point cloud and the acoustic data, it is possible to solve the problem jointly. A signal model that includes both types of information and solves for reflective surfaces could be promising to improve the performance and robustness.

- **Detection of the type of wall** In the point cloud, there is more information present than the planar surfaces. For example, the structure in the surfaces can be detected. Some surfaces have more absorbing or refracting properties, such as bookcases. If this is detected with image recognition techniques, this could be exploited using the digital filters in the loudspeaker.

Bibliography

[1] S. Pelzer, L. Aspöck, D. Schröder, and M. Vorlaender, "Integrating Real-Time Room Acoustics Simulation into a CAD Modeling Software to Enhance the Architectural Design Process," *Buildings*, vol. 4, pp. 113–138, Apr. 2014.

[2] H. W. Yoo, N. Druml, D. Brunner, C. Schwarzl, T. Thurner, M. Hennecke, and G. Schitter, "MEMS-based lidar for autonomous driving," *e & i Elektrotechnik und Informationstechnik*, vol. 135, no. 6, pp. 408–415, Oct. 2018.

[3] V. Zacca, P. Martinez-Nuevo, M. Moller, J. Martinez, and R. Heusdens, "Inferring the location of reflecting surfaces exploiting loudspeaker directivity," in *2020 28th European Signal Processing Conference (EUSIPCO)*. Amsterdam, Netherlands: IEEE, Jan. 2021, pp. 61–65.

[4] A. Kramer and K. Z. Kramer, "The potential impact of the Covid-19 pandemic on occupational status, work from home, and occupational mobility," *Journal of Vocational Behavior*, vol. 119, p. 103442, Jun. 2020.

[5] A. Bick, A. Blandin, K. Mertens, K. Mertens, Federal Reserve Bank of Dallas, and Federal Reserve Bank of Dallas, "Work from Home Before and After the COVID-19 Outbreak," *Federal Reserve Bank of Dallas, Working Papers*, vol. 2020, no. 2017, Feb. 2021.

[6] P. Associates, "Parks Associates: Headphones and Earphones Likely to Experience Initial Sales Spike With Work at Home, Home Schooling, and Entertainment-In-Place," https://www.prnewswire.com/news-releases/parks-associates-headphones-and-earphones-likely-to-experience-initial-sales-spike-with-work-at-home-home-schooling-and-entertainment-in-place-301036654.html.

[7] E. Bary, "Sonos stock rockets 27% early Thursday as COVID-19 spurs surge in speaker purchases," http://www.marketwatch.com/story/sonos-beats-sales-expectations-as-covid-19-spurs-continues-surge-in-purchases-for-the-home-11605733675.

[8] H. V. Fuchs and X. Zha, "Requirement for low-frequency reverberation in spaces for music: Part 1: Smaller rooms for different uses." *Psychomusicology: Music, Mind, and Brain*, vol. 25, no. 3, pp. 272–281, 2015.

[9] J. S. Bradley, R. D. Reich, and S. G. Norcross, "On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility," *The Journal of the Acoustical Society of America*, vol. 106, no. 4, pp. 1820–1828, Oct. 1999.

[10] S. Spors, H. Buchner, R. Rabenstein, and W. Herbordt, "Active listening room compensation for massive multichannel sound reproduction systems using wave-domain adaptive filtering," *The Journal of the Acoustical Society of America*, vol. 122, no. 1, pp. 354–369, Jul. 2007.

[11] R. Rabenstein, M. Renk, and S. Spors, "Limiting Effects of Active Room Compensation using Wave Field Synthesis," in *Audio Engineering Society Convention 118*. Audio Engineering Society, May 2005.

[12] Y. Huang, S. C. Busbridge, and D. S. Gill, "Distortion and Directivity in a Digital Transducer Array Loudspeaker," *Journal of the Audio Engineering Society*, vol. 49, no. 5, pp. 337–352, May 2001.

[13] S. K. A. Nair, S. Joladarashi, and N. Ganesh, "Evaluation of ultrasonic sensor in robot mapping," in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, Apr. 2019, pp. 638–641.

[14] S. Dogru and L. Marques, "Through-Wall Mapping Using Radar: Approaches to Handle Multipath Reflections," *IEEE Sensors Journal*, vol. 21, no. 10, pp. 11 674–11 683, May 2021.

[15] S. Liaquat, U. S. Khan, and Ata-Ur-Rehman, "Object detection and depth estimation of real world objects using single camera," in *2015 Fourth International Conference on Aerospace Science and Engineering (ICASE)*, Sep. 2015, pp. 1–4.

[16] Y. Wang, Z. Lai, G. Huang, B. H. Wang, L. van der Maaten, M. Campbell, and K. Q. Weinberger, "Anytime Stereo Image Depth Estimation on Mobile Devices," in *2019 International Conference on Robotics and Automation (ICRA)*, May 2019, pp. 5893–5900.

[17] S. Royo and M. Ballesta-Garcia, "An Overview of Lidar Imaging Systems for Autonomous Vehicles," *Applied Sciences*, vol. 9, no. 19, p. 4093, Jan. 2019.

[18] W. Sun, Y. Hu, D. G. MacDonnell, C. Weimer, and R. R. Baize, "Technique to separate lidar signal and sunlight," *Optics Express*, vol. 24, no. 12, pp. 12 949–12 954, Jun. 2016.

[19] A. Krokstad, S. Strom, and S. Sørsdal, "Calculating the acoustical room response by the use of a ray tracing technique," *Journal of Sound and Vibration*, vol. 8, no. 1, pp. 118–125, Jul. 1968.

[20] 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, Apr. 1979.

[21] J. Borish, "Extension of the image model to arbitrary polyhedra," *The Journal of the Acoustical Society of America*, vol. 75, no. 6, pp. 1827–1836, Jun. 1984.

[22] Y. Zhang, M. Ye, D. Manocha, and R. Yang, "3D reconstruction in the presence of glass and mirrors by acoustic and visual fusion," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 8, pp. 1785–1798, 2017.

[23] G. Hwang, S. Kim, and H.-M. Bae, "Bat-g net: Bat-inspired high-resolution 3D image reconstruction using ultrasonic echoes," in *Advances in Neural Information Processing Systems*, 2019, pp. 3720–3731.

[24] H. E. Bass, L. C. Sutherland, A. J. Zuckerwar, D. T. Blackstock, and D. Hester, "Atmospheric absorption of sound: Further developments," *The Journal of the Acoustical Society of America*, vol. 97, no. 1, pp. 680–683, 1995.

[25] J. H. Yee, R. Alvarez, D. Mayhall, D. Byrne, and J. DeGroot, "Theory of intense electromagnetic pulse propagation through the atmosphere," *The Physics of fluids*, vol. 29, no. 4, pp. 1238–1244, 1986.

[26] G. E. Smith and B. G. Mobasseri, "Analysis and exploitation of multipath ghosts in radar target image classification," *IEEE transactions on image processing*, vol. 23, no. 4, pp. 1581–1592, 2014.

[27] R. Cole, B. Jameson, D. Garmatyuk, and Y. J. Morton, "Simultaneous indoor localization and detection with multi-carrier radar," in *2014 IEEE Radar Conference*. IEEE, 2014, pp. 0881–0886.

[28] R. Berkvens, A. Jacobson, M. Milford, H. Peremans, and M. Weyn, "Biologically inspired SLAM using wi-fi," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014, pp. 1804–1811.

[29] Z. Li, P. C. Gogia, and M. Kaess, "Dense surface reconstruction from monocular vision and LiDAR," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 6905–6911.

[30] V. Vaish, M. Levoy, R. Szeliski, C. L. Zitnick, and S. B. Kang, "Reconstructing occluded surfaces using synthetic apertures: Stereo, focus and robust measures," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2. IEEE, 2006, pp. 2331–2338.

[31] R. O. Duda and P. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Communications of the ACM*, vol. 15, no. 1, pp. 11–15, Jan. 1972.

[32] I. Dokmanić, R. Parhizkar, A. Walther, Y. M. Lu, and M. Vetterli, "Acoustic echoes reveal room shape," *Proceedings of the National Academy of Sciences*, vol. 110, no. 30, pp. 12 186–12 191, Jul. 2013.

[33] I. Jager, R. Heusdens, and N. D. Gaubitch, "Room geometry estimation from acoustic echoes using graph-based echo labeling," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2016, pp. 1–5.

[34] M. Coutino, M. B. Møller, J. K. Nielsen, and R. Heusdens, "Greedy alternative for room geometry estimation from acoustic echoes: A subspace-based method," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2017, pp. 366–370.

[35] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.

[36] R. Schnabel, R. Wahl, and R. Klein, "Efficient RANSAC for Point-Cloud Shape Detection," *Computer Graphics Forum*, vol. 26, no. 2, pp. 214–226, 2007.

[37] N. Kiryati, Y. Eldar, and A. M. Bruckstein, "A probabilistic Hough transform," *Pattern Recognition*, vol. 24, no. 4, pp. 303–316, Jan. 1991.

[38] A. Yla-Jaaski and N. Kiryati, "Adaptive termination of voting in the probabilistic circular Hough transform," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 9, pp. 911–915, Sep. 1994.

[39] L. Xu, E. Oja, and P. Kultanen, "A new curve detection method: Randomized Hough transform (RHT)," *Pattern Recognition Letters*, vol. 11, no. 5, pp. 331–338, May 1990.

[40] E. Vera, D. Lucio, L. A. F. Fernandes, and L. Velho, "Hough Transform for real-time plane detection in depth images," *Pattern Recognition Letters*, vol. 103, pp. 8–15, Feb. 2018.

[41] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1, pp. 37–52, Aug. 1987.

[42] D. Borrmann, J. Elseberg, K. Lingemann, and A. Nüchter, "The 3D Hough Transform for plane detection in point clouds: A review and a new accumulator design," *3D Research*, vol. 2, no. 2, p. 3, Nov. 2011.

[43] D. S. Johnson, C. H. Papadimitriou, and M. Yannakakis, "How easy is local search?" *Journal of Computer and System Sciences*, vol. 37, no. 1, pp. 79–100, Aug. 1988.