Karl Henrik Olof Ejdfors

Design of an AR Based Framework for Acoustic Simulation

Master's thesis in Electronics Systems Design and Innovation
Supervisor: Andrew Perkis
June 2021

Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Electronic Systems

NTNU
Norwegian University of Science and Technology

Karl Henrik Olof Ejdfors

Design of an AR Based Framework for Acoustic Simulation

Master's thesis in Electronics Systems Design and Innovation
Supervisor: Andrew Perkis
June 2021

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Electronic Systems

NTNU
Kunnskap for en bedre verden

Abstract

Proper room acoustics is vital for a holistic experience and comfort but can be challenging to achieve. Classic acoustic simulation software can present properties about a room's acoustics but does not let a client "experience" the results. This thesis looks into how immersive technologies can let clients experience different acoustical designs and what effects augmented reality (AR) contribute to spatial presence and sound perception for acoustic simulations.

An Android application is developed as a framework for real-time acoustic room simulations in AR. This framework aims to give the user an arena for experiencing a virtual room's acoustics and perceiving how changes to the design affect the sound. The acoustic replication of the room is based on the image-source model for generating room impulse responses to be convolved with anechoic sounds. The system is tested by a focus group, whose profession is acoustics, for evaluating the concept and indicate the immersive effects AR provides to acoustic simulations.

The results from the experiment with the focus group suggest that real-time acoustic room simulation in AR provides a client a sense of being present in an acoustical room. It was also evident that AR technology enhances the perception of small changes in sound. However, the acoustic representation needs further improvements to give the user a more realistic feeling.

Sammendrag

Romakustikk er viktig for komfort og en helhetlig opplevelse, men kan være vanskelig å oppnå. Vanlige akustikk-simuleringsprogrammer kan få frem romakustiske egenskaper, men legger ikke til rette for at en klient kan "oppleve" resultatene. Denne rapporten ser nærmere på hvordan teknologi som bruker immersjon kan få en klient til å oppleve forskjeller i akustiske design og hvilke effekter utvidet virkelighet (AR) bidrar til romfølelse og lydoppfatning i akustiske simuleringer.

En Android-applikasjon er utviklet som et rammeverk for romakustiske simuleringer i sanntid med AR. Dette rammeverket skal gi brukeren en arena for å oppleve akustikken i et virtuelt rom og oppfatte hvordan endringer i designet påvirker lyden. Den akustiske gjengivelsen av rommet er basert på bildekildemetoden for å generere romimpulsresponser som kan foldes med anekoiske lyder. Systemet er testet av en fokusgruppe som jobber innen akustikk for å evaluere konseptet og indikere de immersive effektene AR tilføyer akustiske simuleringer.

Resultatene fra eksperimentet med fokusgruppen indikerer at akustisk romsimulering i sanntid med AR gir en klient en følelse av å være til stede i et akustisk rom. Det var også tydelig at AR-teknologi forbedrer oppfatningen av små lydendringer. Derimot trenger den akustiske gjengivelsen ytterligere forbedringer for å gi brukeren en mer realistisk følelse.

Preface

This thesis marks the completion of the Master's degree program in Electronics Systems Design and Innovation at the Department of Electronic Systems (IES) at the Norwegian University of Science and Technology (NTNU). The Master's thesis was performed and completed during the spring semester of 2021 and is a part of the Signal Processing and Communication specialization.

The thesis' topic was proposed in collaboration with Norsonic AS. The layout of the project was discussed in collaboration with Professor Andrew Perkis at NTNU.

I want to thank Professor Andrew Perkis for his contribution as a supervisor and his valuable support during the work. A special thanks to Electroacoustic Engineer Erlend Fasting at Norsonic AS for feedback, guidance, and fruitful conversations. I would also like to acknowledge the experiment participants at Norsonic AS for evaluating the system. In addition, I would like to thank the piloting participants for their time and for contributing to improvements in the application.

As a final note, I would like to thank my family and friends for their encouragement and support throughout my years of study.

Karl Henrik Olof Ejdfors

Table of Contents

Acronyms

AR Augmented Reality

AVR Acoustic Virtual Reality

BRIR Binaural Room Impulse Response

FFT Fast Fourier Transform

HMD Head-Mounted Display

HRIR Head-Related Impulse Responses

HRTF Head-Related Transfer Function

IMU Inertial Measurement Unit

LTI Linear Time-Invariant

MR Mixed Reality

NRC Noise Reduction Coefficient

RIR Room Impulse Response

SDK Software Development Kit

SLAM Simultaneous Localization and Mapping

VR Virtual Reality

XR Extended Reality

Chapter 1

Introduction

Acoustics can be a priority when designing facilitates requiring high performance of sound reproduction. Cathedrals, for instance, are known to be massive and beautiful constructions that have a very generous acoustic inside. Opera houses and theatres go into the same category, but more common facilities such as offices, educational facilitates, or restaurants may overlook a proper acoustic design [1, 2].

Improper room acoustics can be an attribute for dissatisfaction in, e.g., restaurants. A restaurant with a rustic style, where the building materials are bricks or stones, can look appealing, but the acoustics can be disturbing. This effect is unwanted both for the guests and the restaurant owner. Sometimes it is sufficient to reduce the reverberation in the room, and it can be beneficial to engage architects and engineers for consultation. After a consultation, they can perform and provide a simulation of the room's acoustics.

Acoustic simulation software have proved its position in the industry over the last 30 years [1, 3]. The software are highly accurate for resolving a broad range of structural problems, and new approaches and improvements are applied to implement specific acoustic phenomena. Such simulations can require expert use as the results often are presented as bar charts, numbers, tables, or color maps. These presentations can be challenging for an oblivious client to understand and imagine how they will adapt to reality. A way to bring down the complexity of such presentations is to auralize the result such that one can perceive how a sound or soundscape will fit in a room.

Over the last years, a new trend of "experiencing" the results from a room acoustic simulation has emerged [4–6]. This trend is accomplished by multisensory perception from both visual and auditory stimuli simultaneously, which can adapt to user placement and rotation. This approach can be experienced within the extended reality (XR) domain (i.e., virtual (VR), augmented (AR), and mixed reality (MR)) [5–9] and have several additional benefits to classic acoustic simulations.

XR technology provides an extra dimension of immersiveness in a user's experience [10–13]. This immersive feeling can give the user a sense of "being present" within the system and feel like he takes an active part in the scene. In this thesis, we will create a framework for real-time room acoustic simulations and base the user experience on perceived sound and immersive effects with AR. The framework will be developed as an application for mobile devices, and the sound will be played from a set of headphones. This design choice aims at enhancing the flexibility and accessibility of acoustic simulations and can be an important tool for consultants or advisors to bring on field consultations.

The framework will facilitate for effortlessly experiencing real-time changes in a room's acoustics and different acoustical designs. Communicating acoustical properties of materials for non-acousticians will then be less challenging as it breaks down the complexity of a simulation in an immersive and intuitive way.

In contrast to classic simulation software, there will, in this case, be limitations to computation time for a real-time approach. These limitations will influence how accurate the simulation will perform. This thesis will emphasize the user experience and possibilities of perceiving different acoustical designs and materials in a compelling and immersive way and look further into the effects AR can provide to acoustic simulations. A focus group of eight participants working within acoustics at Norsonic AS will test the prototype application. This experiment will be the foundation for a qualitative analysis of the concept.

To gain context for the framework, we first present related work and use cases for acoustic simulations in the XR domain.

1.1 Related Work

Over the last couple of years, there has been an emerging interest in XR applications in society. The excitement has led to a broader uptake for science, education, industry, and entertainment disciplines [7]. The vast application sector within XR facilitates the users to experience spatial environments and interact with augmented information for an immersive and compelling experience. Previous research has shown that the sense of being present in the environment gets improved in combination with spatial audio [5, 14–16].

Advances in computer vision have led to fields within 3D reconstruction for capturing and mapping real-world scenes from images and video. Schissler et al. [10] proposed a novel algorithm to generate virtual acoustic effects to such scenes by using machine learning algorithms and estimation of acoustic material properties. They classified and optimized the material absorption coefficients to match measured impulse responses to apply acoustics to the scene. After reconstructing the 3D scene with applied acoustics, it can be rendered and explored as a multimodal XR environment.

Kim et al. [6] proposed in their work a similar approach to simulate acoustics in reconstructed scenes and assign those to Google's spatial audio software development kit (SDK) Resonance Audio[1] for immersive audio representation. In both cases, the results are virtual environments of predefined scenes with close to reality acoustics.

[1] `https://resonance-audio.github.io/resonance-audio/`

Acoustic virtual reality (AVR) is a technique for exploring auralizations in a VR environment. Pind et al. [4] showed a use case of this technology in their experiment, where a user can explore AVR mock-up scenes of classrooms in Carl H. Lindner College of Business. The purpose of the experiment was to experience the acoustics in two different designs of a classroom before making permanent design decisions. It proved to be valuable for making objective room acoustic parameters more accessible and understandable, even though the acoustic representation was naïve and not very accurate.

Within the video games and entertainment sector, the audio representation closely links to that of AVR. The objective is often the same, i.e., to have realistic virtual acoustics and improve immersion [4]. Computer games such as *Overwatch* and *Counter-Strike: Global Offensive* takes advantage of this and use spatial audio to provide information to the players, whether it be locating incoming gunfire, recognizing footsteps, or warning the player that their health is low [17]. The emerging use of 3D audio effects in games and XR have led to several audio plugins for game engines.

Steam Audio[2] is a plugin one can use to provide immersive audio solutions for games and XR. The plugin provides audio that will fit the in-game geometry and combine occlusion, reflection, revert, and head-related transfer function (HRTF) effects and is actively developed.

A similar plugin is Google's Resonance Audio, whose aim is "to replicate how real sound waves interact with human ears and with the environment"[18]. Resonance Audio replicates the real-world sound wave interactions with human ears to determine a sound's horizontal location and elevation and provides immersive audio to the scene. Another similar plugin is AudioKinect Wwise[3] which provides interactive audio solutions and simulates audio environments in games and XR.

The provided research of spatial audio in XR environments has many similarities with this project. The combinations of spatial presence and acoustic simulation of environments are the main considered topics when developing the simulation framework. The audio plugins can be used to spatialize audio. However, they cannot be included in our project because they are restricted to predefined scenes and do not support arbitrary absorption coefficients.

1.2 Structure of the Report

The report is divided into several chapters, where each presents different aspects of the system. Chapter 2 presents a theoretical approach to media technology and acoustic virtual reality. Chapter 3 provides the test procedure and method for implementation of the system. Chapter 4 provides the test results and observations from the experiment session. Chapter 5 goes more in depth about the test results, system, and further work, and finally, a conclusion is given in Chapter 6.

[2] `https://valvesoftware.github.io/steam-audio/`
[3] `https://www.audiokinetic.com/products/wwise/`

Chapter 2

Theory

The acoustic framework relies on both media technology and acoustic phenomena to be a persuasive and immersive simulation tool. Different media give the application a base for the user's experience in terms of immersion. At the same time, the acoustic aspects provide the user with a perception of how the sound adapts to changes as he interacts with the application. The following sections describe the use of this theory in the development of the tool.

2.1 Media Technology

This section will present the aspects around the multimodal composition of the application in terms of environment, interactivity, immersion, and audio.

2.1.1 Spatial Aspects of Augmented Reality (AR)

AR technology is the core technology used in our simulation tool for tracking and making it possible for the device to follow the user's orientation. Additionally, it superimposes virtual objects to the real world and creates an augmented environment for the user to interact with. AR has become a widely used technology many smartphone users utilize daily during the last couple of years. Examples of use cases are the translation of foreign language signs by using the mobile's camera and Google Translate, AR GPS navigation on streets, or showing the solar path and other relevant information as an overlay to a camera image.

In our case, we want to enhance the user's understanding of how a room's acoustics is affected by materials and objects in the room. In an AR environment, we can interact with a virtual room and experience how we can affect the acoustics by walking around and changing materials or add objects. Svensson [8] defines AR technology as "aims at enhancing our perception and understanding of the real world by superimposing virtual information on our view of the real world". Hence, the additional information from the augmented environment will enhance the user's perception and understanding of how acoustics can be affected in the real world.

The virtual room in the environment will get its local coordinates and rotation from how Simultaneous Localization and Mapping (SLAM) scans the environment and how the device's Inertial Measurement Unit (IMU) updates. IMU is a combination of the device's accelerometer, gyroscope, and magnetometer and is a typical combination of how modern smart devices handle AR content. With the device's camera, the IMU makes it possible for the device to map the environment and know its position and motion. This sensor technology is referred to as SLAM and enables interactions from the user.

2.1.2 Immersion Within the System

Immersion can be divided into two major perspectives: an individual's psychological state and an objective property of a technology or system [11]. To gain an advantage of an immersive experience, we design the application to unfold around the user. The virtual objects on top of the real environment will be intuitive and familiar, and interactions with the application will give the user a real-time response.

Immersion can refer to the capability a system has to occupy our perceptual system and simulation environments that evoke a feeling of "being there" [12]. To achieve such immersive effects, we include several elements. According to Hameed and Perkis [12], immersion shall include system immersion, absorption and engagement, strategic and tactical immersion, imaginative immersion, challenge-based immersion, ludic immersion, and narrative immersion. Hence, a persuasive system is essential for enhancing immersion, which can be achieved through a well-designed virtual or augmented experience. We emphasize system and challenge-based immersion as well as absorption and engagement through the multimodal structure of the application. This structure contains media, interactions, and real-time feedback and will be an engaging arena for the user.

Sound is an essential component for our experience in virtual environments and can help to, e.g., create a realistic sense of the world, take a part of the environment and help get emotionally engaged [14]. Spatial sound is sound positioned in a 3D space around the listener and can add a sense of presence. Our system will replicate the acoustics in a 3D room and use omnidirectional 3D sound where the perceived sound pressure level reduces with the distance from the source.

Binaural rendering, which requires convolution of source signals with head-related impulse responses (HRIRs) or binaural room impulse responses (BRIRs), aims to evoke an extra immersive experience for the listener [15]. These methods utilize the human auditory system to perceive sound from a particular direction and are of great interest in the fields of XR and virtual acoustics. However, this often requires extra plugins or specialized equipment and will be left out of this prototype framework.

2.2 Acoustic Virtual Reality

Room impulse response (RIR) measurements are widespread for capturing the acoustical characteristics of a real room [19]. The measurements require a physical setup of sound sources and receivers in the room of interest, but in our dynamic virtual room, we need to calculate and generate the RIR. For simulating acoustic phenomena in the room, we need a set of parameters for characterizing it. The following section will discuss an algorithm for generating an RIR.

2.2.1 Room Impulse Response Generation

To simulate room acoustics, we want to generate an RIR for a desired virtual room. There are several approaches for generating RIRs, including wave-based, ray-based and statistical modeling [20]. The wave-based methods are more computational demanding than ray-based, which works better for real-time simulations. This statement, however, is because of simplifications in the ray-based methods. The statistical methods are more used for high-frequency noise analysis and acoustic designs. The ray-based methods, mainly “Ray Tracing” and “Image-Source”, are based on sound particles traveling as sound rays [21]. We will base our generation of RIR on the image-source method.

Allen and Berkley’s image-source method can be used to simulate the reverberation in a rectangular room for a given source and receiver location [22]. This model is based on the direct path between the source-receiver pair and the pathways with corresponding delays between source images and the receiver. The direct path length can be calculated from the known locations of the source and receiver, whereas the source images are used for calculating the reflected paths. Figure 2.1 shows an image source located behind a wall with an equal distance from the wall as the source. Because of symmetry, we can compute the reflected pathway by the distance between the image source and the receiver. The time delay of the reflected signal is thus corresponding to the distance given as

$$\tau = \frac{d}{c}, \tag{2.1}$$

where d is the distance and c is the speed of sound. This model can again be made more complex by introducing more reflections and hence images.

The six reflecting surfaces in a room have their reflection coefficient β, or absorption coefficient α by the relation

$$\alpha = 1 - \beta^2. \tag{2.2}$$

These absorption coefficients are found as noise reduction coefficients (NRC). NRC can be expressed as the average rating of sound absorbed by an acoustic material and ranges from 0 to 1 [23]. An NRC rating of 0 means that the material absorbs no sound, whereas an NRC rating of 1 means that all sound is absorbed. We can express the NRC as the arithmetic average of absorption coefficients of a set of frequencies. The octave band center

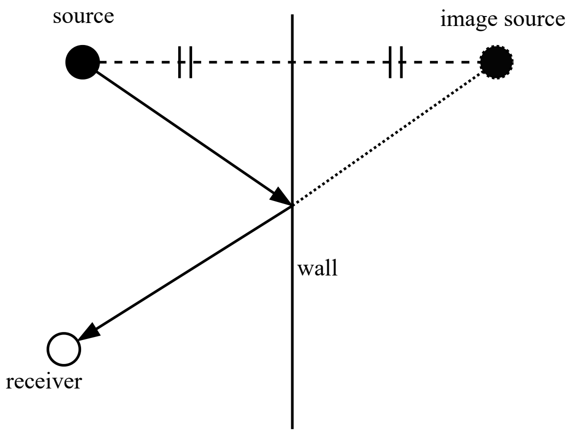


Figure 2.1: One reflection path for the image-source method.

frequencies 125, 250, 500, 1000, 2000, and 4000Hz can determine the NRC as

$$\text{NRC} = \frac{\alpha_{125} + \alpha_{250} + \alpha_{500} + \alpha_{1000} + \alpha_{2000} + \alpha_{4000}}{6}. \tag{2.3}$$

Even though the NRC indicates how well the material absorbs sound, it does not reflect the frequency dependence of the material. For materials with an absorption coefficient that vary much with frequency, the NRC rating is less accurate for representing the material's acoustic properties. A sample of common materials and their NRC rating is shown in Figure 2.2. Here, we can easily see that some materials absorb sound differently even though they have similar absorption coefficients. We can see that glass, on the one hand, reduces its absorptive properties for higher frequencies, whereas concrete, on the other hand, increases its absorptive properties for higher frequencies. This contrast in frequency dependence makes the two materials absorb sound very differently, even though they have a similar NRC rating of 0.06 and 0.05.

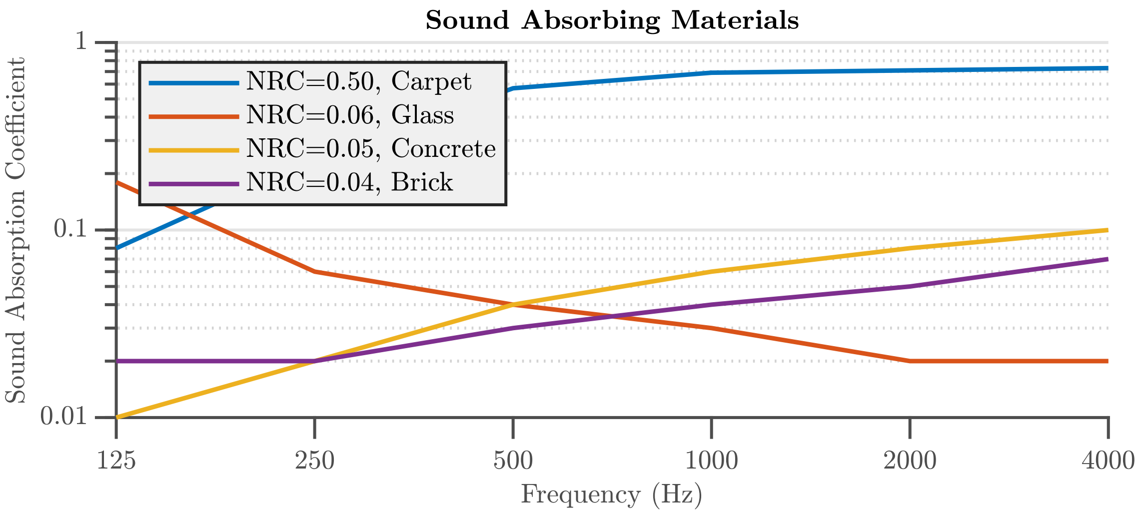


Figure 2.2: Sound absorption coefficient and NRC rating for a sample of common materials [24].

Allen and Berkley proposed in their work a method for generating an RIR in the time domain with reflection coefficients as

$$h\left(t,\mathbf{X},\mathbf{X}'\right)=\sum_{\mathbf{p}\in\mathcal{P}}\sum_{\mathbf{m}\in\mathcal{M}}\beta_{x_1}^{|m_x-q|}\beta_{x_2}^{|m_x|}\beta_{y_1}^{|m_y-j|}\beta_{y_2}^{|m_y|}\beta_{z_1}^{|m_z-k|}\beta_{z_2}^{|m_z|}\frac{\delta\left(t-\tau\right)}{4\pi d},\tag{2.4}$$

for point source position $\mathbf{X}=[x,y,z]$ and receiver position $\mathbf{X}'=[x',y',z']$. The set $\mathcal{P}$, which denotes all desired triples of $\mathbf{p}$ is given as $\mathcal{P}=\{(q,j,k):q,j,k\in\{0,1\}\}$, $\mathcal{M}$ which denotes all desired triples of $\mathbf{m}$ is given as $\mathcal{M}=\{(m_x,m_y,m_z):-n\leq m_x,m_y,m_z\leq n\}$, where n is the number of samples, and the reflection coefficients of the six surfaces are $\beta_{x_1},\beta_{x_2},\beta_{y_1},\beta_{y_2},\beta_{z_1},\beta_{z_2}$ [20, 22]. The number of samples can be set by the desired sampling frequency, f_s, and the reverberation time, RT_{60}, in the room as

$$n=RT_{60}\cdot f_s.\tag{2.5}$$

However, since the triple $\mathbf{m}$ range from $-n$ to n, the RIR algorithm gets a complexity of $\mathcal{O}(n^3)$, and an upper limit should be considered to prevent long computation times. The reverberation time is defined as the time it takes for sound to decay by 60dB and can be empirically estimated with Sabin-Franklin's formula

$$RT_{60}=\frac{24\ln\left(10\right)V}{c\sum_{i=1}^{6}S_i\alpha_i},\tag{2.6}$$

where V is the volume of the room, S_i, and α_i are the surface area and the absorption coefficient of the i^{th} surface, respectively.

2.2.2 Sound Replication and Considerations

A generated RIR provides a unique characterization for an acoustic space. We consider our room to be a linear time-invariant (LTI) system to emulate the reverberation in the room by convolving a dry anechoic audio signal with the RIR. In this way, a listener will perceive the sound as it is originating from the virtual room. Since the impulse response generated by the image-source method is for a source-receiver pair, the sound replication will be "correct" for one point in the room. However, as a naïve approach, it is possible to emulate the sound for small changes in the receiver position by decreasing the sound pressure level with distance from the source. This relation can be accomplished by the spreading for spherical pressure waves p, with distance r as

$$p_{spherical}\propto\frac{1}{r}.\tag{2.7}$$

This naïve approach can be helpful in some contexts, e.g., when the accuracy of the sound replication is less important than the system's flexibility.

In the case of real-time approaches for replicating changes in a room's acoustics, one needs to accept lower accuracy on the output. The tradeoff of accuracy gains flexibility such that, e.g., moving sources, changing geometry, and surface materials can be handled in real-time. This flexibility can be advantageous in the early design stages [4] as it indicates how different designs will be perceived but will not work properly as a high-quality acoustic simulation.

Our generation of the RIR is a ray-based method. As the ray-based methods assume that sound travels along straight lines or rays, we have some wave phenomena that cannot be modeled, such as diffraction and interference [25]. Diffraction appears when the wavelength of a sound wave is smaller or equal to the size of an obstacle and tends to bend around it or spreads out waves beyond small openings. Diffraction is thus most evident at low frequencies and can increase accuracy tremendously. Boundary conditions are also challenging to include in the generation of RIRs and are an essential factor that can affect accuracy [25].

The following chapter will provide the research method and implementation for developing the application.

Chapter 3

Method

This chapter will first provide the research procedure of the system. Furthermore, it will present the implementation, specifications, and technical requirements of the application. In the end, it will provide a pilot test with improvements to the system and the setup of the experiment.

3.1 Research Procedure

We want to understand opinions and experiences from the test group to gain insightful information about the developed acoustic framework. The participants are a focus group of eight professional workers within acoustics who have many years of experience developing, advising, or marketing acoustic equipment. They will take the acting role as a client, such as a restaurant owner, who wants to experience differences in room acoustics in their restaurant. Such clients can gain an advantage of experiencing different acoustical designs to understand the effect reverberation can have on customers and what measures can be taken. The participants' profession is vital for evaluating the system because they will be critical to the sound replication and the simulation tool. However, they can be confused by the acting role, which can bias the answers, but the results will still indicate how a bigger group will perceive the system.

The procedure for research and evaluation of the system is formulated as a protocol. This protocol will ensure a consistent testing scheme and is attached Appendix B, Section B.1.

In terms of research type, we desire to analyze the experiment qualitatively. In this way, we can look into anomalies and differences in individual implementations and answers for comparing them with other interesting observations and results. We use surveys, questionnaires, and open-ended questions for additional comments to collect data about the experiment. These surveys are attached in Section B.5. Observations and discussions with the participants are also vital data collecting sources. This data collection facilitates expressive information, which can capture the participants' experiences, meanings, and perspectives and are essential for studying the system.

The demographic survey and background information about the participants will give context to the analysis of the results. The questionnaires are inspired by Temple Presence Inventory (TPI) [26] for evaluating telepresence and are adapted to fit this experiment. The questions that do not require an open answer are answered by a 7-point Likert scale, where the scale ranges from 1 (negative response) to 7 (positive response). This scale will give a measurable response from each participant, which will be compared individually with the rest of the answers and the other participants' answers.

The open-ended questions will give the participant more freedom to express their experience and provide context to their answers. This data is essential when analyzing the results and will be used together with observations, discussions, and answers to evaluate the proposed hypotheses.

3.2 Hypotheses

We expect the participants to be immersed in the system and get a feeling of being present in the AR simulation. The adaptive acoustics is also expected to be more emphasized when the participant is in an AR environment. Therefore, two hypotheses have been made:

- H_1: Real-time acoustic room simulation in AR provides the client a sense of being present in the acoustic room.
- H_2: AR technology enhances the perception of small changes in sound.

The following section will go further into the system materials and implementation of the application.

3.3 System Materials

There are two sides of the system materials that are used for developing the system. This section will provide the hardware that are used and the software required to build the application.

3.3.1 Hardware

Modern Android devices supporting AR functionality through Google's ARCore SDK can use the developed application. In the implementation, we use a Samsung Galaxy tablet with specifications listed in Table 3.1.

Google's ARCore SDK is a platform for building AR experiences and uses SLAM for real-time motion tracking. SLAM is used to understand where the device is located relative to the world around it by extracting feature points from the rear camera [28]. When the user moves the device around in the real world, the device's IMU and the visual information

Table 3.1: Specifications for Samsung tablet [27].

Property	Value
Model	SM-P610
Sensors	Accelerometer, gyroscope, hall-effect, RGB light
Screen size	10.4”
Camera resolution (back/front)	8.0 MP / 5.0 MP
OS	Android version 10
RAM	4 GB
Space	64 GB
CPU-type	Octa-Core

estimate where the device is relative to the real world over time. These sensor data then work as input to be analyzed and processed by the various software used.

3.3.2 Software

Unity, a game engine made by Unity Technologies, is used to develop the AR application. It is a cross-platform engine widely used for creating games and interactive experiences in 2D, 3D, and XR. Android SDK is linked to Unity and makes it possible to compile to Android devices.

For controlling application behavior in Unity, we use scripts written in C#. These scripts are developed and debugged in Visual Studio Community and are found in Appendix C.

To prepare, pre-process, and analyze the data used in the application, we use MATLAB by MathWorks. MATLAB is a professionally developed, fully documented, and powerful programming- and numeric computing platform. The listed software with their corresponding versions are given in Table 3.2.

Table 3.2: Software used for development.

Software	Version
Unity	2019.4
Android SDK	30
Visual Studio Community	2019 16.9
MATLAB	R2020a

The listed hardware and software are required to implement the application. The following section will provide a system description and implementation of the theoretical concepts from Chapter 2.

3.4 System Description

The framework for the acoustic room simulator with AR technology is, among other things, designed to be an immersive and innovative tool for experiencing changes in perceived room acoustics. The user will perceive the changes in real-time as the acoustic model updates on behalf of the user's inputs.

For human-computer interactions in this framework, we use AR technology. This technology facilitates immersiveness and makes it possible for the user to feel more present in the scene. The user of the framework will experience how he can affect acoustics by superimposing and changing virtual objects to the real environment and changing his position. Walking around will change the auralization of the sound because of the perceived sound level changes and the position-dependent RIR.

The implemented mathematical model for calculating the image-source RIR includes reflections from surfaces, dimensions of the room, player position, and absorption coefficients, as stated in Section 2.2.1. This model processes the input data from the sound source with regard to the placement of the source and the user. Then, it calculates the reverberation time in the room and generates an impulse response. The generation of the impulse response is hence dependent on the length of the reverberation time. This relation will again affect the computation time for the model because long reverberation times require more samples for the impulse response.

Because a group of participants will test and evaluate the system, we have divided the application into three different scenes. The following section will provide the scene composition and what is included in each scene.

3.5 Scene Overview

When a user is testing the system, he should test all the application features in a guided and intuitive way. We accomplish this intuitive flow by assigning smaller tasks to each scene. The scenes are designed to make the user gradually explore more challenging and complex features and are hence divided into three scenes: *Simple*, *Complex*, and *Challenge*. Each scene has a corresponding survey which can be found in Section B.5 in Appendix B.

3.5.1 Scene 1: Simple

The first scene will give the user the experience of being in two very different acoustical rooms and works as a “warm-up” for the other scenes. The first room in the scene has a very long reverberation time, whereas the second room has a much shorter reverberation time. The technical data for the designs of the two rooms is listed in Table 3.3.

The tasks in this scene are considered to be very simple and easy to understand for people with no previous XR experience. The first encounter for the user is an information

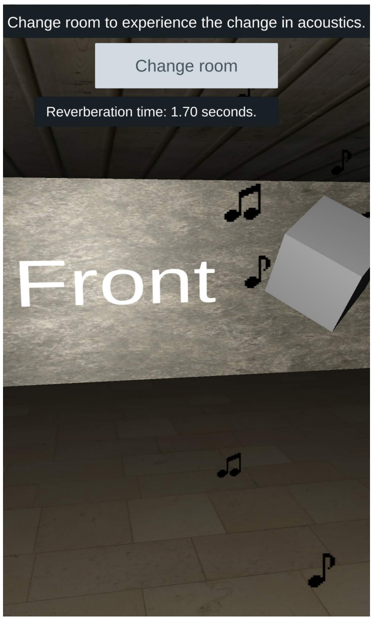
Table 3.3: Technical data for Scene 1.

Attribute	Room 1	NRC	Room 2	NRC
Dimensions L × W × H (m)	8 × 10 × 3		8 × 10 × 3	
Wall material	Concrete	0.05	Wood panel	0.40
Floor material	Marble	0.01	Plywood	0.23
Ceiling material	Wood panel	0.40	Plaster	0.05
Reverberation time (s)	1.70		0.42	

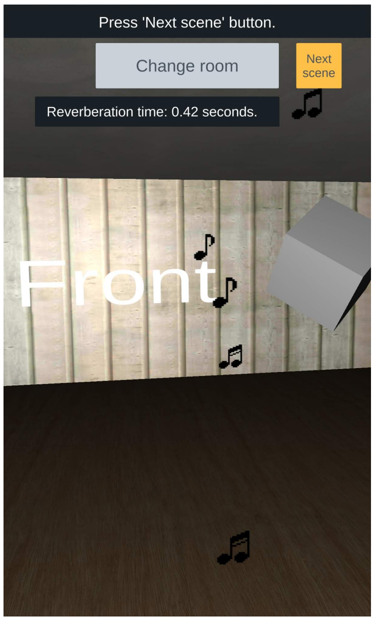
screen, which informs the user about what he is expected to encounter in the scene. The information screen states:

> In this scene, you will experience changes in room acoustics between a "hard" and a "soft" room.
>
> After you have successfully completed the scene, you will be asked to answer a short survey about sound perception.
>
> *You will soon enter an AR environment. Keep in mind that you can walk around and explore the scene.*

After reading the information, the user will be placed inside the virtual room, which is designed to have a long reverberation time. The materials used have hard surfaces that provide little absorption and gives the user a feeling of being inside, e.g., a tunnel or bunker. This room is shown in Figure 3.1a.



(a) First room in Scene 1.



(b) Second room in Scene 1.

Figure 3.1: The two rooms in Scene 1. Surface materials and the reverberation times are different in the two rooms.

When the user is inside the room, he will experience the application's features in a guided way. Figure 3.2 summarizes the flow with additional details about the processes in the application.

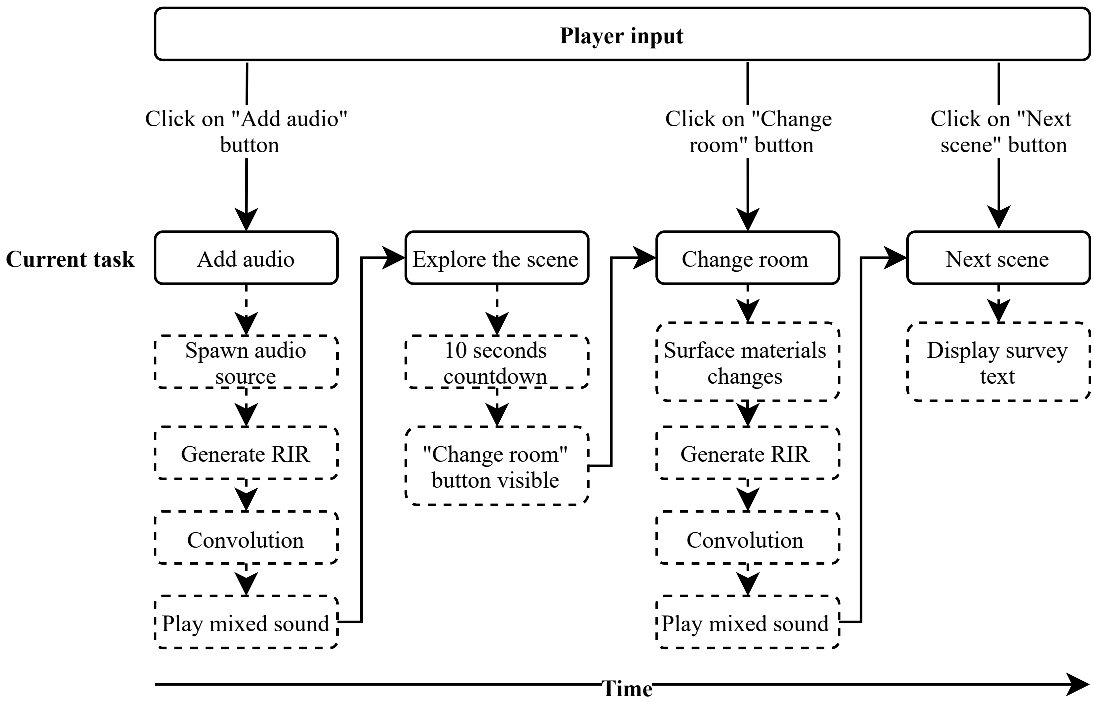


Figure 3.2: Flow diagram of Scene 1.

The first task is to add audio to the scene. When clicking on the “Add audio” button, there will be spawned an audio source playing an excerpt from the audiobook “Ut av det blå”[29]. This excerpt has a woman’s voice who speaks very clear without noise and is mixed with the room’s impulse response to replicate the acoustics in the room. After the audio is added, a timer will count down from 10 seconds and encourage the user to explore the current room. When the countdown is finished, the “Change room” button becomes visible. When the user changes the room, the “Next scene” button becomes visible and blinking, and the surface materials of the room change. The application then generates a new RIR, and the sound updates accordingly. In the second room, the building materials are softer, as shown in Figure 3.1b, and we expect the user to hear that the echo from the first room disappears clearly. The user can now change between the two rooms until he clicks on the “Next scene” button. The scene is then finished, and he will be asked to answer a survey and proceed to the next scene.

3.5.2 Scene 2: Complex

The second scene is designed to be more complex than the first scene. Here, the user will experience a virtual restaurant with several noise sources to make a soundscape for a more authentic feeling. There are two noise sources of excerpts from cocktail party sounds from ODEON [30], and one noise excerpt of guitar play [31], used to replicate the music from a live band. In addition, the same woman’s voice from the previous scene is now attached to a 3D character[1], which is placed inside the restaurant. This character is further referred to

[1] 3D figure “Meghan” downloaded from Mixamo [32].

as the "guest". The technical data for the design of the whole scene is listed in Table 3.4.

Table 3.4: Technical data for Scene 2

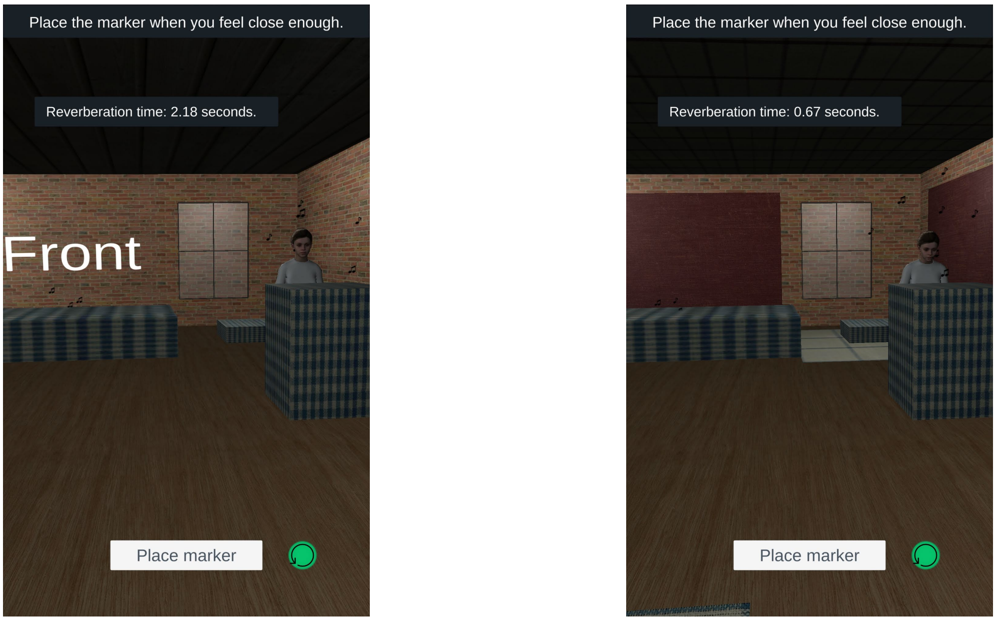
Attribute	Room 1	NRC	Room 2	NRC
Dimensions L × W × H (m)	16 × 12 × 4		16 × 12 × 4	
Wall material	Brick	0.05	Brick	0.40
Floor material	Plywood	0.01	Plywood	0.23
Ceiling material	Wood panel	0.40	Acoustic roof panel	0.95
Items	Tables×5	0.80	Tables×5	0.80
	Windows×4	0.02	Windows×4	0.02
	Wood door×1	0.08	Wood door×1	0.08
	Scene×1	0.80	Scene×1	0.80
			Carpets×3	0.50
			Acoustic tiles×5	0.95
Reverberation time (s)	2.18		0.67	

This scene is divided into two parts: one where the user will explore the restaurant inside an AR environment by walking around in the test location. In the other part, the user will stand still and move a character by interacting with an on-screen joystick and spectating from a bird-view camera of the restaurant. In this way, the user will interact with an immersive AR environment and interact with the same environment without AR functionality enabled. The user will then experience what differences AR does to the perceived acoustics in the room regarding how his senses are engaged and his feeling of being present in the AR environment. To test how the user perceives the sound in the different parts, he is asked to place a "marker" inside the restaurant. This marker shall be placed a distance from the guest where the user feels he can have an undisturbed conversation.

The user will first conduct the AR experiment and be met by an information screen to prepare him for the tasks. The information states:

> In this scene, you will experience an AR restaurant. The environment is noisy, and your main task is to submit when you feel close enough to your guest in order to have an undisturbed conversation.
>
> You will perform this task two times: one time without acoustic absorbing materials and one time with acoustic absorbing materials present.
>
> Afterward, you will be asked to answer a short survey about sound perception in the environment.
>
> *Please remember to use the green button frequently to update sound effects.*

After reading the information, he will be placed inside the AR restaurant. The two rooms in this scene are designed to replicate the acoustics in a rustic restaurant, where the interior is mostly made out of bricks. The first room, as shown in Figure 3.3a, has a long reverberation time, and the echo effect is very present. In the second room, there are acoustic tiles attached to most of the walls' surface, there are carpets on the floor, and the ceiling is also covered with acoustic tiles. This is shown in Figure 3.3b. These materials bring the reverberation time down to comfortable levels, and the user is expected to experience a difference between the two rooms.



(a) First room in Scene 2.

(b) Second room in Scene 2.

Figure 3.3: The two rooms in Scene 2. The second room has sound absorbing materials on the walls and carpets on the floor. The ceiling is also made out of a sound absorbing material.

The flow of the current tasks with additional details in the scene is shown in Figure 3.4. The first task is to add audio to the restaurant. This task initiates a generation of separate RIRs for each sound source and plays the corresponding mixed sound from each source. A green button for updating the sound effect will then become visible, as shown in Figure 3.3. This feature prevents unintended lags in the application that comes from the heavy computation power needed to generate new RIRs and convolving them with their corresponding sound source. However, it is emphasized to the user that he must update the sound effect frequently.

With the audio present, the user is encouraged to move closer to the guest. When the user gets closer than 6 meters, the “Place marker” button becomes visible. When clicking the button, there will be spawned a marker at the current position in the application. The coordinates relative to the guest are then stored in the application. Simultaneously, the “Next room” button becomes visible. The user can still update the location of the marker by clicking the same button.

When clicking the “Next room” button, the surface materials of the room change, and sound-absorbing materials are added. The application then generates new RIRs and updates the sound accordingly. The button then becomes disabled, and the user is urged to place a new marker. After the second marker is placed, the “Next room” button becomes enabled, and by clicking it, the user will be asked to answer a survey.

When the user has finished the survey, he will be met by a new information screen about

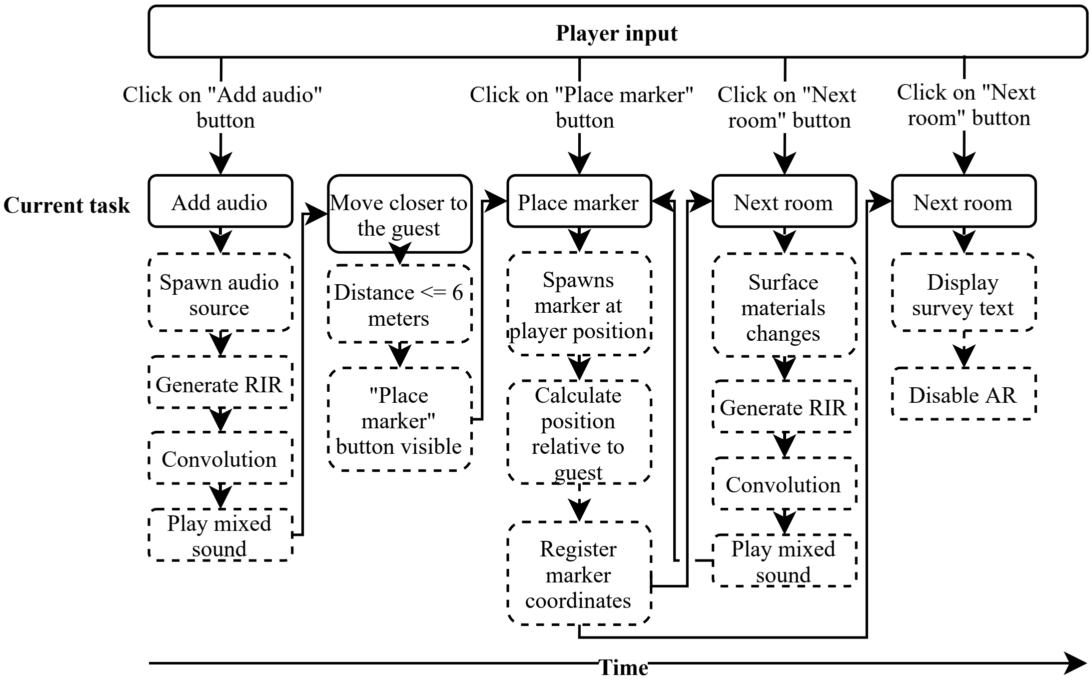


Figure 3.4: Flow diagram of Scene 2 with AR enabled.

the scene with AR functionality disabled. This information states:

> You will now do the same tasks, but this time with AR functionality disabled. Use the on-screen joystick to move around.
>
> After completing the tasks, you will be asked to answer a short survey about sound perception and how you experienced the difference between AR functionality enabled and disabled.

The flow of the scene with AR functionality disabled is shown in Figure 3.5. The tasks are similar to the scene with AR functionality enabled, but this time the user interface is different, as shown in Figure 3.6.

After completing the scene, the user will be asked to answer a survey and then proceed to the final scene. The coordinates for the marker positions will also be displayed, and the user is asked to write them down on the survey.

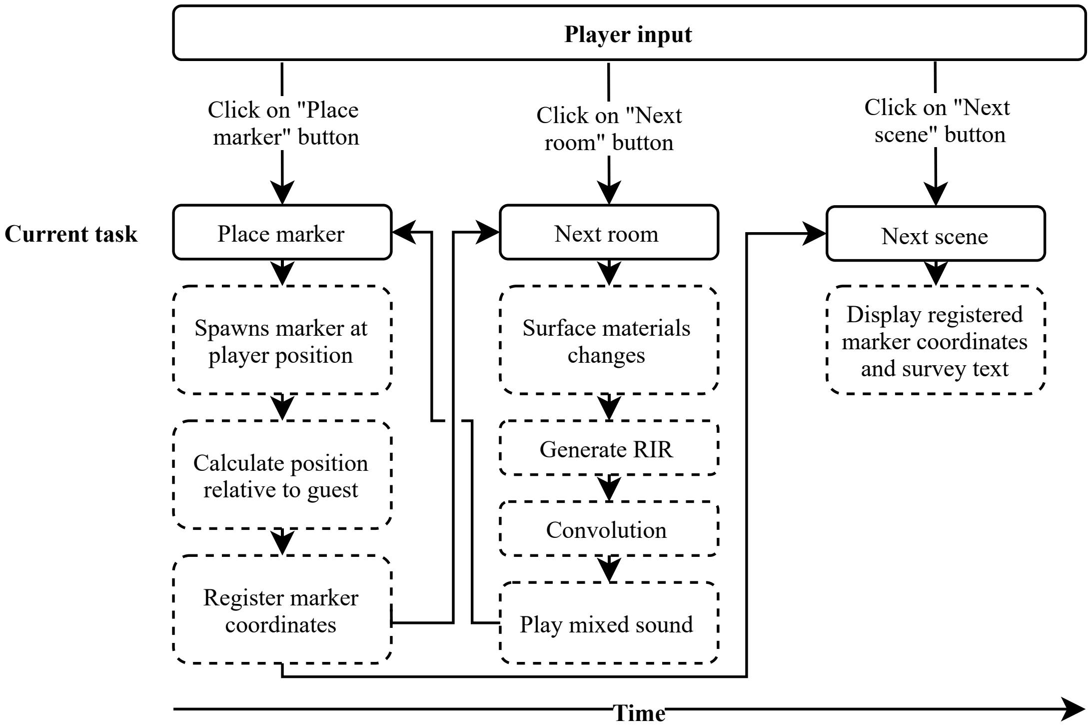


Figure 3.5: Flow diagram of Scene 2 with AR disabled.

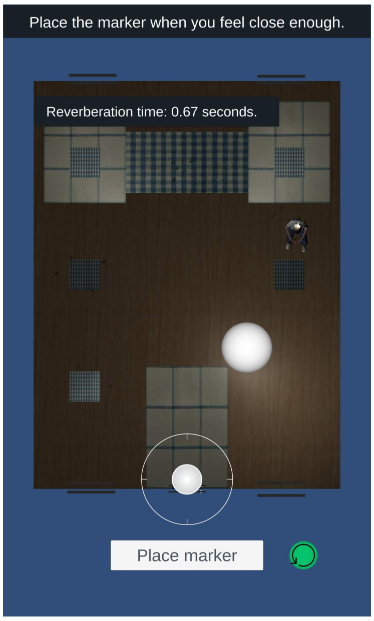


Figure 3.6: Second room in Scene 2. The on-screen joystick will move the white sphere.

3.5.3 Scene 3: Challenge

The third scene is an environment where the user can experience the different features of the framework in a guided way. To make the user try out the different features, they will complete three small tasks by following a walkthrough:

1. Add sound source to the scene.
2. Change front- and back wall material to brick.
3. Reduce the reverberation time by 1/3 by adding objects to the room.

The technical data for the room is listed in Table 3.5. The extra "initial" attribute in the table indicates the pre-set values in the room, and all the other options are optional for the user.

Table 3.5: Technical data for Scene 3

Attribute	Value	NRC	Comment
Dimensions L × W × H (m)	8 × 8 × 4		Initial
Wall material	Wood panel	0.40	Initial
	Plaster	0.05	
	Concrete	0.05	
	Brick	0.05	
Floor material	Marble	0.01	Initial
	Plywood	0.23	
	Concrete	0.05	
	Carpets	0.50	
	Metal	0.10	
Ceiling material	Plaster	0.05	Initial
	Concrete	0.05	
	Acoustic roof panel	0.95	
Items	Guitar play		Initial
	Cocktail noise		
	Carpets	0.50	
	Acoustic tile	0.95	
Reverberation time (s)	0.52		Initial

When the user encounters the last scene, he will first be met by an information screen stating:

> In this scene, you will explore more features in the application. You will be able to change room materials yourself and add sound absorbing objects to the room.
>
> Afterward, you will answer the last surveys about your sound perception, your experience with this application, and your impression.
>
> *Please remember to scale, rotate, and move the objects as it fits you.*

After reading the information, the user will be placed inside the AR room, and the walk-through of the small tasks begins. This room is initially empty, and the user is encouraged to add objects and change building materials. A detailed flow of this scene is illustrated in the diagram shown in Figure 3.7.

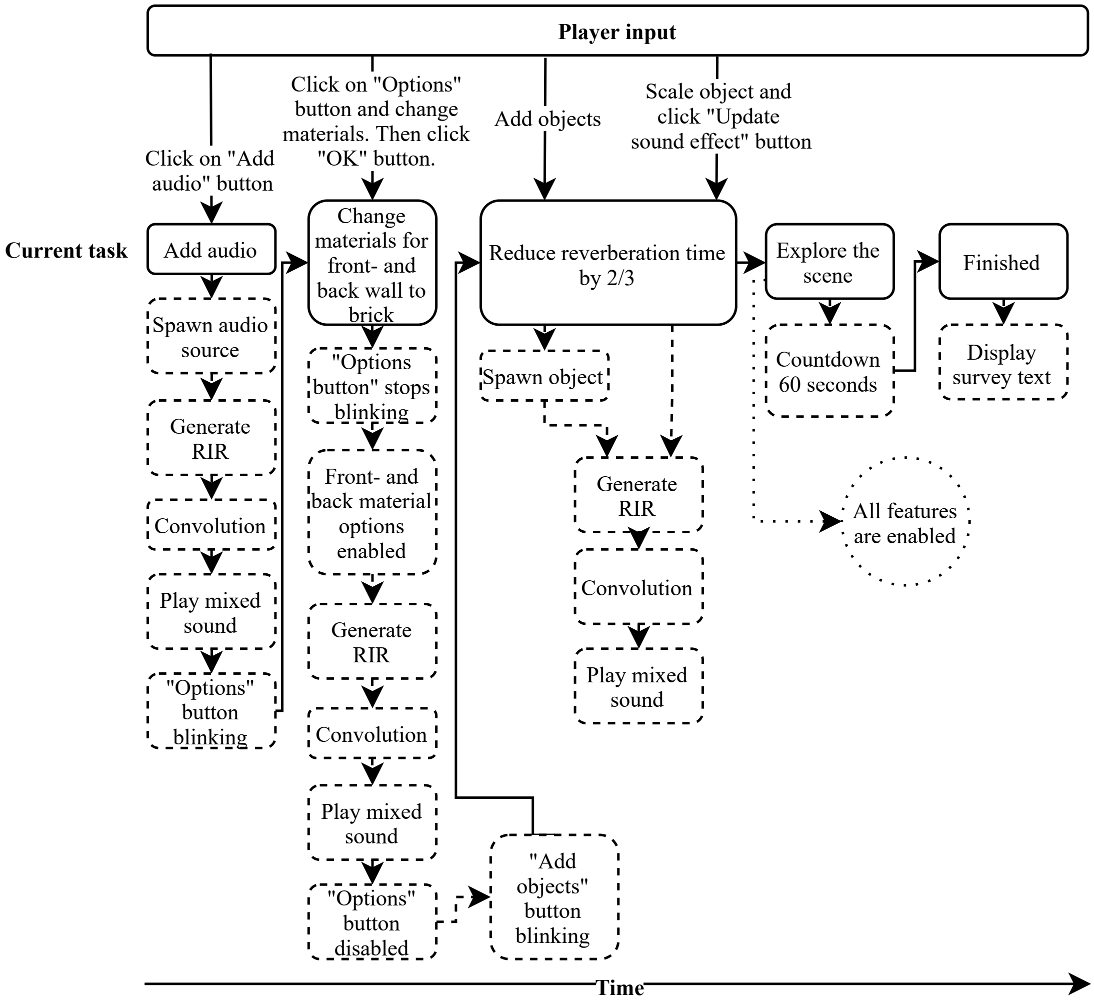


Figure 3.7: Flow diagram of Scene 3.

The user's first task is to add an audio source to the room. The audio clip playing is the same guitar excerpt from Scene 2. When clicking the "Add audio" button, it will disappear, and a "Update audio effect" button, together with an orange blinking "Options" button, will appear. The user will open an options panel by clicking this button and will be able to change materials on the surfaces. When the user has changed the front- and back walls to bricks, the "Options" button will disappear, and the sound updates to the new design. At the same time, the green "Add objects" button becomes visible and is blinking. The user is now encouraged to add objects from the menu to reduce the reverberation time in the room. This menu is illustrated in Figure 3.8a. After adding an object, the user can move and scale it. By clicking the "Update audio effect" button, there will be generated a new RIR, and the sound adapts to the changes. The user can still add other objects to the room until he has reduced the reverberation time by 2/3. After he has reduced the reverberation time, the "Options" button will be enabled again, and all the options in the menu will be enabled. The user now has a minute to explore the scene until a survey text appears. An illustration of a designed room is shown in Figure 3.8b.

(a) Illustration of the room in Scene 3 with menu opened.

(b) Illustration of the room in Scene 3 with acoustic tile on the wall and carpet on the floor.

Figure 3.8: Screenshots from Scene 3.

We now have more insight into how the scenes in the application are built up. The following section will provide what measures are taken for great usability and user interface.

3.6 Usability and User Interface

The design of the application will offer a great experience through interactions with the device. These interactions are made intuitive for the user, such that the available information on the screen is concise and useful. In terms of usability and user interface, several aspects are taken into account.

The integration of virtual objects and their display in the physical world facilitates the flow between reality and virtuality. The degree of emplacement the user experiences through this flow will increase the immersiveness of the experience. By placing the user near the real environment on the reality-virtuality continuum, as illustrated in Figure 3.9, he gets a low degree of emplacement and gains an advantage from being in the physical world and interact with superimposed artificial objects and sound.

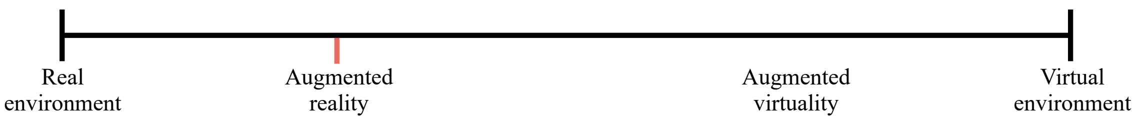


Figure 3.9: The reality-virtuality continuum.

The application is designed for the user to test and interact with all the available features. Hence, each scene is developed as a walkthrough of small tasks where the feature they are testing is highlighted on the screen, and other features are disabled. An example where the current task is to *"change the front- and back wall materials to brick"* is illustrated in Figure 3.10. Here, we can see that the user can only change the front- and back wall material. When they have completed the task and can explore the application freely, they will be able to select all the options. A complete list and information about the system's features are described in detail in the manual in Appendix B, Section B.3.

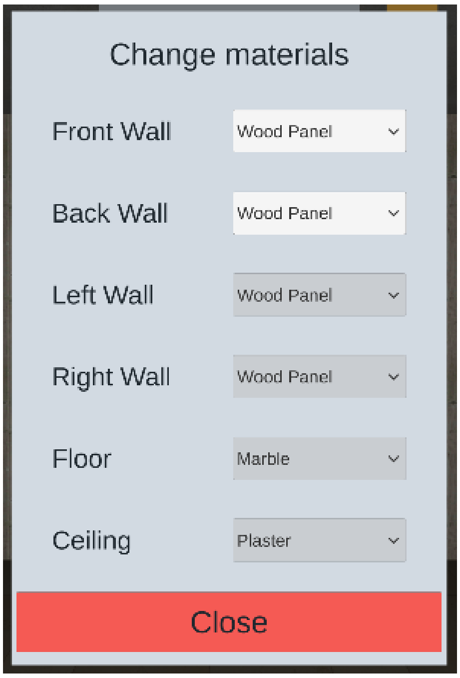


Figure 3.10: Example of enabled and disabled features.

The acoustics in the room is a result of several different factors, such as sample size, reverberation time, reflection order, and human ear sound perception. The following section will go over the implementation of the acoustical algorithms used in this application.

3.7 Implementation of Algorithms

The perceived sound and computation time have been in focus when developing the acoustic system of the application. To make the user perceive changes in the acoustics close to real-time, we need a fast and accurate algorithm for reproducing the acoustics. To accomplish this, we need to account for several restrictions.

As mentioned in Section 2.2.1, the process of changing, e.g., a wall's material in the room until the sound is reproduced for the user, is complex and time-consuming. The first part of the process is to generate an RIR for each sound source in the room. Trial and error testing suggests a combination of the variable values for the image-source method as listed in Table 3.6 for minimal computation time and adequate sound quality.

Table 3.6: Room impulse response values used in implementation.

Property	Value
Sampling frequency, fs (Hz)	16000
Number of samples	$\min\left(RT_{60} \cdot f_s, 2^{14}\right)$
Number of reflections	Auto
Microphone type	Omnidirectional

To get a detailed sample of the RIR while not restricting the number of samples too much, we set the sampling frequency to 16kHz. The maximum number of samples of 2^{14} prevents a too long computation time for generating the RIR and computing the convolution afterward. This restriction will distort the output signal for long reverberation times (> 1s), but the “echo” effect will still be evident. The algorithm for generating the RIR sets the number of reflections automatically. With a manually set reflection order, the representation of the sound was evaluated to be “digital” and “fake”. The virtual omnidirectional microphone provides contributions from all reflecting angles when recording the impulse response and is evaluated as the best solution for this setup.

Because each sound source needs a separate RIR, it is desired to compute them in parallel on the tablet’s available processor cores. Unity’s “Burst Compiler” for parallel jobs makes it possible to generate the impulse responses and convolve them with the input signals in parallel. However, “Burst Compiler” has restricted support of data types and does not support common implementations such as fast Fourier transform (FFT) or convolution. To perform convolution in the frequency domain, we need to implement and adapt these algorithms to be supported by the “Burst Compiler”. The implementations of the algorithms and the application are attached in Appendix C.

Parallel jobs reduce the computation time significantly compared to serial computation when there is more than one sound source in the scene. In addition to computing in parallel, we also pre-process the sound signals by down-sampling them by a factor of two to save computation time. The decimation factor is set by trial and error, where a higher decimation factor leads to too much distortion in the signal. The down-sampling hence reduces the computation time of the convolution algorithm, making the application run faster but at the cost of lower sound quality.

An example of a resulting output signal with the provided configuration is given in Figure 3.11. The upper tile shows the decimated excerpt of the input signal “Ut av det blå”, the middle tile shows the generated RIR for the first room in Scene 1 where the reverberation time is 1.70s. The number of samples is restricted because of computation time limits, and the bottom tile shows the convolution of the two signals. We can see from the waveform that the convolution has added extra echo to the signal as it looks more stretched out than the input signal.

With the implementation ready, it is important to test the system. The following section will go through the procedure of the pilot testing.

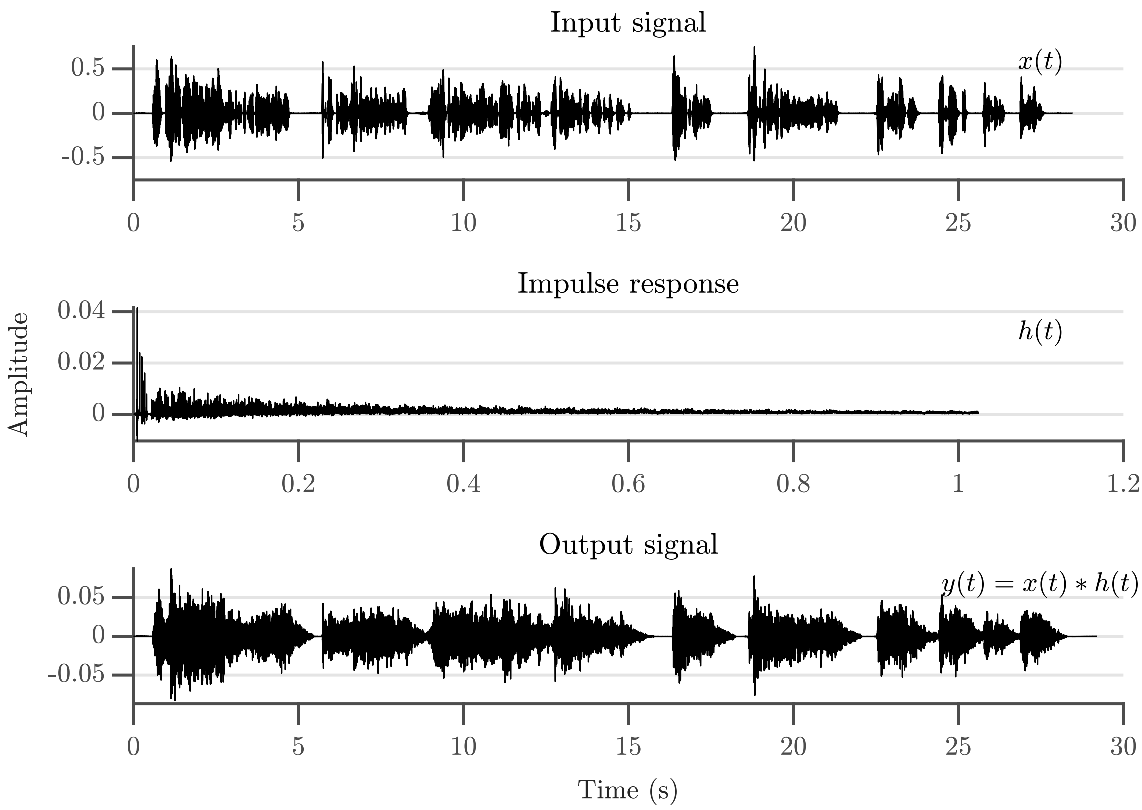


Figure 3.11: Example of resulting decimated input signal, generated impulse response, and output signal.

3.8 Pilot Testing

To fine-tune usability and validate the prototype application before the actual experiment, we tested the application on six students in the weeks before the experiment. The test aimed to understand the time necessary for a session and reveal unclear segments of the application and question formulations in the surveys.

3.8.1 Demographic

The pilot test participants have a gender composition of 50/50 females and males, with ages ranging from 22 to 24, as shown in Figure 3.12. They have little or no prior experience with XR systems and acoustic simulations and were selected to validate the usability. They are students at the Norwegian University of Science and Technology (NTNU) and study chemistry, electronics, or geology.

3.8.2 Procedure

The pilot test sessions were conducted in the ULTIMATE lab at NTNU Gløshaugen. The room was empty to replicate the actual test location. The pilot tests were executed over two weeks, and the application was enhanced after each session. The sessions started with

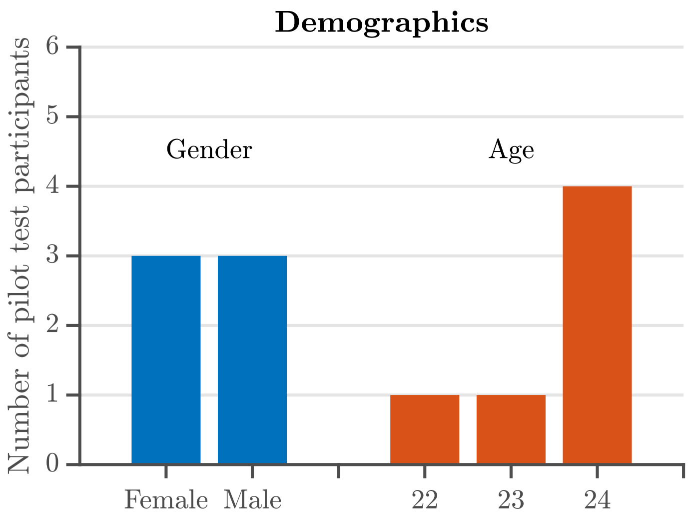


Figure 3.12: Age and gender composition of the pilot test participants.

the experimenter explaining the experiment, and the participants were then provided with an information sheet. When they were finished reading the sheet, they signed a consent form and answered the demographic survey. Afterward, they were provided with the tablet and headset, then the application and experiment started. The experimenter observed and took notes about the participants' execution and was available if they needed assistance.

3.8.3 Observations, Feedback and Changes

The experimenter observed the participants as they conducted the pilot tests and made changes on behalf of this. In this section, we provide some selected observations and feedback.

An important observation was that several of the participants rushed through the information sheet and was later confused while experimenting. To prevent this in the actual experiment, the experimenter sent the information sheet to the participants two days before the experiment and made sure everybody had read and understood the sheet. The information sheet consists of important information about the execution and a manual for the system and is attached in Appendix B, Section B.2, and B.3.

The experimenter also observed that several of the participants had difficulties with the application as it did not recognize the tablet's movements. This unresponsiveness happened because the participants got too close to a white wall without textures in the test location. The application uses the tablet's rear camera to register movements and requires differences in textures for SLAM to work. This was further prevented by ensuring that the participants did not get too close to walls by downscaling the virtual rooms.

It was observed, and feedback was given that some of the participants were confused about their current task during the experiment. An additional information pane was displayed on the top of the screen to cope with this, catching the participants' attention. Blinking neon green text was later introduced to make sure that it was eye-catching.

The experimenter had gained valuable information about usability and timing from the tests. After the piloting tests, there were made changes to the application layout, the research protocol, and some rephrasing of questions in the surveys. In the following section, we will describe the experiment setup of the test with the focus group.

3.9 Experiment Setup

The experiment session is set to the focus group's head office at Norsonic AS in Lier. Each participant has a 30-minute slot for conducting the experiment and answer the surveys. A 5-minute slack is set between each session and a 15-minute break halfway. The participants will first sign a consent form and be provided with a copy of the information they had received before the experiment. This setup is photographed as shown in Figure 3.13. Afterward, they will start the experiment with the tablet and headphones. The experimenter then waits in the background and takes notes, and is available for assistance.

Figure 3.13: Setup of information sheet, manual, and consent form.

The test location has a big empty space, such that the participants can walk around freely without hitting obstacles. The room is photographed as shown in Figure 3.14. In Figure 3.15 we can see two of the participants conducting the experiment.

Figure 3.14: Experiment location at Norsonic's office in Lier.

Figure 3.15: Pictures of two participants using the system.

With the provided method for experimenting, we will provide the results from the experiment in the following chapter.

Chapter 4

Results

This section will present relevant observations and results from the questionnaires and open answers for the participants in the experiment. Eight participants completed the experiment, with the demographic composition as shown in Figure 4.1. As a group, it was revealed by the background information survey that they have bachelor's or master's degrees within the technological field, have much experience within acoustics, and have minimal prior experience with XR systems.

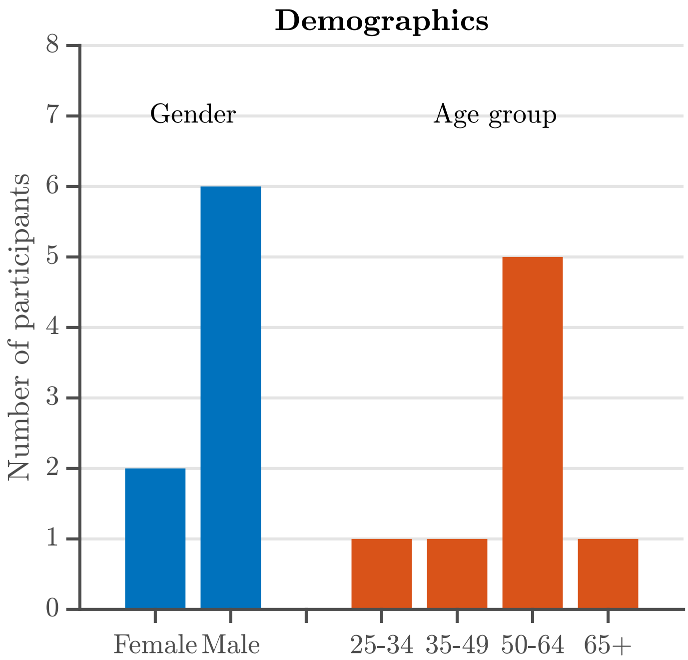


Figure 4.1: Age and gender composition of the participants.

Selected results for each section of the questionnaires in the survey are presented as bar charts. A selection of the open answers linked to the current section is presented as text. Observations and comments about the results are given after presenting the results for each scene. All the participant's answers to the questions can be found in Appendix A.

4.1 Scene 1: Simple

This scene is considered to be a simple "warm-up" for the session. Since most of the participants answered similarly in the questionnaire, we present the mean opinion scores in Figure 4.2 to the questions listed in Table 4.1.

Table 4.1: Questions for Scene 1

No.	Question / Statement
Q1	How realistic did the sound feel in the first room?
Q2	How realistic did the sound feel in the second room?
Q3	There was a clear difference in the sound in the two rooms.
Q4	To what extent did you experience a sense of "being there" inside the environment?

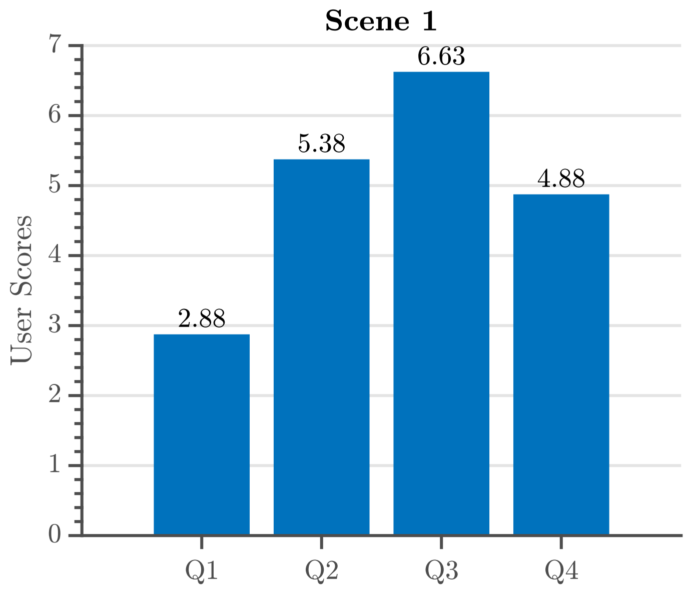


Figure 4.2: Mean opinion scores for the first scene's questions.

During the experiment, it was observed that the participants found it easy to interact with the application from the start. They moved around in the test location and moved closer and further away from the playing virtual sound source. In the first room, where the reverberation time is 1.70s, many participants felt that the acoustic representation was less realistic. This is reflected by the low mean opinion score for *Q1*, which resulted in 2.88. This result may come from the apparent distortion effect of long reverberation times and the participant's bias as they have acoustics as their profession. On the other hand, we can see from the mean opinion score for *Q2* of 5.38 that a much shorter reverberation time of 0.42s gives the participants a more realistic sound feeling.

Participant number two in the experiment evaluated her realistic feeling of the sound in the two rooms to be opposites, with a score of 1 and 7, respectively. This evaluation clearly shows that she found it easy to experience two very different acoustic rooms. However, participant seven found both rooms equally realistic and did not experience a big difference in the two rooms. The majority of the participants found a clear difference which is also reflected in the mean opinion score for *Q3* of 6.63.

The fourth participant in the experiment was observed to have difficulties understanding and operating the test system. The difficulties can be related to a lack of previous experiences with AR technology, which also reflects the participant's answer of *Q4*, being the only one to answer with a rate of 3.

An additional comment in the open text field was: *"Well designed and immersive visual experience. Acoustic simulation has demonstrated its potential and can indeed be made more immersive"*. It was observed that this participant was very excited by the experience and is also the youngest participant in the group. This excitement can suggest that he felt more immersed than the others and was more confident in the execution of the experiment.

4.2 Scene 2: Complex

This scene is more complex than the first scene, and the participants had different experiences with the system. We present answers to the questions listed in Table 4.2 in Figure 4.3 for three selected participants.

Table 4.2: Questions for Scene 2

No.	Question / Statement
Q1/Q6	There was a clear difference in the sound in the two rooms.
Q2/Q7	It was easy to determine when to place the marker in the first room.
Q3/Q8	It was easy to determine when to place the marker in the second room.
Q4	To what extent did you experience a sense of "being there" inside the environment?
Q5	How much did it seem as if you could reach out and touch the objects you saw?
Q9	It was easier to determine when to place the marker with AR functionality enabled.

It was observed during the experiment that there were different competence levels in the group of participants regarding AR experience. Some participants needed guidance from the experimenter to navigate in the virtual room and complete the tasks. The selected participants in Figure 4.3 show three different answers regarding how they perceived the AR environment. The first participant favors the scenario with AR enabled (blue bar) a bit more than with AR disabled (orange bar). This favoring is also reflected in *Q4*, *Q5*, and *Q9* of 6, 5, and 6, respectively.

Figure 4.4 shows the distances of the marker from the guest for the three participants. We can see that the first participant placed the marker further away from the guest when AR was enabled compared to when AR was disabled. This difference can interpret that he used his senses more actively in the AR environment. The distance is especially significant in

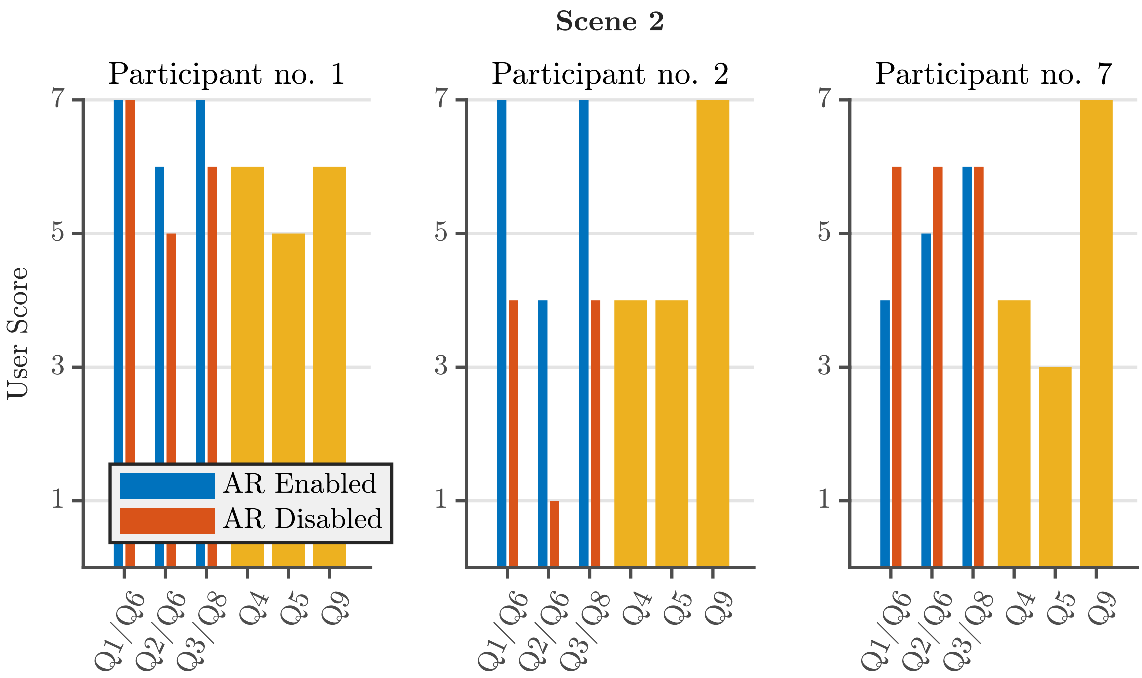


Figure 4.3: Selected answers for the second scene's questions.

Room 2, where there is a bigger difference. An additional comment from the participant is *"It was easier to remember to push the 'update' button in the AR disabled scenario, probably because I was busy being 'immersed' "*, which emphasizes the immersive effect of AR.

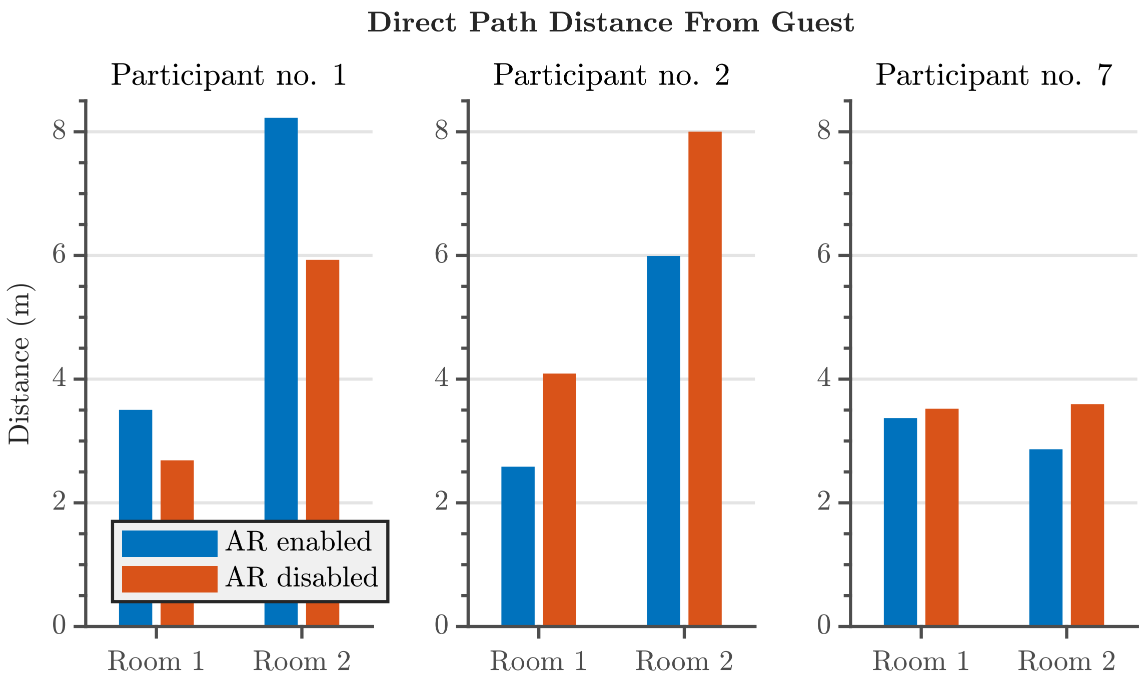


Figure 4.4: Distances to the guest in the four scenarios.

The second participants also favored the scenario with AR enabled. We can see from the gap between the blue- and orange bars in Figure 4.3 that this participant found it much more challenging to place the marker when AR was disabled, especially in the first room where the reverberation time is long (2.18s). In the second room, where the reverberation time is shorter (0.67s), she found it easier to place the marker. On behalf of the answers from the three first questions, it is expected that the participant's spatial presence in the environment would score higher. However, as an additional comment, she stated: *"Unfortunately, I focused more on the sound experience than the 'sense of being there'"*. We can see from Figure 4.4 that she placed the marker further away from the guest in both rooms when AR was disabled. This is again not expected based on her previous answers but can be explained by her additional comment.

The seventh participant, on the other hand, favored the scenario with AR disabled. We can see from Figure 4.3 that he did not hear a clear difference in the sound between the two rooms when AR was enabled and that he found it equally difficult to place the marker in both rooms when AR was disabled. This result is also reflected by the low scores in *Q4* and *Q5* of 4 and 3, respectively. The score for *Q9* of 7 stands out and can come of a more thoughtful decision than *Q2/Q7* and *Q3/Q8* because the participant answered this question last. Figure 4.4 shows that the participant had no significant differences between the two rooms or when AR was enabled and disabled. This result stands out from the rest of the participants but reflects his previous answers in the questionnaire of being neutral. However, the participant commented: *"Would have been nice to include several guests in the room"*, implying that he did not feel so immersed in the AR scene because it lacked a more realistic design.

4.3 Scene 3: Challenge

This scene is the last in the session and is more open-ended than the two previous scenes. We present the mean opinion scores to the questions listed in Table 4.3 in Figure 4.5.

Table 4.3: Questions for Scene 3

No.	Question / Statement
Q1	How much did it seem as if you could reach out and touch the objects you saw?
Q2	To what extent did you experience a sense of "being there" inside the environment you saw?
Q3	How confusing or clear was the experience?
Q4	How easy was it to listen to the changes in the room's acoustics?

During the experiment, it was observed that the participants had become more confident with interacting with the system in the last scene compared to the two previous scenes. This growing confidence was expected as the participants, in general, had little prior experience with AR technology. This is also reflected in the mean opinion scores for *Q1* and *Q2* of 4.50 and 5.38, respectively, as shown in Figure 4.5. This score is slightly better than the similar questions from Scene 1 and Scene 2, and the whole group, except the fifth participant, evaluated these questions equal or higher in the last scene. The fifth

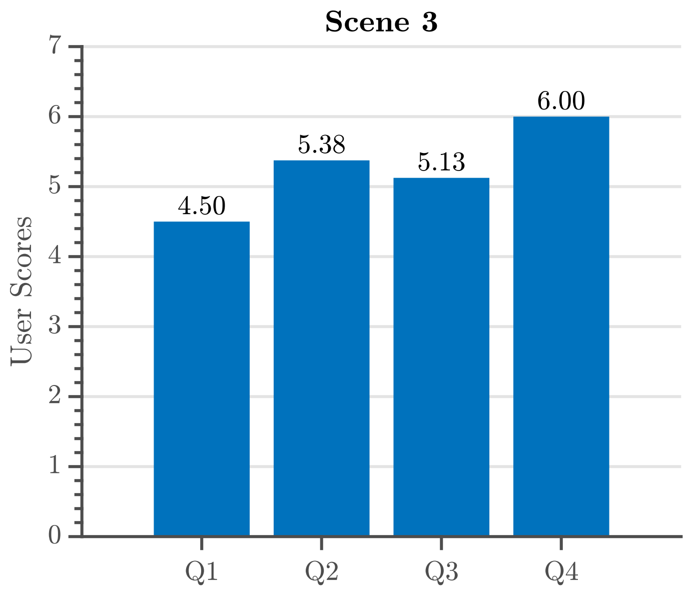


Figure 4.5: Average scores for third scene's questions.

participant evaluated *Q1* to 3, while the corresponding result for the previous scene was 5. This evaluation may suggest that he experienced the restaurant scene to be more realistic because of the additional details. However, his score for *Q2* of 5 indicates that the lack of realism did not influence his sense of being present in the scene.

The mean opinion score of *Q3* of 5.13 indicates that most of the participants found it easy to follow the instructions in the experiment. This is evident even though some of the participants needed guidance from the experimenter during the session. *Q4* has a mean opinion score of 6.00, which indicates that most of the participants found it easy to experience changes in the room's acoustics. On the other hand, the second participant was the only participant to rate this question to 4, indicating a neutral feeling in the changing acoustics. This rate may come because she was more focused on the sound experience and the realism of the sound rather than the immersive experience, as she stated in Scene 2. The other participants who felt more immersed also scored higher on *Q4*, indicating that the immersiveness may influence the sound perception.

The additional comments in the open text field were:

- *"Very convincing, AR works great. The current state of the app would not need loads of work to be a product."*
- *"Scene 1 was more confusing than Scene 3."*

These comments emphasize the participants growing confidence in the system as they used it.

4.4 Room Acoustic Feeling

In the overall evaluation, the answers from three selected participants to the questions about presence in the environment and impression listed in Table 4.4 are shown in Figure 4.6.

Table 4.4: Questions about the overall user experience.

No.	Question / Statement
Q1	It was easy to experience small differences in the room's acoustics.
Q2	How often did you want to try to touch something you saw/heard?
Q3	How completely were your senses engaged?
Q4	To what extent did you experience a sensation of reality?
Q5	Overall, how much did the things you saw/heard sound like they would if you had experienced the directly?
Q6	You perceived this possibility for obtaining information as ...
Q7	Overall impression of this technology.

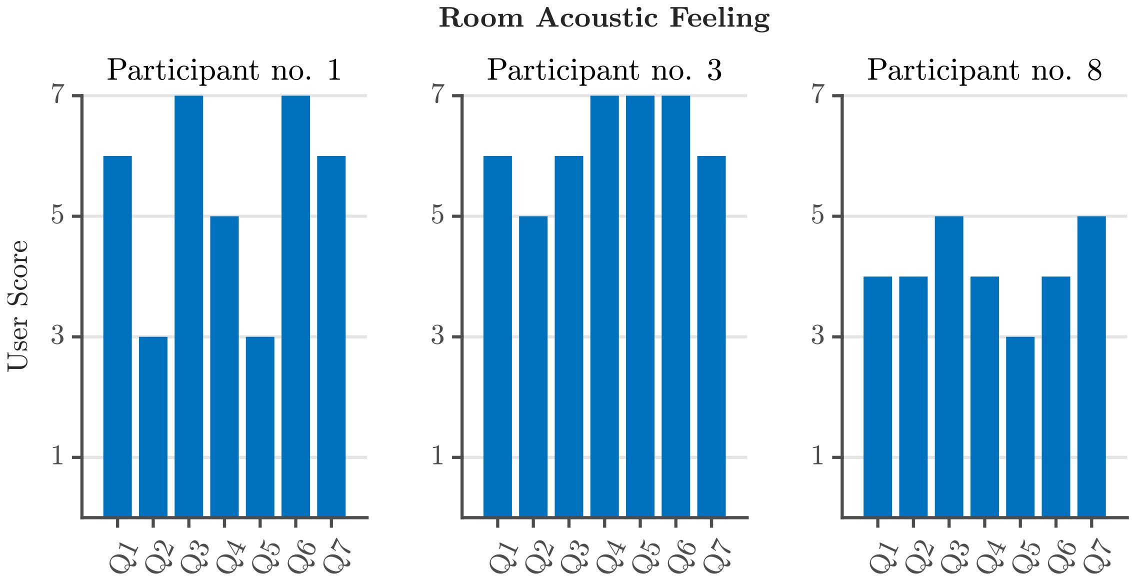


Figure 4.6: Selected answers for the participant's room acoustic feeling.

The selected participants in Figure 4.6 show three different answers regarding how they perceived the overall AR and acoustic experience. The first participant felt that his senses were very engaged, which is also reflected by his previous answers about spatial presence and his comment to why he likes the tool: *"(...) The immersive experience gives a whole new way of exploring how a room sounds"*. The low score for *Q2* of 3 can come from his low score of realism in sound reproduction, as we can interpret from the score for *Q5* of 3 in addition to his comment about his room acoustic feeling: *"Lacks a good share on realism, but indeed proves the concept (...)"*. Despite his low degree of realistic perception, he experienced small differences in acoustics, as we can see from the score for *Q1* of 6. In addition, he felt a sense of presence in the application, which we can interpret from his comment about spatial presence: *"(...) The visuals combined with the movement is very helpful. The acoustic response to distance also contributes significantly"*.

Participant number three scored higher on questions *Q2*, *Q4*, and *Q5* than the first participant. This difference can interpret that a high degree of spatial presence in the AR environment can remove the focus away from the degree of unrealism in the acoustic simulation. In addition, the participant commented: *"You can 'hear' the material"*, which is one of the main intentions with the simulation tool.

On the other hand, the last participant was more neutral in her response than the third participant. From her results in the questionnaire, we can see from Figure 4.6 that she felt immersed to some degree in the experiment but was not very convinced about how realistic the representation of the acoustics felt. Her profession as an acoustic advisor can bias this, making it difficult to evaluate how realistic the sound representation is in a virtual room. This is also reflected in her comment: *"The envelope is hard to simulate"* when she was asked about her spatial presence in the application. Despite her perception of lacking realism in acoustics, she commented that *"Simulation of acoustics is always good. Acoustics is so difficult to explain"* when she was asked why she liked the tool.

When we look at the mean opinion score for all the participants, as shown in Figure 4.7, we can see that most of the participants felt immersed, but they did not get a very realistic feeling from the acoustic simulation. However, the mean opinion score for *Q1* of 5.13 indicates that despite the unrealistic sound representation, they heard a clear difference in the acoustics when the reverberation in the room changed.

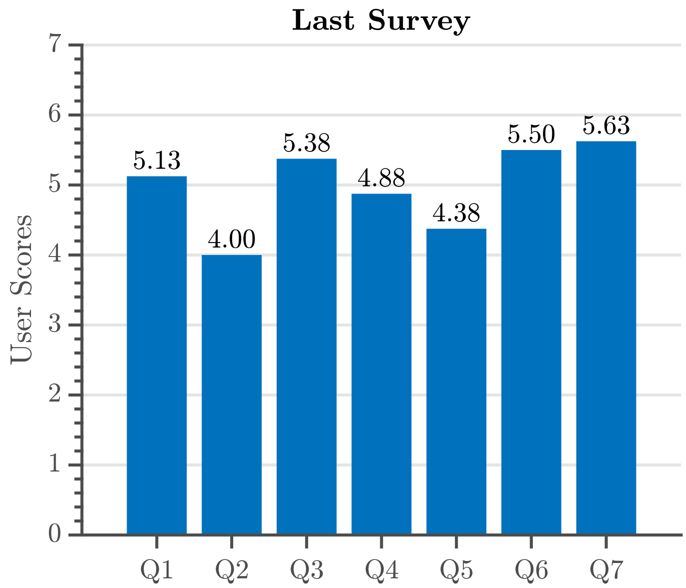


Figure 4.7: Mean opinion scores for the participant's room acoustic feeling.

The following chapter will further discuss the meaning of the results, system, and further work.

Chapter 5

Discussion

There are several focus areas within the system, and in this chapter, we will discuss the prototype in terms of the experiment, system, and further work.

5.1 Experiment Session

The focus group is inexperienced with XR systems and might have felt challenged by or unfamiliar with the technology. Inexperience can be an influencing factor for their immersive experience, making it more challenging to experience the feeling of presence [9], and was observed to be very clear at the beginning of the session. However, as the participants used the system, they became more confident and could interact more independently. This was intended with the composition of the three different scenes in the experiment, where they first encountered a simple scene, then a more complex, and lastly, a more open-ended scene with a challenge.

The analysis and comments about the participants' answers in Chapter 4 indicate that the AR experience affected how immersed they felt in the system. There was a consensus among the participants that they were able to distinguish different acoustical rooms or designs. The results suggest that a combination of immersion and sound representation helps the user to take a part of the virtual scene and feel present to some degree. Even though the system is not fully developed with a high-quality simulation of room acoustics, it is evident that a concept for AR simulation can help users to understand acoustical properties in different materials and get a sense of being present in the experience. This result implies that the proposed hypothesis H_1 is accepted.

The mean opinion score for *Q1* in Figure 4.7 of 5.13 emphasizes that most of the participants in the group found it easy to experience small differences in the changing acoustics. This result may be a combination of multimodal perception from the other senses, which we saw from the mean opinion score for *Q3* of 5.38. This combination indicates that their senses were actively engaged during the session. Even though some participants found it more challenging to interact with the system when AR was enabled in Scene 2, compared to when it was disabled, we can interpret that the participants who felt higher degrees of immersiveness also found it easier to experience small acoustical changes. Since the participants were unfamiliar with XR technology, they may have focused more on interacting with the system. However, it is still evident that they experienced differences in the room acoustics. This result indicates that the proposed hypothesis H_2 is accepted.

5.2 System

The system is designed to be a framework for experiencing room acoustics in an immersive way. Mobile AR applications are very portable and accessible for users and need minimal setup. One could argue that a VR application would fit better for this application as it immerses the user even more with a head-mounted display (HMD). This, on the other hand, requires more equipment and can hence make the system less accessible. As we want flexibility in our system, the obvious choice of technology landed on AR.

It is emphasized that the application is a prototype rather than a fully developed simulation product. The concept of virtual sound replication is, for that reason, more in focus than correct sound replication and design. The image-source method for computing RIRs has proved to be an easy implementation and provided room acoustic effects for the user. Even though the sound felt less realistic for longer reverberation times, it is clear that it distorted the dry sound. For shorter reverberation times, the replication was evaluated to be better. This evaluation can come from the restriction of maximum samples when generating the RIRs, which was necessary for limiting computation time.

Unity's "Burst Compiler" introduced multithreading to the application and reduced computation time significantly when more than one sound source was present. However, we had to implement our own adapted FFT and convolution algorithms to be supported by the compiler because of the restricted support of data types. These rewritings may have introduced suboptimal implementations, leading to longer computation time and less accurate calculations than other high-end implementations.

5.3 Further Work

The quality of the acoustic representation in this framework is not optimal. The simulation takes a naïve approach of presenting the user with an output sound for a fixed source and receiver pair. To have a real-time update of acoustics for all source and receiver pairs, we need the system to run faster or investigate other methods for implementation. A possible implementation can be to introduce a grid-based pre-calculated approach. For all points in the grid, each channel is convolved with an anechoic sound. The user will then always be followed by an array of virtual speakers, presenting the ambisonics from the nearest grid point. This idea takes inspiration from the work of Pind et al. [4].

The image-source method for generating an RIR does not take all acoustic phenomena, such as frequency dependence, diffraction, or interference, into account. This model works great to indicate the acoustics in the room but is not a very accurate approach. An implementation for a more robust RIR generator can be further investigated for increasing the system's quality.

Auditory environments and spatial sound are factors for creating an impression for the listener to be surrounded in a 3D space [13]. Spatial sound has proven to be a great tool for locating and navigating within AR environments [16]. Binaural and HRTF implementations can further improve the user's spatial presence and should be considered in future implementations.

Introducing machine learning for segmentation and classifying objects in a room detected by the device's camera can help create a 3D model of the current room. This model can later be used to apply different acoustic properties to the objects or change the room's design and experience the difference in acoustic. Automatic creation of the current room will facilitate a seamless and accurate representation of the room in AR and reduce the required setup of a scene. Kim et al. [6] used SegNet[1] in a similar approach for a VR application used for spatial audio reproduction.

In the following chapter we will conclude the work of this report.

[1] http://mi.eng.cam.ac.uk/projects/segnet/

Chapter 6

Conclusion

A framework of an acoustic simulation tool for experiencing real-time differences in room acoustics has been proposed in this project. The application emphasizes the concept of sound perception for the user and utilizes AR technology to combine multiple modalities for immersive effects. The system is evaluated and tested by a focus group of eight participants working within acoustics.

The research had a qualitative approach, and the data foundation contains answers from questionnaires, open answers, and observations. The focus group had minimal prior experience with XR technology, but it was evident that they became more confident as they used the system. The results suggested that the participants who felt higher degrees of immersiveness also found it easier to experience small changes in the acoustics, and the following hypotheses were accepted:

- H_1: Real-time acoustic room simulation in AR provides the client a sense of being present in the acoustic room.
- H_2: AR technology enhances the perception of small changes in sound.

The system has proved to work as a prototype for acoustic room simulations in AR. The immersive effect of AR combined with acoustic simulations provides an engaging arena for experiencing sound in different acoustic designs. However, the acoustic model needs further improvements to give the user of the framework a more realistic sound representation.

Bibliography

[1] Chengde Wu and Mark Clayton. BIM-Based Acoustic Simulation Framework. In *30th CIB W78 International Conference*, pages 99–108, 10 2013.

[2] Wai Ming To and Andy W. L. Chung. An innovative approach in data collection for restaurant soundscape study. *Proceedings of Meetings on Acoustics*, 36, 05 2019. doi: 10.1121/2.0001019.

[3] Alfonso Rodríguez-Molares and Manuel Seoane. Benchmarking for acoustic simulation software. *The Journal of the Acoustical Society of America*, 123:3601–3606, 05 2008. doi: 10.1121/1.2934429.

[4] Finnur Pind, Cheol-Ho Jeong, Hermes Sampedro Llopis, Kacper Kosikowski, and Jakob Strømann-Andersen. Acoustic Virtual Reality - Methods and challenges. 04 2018.

[5] Pontus Larsson, Aleksander Väljamäe, Daniel Västfjäll, Ana Tajadura-Jiménez, and Mendel Kleiner. *Auditory-Induced Presence in Mixed Reality Environments and Related Technology*, pages 143–163. Springer International Publishing, London, England, 10 2010. doi: 10.1007/978-1-84882-733-2_8.

[6] H. Kim, L. Remaggi, P. J. B. Jackson, and A. Hilton. Immersive Spatial Audio Reproduction for VR/AR Using Room Acoustic Modelling from 360° Images. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 120–126, 2019. doi: 10.1109/VR.2019.8798247.

[7] Arzu Çöltekin, Ian Lochhead, Marguerite Madden, Sidonie Christophe, Alexandre Devaux, Christopher Pettit, Oliver Lock, Shashwat Shukla, Lukáš Herman, Zdeněk Stachoň, Petr Kubíček, Dajana Snopková, Sergio Bernardes, and Nicholas Hedley. Extended Reality in Spatial Sciences: A Review of Research Challenges and Future Directions. *ISPRS International Journal of Geo-Information*, 9(7). doi: 10.3390/ijgi9070439.

[8] Torbjörn Svensson. Using VR to Simulate Interactable AR Storytelling. In Rogelio E. Cardona-Rivera, Anne Sullivan, and R. Michael Young, editors, *Interactive Storytelling*, pages 328–332, Cham, Switzerland, 2019. Springer International Publishing. doi: 10.1007/978-3-030-33894-7_33.

[9] Camille Sagnier, Emilie Loup-Escande, and Gérard Valléry. Effects of Gender and Prior Experience in Immersive User Experience with Virtual Reality. In Tareq Ahram and Christianne Falcão, editors, *Advances in Usability and User Experience*, pages 305–314, Cham, Switzerland, 2020. Springer International Publishing. doi: 10.1007/978-3-030-19135-1_30.

[10] C. Schissler, C. Loftin, and D. Manocha. Acoustic Classification and Optimization for Multi-Modal Rendering of Real-World Scenes. *IEEE Transactions on Visualization and Computer Graphics*, 24(3):1246–1259, 2018. doi: 10.1109/TVCG.2017.2666150.

[11] Sarvesh Agrawal, Adèle Simon, Søren Bech, Klaus Bærentsen, and Søren Forchhammer. Defining Immersion: Literature Review and Implications for Research on Audiovisual Experiences. *Journal of the Audio Engineering Society. Audio Engineering Society*, 68:404–417, 07 2020. doi: 10.17743/jaes.2020.0039.

[12] Asim Hameed and Andrew Perkis. Spatial Storytelling: Finding Interdisciplinary Immersion. In Rebecca Rouse, Hartmut Koenitz, and Mads Haahr, editors, *Interactive Storytelling*, pages 323–332, Cham, Switzerland, 2018. Springer International Publishing. doi: 10.1007/978-3-030-04028-4_35.

[13] Rolf Nordahl and Niels Christian Nilsson. *The Sound of Being There: Presence and Interactive Audio in Immersive Virtual Reality*, chapter 13. Oxford Handbooks. Oxford University Press, United Kingdom, May 2014. doi: 10.1093/oxfordhb/9780199797226.013.013.

[14] Mirza Beig, Bill Kapralos, Karen Collins, and Pejman Mirza-Babaei. An Introduction to Spatial Sound Rendering in Virtual Environments and Games. *The Computer Games Journal*, 8:199–214, 12 2019. doi: 10.1007/s40869-019-00086-0.

[15] Markus Zaunschirm, Christian Schörkhuber, and Robert Höldrich. Binaural rendering of Ambisonic signals by head-related impulse response time alignment and a diffuseness constraint. *The Journal of the Acoustical Society of America*, 143:3616–3627, 6 2018. doi: 10.1121/1.5040489.

[16] Dariusz Rumiński. An experimental study of spatial sound usefulness in searching and navigating through AR environments. *Virtual Reality*, 19:223–233, 10 2015. doi: 10.1007/s10055-015-0274-4.

[17] J. Broderick, J. Duggan, and S. Redfern. The Importance of Spatial Audio in Modern Games and Virtual Environments. In *2018 IEEE Games, Entertainment, Media Conference (GEM)*, pages 1–9, 08 2018. doi: 10.1109/GEM.2018.8516445.

[18] Rosonance audio. URL `https://resonance-audio.github.io/resonance-audio/`. Accessed: 13.05.2021.

[19] Andres Perez-Lopez and Julien De Muynke. Ambisonics Directional Room Impulse Response as a new Convention of the Spatially Oriented Format for Acoustics. Zenodo, 05 2018. doi: 10.5281/zenodo.1299893.

[20] Emanuël A.P. Habets. Room Impulse Response Generator*. Technical report, Technische Universiteit Eindhoven, 09 2010.

[21] Adil Alpkocak and Kemal Sis. Computing Impulse Response of Room Acoustics Using the Ray-Tracing Method in Time Domain. *Archives of Acoustics*, 35:505–519, 12 2010. doi: 10.2478/v10168-010-0039-8.

[22] Jont B. Allen and David A. Berkley. Image method for efficiently simulating small-room acoustics. *The Journal of the Acoustical Society of America*, 65:943–950, 5 1979. doi: 10.1121/1.382599.

[23] D.G.K. Dissanayake, D.U. Weerasinghe, L.M. Thebuwanage, and U.A.A.N. Bandara. An environmentally friendly sound insulation material from post-industrial textile waste and natural rubber. *Journal of Building Engineering*, 33:101606, 2021. doi: 10.1016/j.jobe.2020.101606.

[24] Maple Integration. Sound Absorption Data for Common Building Materials and Furnishings, 2012. URL `http://mapleintegration.com/sound_ab.php`. Accessed: 11.05.2021.

[25] Samuel Siltanen, Tapio Lokki, and Lauri Savioja. Rays or Waves? Understanding the Strengths and Weaknesses of Computational Room Acoustics Modeling Techniques. In *the International Symposium on Room Acoustics (ISRA2010)*, 01 2010.

[26] Matthew Lombard, Theresa B. Ditton, and Lisa Weinstein. Measuring presence: The Temple Presence Inventory (TPI), 9 2013. URL `http://matthewlombard.com/research/p2_ab.html`. Accessed: 19.03.2021.

[27] Samsung Norway: Galaxy Tab S6 Lite, 06 2020. URL `https://www.samsung.com/no/tablets/galaxy-tab-s/galaxy-tab-s6-lite-10-4-inch-gray-64gb-wi-fi-sm-p610nzaanee/`. Accessed: 24.04.2021.

[28] Google ARCore: Fundamental concepts, 07 2020. URL `https://developers.google.com/ar/discover/concepts`. Accessed: 24.04.2021.

[29] Kristin Valla. Ut av det blå, Jul 2020. URL `https://www.storytel.com/no/nn/books/1476733-Ut-av-det-bl%C3%A5`. Kagge Forlag, Audiobook, Accessed: 19.03.2021.

[30] Odeon and Dipartimento di Ingegneria Università di Ferrara. Anechoic recordings. URL `https://odeon.dk/download/Anechoic/Cocktail.zip`. Accessed: 19.03.2021.

[31] Benjamin Bernschütz. *Anechoic Recordings*. Cologne University of Applied Sciences, Institute of Communication Systems, Cologne, Germany, 01 2013. URL `http://audiogroup.web.th-koeln.de/FILES/anechoicRecordings.pdf`. Accessed: 19.03.2021.

[32] Adobe Systems Incorporated: Mixamo. URL `https://www.mixamo.com/`. Accessed: 19.03.2021.

Appendix A

Survey Answers

A.1 Demographic and Background Information

Table A.1: Participant background information.

Participant no.	Gender	Age group	Education	Primary working tasks	Years within acoustics	Use AR
1	Male	25-34	Master's in electronics	Developer	1-5	Once a year
2	Female	50-64	Bachelor's in data	Developer/ Leader	20+	Once a year
3	Male	50-64	Data engineering	Developer	5-10	Never
4	Male	65+	Master's in cybernetics	Application/ Sales/ Test	20+	Once a year
5	Male	50-64	Electronics/ cybernetics engineering	Marketing	20+	Never
6	Male	50-64	Master's	Developer	5-10	Once a month
7	Male	35-49	Master's in computer application	Developer	5-10	Never
8	Female	50-64	Master's	Advisor	20+	Once a year

A.2 Scene 1 Participant Answers

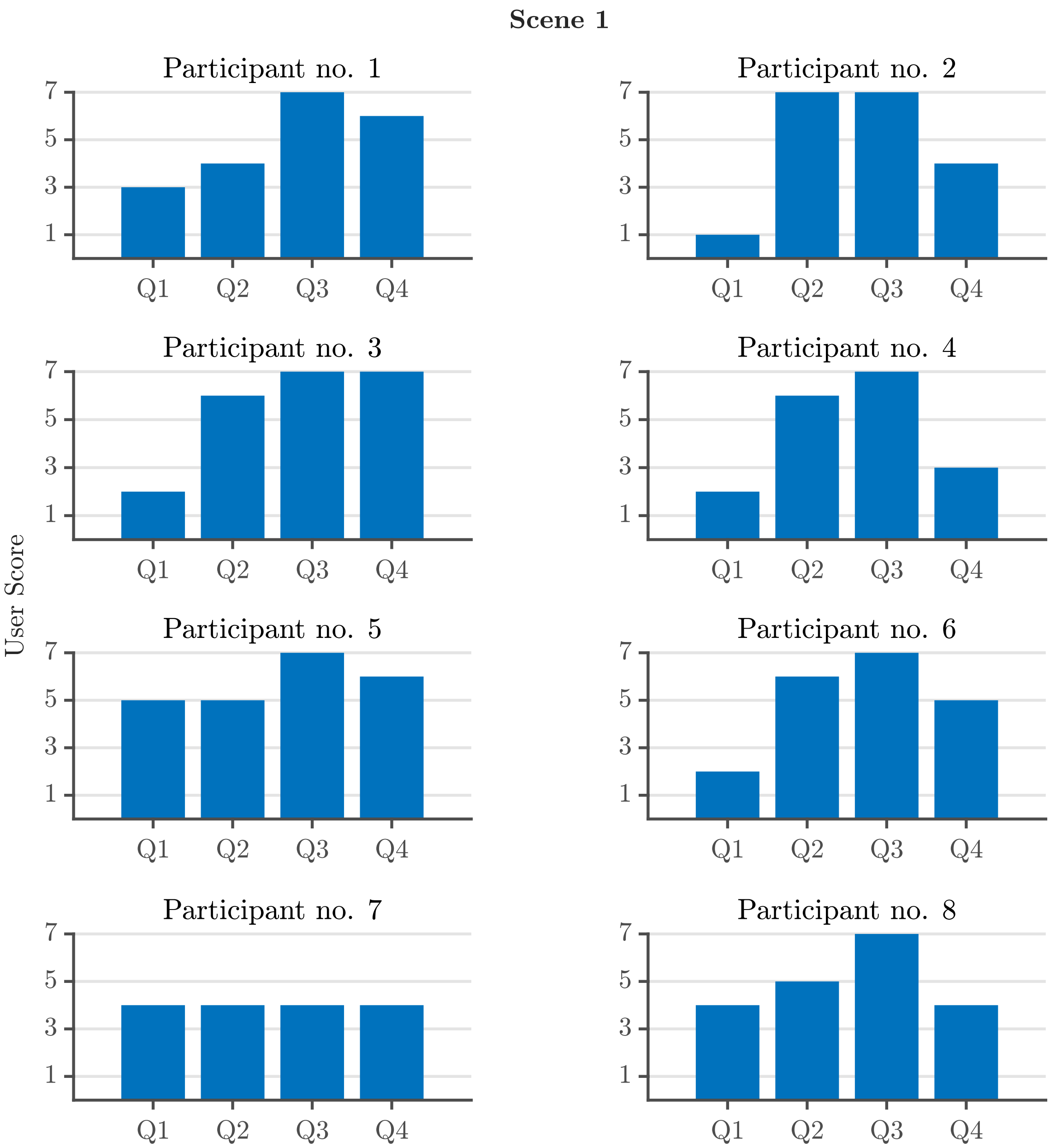


Figure A.1: Participant answers to Scene 1

Table A.2: Additional comments to Scene 1.

Participant no.	Comment
1	Well designed and immersive visual experience. Acoustic simulation has demonstrated its potential and can indeed be made more immersive.

A.3 Scene 2 Participant Answers

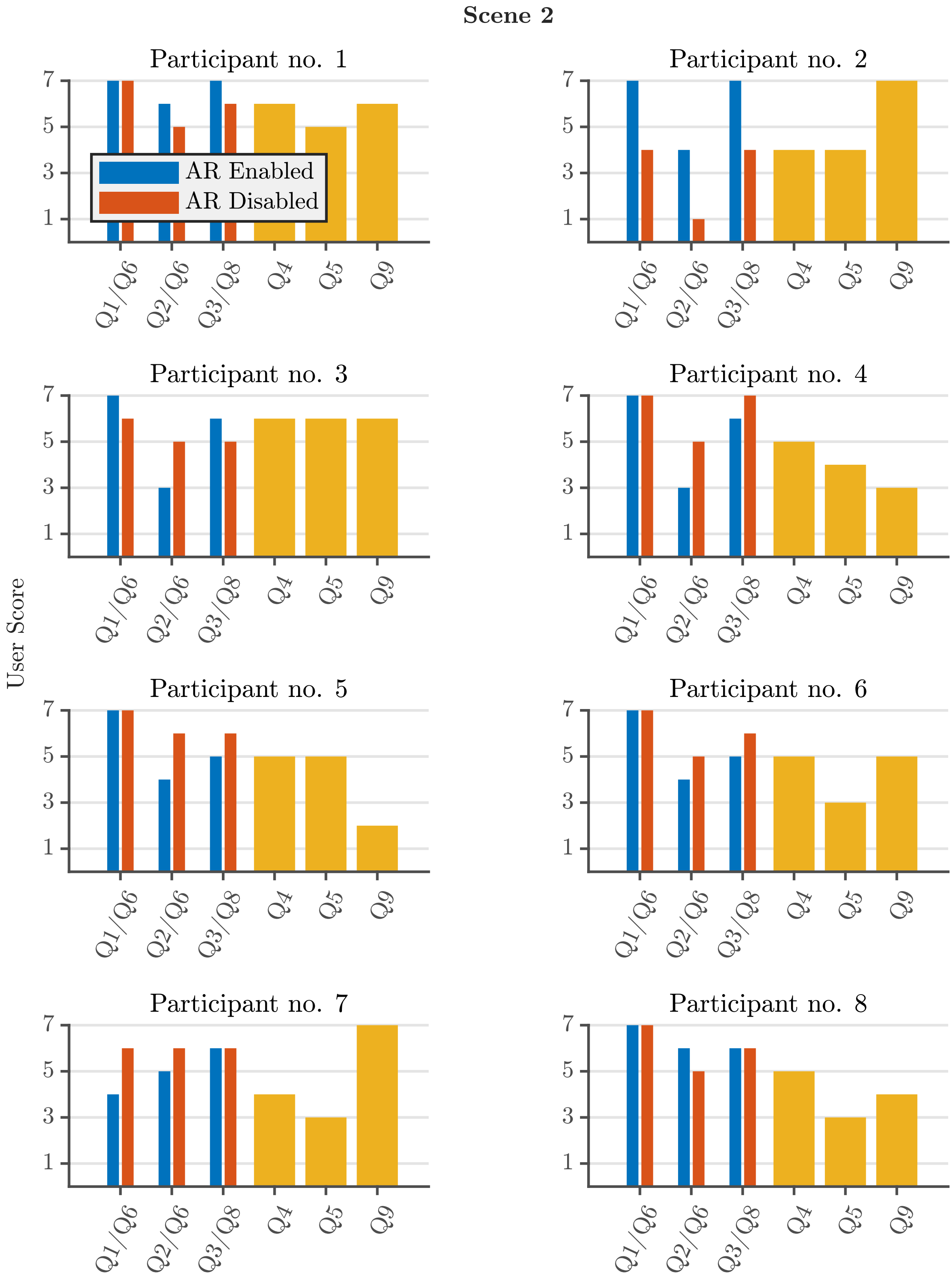


Figure A.2: Participant answers to Scene 2

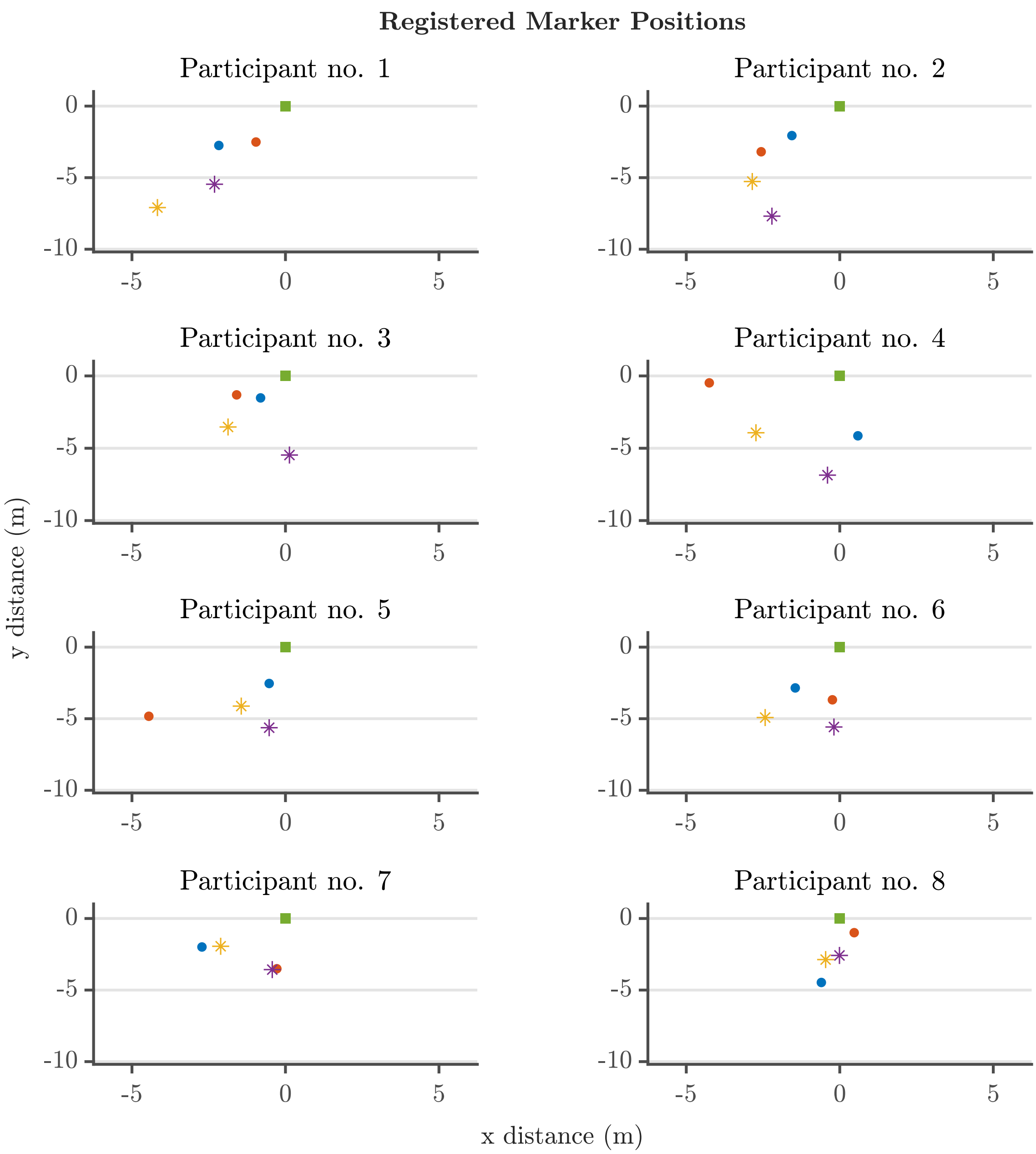


Figure A.3: Registered positions to Scene 2. Blue: Room 1 with AR enabled. Red: Room 1 with AR disabled. Yellow: Room 2 with AR enabled. Purple: Room 2 with AR disabled. Green: Guest.

Table A.3: Additional comments to Scene 2.

Participant no.	Comment
1	It was easier to remember to push the “update” button in the AR disabled scenario, probably because one is busy being ‘immersed’.
7	Would have been nice to include several guests in the room.

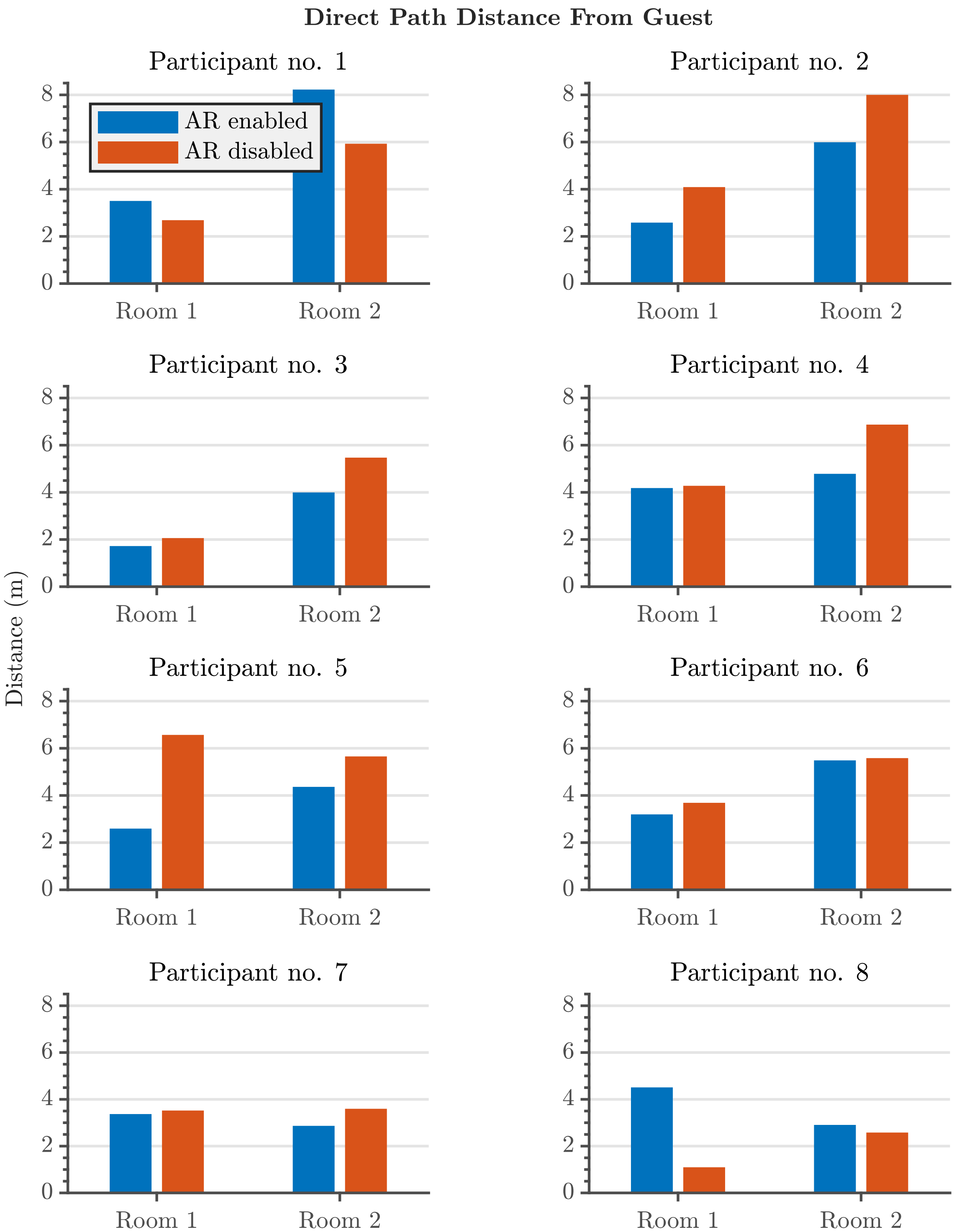


Figure A.4: Calculated distances in Scene 2

A.4 Scene 3 Participant Answers

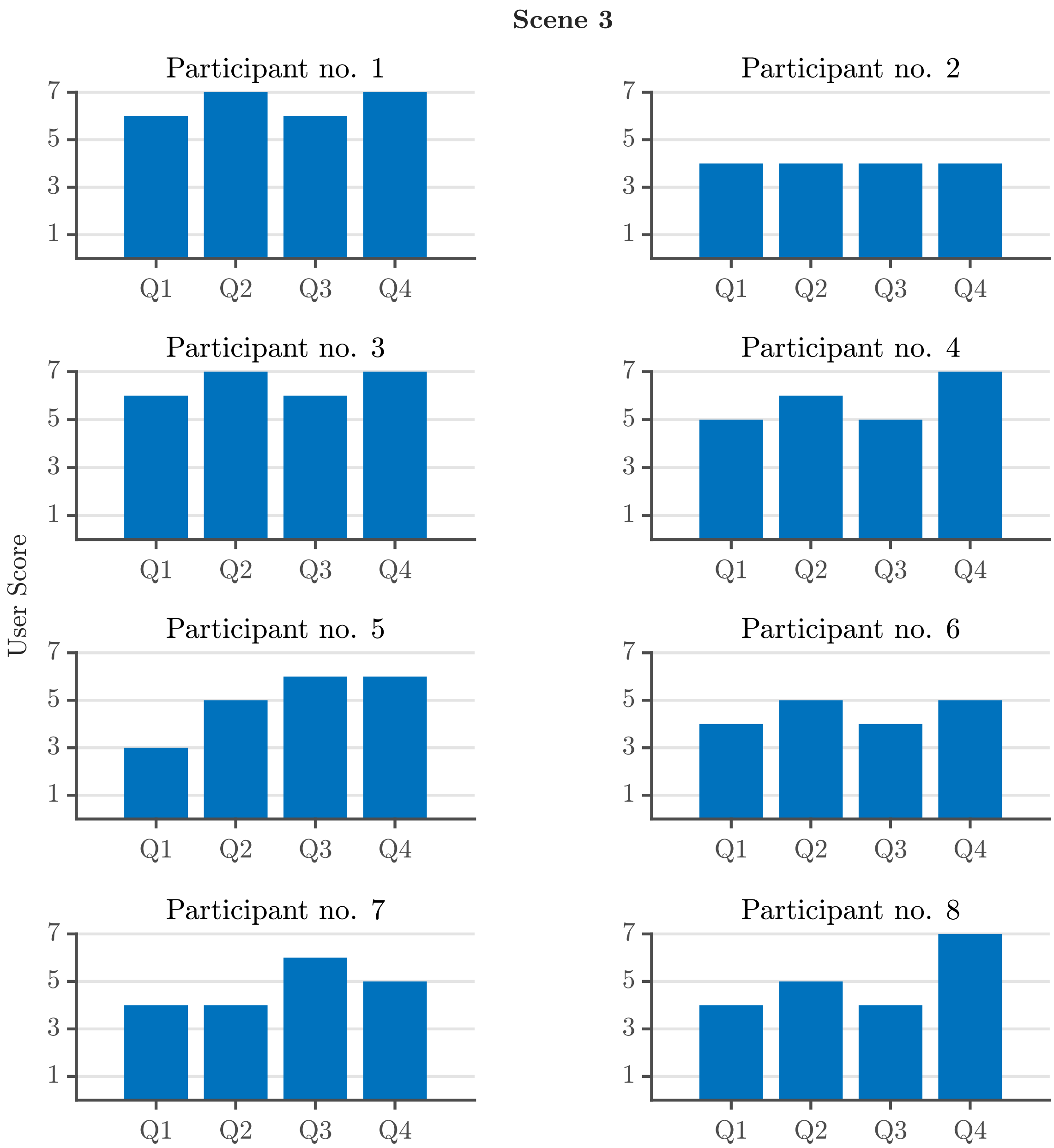


Figure A.5: Participant answers to Scene 3

Table A.4: Additional comments to Scene 3.

Participant no.	Comment
1	Very convincing, AR works great. The current state of the app would not need loads of work to be a product.
2	Unfortunately I focused more on the sound experience than the “sense of being there”.
4	Scene 1 more confusing than Scene 3
5	I have probably misinterpreted the task with AR enabled. (Translated by the author from Norwegian)

A.5 Room Acoustic Feeling

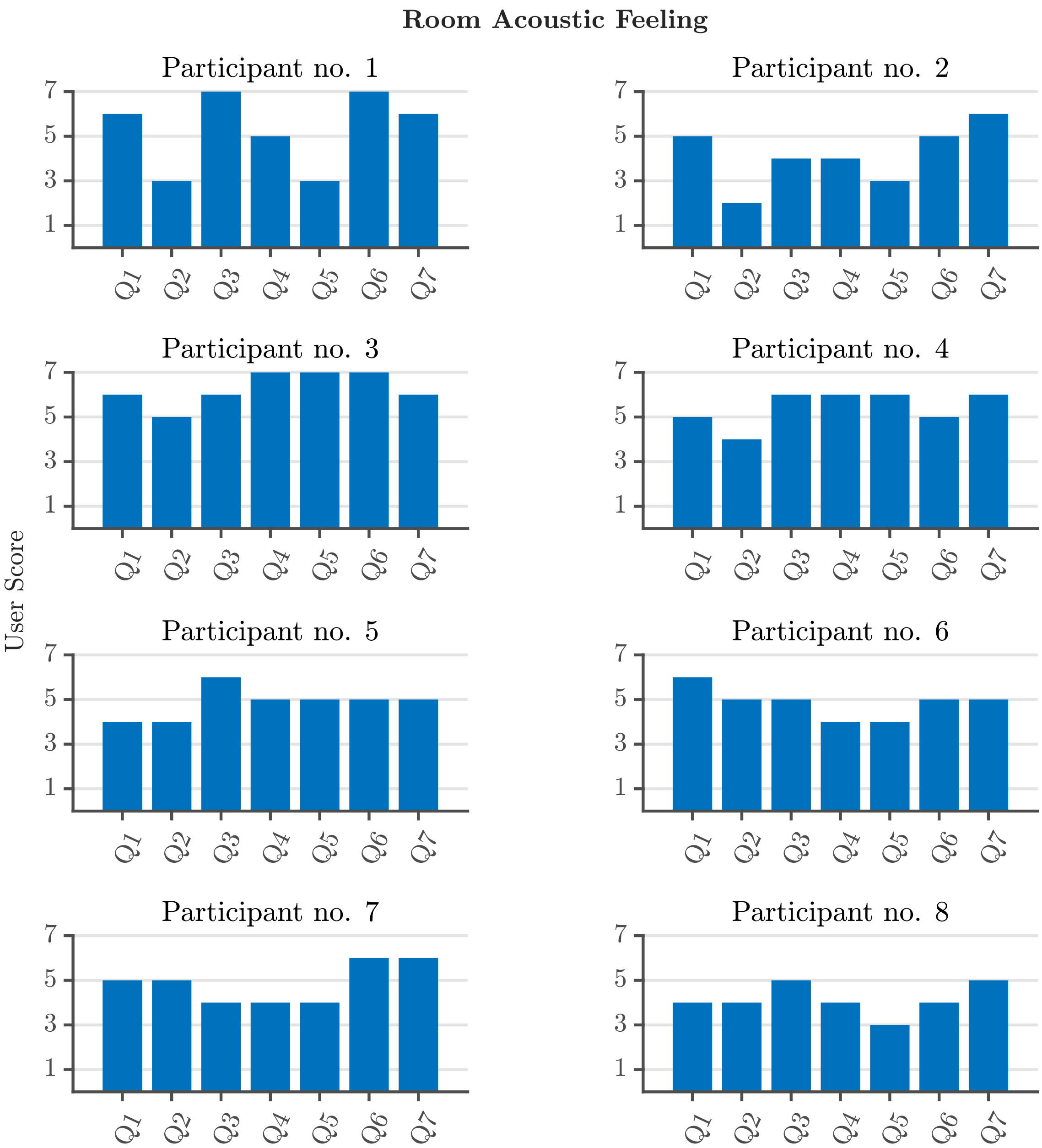


Figure A.6: Participant answers to questionnaire about their room acoustic feeling

Table A.5: Questions and open text answers about the overall experience.

Q8: How would you describe your room acoustic feeling using the simulation tool?

Participant no.	Answer
1	Lacks a good share on realism, but indeed proves the concept. "Binaural" processing would be a massive upgrade.
2	It was exciting to hear/feel the differences.
3	Engaging.
4	Engaging.
5	OK. I feel to some extent to be in the room.
6	Fun, realistic, engaging.
7	Exciting.
8	Engaging.

Q9: Did you get a feeling of being present in this acoustic simulation? Why/why not.

Participant no.	Answer
1	Yes. The visuals combined with movement is very helpful. The acoustic response to distance also contributes significantly.
2	As mentioned before, I didn't focus on the sense of being there. Maybe I should have known that before the test, that this is an important part of the tool.
3	You could "hear" the material.
4	Yes.
5	Yes, I did feel to be present - not perfect, but OK.
6	Yes, to some extent. Graphics could be better.
7	Yes, felt like sitting in a restaurant with AR.
8	The envelope is hard to simulate.

Q10: Please explain why you like/dislike this tool.

Participant no.	Answer
1	I believe this could be a very powerful tool in the hands of acousticians demonstrating upgrades. The immersive experience gives a whole new way of exploring how a room sounds.
2	I think it is a cool tool, and it may have a potential for being useful with improved acoustic simulation.
3	Really help to understand.
4	Positive: real feeling. Negative: hard to understand how to operate.
5	I like it. It is a good tool for simulation/try out different materials etc. Maybe a good tool for convincing fancy architects that concrete and glass is not good for restaurants.
6	I liked it. Easy to use, could hear changes I made.
7	Good for simulation.
8	Simulation of acoustics is always good. Acoustics is so difficult to explain.

Appendix B

Documents

Description

The accompanying documents are relevant to this project. They are presented in the following order.

1. Research Protocol
2. Information Sheet
3. Application Manual
4. Consent Form
5. Demographic Questionnaire and Survey

B.1 Research Protocol

Research Protocol

Master's Thesis Spring 2021

Karl Henrik Ejdfors

Synopsis

Mobile augmented reality (AR) is an evolving technology that can bring new technological opportunities for real-time simulations. Complex systems can be made intuitive and more available by combining the real environment with desired augmented information.

This study aims to test whether a prototype AR application can become an immersive simulation tool for room acoustics. The user will interact with the artificial environment by walking around in it and changing the object's absorption coefficients, placing objects in the environment, and auralize the change. The prototype will emphasize real-time feedback and interactivity rather than being a correct simulation tool.

The prototype's ability will be tested on a group of participants. Every participant will fill out survey forms for feedback. These participants will be professional workers within acoustics and will take an acting role as a fictitious client.

Introduction / Background

Technological advances have made smart devices common for everyone. Modern devices can utilize augmented reality (AR) content to enhance the real environment. AR is often mentioned for entertainment purposes and game settings but can also be utilized to superimpose relevant information and provide the user with new insight. This technology facilitates immersiveness and interactive experiences, which can be highly relevant for making complex systems more accessible. Such a complex system can be acoustic room simulations. They often require specialized equipment, setup, post-processing and can be very time-consuming operations.

This study will investigate how AR can be utilized as a simple real-time simulation tool for room acoustics. A prototype application will illustrate the concept of how such a simulation can be accomplished and lays the fundamental ground for further development. The prototype users will sense how the virtual acoustics in a room are adapted to interactive changes in the object's absorption coefficients. The simulation will give the user an indication of how and why acoustic materials are essential.

Hypotheses

This study will investigate the hypotheses:

1. “Real-time acoustic room simulation in AR provides the client a sense of being present in the acoustic room.”
2. “AR technology enhances perception of small changes in sound.”

Methodology and Design

The ULTIMATE lab at NTNU Gløshaugen will be used to prepare the prototype application before testing on participants. We will conduct tests on six fellow students to pilot and evaluate the system before testing on the professional participants.

Eight participants will conduct the tests, and they will answer surveys between each scene. The scenes are described below. Some of the questions in the surveys are weighted from 1-7 on a Likert scale. Some questions require measurements made in the application, and some questions require open answers. This composition of questions will provide feedback with both measurable data and perspective for the answers, which can be desirable for studies with a small number of participants.

If we only used measurable questions, we would get a low degree of insight because of the small group. The results would be objective but not representative for a bigger group. On the other hand, if we only based our study on observations and open answers, it would be challenging to have measurable data and evaluate our hypotheses. Our research method will contain sources of error regarding lack of universality as our participants may not be a representative selection. They will also take the role of a fictitious client, which can be confusing as they take an acting role. The results will still indicate how a bigger group will perceive our tool because of the participant’s profession.

The mean opinion score of the Likert scale weighted questions will be one part of what extent we can accept or reject our hypotheses. Another part will be an analysis of the open answers. Our second hypothesis will also be evaluated based on the differences in distance between AR enabled and disabled for Scene 2.

For evaluation of the framework, the application will be individually tested on professional workers within acoustics. They will take the role of a fictitious client interested in changing or trying out different acoustic elements and will use our framework for this purpose.

The participants are employees at Norsonic AS, and the test location will be set to a room in the company’s building in Lier. The participants will be provided with the development tablet and an over-ear headset. First, they will complete a simple predefined scene, then they will complete a more complex predefined scene, and lastly, they will complete a challenge of reducing the reverberation time in a more open scene.

Scene 1: Simple

Aim: Listen to changes in room acoustics by changing building materials.

Initial properties:

Table B.1: Initial properties.

Attribute	Material
Walls	Concrete
Floor	Marble
Ceiling	Wood Panel

Room dimensions (L $\times$ W $\times$ H (m)): $8 \times 10 \times 3$

Reverberation time: 1.70 seconds

Changed properties:

Table B.2: Changed properties.

Attribute	Material
Walls	Wood Panel
Floor	Plywood
Ceiling	Plaster

Room dimensions (L $\times$ W $\times$ H (m)): $8 \times 10 \times 3$

Reverberation time: 0.42 seconds

Challenge: The test participant will listen to a recording of a woman talking with applied room effects from the initial properties. Later, he will change materials in the room and experience the changes in room acoustics.

Scene 2: Complex

Aim: Experience how close you must be to have an undisturbed conversation at a restaurant with and without sound-absorbing materials and evaluate the impact AR has on small changes in sound perception.

Scenario: You and a guest of yours are at a restaurant. There are two noisy conversations in the same room as well as a live band playing guitar. The restaurant is old and has a rustic appearance. The main materials in the room have hard surfaces, which does not favor short reverberation times. You and your guest are trying to have an undisturbed conversation.

Initial properties:

Table B.3: Initial properties.

Attribute	Material
Walls	Brick
Floor	Plywood
Ceiling	Wood Panel
Lightly furnished, evenly spaced	

Room dimensions (L × W × H (m)): 16 × 12 × 4

Reverberation time: 2.18 seconds

Changed properties:

Table B.4: Changed properties.

Attribute	Material
Walls	Brick and acoustic elements
Floor	Plywood and carpets
Ceiling	Acoustic tiles
Lightly furnished, evenly spaced	

Room dimensions (L × W × H (m)): 16 × 12 × 4

Reverberation time: 0.67 seconds

Challenge: With initial properties, the participant will evaluate how close he must be to have an undisturbed conversation. This is accomplished by getting closer to the guest and press a button in the application when he evaluates himself as close enough. He will then experience the same scenario with changed properties. Later, he will do the same experiments with AR functionality turned off and a simple 2D graphical user interface to interact with.

Scene 3: Challenge

Aim: Evaluate the interactive framework as a simulation tool for room acoustics.

Description: The participant will interact with the simulation tool to evaluate its potential and usefulness in real-life settings. In this scene, the participant will be able to apply different acoustic materials to the room as well as sound sources, obstacles, and objects.

Challenge: The participant will be presented with the initial room as described below. He will first add an audio source to the room. His next task is to change materials of the front- and back wall to bricks (reverberation time will be 1.22 seconds). Afterward, he will add a sound-absorbing object(s) to the room until the room's reverberation time is reduces by 2/3 or more (i.e., reverberation time must be below 0.81 seconds).

Initial properties:

Table B.5: Initial properties.

Attribute	Material
Walls	Wood Panel
Floor	Marble
Ceiling	Plaster

Room dimensions (L $\times$ W $\times$ H (m)): $8 \times 8 \times 4$

Reverberation time: 0.52 seconds

Covid-19 Backup Plan

If Covid-19 regulations require that the above plan can not be successfully accomplished, one of the plans will be used for evaluating the framework.

1. The parking space outside of Norsonic AS will be used as a test location.
2. The ULTIMATE lab at NTNU Gløshaugen or a room in the experimenter's home (if necessary) will be used as a test location. The experimenter will test the prototype in this room, and the whole session will be audio-recorded and filmed. There will be composed a video which will be distributed to the contemplated test group together with the surveys. The participants will answer the survey after they have watched each scene in the experiment.

Resources

For safe execution of the tests, we will follow local Covid-19 regulations for infection control.

Table B.6: Initial properties.

Equipment	Description
Test device	Samsung Galaxy Tab S6 Lite
Over-eat headset	Sony WH-1000XM4
Disinfection wipes	Any
Test location	(L $\times$ W (m)) 6×4 free space
Analyzation tool	MATLAB

Results, Analysis and Discussion

The results will be generated by answered surveys from the participants and analyzed by comparing the answers and evaluating their open answers. Observations made by the experimenter will also be included. Regarding hypothesis two, we will analyze the participant's position relative to their "talking guest" and analyze differences between AR enabled and disabled.

The hypotheses will be evaluated by the measurable questions in the surveys together with analysis and comparison of the open answers. Hypothesis one is challenging to measure with questionnaires alone and requires additional information from open answers. Hypothesis two will be evaluated based on differences in answers for AR functionality turned on and off for Scene 2 as well as measurable questions regarding sound perception.

Priority and Timetable

- 15.03: Start gathering piloting participants.
- 17.03: Finished gathering piloting participants.
- 18.03: Start piloting on students.
- 24.03: Start gathering test participants.
- 25.03: Finished piloting on the students.
- 05.04: Finished gathering the test participants.
- 07.04: Start the testing.
- 09.04: Finished with the testing.
- 03.05: Finished analyzing the results.
- 11.06: Deadline for delivering the thesis.

Covid-19 Backup Plan

If Covid-19 regulations require that the above testing plan will not be successfully accomplished, we can postpone scheduled testing until 28.05.

B.2 Information Sheet

Experience Acoustic Room Simulation in AR

Dear participant,

Thank you very much for your participation in this experiment. This study will last approx. 30 minutes. At the end of this document, you will find a short manual for the application. **Please make yourself familiar with the attached manual before starting the experiment.**

During this experiment, you will take the role of a client who wants to experience differences in room acoustics. Such typical clients can be restaurant owners, interior architects, or construction clients. Please remember your role when you perform this experiment.

During this experiment, you will be working in an augmented reality (AR) environment on an Android tablet. The experimenter will provide you with a set of headphones and a tablet to perform different tasks in the application and experience changes in a room's acoustics. The purpose of this experiment is to evaluate the effect a real-time simulation tool in AR has on a client's feeling of being a part of the acoustic room.

The experiment is divided into four main parts:

1. After signing the consent form, you will fill in a demographic questionnaire that captures statistical data. Afterward, the experimenter will provide you with the tablet and headphones.

2. During the first scene in the experiment, you will experience changes in room acoustics between a "hard" room and a "soft" room. You will follow simple tasks that are shown at the top pane of the screen. When completed, you will be asked to fill out a survey.

3. During the second scene of the experiment, you will experience the effect sound-absorbing materials have in a room. First, you will be placed inside an AR restaurant with several sound sources. One of the sound sources is a guest of yours. Your mission is to find out how close you need to be your guest to have an undisturbed conversation. You will experience this scenario four times in total; two times with AR functionality enabled and two times by interacting with an on-screen joystick. The display will provide all the tasks at the top pane of the screen, and you will be asked to answer surveys.

4. During the third scene of the experiment, you will enter a less restricted environment. Here, you will be able to change the materials and add objects. Again, all the tasks will be provided to you at the top pane of the screen. After you have finished the scene, you will answer the last surveys.

Please note, **you are not getting tested**, but **you are testing the system**!

All the data that you provide and we are recording during this experiment will be pseudonymized.

During the experiment, you always have the chance to leave the study without the need to provide any reason. In case you have questions during the experiment at any point, please feel free to ask the experimenter.

And now: Have fun during the experiment!

Experimenter: Karl Henrik Ejdfors, +47 917 15 510

B.3 Manual

Manual for Acoustic Room Simulation Application

Please keep in mind that this is a prototype application and is not very responsive to sudden movements. The application uses the rear camera to navigate and record movements. Please do not cover the camera and be delicate with the tablet. Please do not shake it too much while experimenting.

Current task: your current task will always be shown at the top pane of the screen.

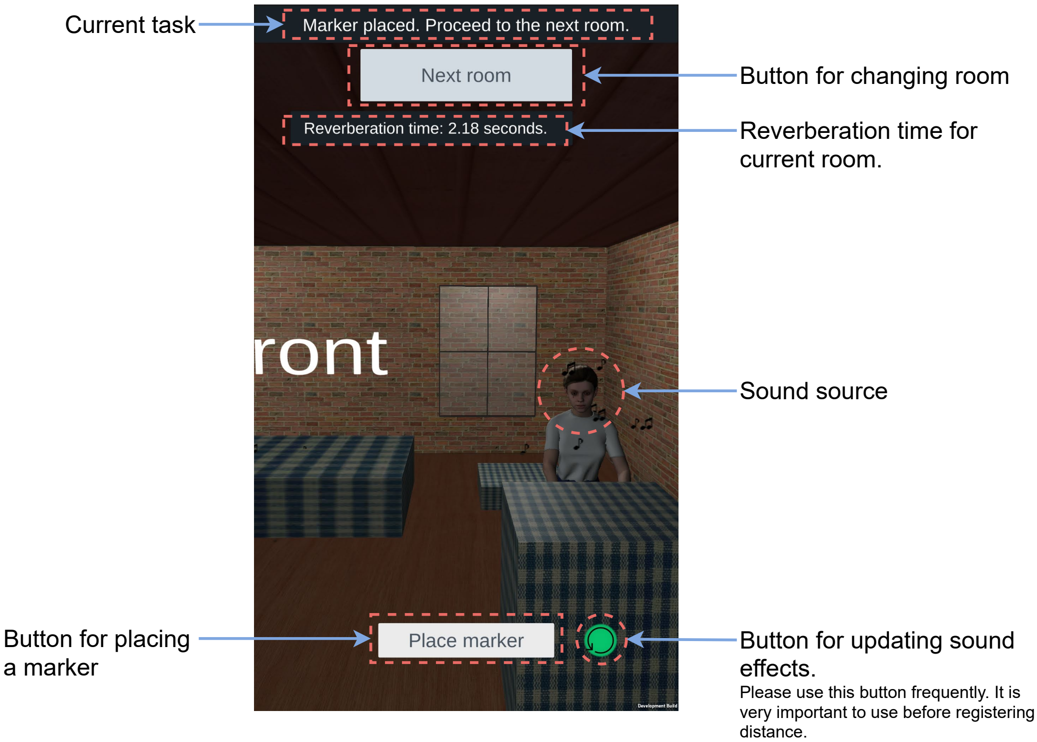
Next room: a button will be displayed if you can proceed to the next room.

Reverberation time: you will always be able to see the total reverberation time in the current room.

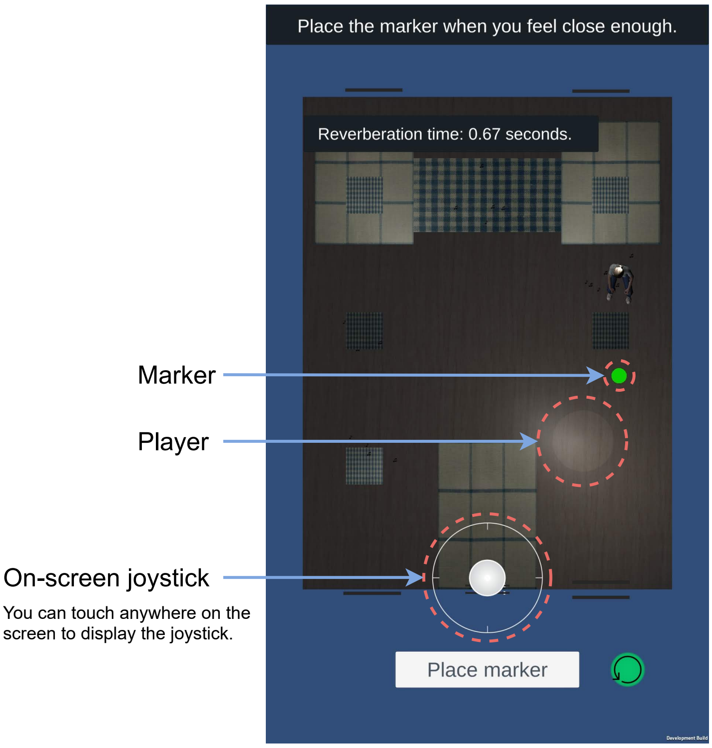
Sound source: every sound source will emit music notes.

Place marker: a button will be displayed if you can place the marker. You can press this button as many times as you like.

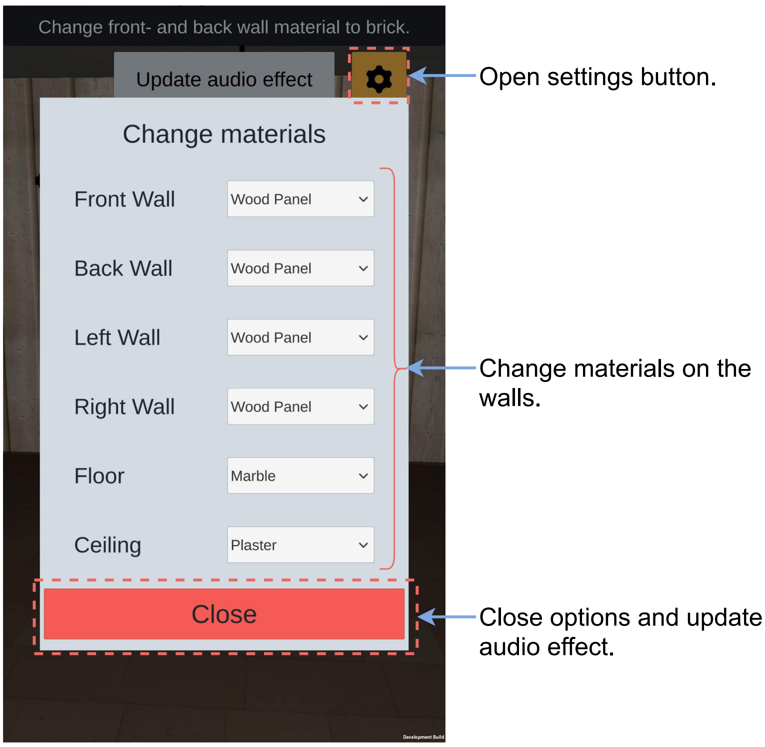
Update sound effect: when clicked, you will update the sound effect at your current position. The sound effects are dependent on your location in the room; hence you should use this button frequently to get the best impression of the acoustics. A freeze in the application of 1-2 seconds is normal.



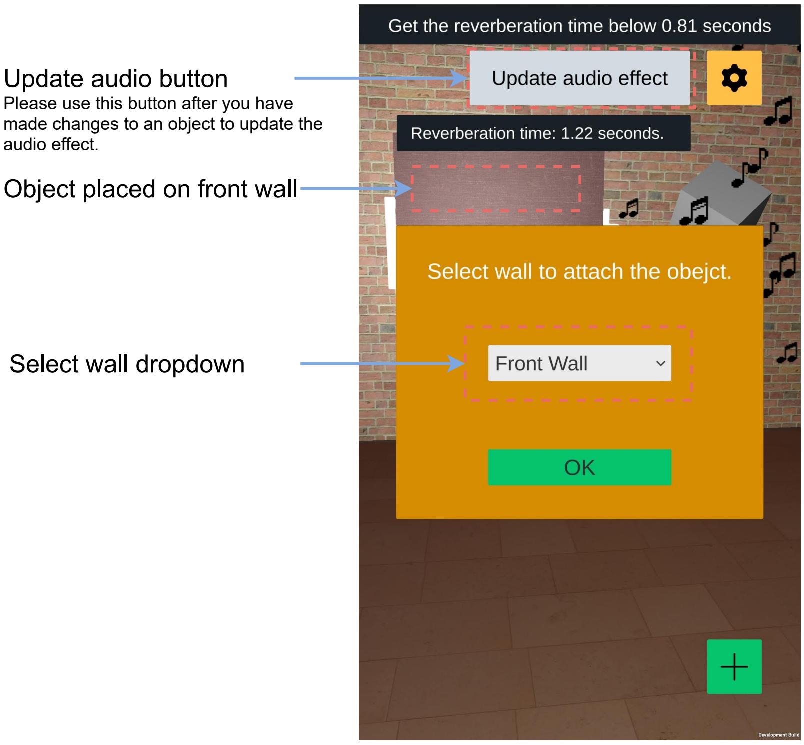
Joystick: the joystick will be displayed when your finger touches the screen. This feature is only available in Scene 2 when AR functionality is turned off. Use this joystick to move around.

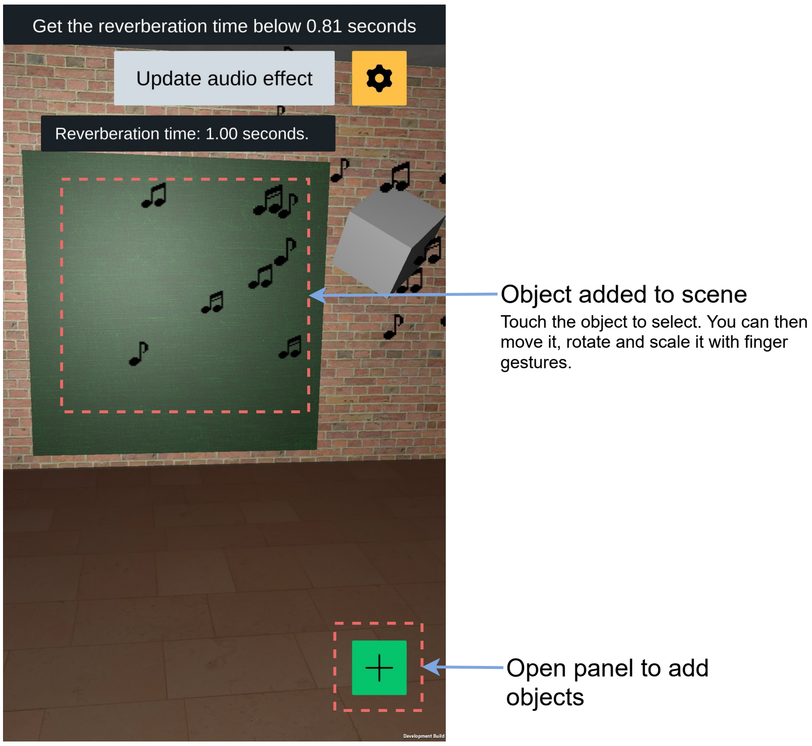


Settings: open the settings panel by clicking the gear button. Here, you can change building materials on the walls. When you close the panel, the application will freeze for 1-2 seconds.



Objects: to add objects to the scene, click on the "plus"-icon and select your desired object. By touching an object in the scene, you will select it and be able to move, scale and rotate it with finger gestures. Remember to update the audio effect after interacting with the objects. To deselect the object, you simply touch anywhere else on the screen.





B.4 Consent Form

Consent Form

I have read the information for the study "Experience Acoustic Room Simulation in AR". I will participate in this study. I was informed that the following data will be obtained today during this study from me: demographic questionnaire, surveys, and feedback provided to the experimenter. I approve that all recorded data will be saved and will be used pseudonymized (e.g., identification data will be stored separately from recorded data and only be accessible to a small circle of authorized personnel) for research analysis. All data I give will be handled confidentially. All information will be used for research purposes only. Personal data will not be given to any third party.

I am aware that participating in this study is voluntary, and I can withdraw anytime without giving any reason. By doing so, I will not suffer from any disadvantage.

Additionally, I am aware that I will handle everything confidentially, I hear and see today, and I will not give any information to other people.

Name: ______________________________

Date: ______________________________

Signature: ______________________________

Experimenter: Karl Henrik Ejdfors, +47 917 15 510

B.5 Survey Questionnaire

Demographic

Gender: ☐ Female ☐ Male

Age group: ☐ Under 18 ☐ 18-24 ☐ 25-34 ☐ 35-49 ☐ 50-64 ☐ 65+

Background Information

Please state your education: ______________________

What are your typical work tasks within acoustics? ______________________

How many years have you worked in the acoustic industry? ☐ 1-5 ☐ 5-10 ☐ 10-15 ☐ 15-20 ☐ 20+

How often do you use AR? ☐ Never ☐ Once a year ☐ Once a month ☐ Once a week ☐ Every day

Survey: Scene 1

How realistic did the sound feel in the first room?	☐ Extremely bad	☐	☐	☐ Fair	☐	☐	☐ Ideal
How realistic did the sound feel in the second room?	☐ Extremely bad	☐	☐	☐ Fair	☐	☐	☐ Ideal
There was a clear difference in the sound in the two rooms.	☐ Strongly disagree	☐	☐	☐ Neutral	☐	☐	☐ Strongly agree
To what extent did you experience a sense of 'being there' inside the environments?	☐ Not at all	☐	☐	☐ Fair	☐	☐	☐ Very much

Additional comments.

Survey: Scene 2

With AR functionality enabled.

There was a clear difference in the sound in the two rooms.	☐ Strongly disagree	☐	☐	☐ Neutral	☐	☐	☐ Strongly agree
It was easy to determine when to place the marker in the first room.	☐ Strongly disagree	☐	☐	☐ Neutral	☐	☐	☐ Strongly agree
It was easy to determine when to place the marker in the second room.	☐ Strongly disagree	☐	☐	☐ Neutral	☐	☐	☐ Strongly agree
To what extent did you experience a sense of 'being there' inside the environments?	☐ Not at all	☐	☐	☐ Fair	☐	☐	☐ Very much
How much did it seem as if you could reach out and touch the objects you saw?	☐ Not at all	☐	☐	☐ Fair	☐	☐	☐ Very much

With AR functionality disabled.

There was a clear difference in the sound in the two rooms.	☐ Strongly disagree	☐	☐	☐ Neutral	☐	☐	☐ Strongly agree
It was easy to determine when to place the marker in the first room.	☐ Strongly disagree	☐	☐	☐ Neutral	☐	☐	☐ Strongly agree
It was easy to determine when to place the marker in the second room.	☐ Strongly disagree	☐	☐	☐ Neutral	☐	☐	☐ Strongly agree

---

It was easier to determine when to place the markers with AR functionality enabled.	☐ Strongly disagree	☐	☐	☐ Neutral	☐	☐	☐ Strongly agree

Write down the numbers on the screen.

AR enabled	Room 1	x:	________	y:	________
	Room 2	x:	________	y:	________
AR disabled	Room 1	x:	________	y:	________
	Room 2	x:	________	y:	________

Additional comments.

Survey: Scene 3

How much did it seem as if you could reach out and touch the objects you saw?	☐ Not at all	☐	☐	☐ Fair	☐	☐	☐ Very much
To what extent did you experience a sense of 'being there' inside the environment you saw?	☐ Not at all	☐	☐	☐ Fair	☐	☐	☐ Very much
How confusing or clear was the experience?	☐ Very confusing	☐	☐	☐ Neutral	☐	☐	☐ Very clear
How easy was it to listen to the changes in the room's acoustics?	☐ Very difficult	☐	☐	☐ Neutral	☐	☐	☐ Very easy

Additional comments.

Survey: Room acoustic feeling.

You are now finished with all the scenes in the simulation tool. Please evaluate the next questions on basis of your total experience.

Please remember your acting role as a client.

Presence in the environment

It was easy to experience small differences in the room's acoustics.	☐ Strongly dis-agree	☐	☐	☐ Neutral	☐	☐	☐ Strongly agree
How often did you want to or try to touch something you saw/heard?	☐ Never	☐	☐	☐ Neutral	☐	☐	☐ Always
How completely were your senses engaged?	☐ Not at all	☐	☐	☐ Fair	☐	☐	☐ Very much
To what extent did you experience a sensation of reality?	☐ Not at all	☐	☐	☐ Fair	☐	☐	☐ Very much
Overall, how much did the things in the environment you saw/heard sound like they would if you had experienced them directly?	☐ Not at all	☐	☐	☐ Fair	☐	☐	☐ Very much

Impression

You perceived this possibility for obtaining information as ...	☐ Not helpful at all	☐	☐	☐ Helpful to some degree	☐	☐	☐ Very helpful
Overall impression of this technology.	☐ Extremely bad	☐	☐	☐ Neutral	☐	☐	☐ Ideal

How would you describe your room acoustic feeling using the simulation tool?
(e.g. engaging, static, exciting, authentic to the real world etc.)

Did you get a feeling of being present in this acoustic simulation? Why/why not.

Please explain why you like/dislike this tool.

Appendix C

Supplementary Code Files

Description

The accompanying files are the controller scripts that are used for developing the application.

Attached file

Code.zip

Project URL: `https://github.com/khejd/masteroppgave-kode`

Filenames

- AcousticElement.cs
- AcousticElementDisplay.cs
- ChallengeScene.cs
- ComplexScene.cs
- ConvolutionJob.cs
- DropdownAcousticElement.cs
- Joystick.cs
- RoomImpulseResponseJob.cs
- SceneChanger
- SimpleChangeRoom.cs
- SimpleScene.cs
- TapToSpawn.cs

- ToggleActive.cs

Figure C.1, Figure C.2 and Figure C.3 show the component hierarchy in each scene.

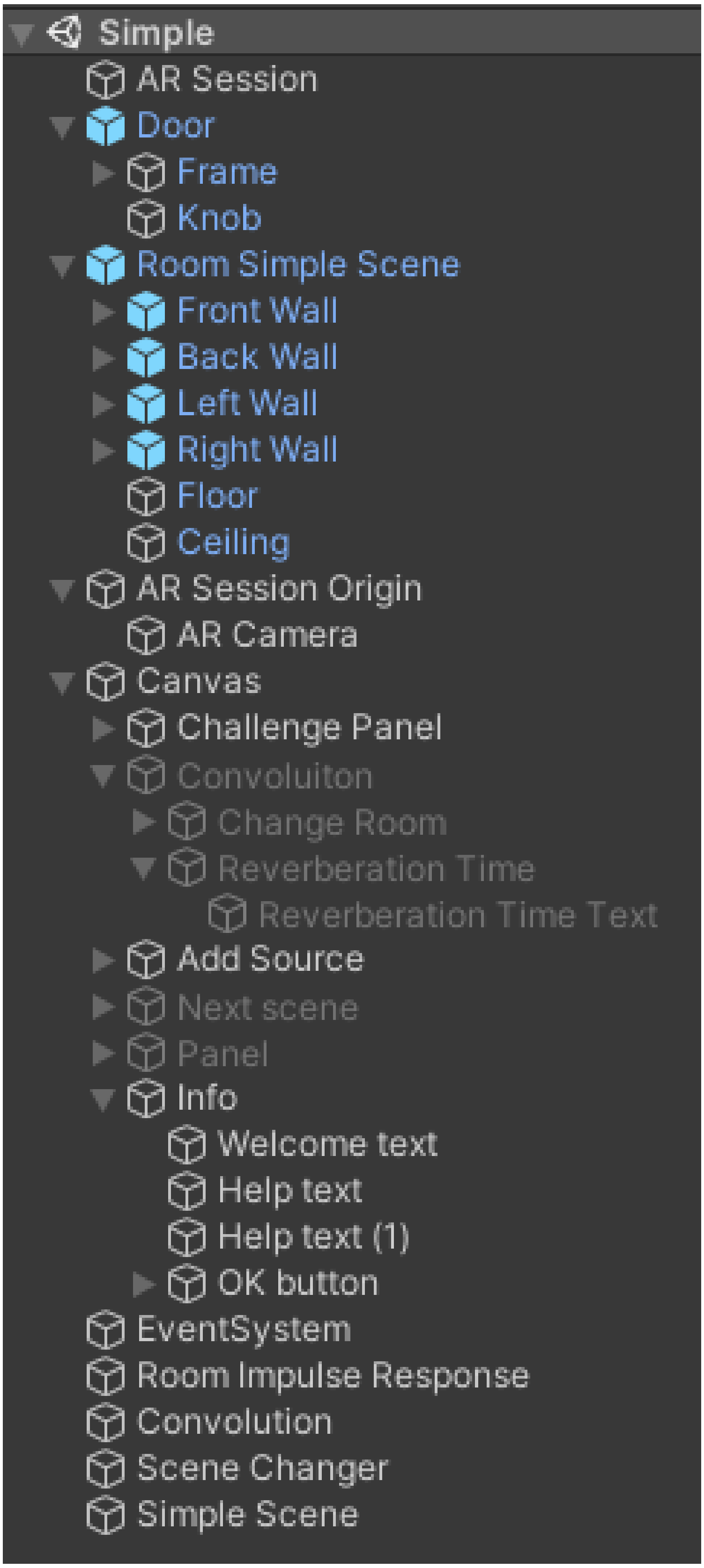


Figure C.1: Component hierarchy for Scene 1

Figure C.2: Component hierarchy for Scene 2

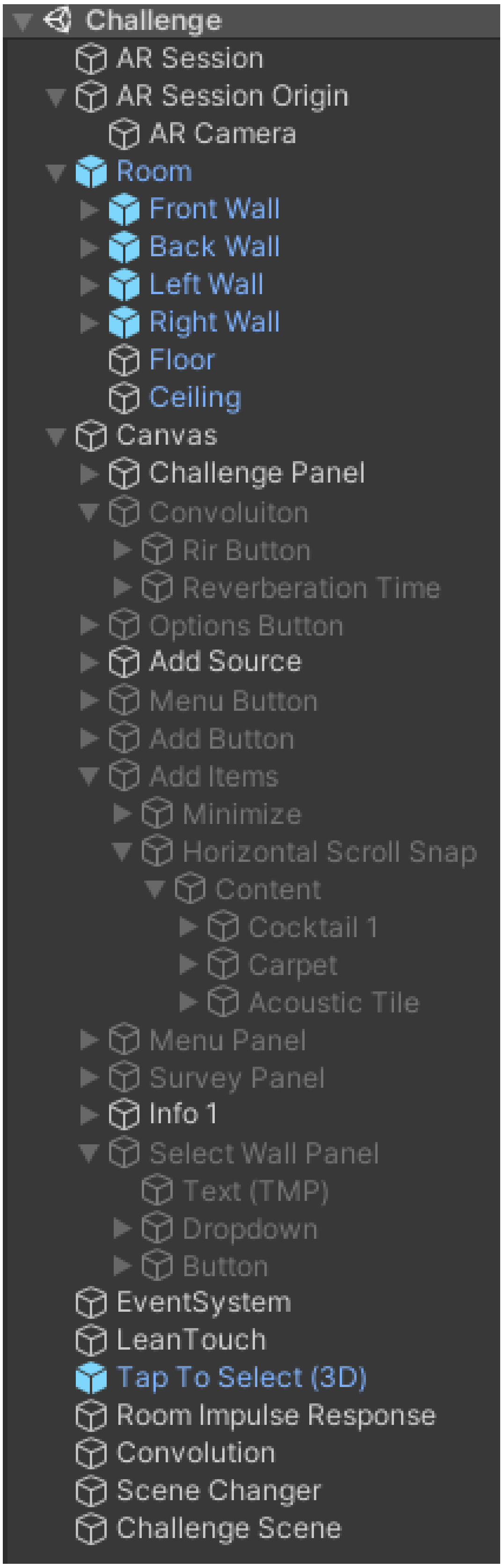


Figure C.3: Component hierarchy for Scene 3

NTNU
Norwegian University of
Science and Technology