



**CHALMERS**  
UNIVERSITY OF TECHNOLOGY

---

# **The Paraspax method applied on loudspeaker arrays**

Loudspeaker array-based synthesis of varying spaces including an investigation on how the sound field changes at different position within the array

Master's thesis in Master Program Sound and Vibration

**HANNA PERSSON**



MASTER'S THESIS ACEX30

# The Paraspax method applied on loudspeaker arrays

Loudspeaker array-based synthesis of varying spaces including an investigation on how the sound field changes at different position within the array

HANNA PERSSON



Department of Architecture and Civil Engineering  
*Division of Applied Acoustics*  
Audio Technology Group  
CHALMERS UNIVERSITY OF TECHNOLOGY  
Gothenburg, Sweden 2022

The Paraspax method applied on loudspeaker arrays  
Loudspeaker array-based synthesis of varying spaces including an investigation on  
how the sound field changes at different position within the array

HANNA PERSSON

© HANNA PERSSON, 2022.

Supervisor: Jens Ahrens, Division of Applied Acoustics  
Examiner: Jens Ahrens, Division of Applied Acoustics

Department of Architecture and Civil Engineering  
Division of Applied Acoustics  
Audio Technology Group  
Chalmers University of Technology  
SE-412 96 Gothenburg  
Telephone +46 31 772 1000

Department of Architecture and Civil Engineering  
Gothenburg, Sweden 2022

The Paraspax method applied on loudspeaker arrays  
Loudspeaker array-based synthesis of varying spaces including an investigation on  
how the sound field changes at different position within the array  
HANNA PERSSON  
Department of Architecture and Civil Engineering  
Chalmers University of Technology

## Abstract

Binaural room impulse responses (BRIRs) describes the transmission from a sound source to a listeners left and right ear, unlike monaural room impulse responses which only contains one channel and therefore sounds the same to both ears. The first method to get BRIRs of a space is by recordings of a sound source using a dummy head with microphones in each ear for different head orientations and positions in the space. This could be both time consuming and costly and therefore research is trying to find new ways that are more practical and includes signal processing.

The Paraspax method is a method for parametric spatial audio rendering with 6 DoF based on one monaural room impulse response. The method encodes monaural and spatial parameters offline into a parametric spatial sound field for arbitrary head orientations and room positions. The most important parameters are the amplitudes of the direct sound and up to 10 early reflections with corresponding times and directions of arrival (TOAs, DOAs). The TOAs are simulated from a reflection detection algorithm and the image source model provides the DOAs. These, together with the rest of the parameters, forms the basis of BRIRs synthesized for audio reproduction using headphones. The work of this thesis contains an extension of the BRIR synthesis into a loudspeaker array-based synthesis where the parametrized direct sound and early reflections are distributed over some loudspeakers arranged over a sphere. The resulting sound field is estimated for a listener positioned at different positions inside the loudspeaker array. The authors of the Paraspax have presented the method for a shoebox-shaped room but it is still unknown how it works for other environments and therefore a handful of room impulse responses will be tested.

The thesis will answer what the minimum number of loudspeakers in the loudspeaker array is and how the sound field at different listening positions differs from the sound field created at the center of the array. It will also be shown how some parameters of the loudspeaker array influences the sound. Convolvering an anechoic drums audio file with the synthesized sound field created by the loudspeaker array acts for virtually place the listener at different positions in the loudspeaker array and the resulting sound represents how the drums are perceived in the different environments of test. The results obtained by the loudspeaker array containing a reduced number of loudspeakers are analyzed and compared with a 84-loudspeaker array. It will be shown that the loudspeaker array is highly dependent on the simulated DOAs, and especially the azimuth angles as it will appear that the loudspeakers should be placed around the listener. If the DOAs of the early reflections are varied enough in azimuth, it shows that three loudspeakers are enough. The sound behaves differently depending on the direction in which the listener moves, but by

---

increasing the number of loudspeakers or the radius of the array, the listener can generally move more freely with the exception of when the TOA differences between the different loudspeakers being too large.



## Acknowledgements

I would like to express my gratitude to my supervisor and examiner Jens Ahrens at the Division of Applied Acoustics at Chalmers. Without you this project would not have been possible. Thank you for proposing this project as a master's thesis, for your wise and humble words in guiding me through this work, and thank you for your quick response to emails that did not make it a hindrance at all that you were in the US for almost throughout the course of the project. I would also like to send my thanks to Wolfgang Kropp at the Division of Applied Acoustics who was kind enough to lend me his headphones, and also to the other employees of the Division who taught me their knowledge in acoustics and have been very helpful to me during my two years as a master student.

I'm thankful to my classmates, and my dear partner and roommate Daniel Hall for acting as a sounding board during the course of the project. I also want to thank Daniel for the great support he gave me during my study period. I want to give my last thanks to my friend Christine Jeppsson just for being there.

Hanna Persson, Gothenburg, October 2022





# Contents

<b>List of Figures</b>	<b>xiii</b>
<b>List of Tables</b>	<b>xvii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 Related works . . . . .	2
<b>2 Theory</b>	<b>3</b>
2.1 Echo density profile . . . . .	3
2.2 Image source model . . . . .	4
2.3 Head related impulse responses . . . . .	6
<b>3 Methods</b>	<b>7</b>
3.1 Mixing time . . . . .	8
3.2 Reverberation level . . . . .	10
3.3 Spectral components . . . . .	10
3.3.1 Reflection detection . . . . .	12
3.3.2 Directions of arrival . . . . .	14
3.4 Late reverberation . . . . .	16
3.5 Early diffuse sound . . . . .	18
3.6 Extension on loudspeaker arrays . . . . .	20
3.6.1 Listener at the center . . . . .	21
3.6.2 Virtual loudspeaker array . . . . .	23
3.6.3 Change listener position . . . . .	25
<b>4 Results</b>	<b>27</b>
4.1 At the center of the loudspeaker array . . . . .	27
4.2 Investigation of the sweet spot . . . . .	36
<b>5 Discussion</b>	<b>49</b>
5.1 Methods . . . . .	49
5.2 At the center of the loudspeaker array . . . . .	50
5.3 Investigation of the sweet spot . . . . .	52
5.4 Influence of different parameters . . . . .	54
<b>6 Future research</b>	<b>55</b>

<b>7 Conclusion</b>	<b>57</b>
<b>Bibliography</b>	<b>59</b>
<b>A Tested room impulse responses</b>	<b>I</b>
A.1 Genesis 6 studio . . . . .	I
A.2 Trollers gill . . . . .	II
A.3 Maes howe . . . . .	II
A.4 Arthur sykes rymer auditorium . . . . .	III
A.5 Koli national park . . . . .	III
A.6 Stairway . . . . .	III
A.7 Hoffman lime kiln . . . . .	VI
A.8 Central hall . . . . .	VII
A.9 Helsington church . . . . .	VIII
A.10 Promenadikeskus concert hall . . . . .	VIII
A.11 Innocent railway tunnel . . . . .	X
A.12 Falkland palace royal tennis court . . . . .	XI
A.13 Shrine and parish church of all saints . . . . .	XI
A.14 Hamilton mausoleum . . . . .	XIV
A.15 Terrys factory warehouse . . . . .	XIV
A.16 Table . . . . .	XV
<b>B Results</b>	<b>XVII</b>

# List of Figures

2.1	Path of a first-order reflected sound ray from a sound source $S$ to a receiver $R$ using the image source $S'$ . . . . .	5
2.2	Path of a second-order reflected sound ray from a sound source $S$ to a receiver $R$ using the image source $S''$ . . . . .	5
3.1	Calculated echo density profiles using the Paraspax method of three types of RIRs; a tunnel, a semi outside environment and a warehouse of big volume. . . . .	9
3.2	Reverberation level of the Genesis 6 studio estimated by the RMS method. . . . .	11
3.3	Reverberation level of the Shrine and parish church of all saints estimated by the EDC method. . . . .	11
3.5	Direct sound and early reflections found by the reflection detection algorithm of the Paraspax method in the RIR of the Hamilton mausoleum. . . . .	13
3.4	Direct sound and early reflections found by the reflection detection algorithm of the Paraspax method in the RIR of the Helsington church. . . . .	13
3.6	Direct sound and early reflections found by the reflection detection algorithm of the Paraspax method in the RIR of the Koli national park. . . . .	14
3.7	DOAs of the direct sound and early reflections in azimuth and elevation of Arthur sykes rymer auditorium, found by the Paraspax method. . . . .	15
3.8	The inside (left) and the floor plan marked with different source and receiver positions (right) of the Hoffman lime kiln chamber. . . . .	16
3.9	DOAs of the direct sound and early reflections in azimuth and elevation of the Stairway, found by the Paraspax method. . . . .	17
3.10	DOAs of the direct sound and early reflections in azimuth and elevation of Hoffman lime kiln, found by the Paraspax method. . . . .	17
3.11	The early part of the measured RIR of the Helsington church plotted together with the detected direct sound and early reflections (directional part) and the corresponding weighting function. The directional part and weighting function is found by the Paraspax method. . . . .	19
3.12	The binaural diffuse reverberation built from binaural white noise and based on the measured RIR of the Helsington church plotted together with the inverse weighting function. . . . .	20

3.13	The directional and diffuse components that forms the early part of the synthesized BRIR of the Helsington church. . . . .	21
3.14	A spherical loudspeaker array of 84 loudspeakers positioned at the green dots. . . . .	22
3.15	Early reflections passed to two different loudspeakers of the array. One loudspeaker is assigned one reflection (left) while another is assigned three reflections (right). . . . .	23
3.16	Measured HRIRs from a Neumann KU100 artificial head showing how sound is reaching the left and right ear when the sound source is positioned at $90^\circ$ to the left (left) and at $180^\circ$ right behind (right). . . . .	24
3.17	Sound travelling from the loudspeakers to a listener positioned at the center of the array (black dot) and to a listener at an arbitrary position (red dot) within a simplified loudspeaker array of four loudspeakers in the horizontal plane. . . . .	25
4.1	A loudspeaker array in the vertical plane containing 7 loudspeakers positioned at the green dots right in front of the listener. . . . .	29
4.2	A loudspeaker array in the horizontal plane containing 3 loudspeakers positioned at the green dots around the listener. . . . .	30
4.3	Measured monaural room impulse response of Maes Howe. . . . .	31
4.4	Loudspeaker array synthesis of Maes howe from a loudspeaker array in the horizontal plane containing 3 loudspeakers (left) and from the 84-loudspeaker array (right). . . . .	31
4.5	Interaural coherence of Maes howe for a loudspeaker array in the horizontal plane containing 3 loudspeakers (left) and for the 84-loudspeaker array (right). . . . .	32
4.6	Loudspeaker array synthesis of the Helsington church for a loudspeaker array only changing in the horizontal plane using 3 loudspeakers (left) and its monaural RIR (right). . . . .	33
4.7	Measured monaural room impulse response of the Helsington church. . . . .	33
4.8	Interaural coherence of the Helsington church for a loudspeaker array in the horizontal plane containing 3 loudspeakers (left) and for the 84-loudspeaker array (right). . . . .	34
4.9	Loudspeaker array synthesis of the Promenadikeskus concert hall for a loudspeaker array in the horizontal plane containing 3 loudspeakers (left) and the 84-loudspeaker array (right). . . . .	35
4.10	Interaural coherence of the Promenadikeskus concert hall for a loudspeaker array in the horizontal plane containing 3 loudspeakers (left) and for the 84-loudspeaker array (right). . . . .	35
4.11	DOAs of the direct sound and early reflection of Trollers gill. . . . .	37
4.12	DOAs of the direct sound and early reflection of the Koli national park. . . . .	37
4.13	DOAs of the direct sound and early reflection of the Central Hall. . . . .	38
4.14	Directions of the listener's movements in the 3-loudspeaker array. . . . .	38
4.15	Loudspeaker array synthesis of Maes howe for a 3-loudspeaker array in the horizontal plane at the listener position 1 meter in front of the center (left) and 2 meters in front of the center (right). . . . .	39

4.16	Loudspeaker array synthesis of Maes howe for a 3-loudspeaker array in the horizontal plane at the listener position 3.5 meter in front of the center (left) and 5.5 meters in front of the center (right). . . . .	40
4.17	Interaural coherence of Maes howe for a 3-loudspeaker array in the horizontal plane at the listener position 1 meter in front of the center (left) and 3.5 meters in front of the center (right). . . . .	41
4.18	Loudspeaker array synthesis of Maes howe for a 3-loudspeaker array in the horizontal plane at the listener position 1 meter behind the center (left) and 2 meters behind the center (right). . . . .	41
4.19	Loudspeaker array synthesis of the Genesis 6 studio for a 3-loudspeaker array in the horizontal plane at the sweet spot. . . . .	42
4.20	Loudspeaker array synthesis of the Genesis 6 studio for a 3-loudspeaker array in the horizontal plane at the listener position 1.5 meters aside to the left of the center (left) and 3.5 meters aside to the left (right). . . . .	43
4.21	Loudspeaker array synthesis of the Genesis 6 studio for a 3-loudspeaker array in the horizontal plane at the listener position 1.5 meters aside to the right of the center (left) and 3.5 meters aside to the right (right). . . . .	44
4.22	HRIR from loudspeaker 3 to the listener at the new listener positions 1.5 m aside to the right (left) and 3.5 m aside to the right (right). . . . .	44
4.23	Loudspeaker array synthesis of Maes howe at the listening position 2 m diagonally in front of the center to the left (left) and at the listening position 2 m diagonally in front of the center to the right (right). . . . .	45
4.24	Loudspeaker array synthesis for the listener position 3.5 m diagonally forward to the left of the Promenadikeskus concert hall (left) and of Maes howe (right). . . . .	46
4.25	Loudspeaker array synthesis for the listener position 3.5 m aside to the left of the Genesis 6 studio when the radius is decreased to 5 m (left) and increased to 20 m (right). . . . .	47
A.1	Floor plan with source and receiver positions (left) and photo taken from the control room (right) of the Genesis 6 studio. . . . .	I
A.2	The measured valley of Trollers gill (left) and floor plan with source and receiver positions (right). . . . .	II
A.3	The outside location of Maes Howe (left) with floor plan (right). . . . .	II
A.4	The interior of Maes Howe. . . . .	III
A.5	Floor plan with measurement positions of source and receiver at the Arthur sykes rymer auditorium. . . . .	IV
A.6	The interior of the Arthur sykes rymer auditorium. . . . .	IV
A.7	The Koli national park at summer. . . . .	V
A.8	The floor at which measurements were made at the Stairway (left). The floors below the measurement floor (right). . . . .	V
A.9	Floor plan of Hoffman lime kiln with source and receiver positions. . . . .	VI
A.10	The exterior (left) and interior (right) of the Hoffman lime kiln. . . . .	VI
A.11	Floor plan of the Central hall with source and receiver positions. . . . .	VII

A.12	The interior of Central hall. The speaker used in the measurements is visible on stage (right). The hall is equipped with bleachers at the back and a bunch of chairs at the front (left). . . . .	VII
A.13	Floor plan of the Helsington church with source and receiver positions.	VIII
A.14	The interior of Helsington church. The loudspeaker at the altar (right) and the microphone at position "R6" (left). . . . .	IX
A.15	Floorplan of the Promenadikeskus concert hall with source and receiver positions. . . . .	IX
A.16	The interior of the Innocent railway tunnel. . . . .	X
A.17	Source and receiver positions of the Innocent railway tunnel. . . . .	XI
A.18	The interior of the Falkland palace royal tennis court. . . . .	XII
A.19	Floor plan of the Shrine and parish church of all saints together with source and receiver positions. . . . .	XII
A.20	The interior of the Shrine and parish church of all saints. . . . .	XIII
A.21	The exterior (left) and the interior, showing the microphone and loudspeaker position (right) of Hamilton mausoleum. . . . .	XIV
A.22	The interior of Terrys factory warehouse including microphone position (left) and loudspeaker position (right). . . . .	XV

# List of Tables

3.1	Window lengths in echo density profile when calculating mixing time for three different measured RIRs. . . . .	8
4.1	Numbers of loudspeakers required in the loudspeaker array for different impulse responses. . . . .	28
A.1	Information about the tested room impulse responses. . . . .	XVI
B.1	Various results for all tested room impulse responses. . . . .	XVIII



# 1

## Introduction

Spatial audio can be obtained by convolving audio signals by spatial room impulse responses (SRIR). SRIRs consists of a spatial description of the sound field in a given room in addition to its monaural parameters. The goal is to create a perceptually plausible virtual environment that is coherent with the real environment. Audio in virtual environments is often used together with other things that require high computer load, such as visuals, and therefore it is important to simultaneously maintaining a low computation time. This can be fulfilled by parametrizing the monaural and spatial parameters offline into a parametric spatial description of the sound field. An example of a method that do this is the *Paraspax method* [1]. The Paraspax derives monaural and spatial parameters for a 6 degrees of freedom (DoF) virtual environment which corresponds to arbitrary head orientations and room positions of the listener. This parametric description forms the basis of binaural room impulse responses (BRIRs) which are synthesized by a synthesis algorithm of the Paraspax method.

The work of this thesis is to use the Paraspax method to create parametric SRIRs for a number of different monaural room impulse responses (RIRs) of different environments. Furthermore, the BRIR synthesis in the Paraspax method creates 2-channel (left and right) BRIRs for any head orientation and listener translation applied for head phones and can be used for real-time rendering. Within this work, the BRIR synthesis will instead be extended to a loudspeaker array-based synthesis such that the monaural and spatial parameters of the RIR are distributed over a number of loudspeakers arranged along a sphere that forms the array. The loudspeaker array synthesis is first generated for a listener that is positioned at the center of the array, where the work will investigate the lowest amount of loudspeakers required in the array in order to not reduce the quality of the sound field created by the spherical loudspeaker array used as the starting point. In a next step, the listener changes position inside the array and it will be examined to what extent the listener can move from the center before the sound image will change and not sound as good anymore.

### 1.1 Background

The method of the Paraspax [1] has yet only been tested with one monaural omnidirectional room impulse response for a shoebox-shaped room of dimensions 11.73 m  $\times$  4.74 m  $\times$  4.62 m (length  $\times$  width  $\times$  height). The floor of the room is of concrete and the walls are plain, where one of the two long sides consists of large glass panes.

The reverberation time was measured to 0.9 s. However, to examine any limitations of the method, it is of interest to know how the method behaves with varying environments, for example outside environments with low reverberation times that only contributes to a low number of reflections as well as warehouses or churches of high reverberation times. Descriptions of the tested monaural RIRs together with measurements positions of loudspeaker and microphones can be found in Appendix and they are available online at [2] and [3].

In addition, the method have only been used for headphone reproduction but any research on how it behaves with loudspeaker arrays is yet to be done.

## 1.2 Related works

As the method of this thesis is built upon the Paraspax method [1], the presented theory and parametrization of the monaural RIRs presented in Methods before the extension to loudspeaker arrays takes place is taken from the Paraspax method.

The Paraspax method is inspired by the Binauralization of omnidirectional room impulse responses (*BinRIR*) algorithm which is a method for parametric spatial audio rendering with 6 DoF [4]. Just as for the Paraspax, only a single measured omnidirectional RIR is required to obtain a set of BRIRs of the room of interest. The process of synthesizing the late reverberation in BinRIR is reused in the Paraspax and will be explained in detail in Methods. However, listening tests where synthesized BRIRs was compared with measured BRIRs resulted in uncertainties in the method and especially regarding the incidence direction of the early reflections and therefore, the Paraspax has extension and improvements from the BinRIR algorithm.

# 2

## Theory

### 2.1 Echo density profile

The mixing time of a room impulse response is the moment in time where the specular part meets the diffuse part. The Paraspax method calculates the mixing time according to Abel et al. [5], where the echo density from a reverberation impulse response is measured.

Objects and reflective surfaces in a reverberant environment interacts with sounds to create reflections. These reflections will in turn interact with the environment to create even more reflections. When measuring a RIR, these reflections increases in time until the echo density can be seen statistically, or more specific, the sound pressure amplitudes of an impulse response is assumed to have a Gaussian distribution of zero mean with evolving color and level.

The echo density profile  $\eta(t)$  is measured over time with the property that once an acoustic space is fully mixed it takes on a Gaussian distribution. Over a sliding reverberation impulse response window of length  $2\delta + 1$  in samples, it is defined as the number of impulse response taps laying outside the standard deviation of the window divided by the expected number of samples lying outside a standard deviation  $\text{erfc}(1/\sqrt{2}) = 0.3173$  for a Gaussian distribution

$$\eta(t) = \frac{1/\text{erfc}(1/\sqrt{2})}{2\delta + 1} \sum_{\tau=t-\delta}^{t+\delta} \mathbf{1}\{|h(\tau)| > \sigma\}. \quad (2.1)$$

If the argument of  $\mathbf{1}\{\cdot\}$  is true, it returns one, otherwise zero,  $h(t)$  denotes the reverberant impulse response and the window standard deviation is

$$\sigma = \sqrt{\frac{1}{2\delta + 1} \sum_{\tau=t-\delta}^{t+\delta} h(\tau)}. \quad (2.2)$$

By normalizing by the expected standard deviation of a Gaussian noise, the resulting number of taps outside the standard deviation is a number between 0 and 1. Few prominent reflections separated in time and level will contribute to a larger standard deviation resulting in an echo density profile close to 0. As the reflections occurs more frequently and decreases in amplitude over time, the echo density increases over time. The mixing time is then defined as the time at which the echo density measure reaches 1 the first time.

The choice of sliding window lengths affects the echo density profile. Shorter windows are expected to have a high variance about its local mean as it includes

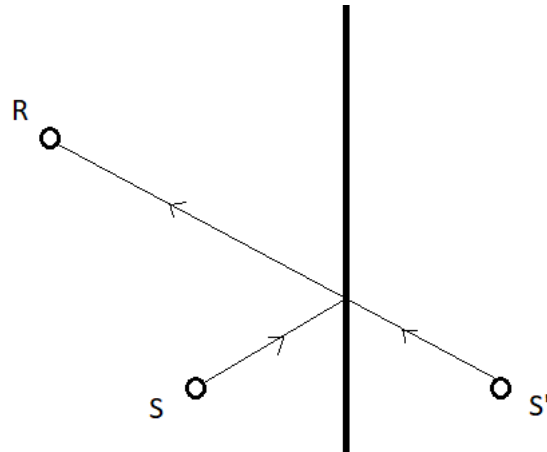
fewer impulse response taps. The window should be long enough so that it covers a few reflections but short enough to provide good statistics. Impulse responses of closely overlapping reflections should have longer windows so that the variation between the different windows should not be so great. Shorter windows is a good choice for impulse responses having only a few prominent reflections. However, too short window lengths can contribute to jumps in the echo density profile when no reflection is within the window. According to Abel et al. [5] a good choice for window length is between 20 and 30 ms. Moreover, Abel et al. presents time varying window lengths with the idea of having shorter windows at the beginning of the impulse response where the echo density haven't had time to increase yet and then let the window increase with the increasing echo density. Within work, however, only constant window lengths will be considered.

## 2.2 Image source model

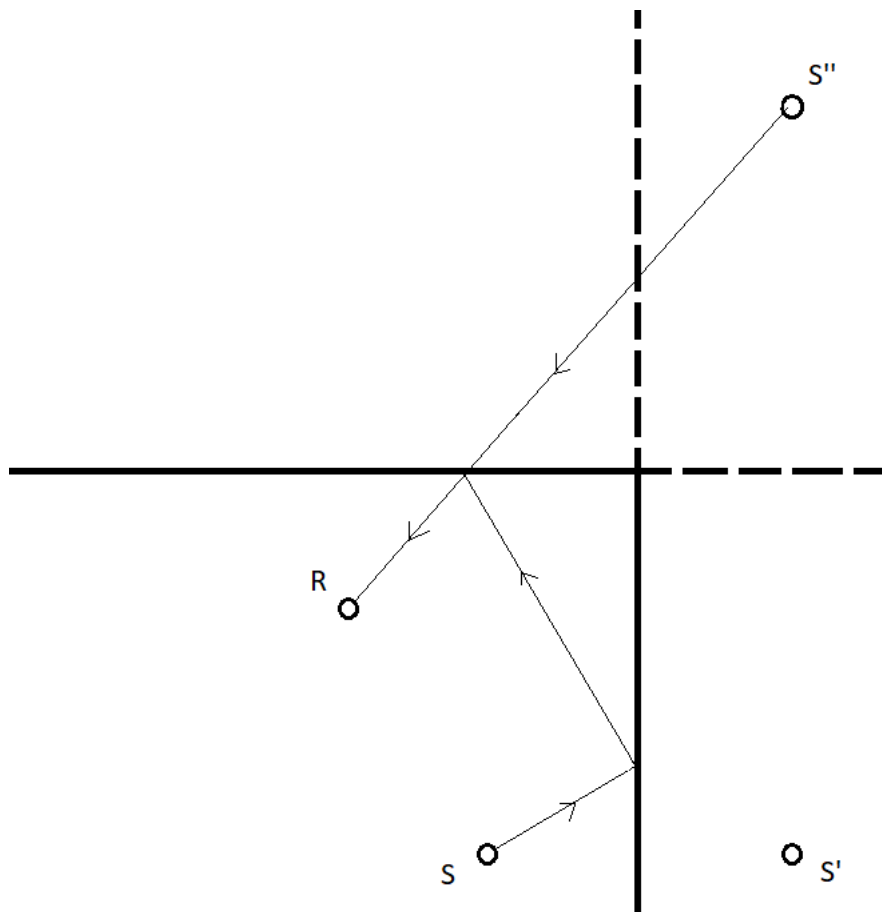
The image source model is used to find the specular reflection pattern from a sound source to a receiver within an enclosure, i.e. a room, and here a shoebox-shaped room will be considered. The method contains some simplifications about the sound field. The sound waves are idealized as sound rays which, when in a homogeneous medium, travels along straight lines and from this diffraction is neglected. In addition, interference is not considered but instead the intensity of the sound field components are added. Sound in rays are perfectly reflected at a boundary, which is not the case in real world where some of the energy is scattered in an omnidirectional pattern. As the reflections are repeated, the scattered energy increases until the majority of sound energy is diffuse. For this reason, the image source model should only be used for predicting the early reflections in a room and not the diffuse sound.

Let  $S$  be a sound source and  $R$  be a receiver, both located in a room of plane and smooth walls. The sound energy travels from  $S$  at a constant speed of sound along the rays and decreases with the distance attenuation of  $1/r^2$ , where  $r$  is the traveled distance. Except from the direct sound ray, the rays are reflected in the walls before they hits  $R$ . A reflected ray can be seen as originating behind the wall from a virtual sound source  $S'$  called image source. The image source is mirrored on the line perpendicular to the wall such that the distance between the wall and  $S$  is the same as the distance between the wall and  $S'$ , as illustrated in Figure 2.1. In this way, the path of the first-order reflected ray from  $S$  to  $R$  corresponds to the path from  $S'$  to  $R$ . The intensity of the reflected ray is reduced by a factor  $1-\alpha$ , where  $\alpha$  is the absorption coefficient of the wall [6].

If the first-order reflected ray hits some wall a second time, the process above is repeated, creating a second-order image source  $S''$ . The second-order image source  $S''$  is in turn mirrored on the line perpendicular to the first-order image source  $S'$ , as seen in Figure 2.2 The intensity of the resulting second-order reflection is then reduced by a factor  $(1-\alpha_1)(1-\alpha_2)$ , where  $\alpha_1$  and  $\alpha_2$  denotes the absorption coefficients of the first and second wall, respectively. This process can then again be repeated to obtain a third-order reflection and so on. The impulse response is then obtained by summing up signals from the source  $S$  and each image source  $S'$ ,  $S''$ , ..



**Figure 2.1:** Path of a first-order reflected sound ray from a sound source  $S$  to a receiver  $R$  using the image source  $S'$ .



**Figure 2.2:** Path of a second-order reflected sound ray from a sound source  $S$  to a receiver  $R$  using the image source  $S''$ .

## 2.3 Head related impulse responses

Humans cannot pick up sound equally from all directions as the omnidirectional microphone can. How sound is received to us depends on the size and shape of our pinnae, head and upper torso, among others. Moreover, the distance from the source to each ear can differ, the head can shadow the ears differently and the sound waves can either directly enter the ear canal or first get reflected in the torso or pinnae. The resulting modified signal received from a sound source is then processed by our auditory system and gives us the ability to localize the source. The location is estimated in our brain by comparing binaural cues, i.e. cues received by both ears, such as interaural time and level differences. Humans can locate sounds in three dimensions (distance, azimuth and elevation). Azimuth and elevation are angles originating from a spherical coordinate system. The elevation angle describes the vertical angle of a sphere, defined from a fixed zenith at  $0^\circ$  and extends to  $180^\circ$ , while the azimuth is the angle, defined up to  $360^\circ$ , of the orthogonal projection on a horizontal plane that is orthogonal to the zenith and goes through the origin. Thus, HRIRs relates the location of the source to the location of the ears and can be used to create virtual sound sources.

The frequency domain counterpart of the HRIR is called the head related transfer function (HRTF). Sets of HRIRs and HRTFs always comes in pairs, corresponding the left and right ear. BRIRs are obtained either by convolution of a RIR with a set of HRIRs in time domain or by filtering the RIR in frequency domain with the analogous HRTFs set.

HRIRs are commonly measured with a dummy head in an anechoic chamber. A dummy head have the shape and size of a human head including pinnae and ear canals where the microphones are placed. In this way the recording is made up from the dummy head's and especially a human's perspective. There are a handful of dummy heads which differs in design, for example KEMAR and Neumann KU100. The measurements are performed by rotating the dummy head in the horizontal plane of high resolution in front of a loudspeaker at a constant distance throughout the measurement. The result is a set of HRIRs that corresponds to how humans pick up sound when the source is positioned at various positions around us.

In the far field, when the distance between the source and dummy head  $r$  is greater than 1 m, the HRIRs is attenuated by a factor  $1/r$ . For distances smaller than 1 m, the measured differences between the left and right ear will increase and it is therefore more common to perform the measurements with at least 1 m between the source and dummy head.

# 3

## Methods

To run the Paraspax method [1] with any monaural RIR the distance from source to receiver as well as source direction in azimuth and elevation are required. The method got its name from the three keywords *parametrization*, *spatialization* and *extrapolation* that forms the basis of the method. In the first part of the parametrization, following standard monaural room acoustic parameters according to ISO 3382-2 [8] are calculated; reverberation time ( $RT_{60}$ ,  $RT_{30}$ ,  $RT_{20}$ ), early decay time ( $EDT$ ), clarity ( $C_{80}$ ,  $C_{50}$ ), definition ( $D_{50}$ ,  $D_{80}$ ), early decay curve ( $EDC$ ) and direct-to-reverberant ratio ( $DRR$ ). The parameters are calculated in both octave bands and broadband spectrum in the frequency range of human hearing (20 Hz - 20 kHz [9]). Then the amplitude and time of arrival (TOA) for the direct sound and early reflections are estimated. As a last step in the parametrization, the reverberation level is calculated, defined as the level of the diffuse sound field at the TOA of the first early reflection.

The spatialization sets directions of arrival (DOAs) to the direct sound and early reflections in spherical coordinates (azimuth and elevation). By doing this a 3 DoF sound field for arbitrary head orientations of the listener is yield. The method can be extended to create a sound field of 6 DoF for arbitrary head orientations as the listener moves through the room. This is done in the extrapolation part where the amplitudes, TOAs and DOAs of the direct sound and early reflections are modified corresponding to a virtual space. The result is a BRIR synthesized for headphone reproduction. By extending the method on loudspeaker arrays, the extrapolation part is replaced by a loudspeaker array setup, where the parameterized directional components of the measured RIR are distributed between the loudspeakers according to their respective DOAs. Each loudspeaker signal is then assigned synthesized early diffuse sound and late reverberation.

The parts of the method described in section 3.1-3.5 is taken directly from the Paraspax method and shows how a RIR is parameterized and spatialized, followed by how the diffuse reverberation is synthesized. It will also be reported how the Paraspax method behaves for a variety of rooms, that are listed in Appendix. In section 3.6 the extension of the method will be explained in detail, where the parts that are re-used from the Paraspax method are noted. Except for these parts, the extension on loudspeaker array is created from scratch.

### 3.1 Mixing time

A RIR is composed of direct sound, early reflections and late reverberation. The direct sound and early reflections are spectral components and appears in the beginning of the RIR. The late reverberation that consists only of diffuse sound belongs to the latter part and its start in time is determined by the mixing time.

The Paraspax method divides and processes the early part from the late part and connects them once all respective components are found and it is therefore important that the mixing time is predicted exactly in the transition between the directional and diffuse sound field of the RIR. The mixing time is a good predictor if it is estimated after all early reflections. If it instead appears too early in time, some of the early reflections will be part of the diffuse sound. Nor must it be estimated too late in time since the diffuse sound in the early part of the synthesized BRIR will depend on when the mixing time starts, and the same applies to each loudspeaker signal of the loudspeaker array.

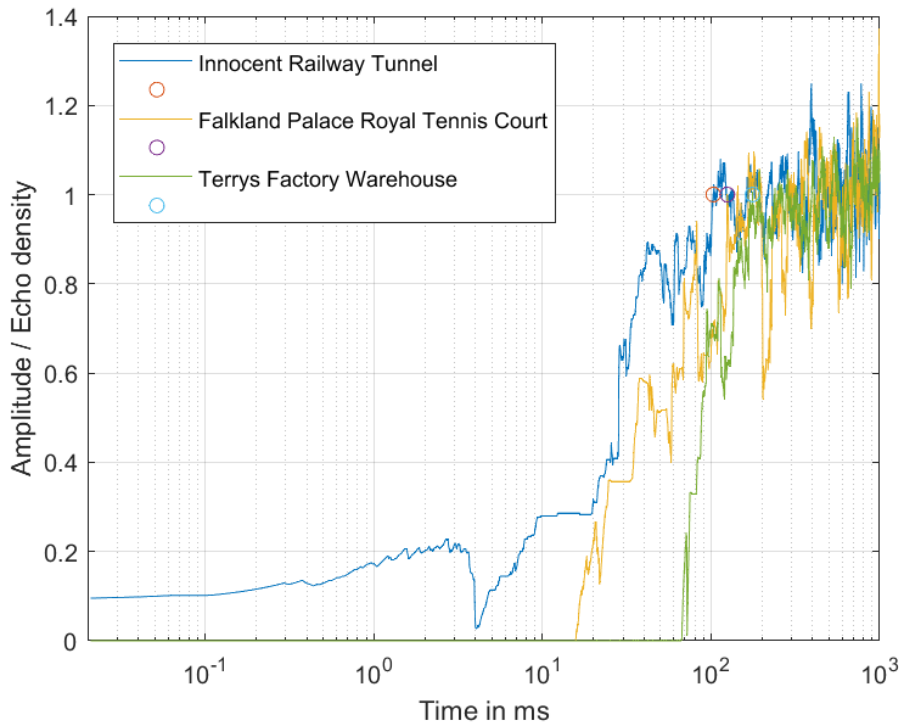
The echo density profile is calculated for estimating the mixing time and by default, the Paraspax method uses a window length of approximately 21.3 ms. By adjusting the window lengths, the mixing time is shifted, where shorter lengths contributes to earlier mixing times and longer lengths gives later mixing times. Table 3.1 shows the window length and mixing time in milliseconds of three selected monaural RIRs. The first is called the *Innocent railway tunnel* and is a tunnel previously accommodated with two railway tracks. It is now used for pedestrians and bicycles as the tracks got replaced by paving. The tunnel have a semicircular cross section of dimensions 4.5 m  $\times$  6 m (height  $\times$  width) and it extends as far as 517 m [2]. The *Falkland palace royal tennis court* is a semi outside environment with the size of 2300 m<sup>3</sup> and no roof. Its walls and floor is made of a concrete-like material. The *Terrys factory warehouse* is an empty industrial building of 4500 m<sup>3</sup>.

Name	Window length in ms	Mixing time in ms
Innocent railway tunnel	42.67	114.88
Falkland Palace Royal tennis court	25	168.56
Terrys factory warehouse	21.3	176

**Table 3.1:** Window lengths in echo density profile when calculating mixing time for three different measured RIRs.

The full list containing the mixing times with respective window lengths of all RIRs is presented in Appendix, showing that the default window length only is ideal for 5 out of 15 tested RIRs. There are no clear pattern between suitable window length and type of RIR, meaning that there are no rule that works in every case. The window lengths used in the end was selected by trial and error so that the estimated mixing time appears in the transition of the early and late part of the RIR. Furthermore, the window lengths changes drastically independent of reverberation time, indicating that the echo density profile measure of a RIR is insensitive to reverberation time, just as stated by Abel and Huang [5].

The echo density profile of the three RIRs listed in Table 3.1, is presented in Figure 3.1, where respective mixing times are marked with circles. The mixing times is set to when the echo density reaches 1 the first time, according to Abel and Huang [5].



**Figure 3.1:** Calculated echo density profiles using the Paraspax method of three types of RIRs; a tunnel, a semi outside environment and a warehouse of big volume.

Typical for a reverberation impulse response is a echo density profile starting around 0 and then increasing towards 1. As seen in Figure 3.1, this is the case for all three RIRs, and also for all tested 15 RIRs. Furthermore, the narrow shape of the tunnel allows the sound to be reflected quickly which results in an echo density profile starting above 0 in contrast to the tennis court and warehouse, and especially the warehouse. Its big, empty space causes the sound to travel longer before it gets reflected, but as soon as it does so the space is quickly fully mixed. This is seen in Figure 3.1 as it begins to grow fast. Both the space of the tunnel and the tennis court are equipped with cavities where the sound disappears without being reflected back. This results in a space that never fully gets mixed. This can be seen in the graphs of their respective echo density profile which jumps up and down. This is especially the case for the tennis court which has a larger opened surface compared to the tunnel.

The mixing time for the tunnel is estimated earliest in time, followed by the tennis court and then the warehouse. This estimation holds for any measurement setup as the echo density profile measure is independent of measurement setup within the same room. The increase in echo density profile only depends on the room's shape and volume.

## 3.2 Reverberation level

Reverberation level contains information about both directional and diffuse reverberation. Within the Paraspax there are three different methods for estimating the reverberation level called the MAX, RMS and EDC method. What differentiates the various methods is how the envelope of the absolute pressure response  $|p|$  is estimated. Just as perceived by the names of the different methods, the MAX and RMS method uses a sliding window of 1 ms and then the maximum respective the root-mean-square of that window is calculated. The EDC method uses the previously calculated early decay curve and transforms it into a level curve.

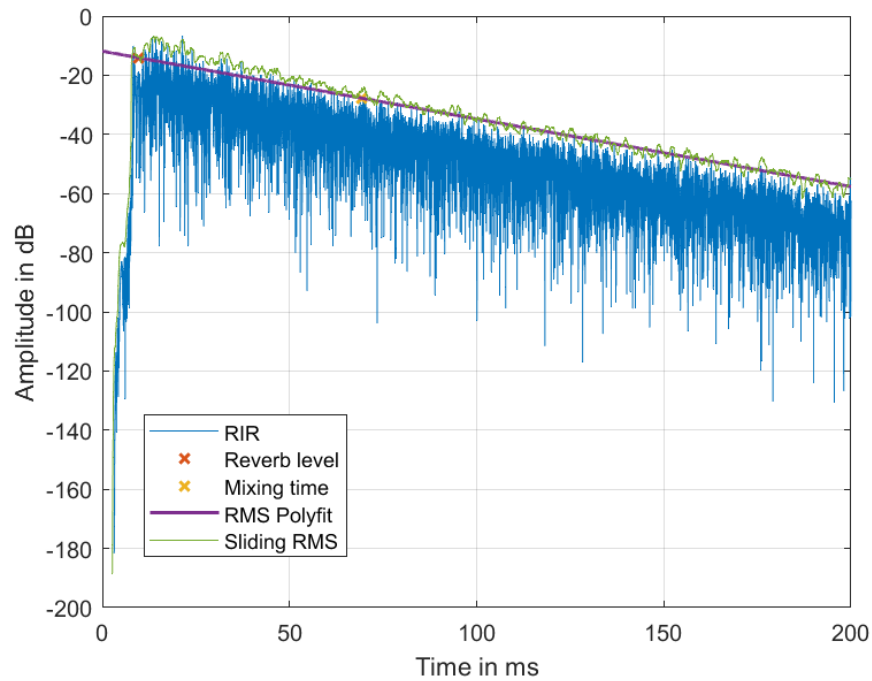
The following steps are the same for all three methods. A first-order polynomial fit is used for the envelope from two to three times the mixing time in order to guarantee that no early reflections distorts the envelope. The remaining values of the decay curve is estimated by linear extrapolation. The reverberation level is defined as the level of the diffuse sound field at the TOA of the first early reflection, but the amplitude of the reverberation can be found from the decay curve at any time, for example at the mixing time.

According to the authors of the Paraspax, the reverberation level can be estimated using any of the three methods, but the MAX method is the one that provides the best estimates. This conclusion holds also for the tested RIRs, where the MAX method was used for all of them except the *Genesis 6 studio* and the *Shrine and parish church of all saints north street* that got the best reverberation level estimates using the RMS and EDC method, respectively. The reverberation level of these spaces as well as sliding window and polynomial fit can be seen in Figure 3.2 and 3.3. These plots shows that respective polynomial fit follows the decay of the amplitude in dB and is thereby a good representation of the diffuse sound in the spectral components. The reverberation level is calculated as -6.7 dB for the Genesis 6 studio and -9.4 dB for the Shrine and parish church of all saints north street. The reverberation level for the other RIRs are listed in the Appendix.

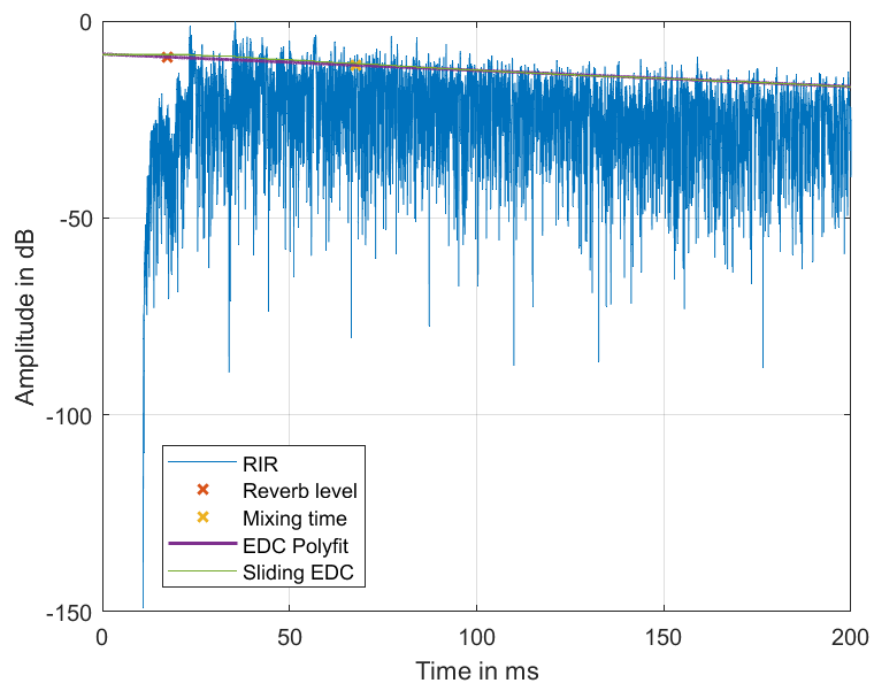
## 3.3 Spectral components

The Paraspax method divides the early part of the RIR into spectral components and early diffuse sound. The spectral components of the direct sound and early reflections are estimated in time, defined by the respective TOAs, with corresponding amplitudes. They are then assigned with DOAs in spherical coordinates which describes the angle at which they reach the listener. The parts of the processing that produces these parameters are described below, followed by the processing that produces the synthesized early diffuse sound. The directional components parameterized by the Paraspax method is used for the loudspeaker array, and a number of decorrelated copies of the early diffuse sound are synthesized, one for each loudspeaker.

The direct sound is one single event in the RIR and is easily found by applying a 1 ms long window to the onset. Then the TOA is defined as the time index of the absolute maximum of the pressure response within this window. To get the



**Figure 3.2:** Reverberation level of the Genesis 6 studio estimated by the RMS method.



**Figure 3.3:** Reverberation level of the Shrine and parish church of all saints estimated by the EDC method.

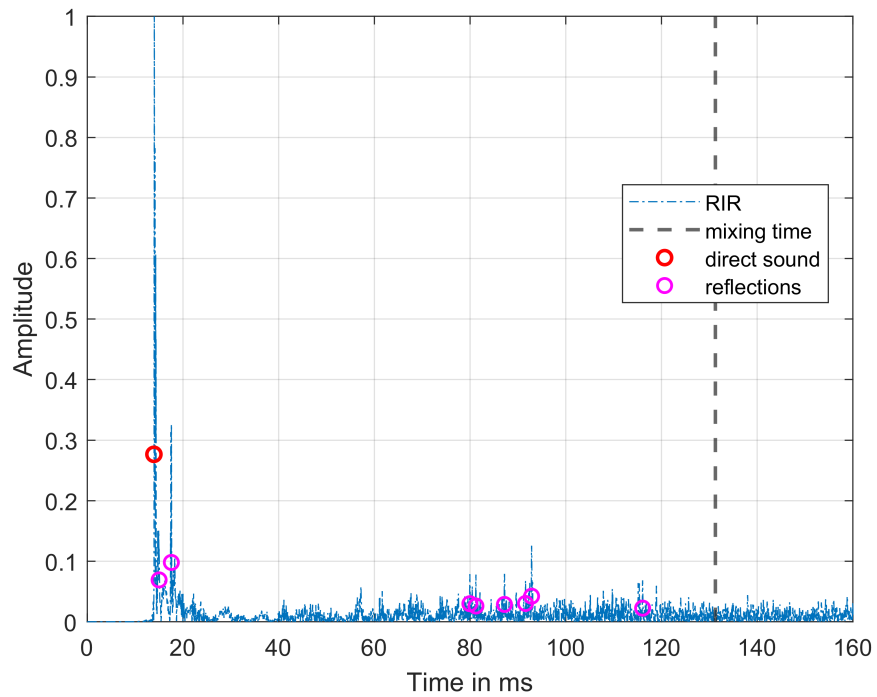
corresponding RMS amplitude a new asymmetric window is applied around the TOA. The window is of length 1.5 ms, starting 0.5 ms before the TOA and ending 1 ms after due to summing localization, i.e. if two or more sound waves arrive within a time interval of 1 ms or smaller then all sound sources contribute to the direction of the perceived total sound [10]. The amplitude is then defined as the RMS average of the window. The method succeeded in finding these components of the direct sound for all tested RIRs. Moreover, the method finds up to 10 early reflections for each RIR which makes the approach a bit more complex. The TOAs and amplitudes will be tracked by a reflection detection algorithm explained in subsequent section.

#### 3.3.1 Reflection detection

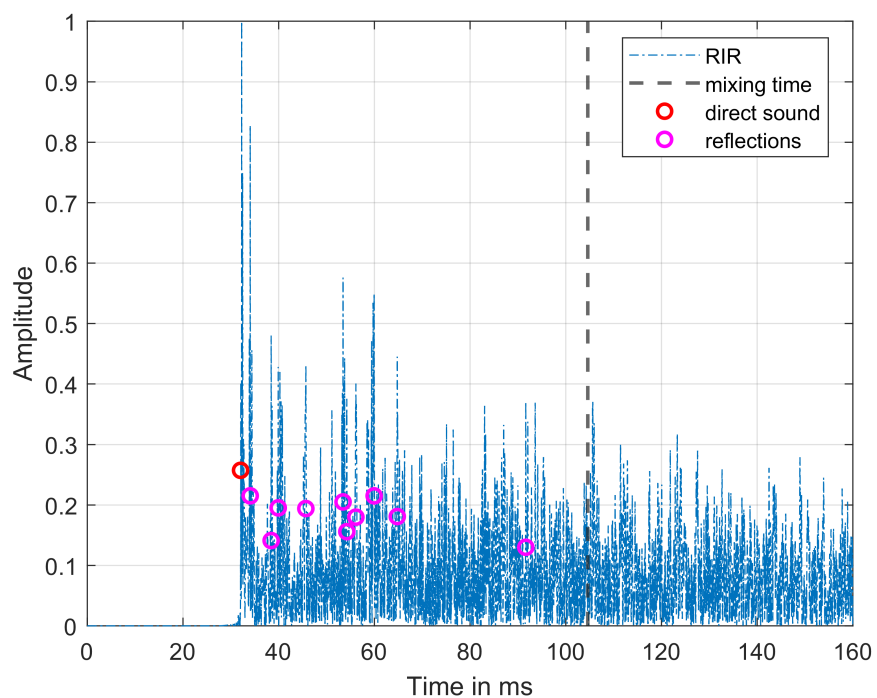
When the TOA and amplitude of the direct sound is found, a reflection detection algorithm is used to find the TOAs and amplitudes of the early reflections. First, the TOAs are found by applying a sliding window of 1 ms to the whole RIR or up to two times the estimated mixing time. If the energy of a time index in the window is three times higher than the median energy of the whole window then a reflection is defined at this time index, i.e. the TOA of the reflection. The RMS amplitudes of the early reflections at the TOAs are calculated in an asymmetrical window in the same way as for the direct sound. The high resolution of the window length is needed in order to capture the ground reflection that is important, especially for outside environments where the ground reflection might be the only reflection. Furthermore, the high resolution normally provides more than 10 early reflections which are first selected according to summing localization; if more than one TOA is found within a time span of 1 ms then the one with the highest RMS amplitude is defined as a reflection while the other(s) are removed from the early reflections.

As there still may be many reflections in the selection list, they are sorted by their amplitudes in descending order such that the early reflections selected in the end corresponds to the loudest reflections. The Paraspax method selects between 6 and 10 most prominent reflections due to interest of lowering the computational load. The goal is to not use an unnecessary number of reflections that slows down the processing but at the same time to use as many as necessary to fully recreate a space. A previous study where the aim was to use a minimal set of salient early reflections showed by listening experiments that 6 reflections are enough to reproduce parametric spatial audio rendering that is indiscernible from a fully-rendered reference for speech content based on the image source model [10]. The image source model is one out of three different approaches for calculating the DOAs of the early reflections in the Paraspax method, and it is also the image source model implemented by the Paraspax method that is used within this thesis in subsequent section.

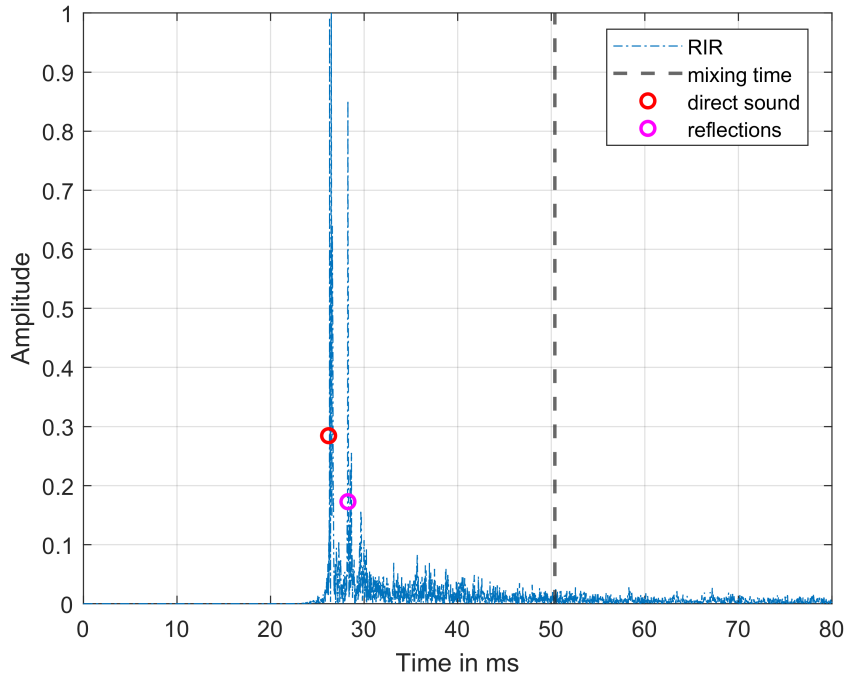
The absolute pressure response of the RIR is plotted together with the TOAs and RMS amplitudes, detected by the Paraspax method, of the direct sound and early reflections of three tested RIRs in Figure 3.4-3.6. For these RIRs the reflection detection algorithm found a different number of reflections.



**Figure 3.5:** Direct sound and early reflections found by the reflection detection algorithm of the Paraspax method in the RIR of the Hamilton mausoleum.



**Figure 3.4:** Direct sound and early reflections found by the reflection detection algorithm of the Paraspax method in the RIR of the Helsington church.



**Figure 3.6:** Direct sound and early reflections found by the reflection detection algorithm of the Paraspax method in the RIR of the Koli national park.

Figure 3.4-3.6 shows that the mixing time appears after all early reflections which means it is a good predictor. The reflection algorithm detected 10 early reflections for the *Helsington church* while it only found 9 for the *Hamilton mausoleum*. The *Koli national park* is an outside environment only contributing to a single reflection.

### 3.3.2 Directions of arrival

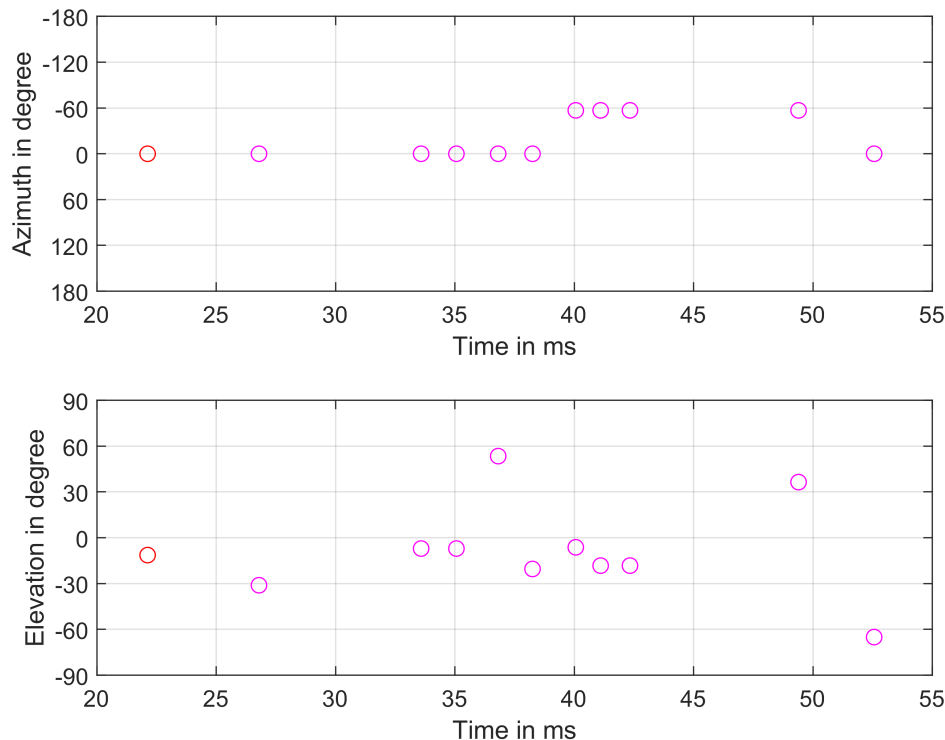
The Paraspax method uses three different approaches for estimating the DOAs of the selected early reflections whose TOAs and amplitudes have been previously found. The approach used here is the image source model (see Section 2.2), but for those interested in the other approaches, based on pseudo-randomized or precomputed DOAs, are referred to [1]. As mentioned before, the method requires predefined source direction and source distance which can be used to estimate the DOA of the direct sound. Alternatively, the direct sound DOA can be found from the image source simulation together with the DOAs of the reflections.

Geometrical data (source and receiver positions and room dimensions) is necessary for spatialization using the image source model. The room dimensions are an approximation of a shoebox-shaped room, an approximation that may differ greatly from reality for some of the tested rooms that, for example, have arched sides, contain small passages or misses one or some of the walls. The order of image sources is set to 2 which would be enough since first-order image sources in an empty shoebox-shaped room contributes to six dynamically reproduced early reflections from the four walls, floor and roof. The Paraspax method also allows for preferring first-order

reflections but this was not the case in this study.

The image source simulation of the Paraspax method derives all second-order reflections from the simulated room, each with a corresponding TOA and DOA. These TOAs are compared with the TOAs obtained from the reflection detection such that those with the smallest TOA differences are defined as the same reflection. Then the azimuth and elevation of the corresponding DOA describes in which direction the reflection will arrive to the listener. In the following, the environments of three tested RIRs are described and their respective DOA pattern found from the simulated image source model of the Paraspax method will be shown.

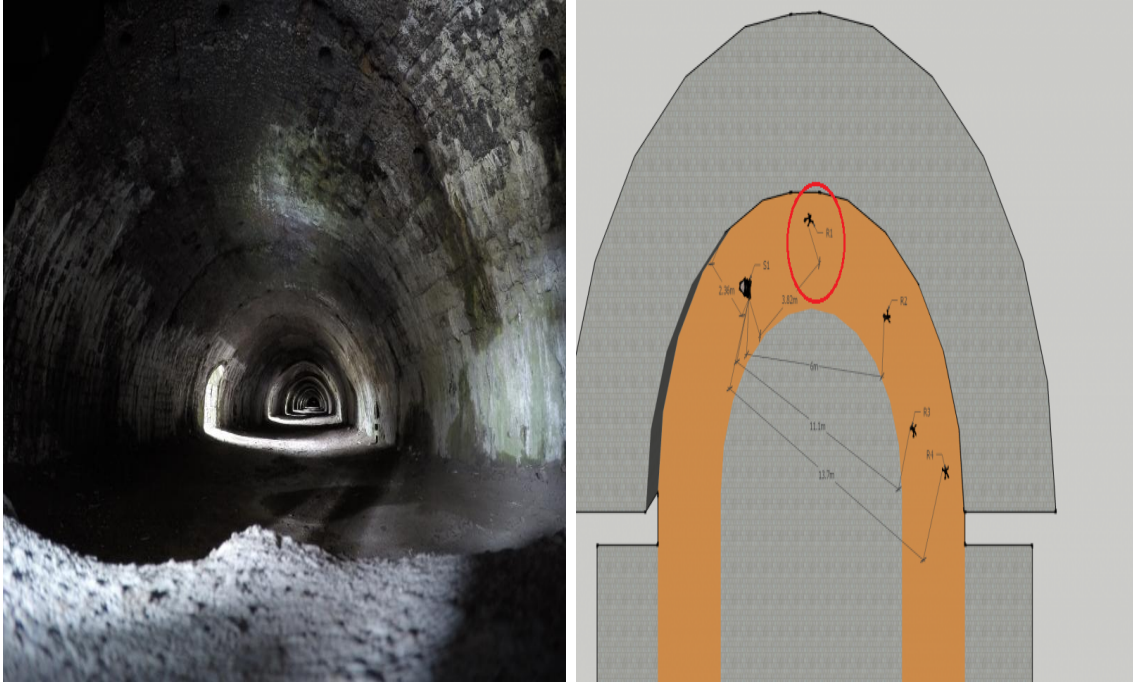
*Arthur sykes rymer auditorium* is said to reproduce sound of high quality thanks to its unique acoustics. Its preferred noise criterion (PNC) is better than the PNC 15 standard, which means it shuts out outside noise. It is a rectangular-shaped auditorium so the space itself is shoebox-shaped. The simulated DOAs of the direct sound (red) and early reflections (fushia) in the auditorium are presented in Figure 3.7, showing that reflections hits the receiver only in two azimuth angles but in various elevation angles. Due to its rectangular shape, the result is more accurate than the following examples, but the image source model does not take into account the inclined medical floor and its interior. However, the image source model is a popular approach for generating early reflections due to its efficient simulation [11].



**Figure 3.7:** DOAs of the direct sound and early reflections in azimuth and elevation of Arthur sykes rymer auditorium, found by the Paraspax method.

The *Stairway* of a university is located in a shoebox-shaped room of which rises in height. The impulse response is assumed measured at one of the mid floors so

that the source and receiver positions are positioned in the middle of the room. Note that reflections in the steps will not be estimated in an empty shoebox-shaped room of the image source model. The tunnel-like chamber *Hoffman lime kiln* has arched side walls and roof. It is a large U-shaped stone construction, whose impulse response was measured at position "R1", seen in Figure 3.8, differs markedly from a rectangular room. The curved area, seen in the floor plan, is used in the image source simulation, approximated as a shoebox-shaped room of dimensions  $25\text{ m} \times 4.72\text{ m} \times 2.3\text{ m}$  (length  $\times$  width  $\times$  height).

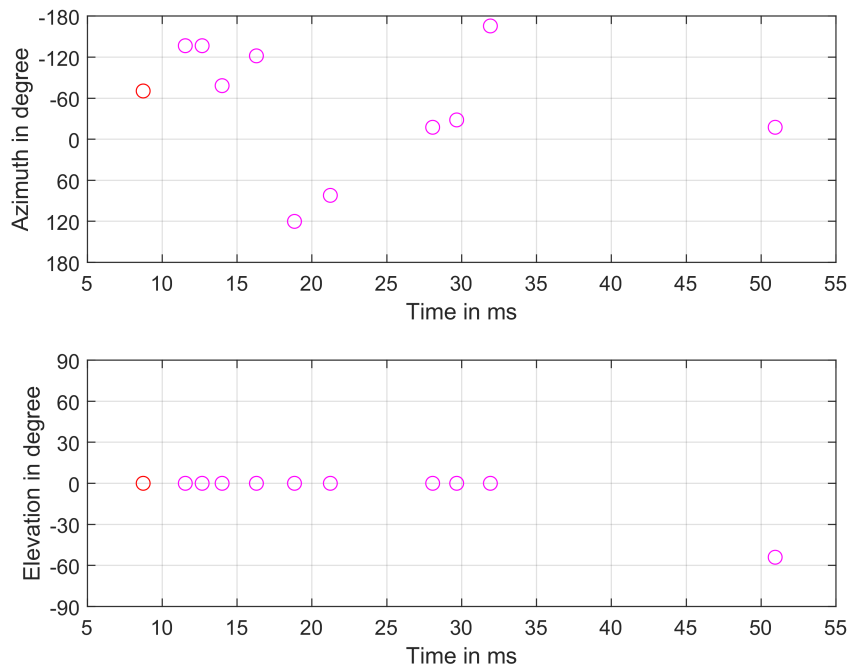


**Figure 3.8:** The inside (left) and the floor plan marked with different source and receiver positions (right) of the Hoffman lime kiln chamber.

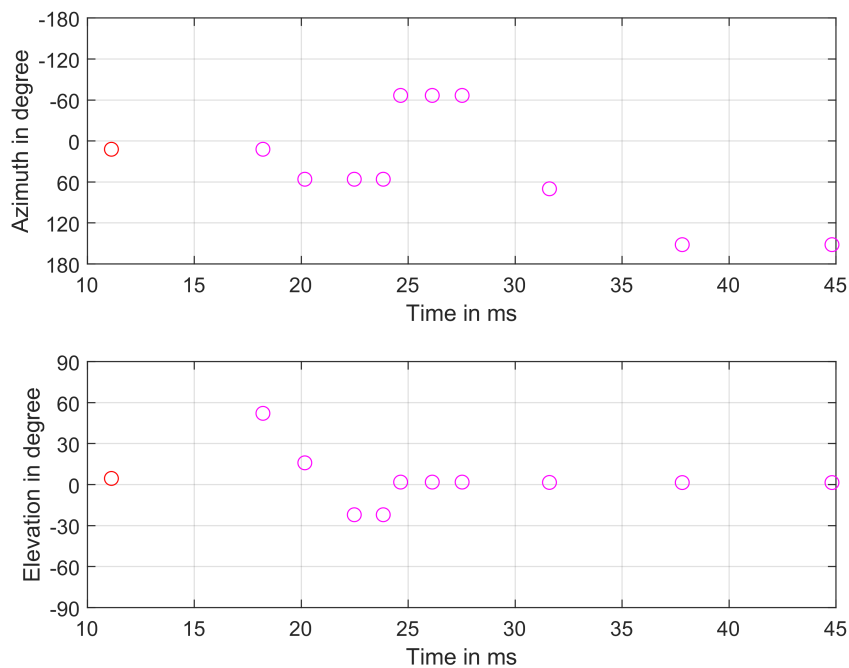
The DOAs of the direct sound and early reflections of the Stairway and Hoffman lime kiln can be seen in Figure 3.9 and Figure 3.10 respectively. Due to where the source and receiver are positioned in the Stairway, the sound waves are more likely to get reflected in the side walls than in the floor and roof and therefore, almost all reflections have  $0^\circ$  in elevation angle but have the greater spread in the azimuth. However, the stairs contribute to floor and ceiling formations where some sound waves would be reflected in reality which will not be included here. The approximation of the space of Hoffman lime kiln allows some of the sound waves to get reflected instead of disappearing through the long parallel corridors. It means that the image source simulation contributes to more reflections than in real world compared to the Stairway where the case is opposite.

### 3.4 Late reverberation

The diffuse reverberation of the Paraspax method is synthesized in the same way as in the BinRIR algorithm [4]. Just as in the Paraspax, the measured RIR is



**Figure 3.9:** DOAs of the direct sound and early reflections in azimuth and elevation of the Stairway, found by the Paraspax method.



**Figure 3.10:** DOAs of the direct sound and early reflections in azimuth and elevation of Hoffman lime kiln, found by the Paraspax method.

separated and divided into direct sound, early reflections and late reverberation, which are treated separately before being put together.

The aim is to create 2-channel late reverberation for a pair of ears that is spatially equally distributed. It means that the late reverberation is omnidirectional, having on average the same proportion of sound energy from every direction. To obtain this, the signals are decorrelated, i.e. the cross-correlation is reduced. Interaural coherence is the measure of similarity between the reverberation received by each of the two ears, where a low value is desirable and creates a more pleasant sound in comparison to correlated diffuse sound which sounds strange and not very diffuse.

In a first step, binaural white noise is generated and filtered with an interaural coherence filter which gives each channel slightly different parameters. The binaural noise is then split into time segments of 2.67 ms [4] and convolved with small chunks of the measured RIR of 0.67 ms to adapt the noise into the energy decay curve. The length of the windows was determined using a listening test. All time sections are windowed with raised-cosine ramps and then finally added all together with the overlap-add method which overlap all time sections so that the cosine ramps entering each other before adding them. This gives a smooth and coherence result.

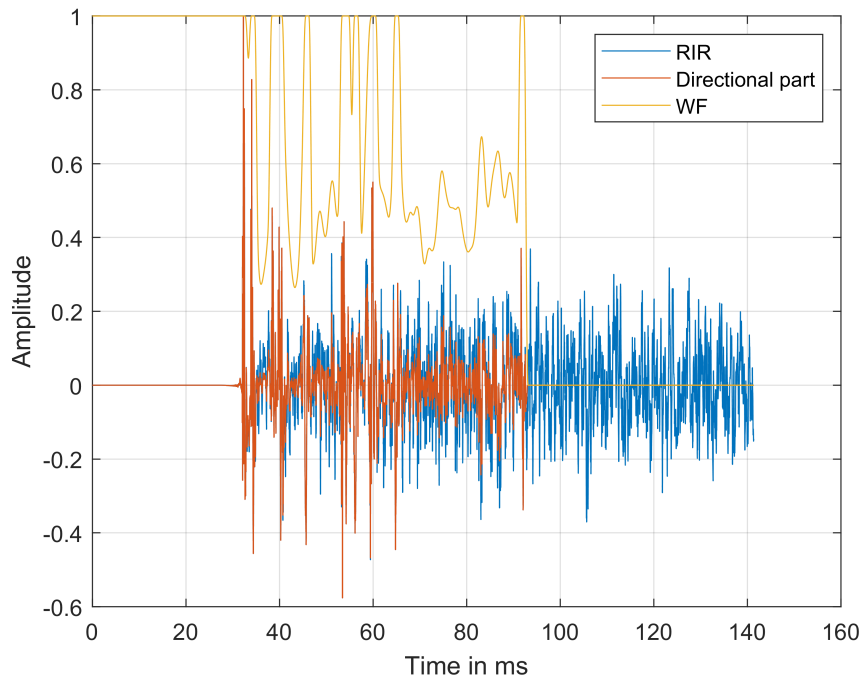
The late reverberation for the loudspeaker array is synthesized in the same way as in the Paraspax method, but a few more decorrelated copies are generated, one for each loudspeaker signal. So instead of creating a 2-channel decorrelated late reverberation for headphone reproduction as in the Paraspax method, a  $N$ -channel decorrelated late reverberation is synthesized, where  $N$  is the number of loudspeakers in the array.

## 3.5 Early diffuse sound

The late reverberation synthesized in the Paraspax method is based on the measured RIR in the whole time range so it extends therefore in both the early and latter part. The diffuse sound in the latter part is used as late reverberation. However, diffuse sound is also used in the early part of the synthesized BRIR and added to the parameterized directional components. The motivation comes from a study based on listening experiments [12], which has shown that by adding diffuse sound to the spectral components in the early part of a BRIR contributes to higher perceptual quality than if only spectral components are used. Therefore, early diffuse sound will also be present in each loudspeaker signal of the loudspeaker array. In the Paraspax method, it is synthesized by taking the early part (up to 2-3 times the mixing time) of the synthesized late reverberation from the previous step and performing some further processing described below. Since the loudspeaker array requires a  $N$ -channel late reverberation where each channel is decorrelated, the early diffuse sound will consist of a  $N$ -channel decorrelated early diffuse sound, one for each loudspeaker in the array, that is built upon the  $N$ -channel late reverberation. The following description of how the early diffuse sound is obtained by the Paraspax method is the same for the loudspeaker array, except that a few more copies are synthesized.

Together with the late reverberation, the early diffuse sound field is estimated from a weighting function based on the selected early reflections obtained from the

reflection detection. It is created by taking a sliding window of 1 ms to the absolute pressure response and convolving it by a Hanning window of 3 ms. Its values at the TOAs of the direct sound and selected early reflections are windowed with a 1.5 ms window and set to 1. The strong edges that arisen are smoothed out with a 1 ms window. The weighting function of the Helsington church, whose direct sound and early reflections are presented in Figure 3.4, can be seen in Figure 3.11 illustrated by the yellow curve.

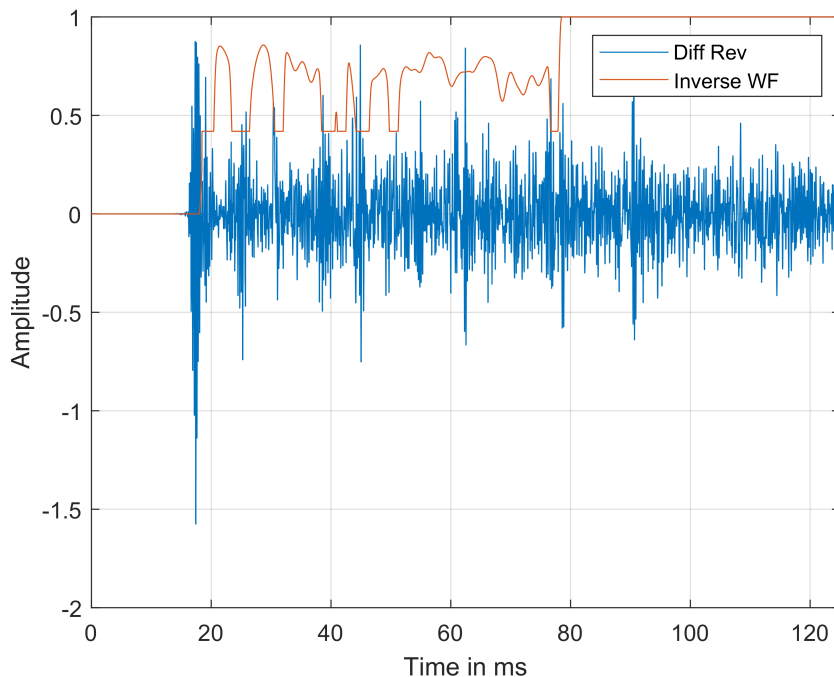


**Figure 3.11:** The early part of the measured RIR of the Helsington church plotted together with the detected direct sound and early reflections (directional part) and the corresponding weighting function. The directional part and weighting function is found by the Paraspax method.

To get the early diffuse sound, the weighting function is inverted as the square root. Let  $wf$  be the weighting function. The inverse weighting function is then

$$(wf)^{-1} = \sqrt{1 - wf}. \quad (3.1)$$

Reverberation level is also used in BRIR synthesis when estimating the early diffuse sound as preserves the diffuse sound in the spectral components. In order for the directional parts to still be prominent and not masked by the diffuse sound, the inverse weighting function is limited to not exceeding the value of the reverberation level. The inverse weighting function of the Helsington church and its diffuse reverberation in the early part of one of the  $N$  channels is plotted in Figure 3.12. The binaural diffuse sound in the early part extends up to the 2 times the mixing time.

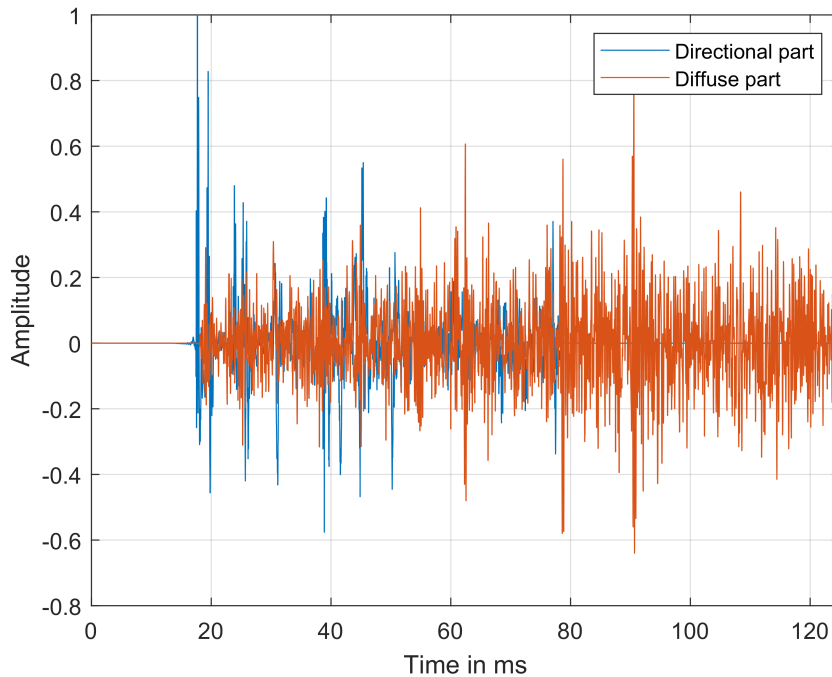


**Figure 3.12:** The binaural diffuse reverberation built from binaural white noise and based on the measured RIR of the Helsington church plotted together with the inverse weighting function.

The early diffuse sound is obtained by multiplying the two functions in Figure 3.12, and the obtained results can be seen in Figure 3.13, where it is plotted together with the directional part obtained by multiplying the weighting function with the measured RIR in order easily show that the early diffuse sound never exceeds the amplitudes of the direct sound and the early reflections.

### 3.6 Extension on loudspeaker arrays

The monaural and spatial parameters calculated in the previous steps can be used for spatial audio reproduction. The extension on loudspeaker arrays uses a number of loudspeakers as the source, instead of headphones that are used in the Parasax method. Instead of using a physical loudspeaker array, head-related impulse responses can be used to create virtual sound sources which makes it possible to virtually place a listener inside the array using headphones. The listener is then able to virtually move within the array by adjusting each loudspeaker signal in relation to the new distance and angle between the listener and the loudspeakers. The simulated sound field at the center and how it is constructed will be explained first followed by the construction of the sound field as the listener changes position.

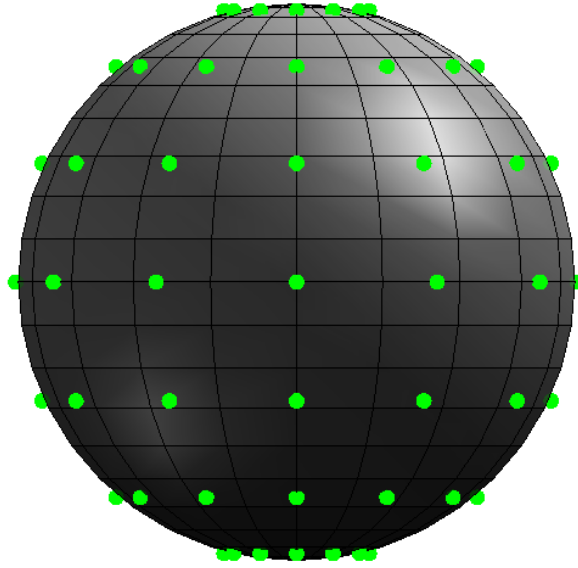


**Figure 3.13:** The directional and diffuse components that forms the early part of the synthesized BRIR of the Helsington church.

### 3.6.1 Listener at the center

The loudspeaker array is defined as a number of loudspeakers placed over a sphere. The listener is placed at the origin of the sphere so the position of each loudspeaker relative the listener can be represented by spherical coordinates (azimuth and elevation). In total, the loudspeaker array consists of 84 loudspeakers and their positions are illustrated in Figure 3.14 by the green dots. The loudspeaker array contains 7 elevation angles with a resolution of  $25^\circ$ , starting at  $75^\circ$  above and ending at  $-75^\circ$  below the listener. The azimuth angle extends around the listener from  $0^\circ$  to  $330^\circ$  with a resolution of  $30^\circ$ . The loudspeakers lying on the circle in the horizontal plane orthogonal to the zenith, at the fourth elevation row, are in line with the ears of the listener. The north and south poles of the sphere are not equipped with any loudspeaker. The paper presented by Müller and Ahrens [13] shows that listeners who performed a listening test could not hear any clear differences in SRIRs containing elevated early reflections or not. Although there are larger perceived spatial differences in loudspeaker-based reproduction compared to reproduction using headphones, the elevated reflection has to be strong in order for a listener to hear clear differences when projecting this reflection onto the horizontal plane.

In a first step, the sound field is rotated such that the DOAs of the direct sound and early reflections are converted into the global coordinate system relative the listener in which the loudspeaker array is defined. By doing this the direct sound will always be played-back from the loudspeaker positioned right in front of the listener that corresponds to  $0^\circ$  in azimuth and elevation. Then the asymmetrical windows of 1.5 ms containing the amplitudes of the early reflections at the respective



**Figure 3.14:** A spherical loudspeaker array of 84 loudspeakers positioned at the green dots.

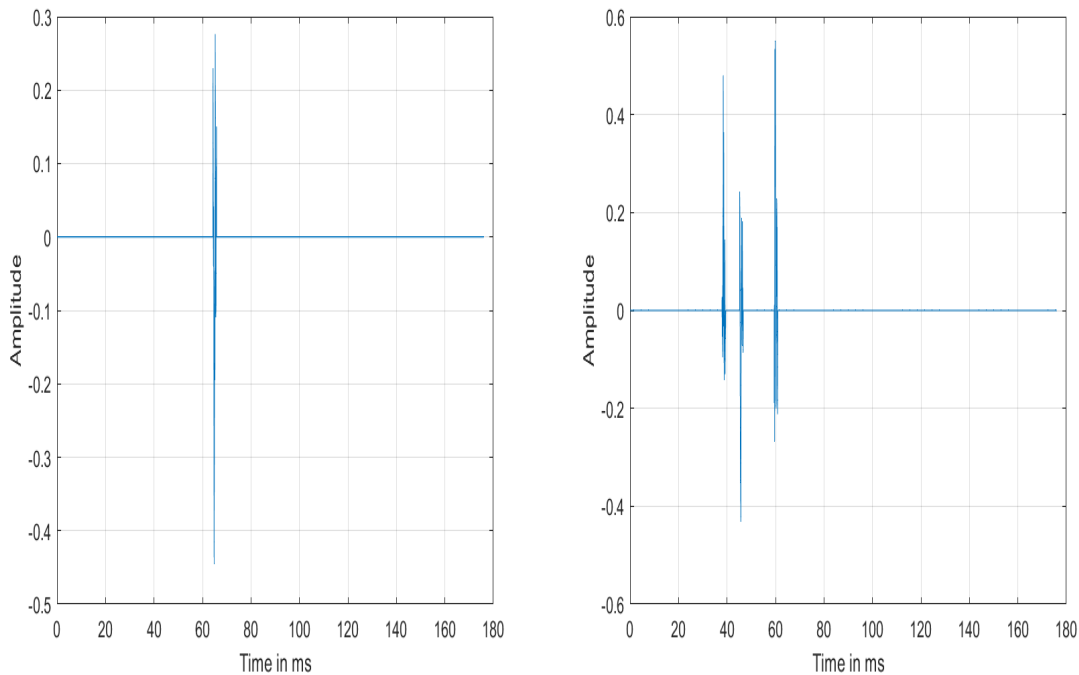
TOAs are distributed over the loudspeakers. The loudspeaker position in spherical coordinates that matches the DOA of a reflection the best is the loudspeaker that will play-back that reflection. Figure 3.15 shows the loudspeaker signals of two of the loudspeakers in the array of the Helsington church when all reflections has been distributed over the loudspeaker array. One early reflection is passed to one of the loudspeakers as its position matches the DOA of that reflection. For the other loudspeaker, the incidence angle of three reflections matches its position.

The time it takes for each loudspeaker signal to reach the listener (the TOA of the loudspeaker array) is determined by the radius of the loudspeaker array,  $r_{LA}$ , which is set to 10 meters,

$$TOA_{LA} = \frac{r_{LA}}{c} \approx 29.2 \text{ ms}, \quad (3.2)$$

where  $c = 343 \text{ m/s}$  is the speed of sound in air. Each loudspeaker signal will therefore be shifted according to this radius so that the time of arrival of the direct sound is the same for all tested RIR as the same loudspeaker array is used to reproduce all rooms.

Both the late reverberation and early diffuse sound are synthesized (and the description of how it is synthesized can be read in section 3.4 and 3.5) such that they consists each of a 84-channel of decorrelated signals that corresponds to each loudspeaker in the array. The method has now simulated a set of spectral components, early diffuse sound and late reverberation, represented by each loudspeaker. The listener can either physically be placed inside a loudspeaker array whose respective



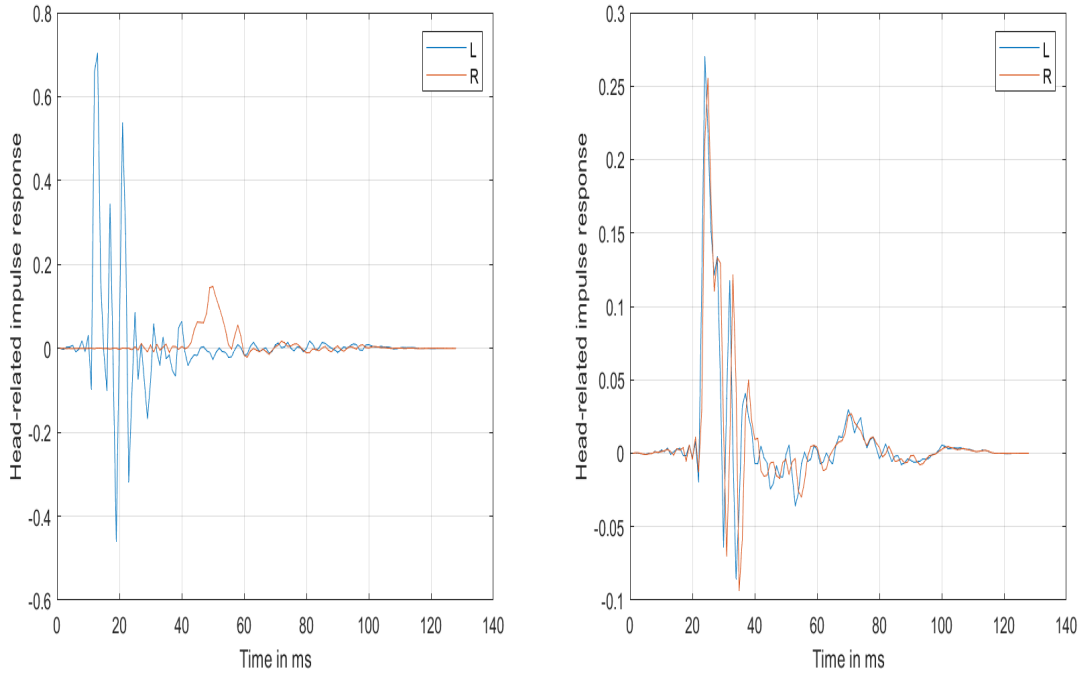
**Figure 3.15:** Early reflections passed to two different loudspeakers of the array. One loudspeaker is assigned one reflection (left) while another is assigned three reflections (right).

loudspeakers plays its assigned loudspeaker signal, or the loudspeaker array can be created virtually by using HRIRs corresponding to each loudspeaker position in the array for creation of virtual sources.

### 3.6.2 Virtual loudspeaker array

When simulating a virtual loudspeaker array it can be play-backed using headphones and the aim is therefore to synthesize a 2-channel BRIR that corresponds to the total contribution of all loudspeaker signals in the array.

In the Paraspax method, HRIRs are simulated as spherical harmonics coefficients at the spatial order of  $M \leq 35$  which corresponds to arbitrary head orientations of the listener. The HRIR set used is measured from a Neumann KU100 artificial head. This set of HRIRs can be used for creating virtual loudspeakers, where only HRIRs of the head orientations that corresponds to the loudspeaker positions relative the listener are used. The HRIRs are used to get the transmission of the loudspeaker signals from each loudspeaker to the listener. The HRIR for two different loudspeaker positions ( $90^\circ$  in azimuth to the left, and  $180^\circ$  in azimuth, right behind the listener) are shown in Figure 3.16. The elevation angle of the two HRIRs is at  $0^\circ$ . As seen in Figure 3.16, the sound coming from the loudspeaker positioned to the left of the listener will be perceived as louder by the left ear for almost the entire time range. The sound reaching the right ear is attenuated as the sound path is obstructed by the listener's head. However, for the loudspeaker placed at  $180^\circ$  right behind the listener, both ears will hear approximate the same amount of the signal



**Figure 3.16:** Measured HRIRs from a Neumann KU100 artificial head showing how sound is reaching the left and right ear when the sound source is positioned at  $90^\circ$  to the left (left) and at  $180^\circ$  right behind (right).

at the same time.

The loudspeaker signals containing the assigned spectral components, early diffuse sound and late reverberation is convolved with its respective HRIR and by doing this, each loudspeaker signal will have a left and right channel. As the desired output of the virtual loudspeaker array is the combined contribution from all loudspeakers, the loudspeaker signals are added to form a 2-channel loudspeaker array-based synthesis representing how the sound is perceived by the listener that is virtually placed inside the array. As the part of the loudspeaker signals that consists of the directional components only contains energy at the TOAs of the direct sound and early reflections, the contributions from all loudspeakers can easily be added. For the early diffuse sound and late reverberation, on the other hand, a further step is required when summing up the various loudspeaker signals.

The early diffuse sound that is played-back from each loudspeaker are incoherent with equal RMS. By adding two such signals, the sound increases by 3 dB. To counteract this, the sum of the early diffuse sound of each loudspeaker is divided by the square root of the number of loudspeakers used in the array,

$$\tilde{p}(t) = \frac{\sum_{n=1}^N \tilde{p}_n(t)}{\sqrt{N}}, \quad (3.3)$$

where  $\tilde{p}(t)$  is the RMS sound pressure amplitudes added from all loudspeakers and  $N$  denotes the number of loudspeakers. The same applies the late reverberation.

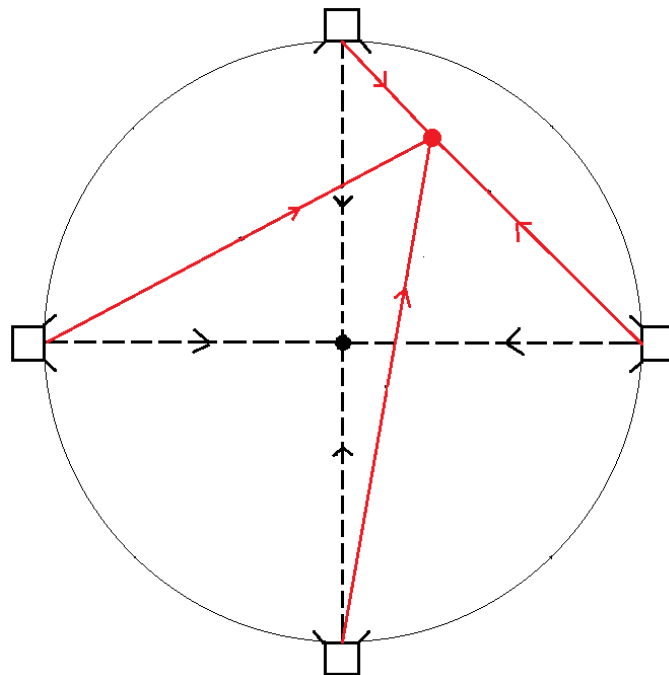
To move the listener inside the loudspeaker array, the listener can either physically move, or the loudspeaker signals in the virtual loudspeaker array can be mod-

ified according to the new loudspeaker positions relative the listener for each new listener position.

### 3.6.3 Change listener position

In the Paraspax method, only the direct sound and early reflections are modified as the listener moves the head or changes its position. The early diffuse sound and the late reverberation is kept constant during the audio reproduction, but the reproduction is still accurate due to changing the directional components in the method also results in a change in the DRR [1]. However, for a listener moving within a loudspeaker array, the distance to the different sources changes with each new listener position and therefore should the whole signal of each loudspeaker change.

The first step is to define the new loudspeaker positions in spherical coordinates relative the new listening position. Each loudspeaker signal still contains the same components, but the incident angle of the sound from each loudspeaker to the listener changes, where certain loudspeaker signals are strengthened in some areas while they are weakened in others. Figure 3.17 shows an example of how the sound rays of the loudspeaker signals reaches a listener, positioned at the new position, marked with a red dot, and how it differs from when the listener is positioned at the center, marked with a black dot. The loudspeaker array is illustrated as 4 loudspeakers in the horizontal plane, positioned at varying azimuth angles at the elevation angle of  $0^\circ$  for simplicity.



**Figure 3.17:** Sound travelling from the loudspeakers to a listener positioned at the center of the array (black dot) and to a listener at an arbitrary position (red dot) within a simplified loudspeaker array of four loudspeakers in the horizontal plane.

As seen in Figure 3.17, when the listener moves inside the array, the distance between the listener and every source changes and is not the same for all loudspeakers, as it was when the listener was positioned at the center. The TOA of each loudspeaker signal therefore has to be modified according to the new distance to the listener. The new distances for each loudspeaker relative the listener are calculated and the loudspeaker signals are shifted accordingly.

A variable  $d$  is defined as the difference between the distance at the sweet spot,  $d_{ss}$ , and the distance at the new position,  $d_{np}$ ,

$$d_n = d_{ss_n} - d_{np_n}, \quad \text{for } n = 1, 2, \dots, N. \quad (3.4)$$

Negative values of  $d_n$  denotes that the new position is closer to loudspeaker  $n$  than before, while greater distances gives positive values. The respective loudspeaker signals are amplified or reduced, depending on  $d_n$ . The Paraspax method uses the inverse-square law saying that the sound energy radiating from a point source decreases proportional to the square of the distance. This distance attenuation, however, is often too extreme for a loudspeaker which cannot really be equated to a point source. A factor for the distance attenuation that works in most cases for loudspeakers is the square root of the distance  $r_n$  between loudspeaker  $n$  and the new listening position, such that

$$\tilde{p}_n(t) \propto \frac{1}{\sqrt{r_n}}, \quad \text{if } d_n > 0, \quad (3.5)$$

$$\tilde{p}_n(t) \propto \sqrt{r_n}, \quad \text{if } d_n < 0, \quad (3.6)$$

where  $\tilde{p}_n(t)$  denotes the RMS sound pressure from loudspeaker  $n$  containing the directional as well as the diffuse sound.

For the virtual loudspeaker array, the loudspeaker signals are convolved with a new set of HRIRs that corresponds to the loudspeaker positions relative the new listener position, as seen in Figure 3.17.

# 4

## Results

The method is tested for the 15 selected monaural RIRs that can be found in [2] and [3]. The aim is to reduce the number of loudspeakers used in the loudspeaker array while maintaining the sound quality. The number of loudspeakers is determined when the listener is positioned at the center of the array. Using the resulting number of loudspeakers, the listener will move within the loudspeaker array and it will be examined how far from the center the listener can move without changing the sound image. Then the influence of different parameters of the loudspeaker array is investigated and it will be examined if these parameters also changes how far the listener can move from the center.

The results are obtained from the virtual loudspeaker array such that the listener is virtually placed inside the loudspeaker array. The resulting loudspeaker array synthesis that represents the total contribution from all loudspeakers is convolved with an anechoic drums audio file. The resulting audio is analyzed by listening to the auralization and conclusions are drawn from the listening along with analysis of corresponding plots.

### 4.1 At the center of the loudspeaker array

When the listener is positioned at the center of the array, the aim is to reduce the number of loudspeakers while maintaining the sound quality. It could be both costly and time consuming to assemble a loudspeaker array containing many loudspeakers. The sound field created by the loudspeaker array of 84 loudspeakers presented in Figure 3.14 will be used as a reference and compared with two simplified versions of a loudspeaker array. The number of loudspeakers in the simplified versions is reduced until the sound quality differs from the sound quality created by the 84-loudspeaker array. The two simplified loudspeaker arrays that will be tested here contains loudspeakers whose positions only varies in azimuth or elevation, unlike the 84-loudspeaker array whose loudspeaker positions have different azimuth and elevation angles. The simplified loudspeaker array having loudspeaker positions with varied azimuth angle is in the horizontal plane and the loudspeakers are positioned at ear level of the listener on the circle at the fourth row in the 84-loudspeaker array in Figure 3.14 when the elevation angle is  $0^\circ$ . In the other simplified version, the loudspeakers are positioned right in front of the listener at the azimuth angle of  $0^\circ$ , and at different elevation angles such that the loudspeaker array is in the vertical plane.

The auralization from the 84-loudspeaker array compared with the auralization

## 4. Results

---

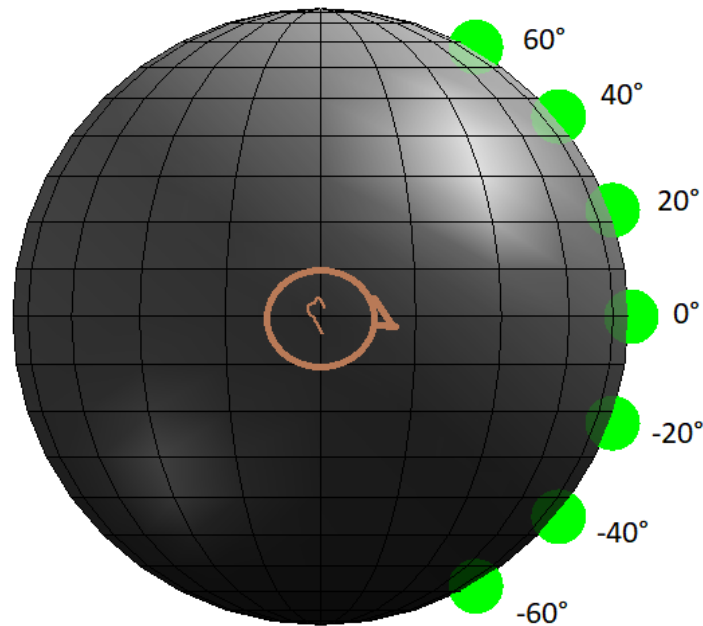
from the simplified loudspeaker arrays resulted in that the quality of the sound field is still maintained for the number of loudspeakers that are presented in Table 4.1 for all of the 15 tested environments. For the environments where the number of loudspeakers are not specified in Table 4.1, the simplified loudspeaker arrays could not be used.

Name	Number of loudspeakers (varying azimuth)	Number of loudspeakers (varying elevation)
Genesis 6 studio	3	-
Trollers gill	-	-
Maes howe	3	-
Arthur sykes rymer auditorium	-	-
Koli national park	-	-
Stairway	3	-
Hoffman lime kiln	3	-
Central hall	-	-
Helsington church	3	-
Promenadikeskus concert hall	3	-
Innocent railway tunnel	3	-
Falkland palace royal tennis court	3	-
Shrine and parish church of all saints	3	-
Hamilton mausoleum	3	-
Terrys factory warehouse	4	-

**Table 4.1:** Numbers of loudspeakers required in the loudspeaker array for different impulse responses.

For the loudspeaker array in the vertical plane, up to seven loudspeakers was used at different elevation angles, but this loudspeaker array setup could not achieve binaural sound at all. The 7-loudspeaker array setup is presented in Figure 4.1 from the side, where the listener is placed at the origin of the sphere and the mid-loudspeaker at  $0^\circ$  in elevation is at the listener's ear level. The loudspeakers are positioned from  $60^\circ$  above the listener to  $-60^\circ$  below the listener with a  $20^\circ$  resolution. A lower amount of loudspeakers in the loudspeaker array in the vertical plane looks as in Figure 4.1, but with greater spacing between the loudspeakers due to the higher resolution. The reason why the simplified version of the loudspeaker array in the vertical plane cannot be used can be seen from Figure 4.1. The loudspeakers are positioned right in front of the listener and each loudspeaker signal will therefore reach the left and right ear of the listener equally and create a mono sound. In order for the loudspeaker array with varying elevation to create binaural sound, there should be variations in the azimuth as well so that the sound reaches the listener from behind and the sides as well.

As seen in Table 4.1, three loudspeakers are enough in 10 cases for the simplified loudspeaker array in the horizontal plane. The corresponding 3-loudspeaker array

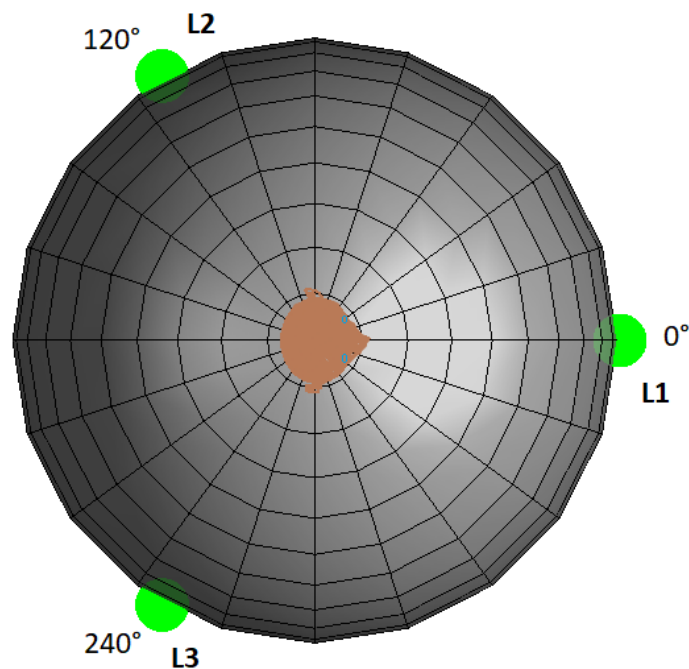


**Figure 4.1:** A loudspeaker array in the vertical plane containing 7 loudspeakers positioned at the green dots right in front of the listener.

is presented from above in Figure 4.2, where the loudspeakers are represented by green dots placed at an equal azimuth angle between each other at  $0^\circ$ ,  $120^\circ$  and  $240^\circ$ . Henceforth these loudspeakers will be called "L1", "L2" and "L3", respectively. The listener is positioned at the origin of the circle, facing the rightmost loudspeaker of azimuth angle at  $0^\circ$ .

For all the environments that required three loudspeakers in the loudspeaker array in the horizontal plane, presented in Table 4.1, the 3-loudspeaker array created a sound field which can be equated to the sound field created by the 84-loudspeaker array. The sound field sounds spacious and dynamic, and it sounds like the sound from the different drums in the audio are coming from different directions. It also sounds wide in comparison to when a 2-loudspeaker array is used. Then the sound quality drops drastically. The loudspeakers are then positioned right in front and behind the listener at  $0^\circ$  and  $180^\circ$  so that the spaciousness in the sound decreases and it sounds more flat and narrow. It is also harder to hear from which direction the sound from the different drums come from. The position of the loudspeakers results in that the loudspeaker signals reaches the listener equally and therefore it sounds more monaural than binaural.

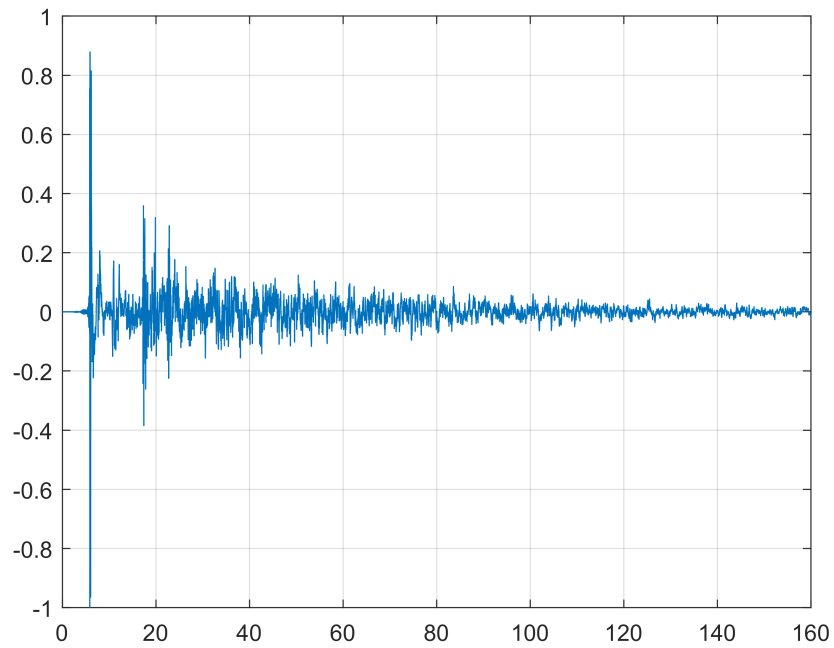
The Terrys factory warehouse is the only environment where it is preferred to use a 4-loudspeaker array in the horizontal plane over a 3-loudspeaker array. The loudspeakers are then placed around the listener with equal azimuth angle between them as in Figure 4.2, but with the loudspeaker positions at  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$  and  $270^\circ$  in azimuth. However, the reflections are distributed only over the three loudspeakers positioned at  $0^\circ$ ,  $180^\circ$  and  $270^\circ$ , while the loudspeaker signal at  $90^\circ$  only contains diffuse sound. When the 3-loudspeaker array is used, only the two loudspeakers at



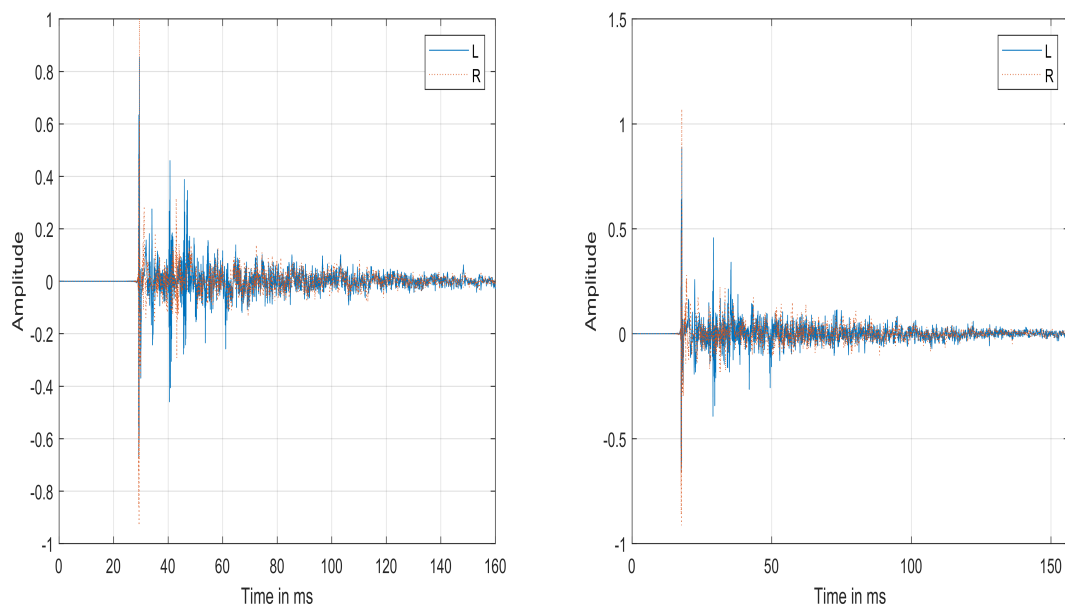
**Figure 4.2:** A loudspeaker array in the horizontal plane containing 3 loudspeakers positioned at the green dots around the listener.

$0^\circ$  and  $120^\circ$  are assigned reflections. What distinguishes the 3-loudspeaker array from the 4-loudspeaker array is that the 3-loudspeaker array creates a more flat and mono sound than the 4-loudspeaker array that sounds richer and more binaural. In addition, the 3-loudspeaker array creates some disturbing echo which could not be heard in the loudspeaker array of 84 and 4 loudspeakers. Moreover, the DOAs of the reflections are more in line with the loudspeaker positions of the 4-loudspeaker array compared to when three loudspeakers are used.

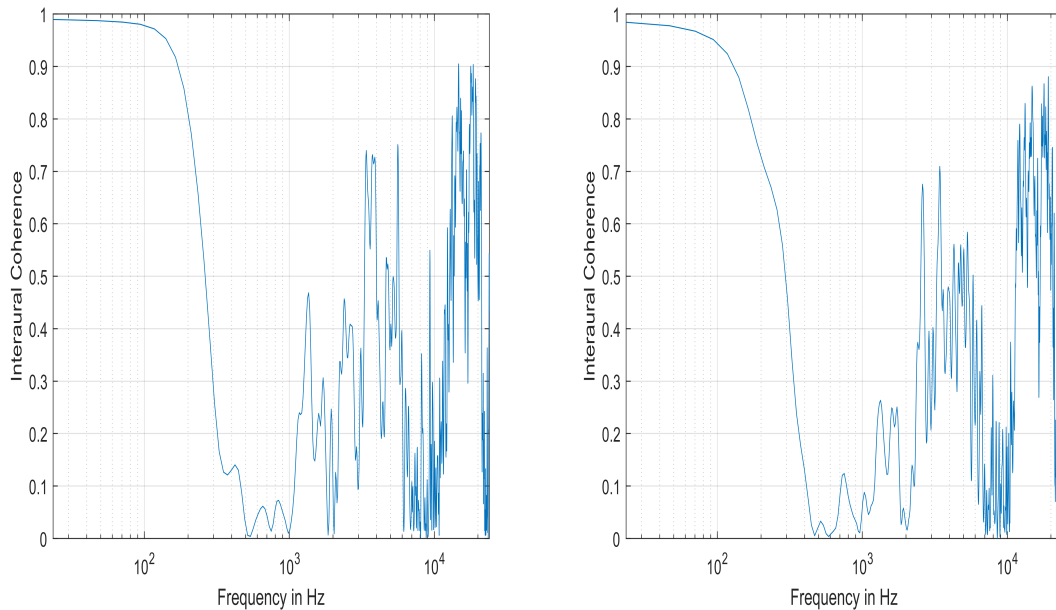
The measured monaural RIR of Maes howe is shown in Figure 4.3 and the synthesis created from the 3-loudspeaker array in the horizontal plane of Maes Howe is shown in Figure 4.4 together with the synthesis from the 84-loudspeaker array for comparison. In these plots, it can be seen that the method successfully recreated the structure of the monaural RIR, and that a reduced number of loudspeakers gives rise to similar plots as for the 84-loudspeaker array. In the 3-loudspeaker array the direct sound is played from loudspeaker L1. The DOAs in azimuth of the 10 early reflections are approximated to the angles of the loudspeaker positions at  $120^\circ$  and  $240^\circ$  such that seven reflections are played-back from loudspeaker L2 and three reflections are played-back from loudspeaker L3. This can be seen in the plot, which is dominated at the left ear because L2 is positioned on the left hand side of the listener. When using an array of 4 loudspeakers instead of 3 for this particular environment, the azimuth of the early reflections are more in line with the azimuth of the loudspeaker positions. The synthesis, both from a 4-loudspeaker array as well as a 3-loudspeaker array, sounds similar as to the synthesis of the loudspeaker array of 84 loudspeakers. However, since the goal is to reduce the number of loudspeakers



**Figure 4.3:** Measured monaural room impulse response of Maes Howe.



**Figure 4.4:** Loudspeaker array synthesis of Maes Howe from a loudspeaker array in the horizontal plane containing 3 loudspeakers (left) and from the 84-loudspeaker array (right).



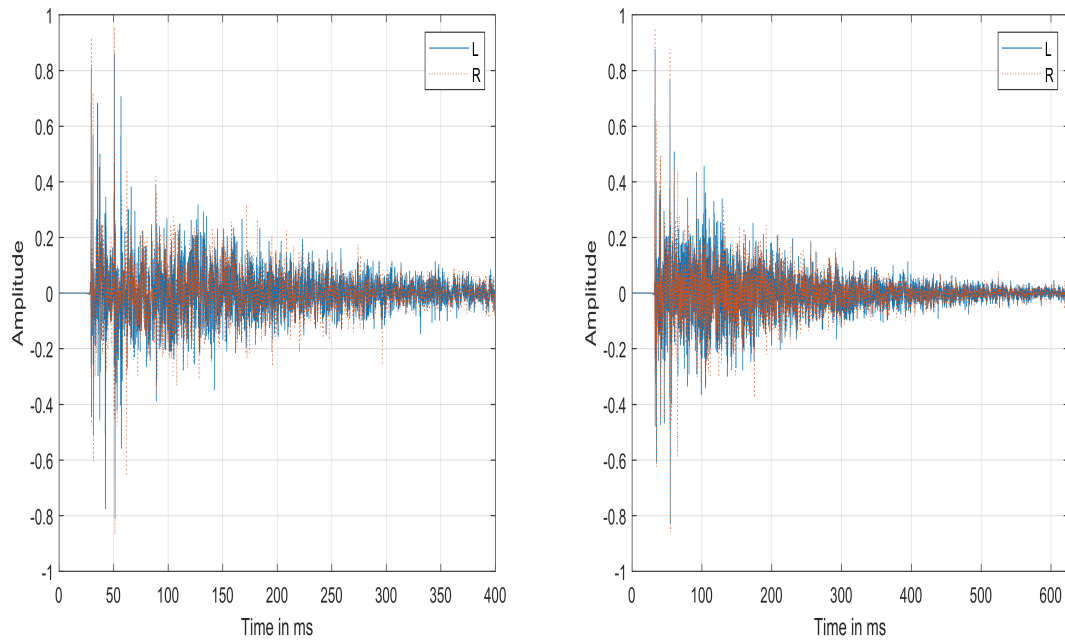
**Figure 4.5:** Interaural coherence of Maes howe for a loudspeaker array in the horizontal plane containing 3 loudspeakers (left) and for the 84-loudspeaker array (right).

in the array, the loudspeaker array containing three loudspeakers is the one to be chosen at the end.

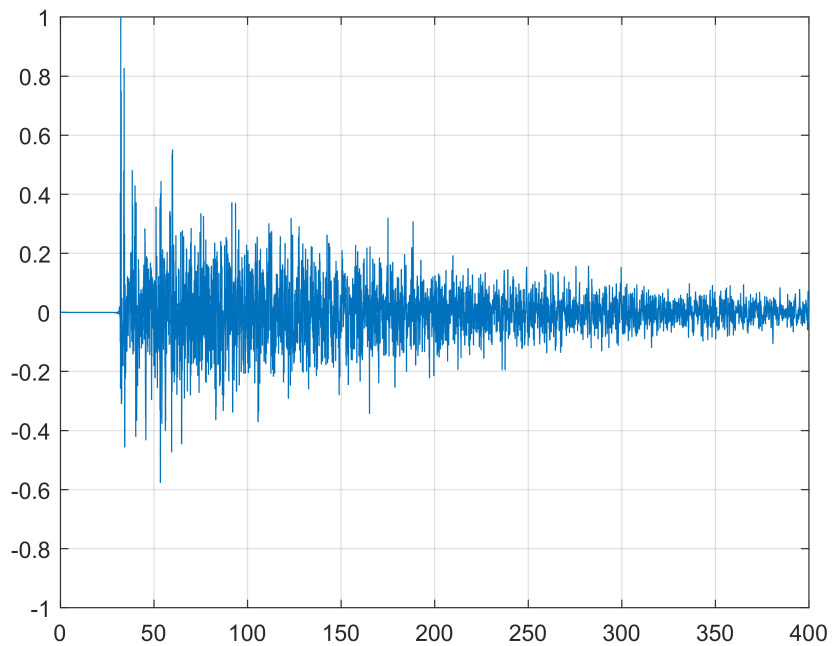
The frequency-dependent interaural coherence (IC) can be seen in Figure 4.5 for the 3- and 84-loudspeaker array. The IC of the 3-loudspeaker array shows values close to 1 for frequencies up to 100 Hz before it decreases for higher frequencies. The IC then continues up and down and is higher than 0.5 at frequencies around 2 500 Hz, between 3 200 Hz and 3 900 Hz and from 5 200 Hz to 5 600 Hz as well as above 12 700 Hz, indicating that the signals in the synthesis are more correlated than decorrelated at these frequencies. The IC of the 84-loudspeaker array shows that the signals are decorrelated to a higher degree in general compared to when a reduced number of loudspeakers are used, and this is especially applied to the mid-frequency region.

For the Helsington church, three and one reflections are assigned to loudspeaker L2 and L3, respectively, while the direct sound and six reflections are played-back from loudspeaker L1. The synthesis from the 3- and 84-loudspeaker array are shown in Figure 4.6, which is dominated at the left ear because most reflections are assigned to L2. However, it is not as left ear dominated as the synthesis of Maes howe and the reason is that most reflections are played-back from the loudspeaker in front of the listener. Its monaural RIR is presented in Figure 4.7, where it can be seen that the space contributes to more and denser reflections than the synthesis that only contains 10 early reflections. However, by comparing the synthesis with the measured RIR, it is clearly audible that it is the same space.

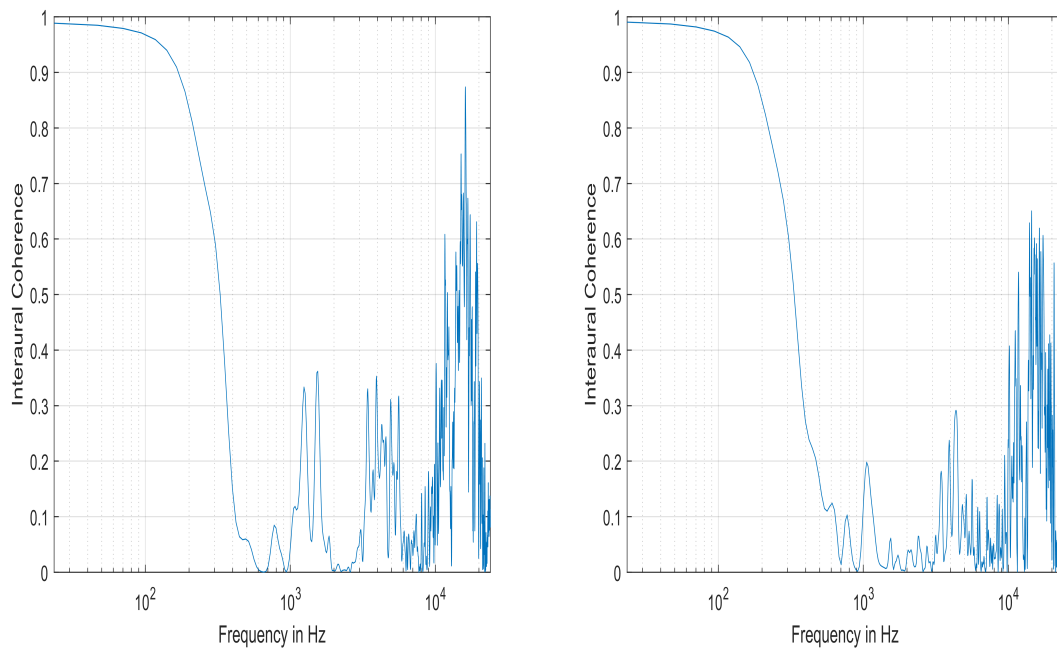
Its IC is presented in Figure 4.8 together with the IC from the 84-loudspeaker array, which shows that the loudspeaker signals have incoherent directional compo-



**Figure 4.6:** Loudspeaker array synthesis of the Helsington church for a loudspeaker array only changing in the horizontal plane using 3 loudspeakers (left) and its monaural RIR (right).



**Figure 4.7:** Measured monaural room impulse response of the Helsington church.



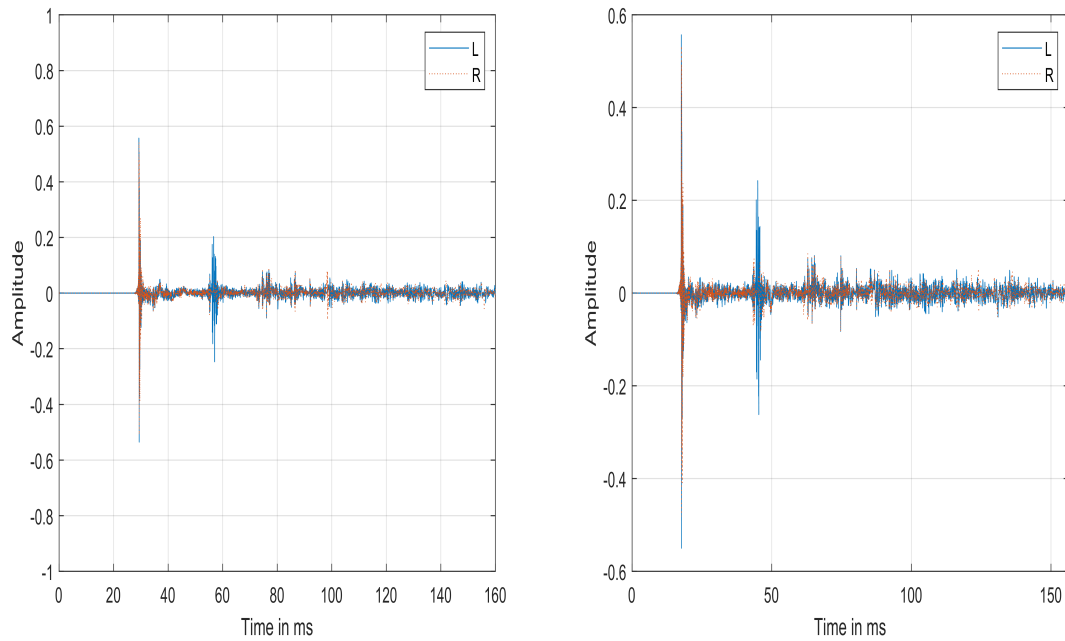
**Figure 4.8:** Interaural coherence of the Helsington church for a loudspeaker array in the horizontal plane containing 3 loudspeakers (left) and for the 84-loudspeaker array (right).

nents in the frequency range between 300 Hz and 14 500 Hz. It can also be seen from the plots in Figure 4.8 that the signals get more coherent when the number of loudspeakers in the array is reduced.

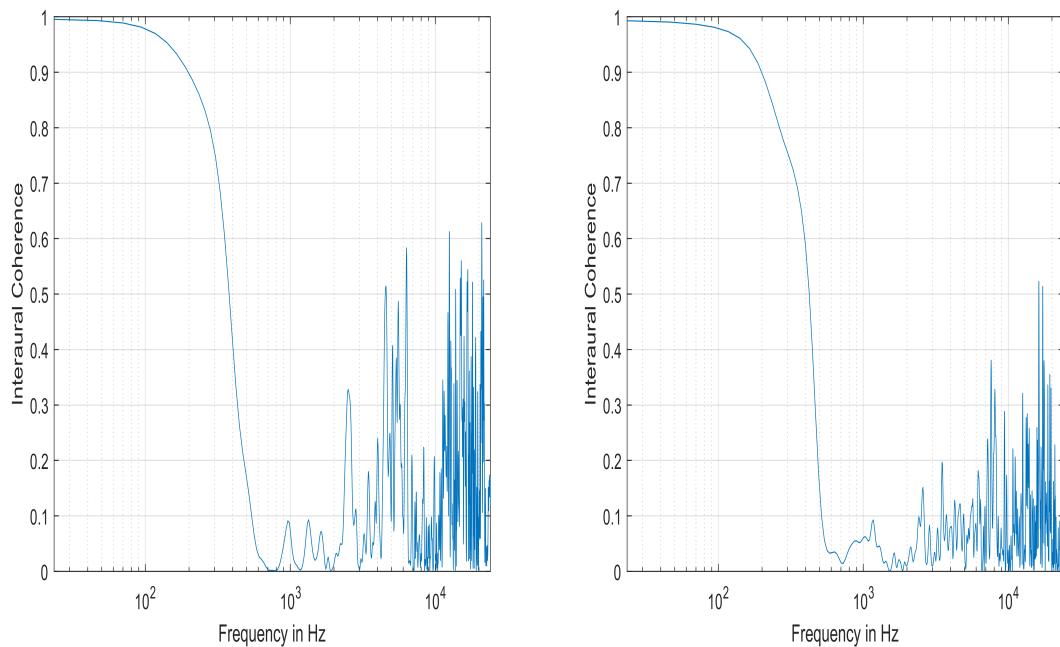
The result of having 3 loudspeakers in the array in the horizontal plane is clear for the Genesis 6 studio, Maes howe, Stairway, Hoffman lime kiln, the Helsington church, the Innocent railway tunnel, Falkland palace royal tennis court and Hamilton Mausoleum, which all have at least four reflections distributed over loudspeaker L2 and L3, and the rest of the reflections are assigned to loudspeaker L1. The Promenadikeskus concert hall and the Shrine and parish church of all saints north street have only 3 reflections assigned to both L2 and L3 and here the result is not that obvious. The synthesis of the Promenadikeskus concert hall from the 3- and 84-loudspeaker array are presented in Figure 4.9, where no clear differences can be pointed out from the plots but they differ more from each other by listening to the results. From the 3-loudspeaker array it is harder to here the spaciousness since the sound arrives mostly from the loudspeaker at the front and dominates the total output while the other two loudspeakers that are necessary for creating binaural sound contributes to a lower sound as only three reflections are distributed over these loudspeakers.

The corresponding IC can be seen in Figure 4.10, showing incoherent directional events in the frequency range from 400 Hz to approximate 10 000 Hz. The signals are more correlated in the mid frequencies in the 3-loudspeaker array compared to the 84-loudspeaker array, but they are still more correlated than decorrelated.

The syntheses in Figure 4.4-4.9 shows that in overall, the method performs well.



**Figure 4.9:** Loudspeaker array synthesis of the Promenadikeskus concert hall for a loudspeaker array in the horizontal plane containing 3 loudspeakers (left) and the 84-loudspeaker array (right).



**Figure 4.10:** Interaural coherence of the Promenadikeskus concert hall for a loudspeaker array in the horizontal plane containing 3 loudspeakers (left) and for the 84-loudspeaker array (right).

The TOAs of the direct sound and early reflections in the loudspeaker array syntheses are in line with the TOAs in the monaural RIRs, apart from that the syntheses are shifted according to the radius of the loudspeaker array. However, the amplitude of the direct sound and early reflections in the syntheses are reduced in comparison to the monaural RIR for some environments, while it is louder in the syntheses for some other environments. Even though a maximum of ten early reflections are found in the method, the RIRs retain their shape when they are recreated for loudspeaker array synthesis, and the resulting binaural audio from the convolution process recreates what it would have sounded like to be in the original location. Depending on how many reflections that are assigned to each loudspeaker in the array, the syntheses can be left or right ear dominated, which is the case for all tested RIRs as most reflections tend to come from one side of the room.

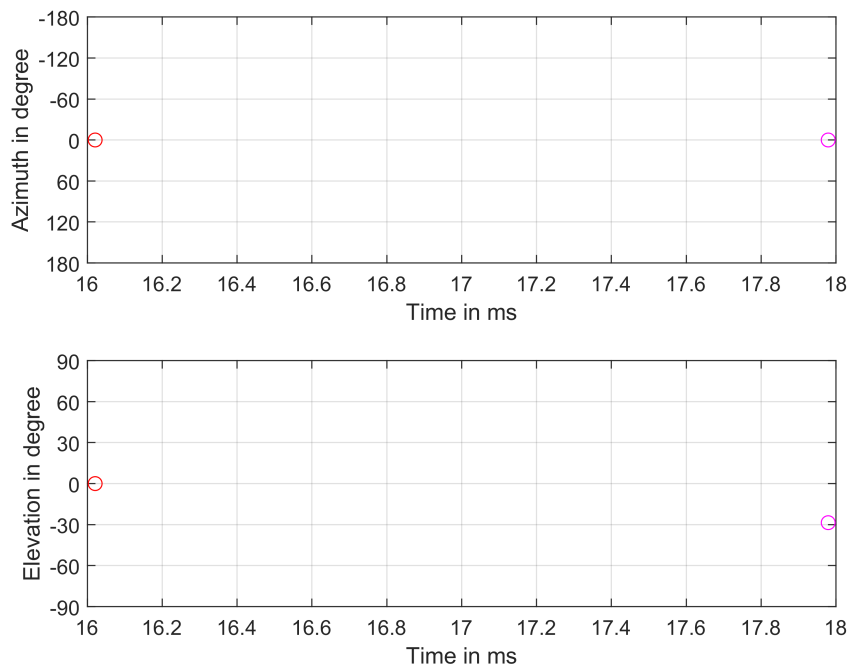
The interaural coherence of the tested environments shows that the signals at the left and right ears are coherent for low frequencies below around 300 Hz and for high frequencies above 10 kHz. In the frequency range between these frequencies the signals are incoherent. For some of the tested RIRs, a reduced number of loudspeakers in the array contributes to the signals getting more coherent than the signals of the 84-loudspeaker array. Also, for those spaces that are not rectangular in shape, but the geometry has been approximated to a shoe-box room, the interaural coherence is overall higher in the whole frequency range, compared to the spaces where the true dimensions could be used.

The DOAs of the direct sound and early reflections of the RIRs that did not perform well on the loudspeaker array are presented in Figure 4.11-4.13, while the DOAs of the Arthur sykes rymer auditorium was already presented in Figure 3.7 in the Methods chapter. What these environments have in common is that the DOAs of the early reflections have little or no variation in azimuth. Therefore, a loudspeaker array that only varies in azimuth will not create binaural sound as it requires some reflections to reach the listener from behind. However, the early reflections of these environments varies in elevation but binaural sound can not be created for a loudspeaker array that varies only in elevation due to the positions of the loudspeakers, which are placed in front of the listener.

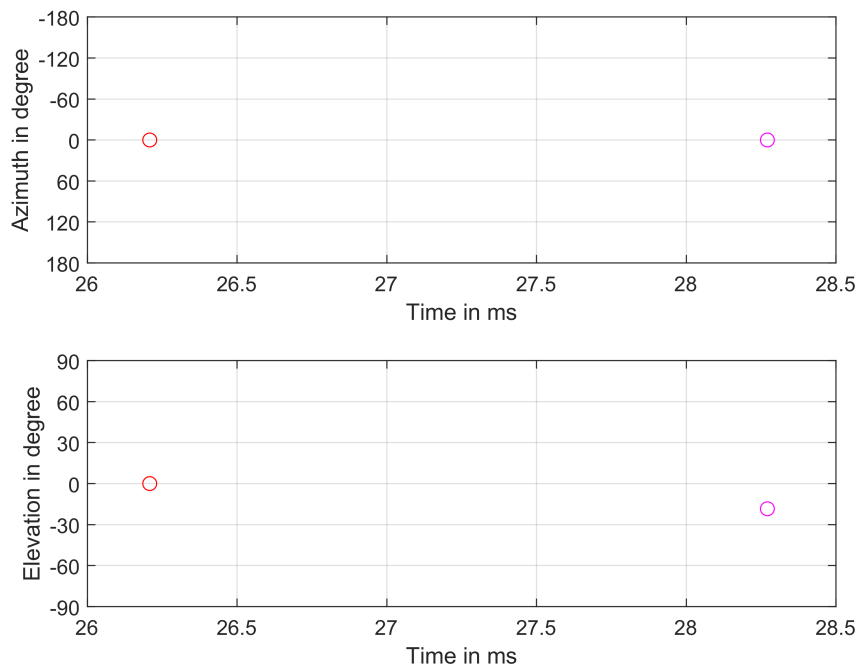
## 4.2 Investigation of the sweet spot

The investigation of the sweet spot described in this section assumes a loudspeaker array in the horizontal plane containing 3 loudspeakers, which is the required number of loudspeakers for most of the environments tested. The sound field created by the loudspeaker array is synthesized every 0.5 meter when the listener moves forward, backwards and straight to the sides as well as diagonally forward and backwards, both to the left and right. The different directions of the listener's movement in the 3-loudspeaker array can be seen in Figure 4.14. The environments in which the method could not reproduce any binaural sound using the loudspeaker array will not be included in this investigation. Also, different parameters of the loudspeaker array will be changed in order to examine its influence on how the sound image changes at different positions.

As the listener moves inside the loudspeaker array, the distance between the



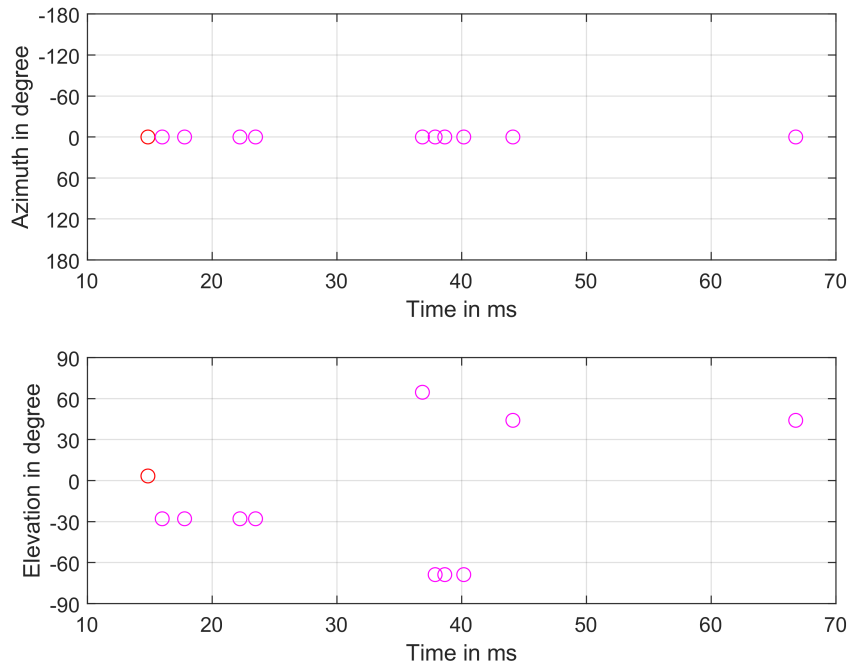
**Figure 4.11:** DOAs of the direct sound and early reflection of Trollers gill.



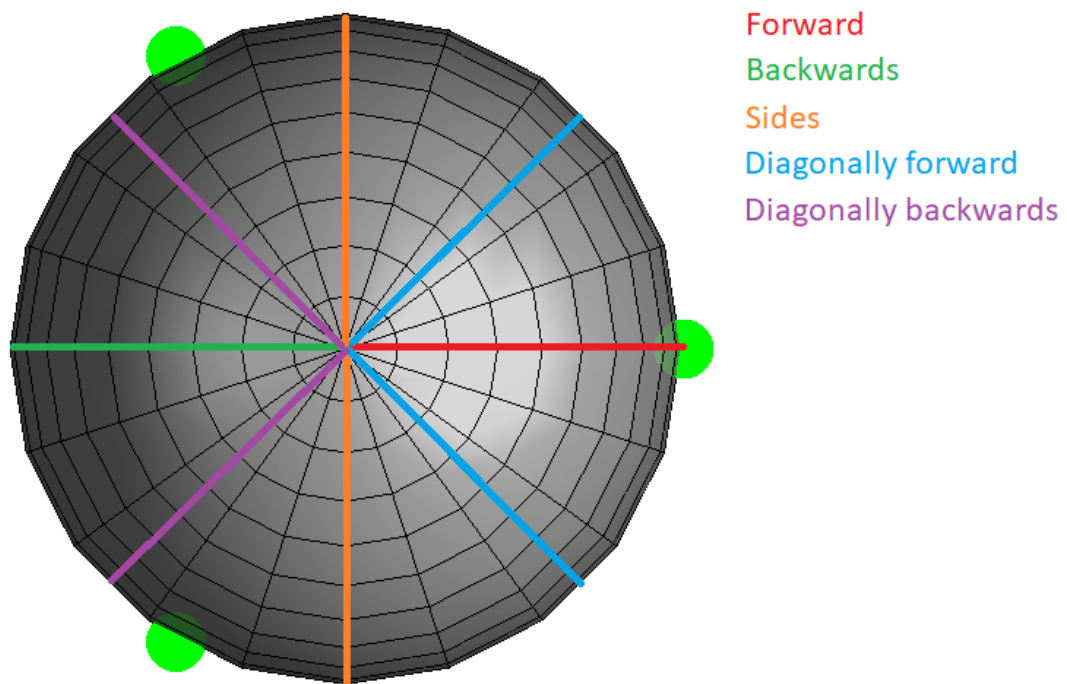
**Figure 4.12:** DOAs of the direct sound and early reflection of the Koli national park.

## 4. Results

---



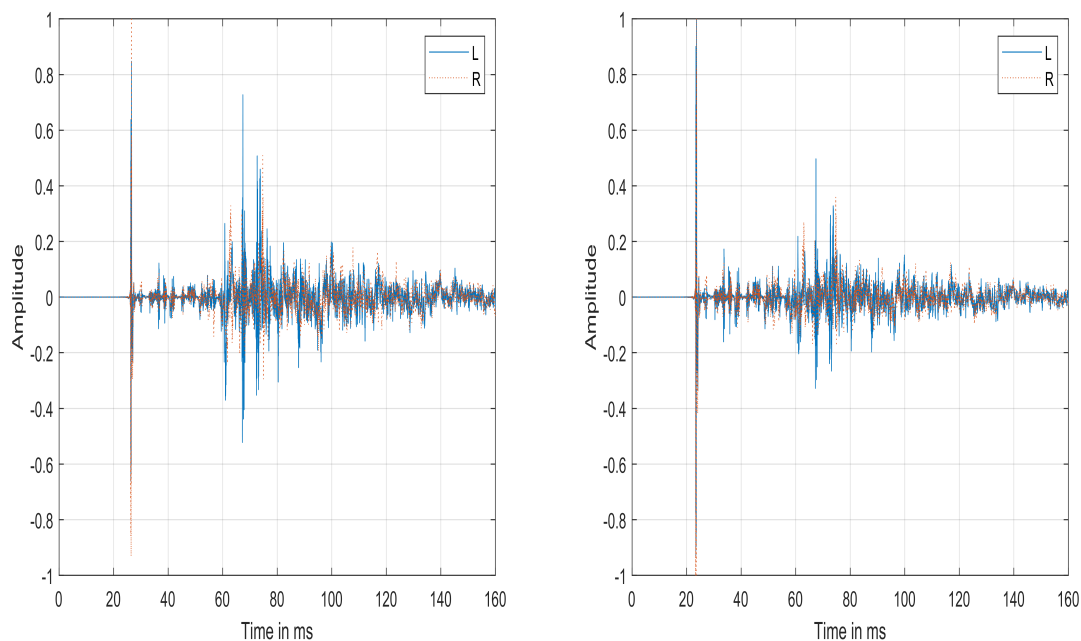
**Figure 4.13:** DOAs of the direct sound and early reflection of the Central Hall.



**Figure 4.14:** Directions of the listener's movements in the 3-loudspeaker array.

listener and each loudspeaker will change. A reduced distance to a particular loudspeaker makes this loudspeaker signal dominating the total output from all loudspeakers as the TOA of that loudspeaker signal decreases as well as the amplitude increases. For each listener position in the loudspeaker array, the angle between listener and loudspeaker changes and thus new HRIRs are created for each and every step. At certain positions, the signal from a particular loudspeaker reaches the listener with a lower level due to how it is angled towards the listener.

By moving straight forward towards loudspeaker 1, the directional components assigned to this loudspeaker are amplified and as the TOA of these directional components decreases, they get separated from the rest of the directional components. This can be seen in Figure 4.15, where the loudspeaker array synthesis of Maes howe is presented for the listener positioned 1 and 2 meters in front of the center, respectively.

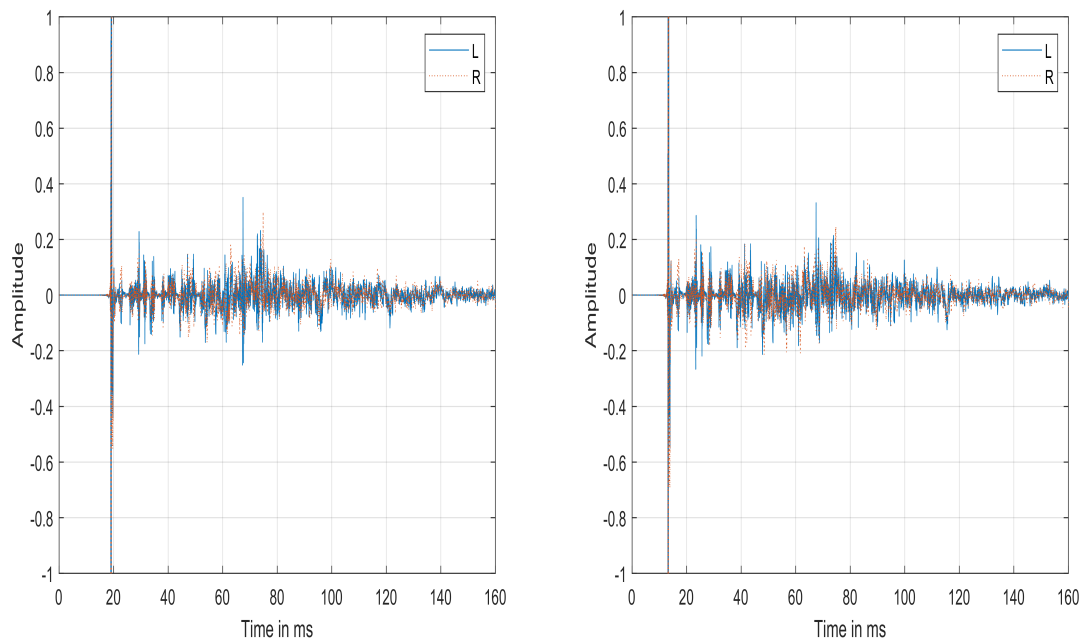


**Figure 4.15:** Loudspeaker array synthesis of Maes howe for a 3-loudspeaker array in the horizontal plane at the listener position 1 meter in front of the center (left) and 2 meters in front of the center (right).

Only the direct sound is played back from loudspeaker 1 and all early reflections are assigned to loudspeakers 2 and 3. As seen in Figure 4.15, the direct sound gets more and more amplified as well as more separated from the rest of the reflections as the listener gets closer to loudspeaker 1. At the same time, the amplitude of the reflections and diffuse sound played back from the loudspeakers behind the listener decreases and they arrive to the listener later in time. The quality of the sound image deteriorates already 0.5 meters from the center as it now sounds more flat and not as rich as it does at the center, and the sound color does not sound as wide as before. It continues to sound worse until the listener is 3.5 meters in front of the center when it starts to sound as it does at the sweet spot again. The syntheses at

## 4. Results

---



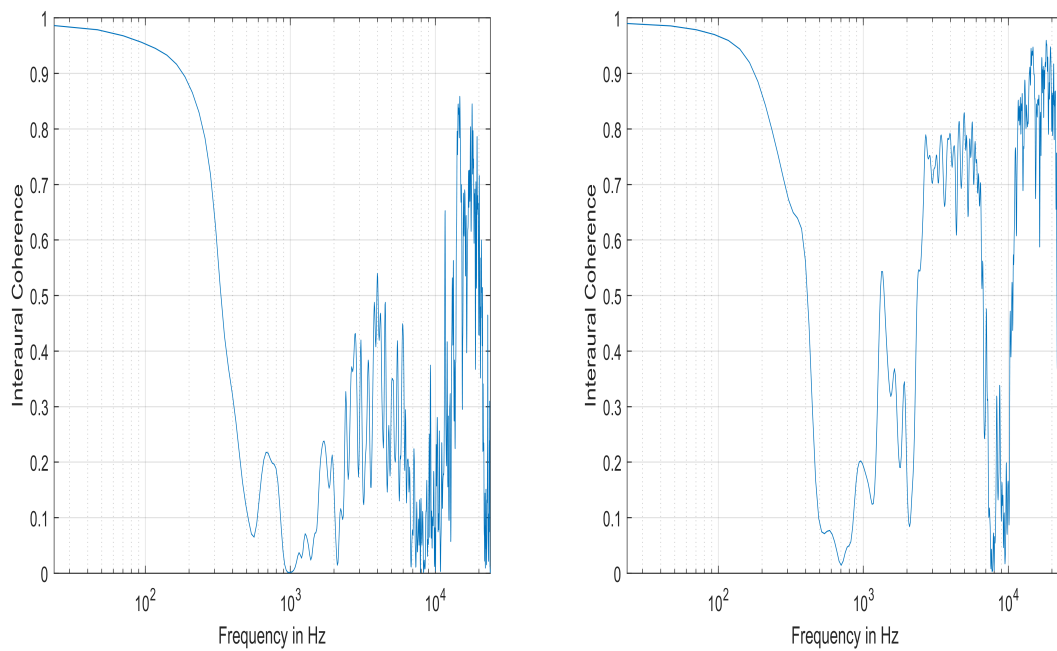
**Figure 4.16:** Loudspeaker array synthesis of Maes howe for a 3-loudspeaker array in the horizontal plane at the listener position 3.5 meter in front of the center (left) and 5.5 meters in front of the center (right).

3.5 and 5.5 meters in front of the center is shown in Figure 4.16, where it can be seen that the reflections gets more and more attenuated until they are a part of the diffuse sound at 5.5 meters from the center.

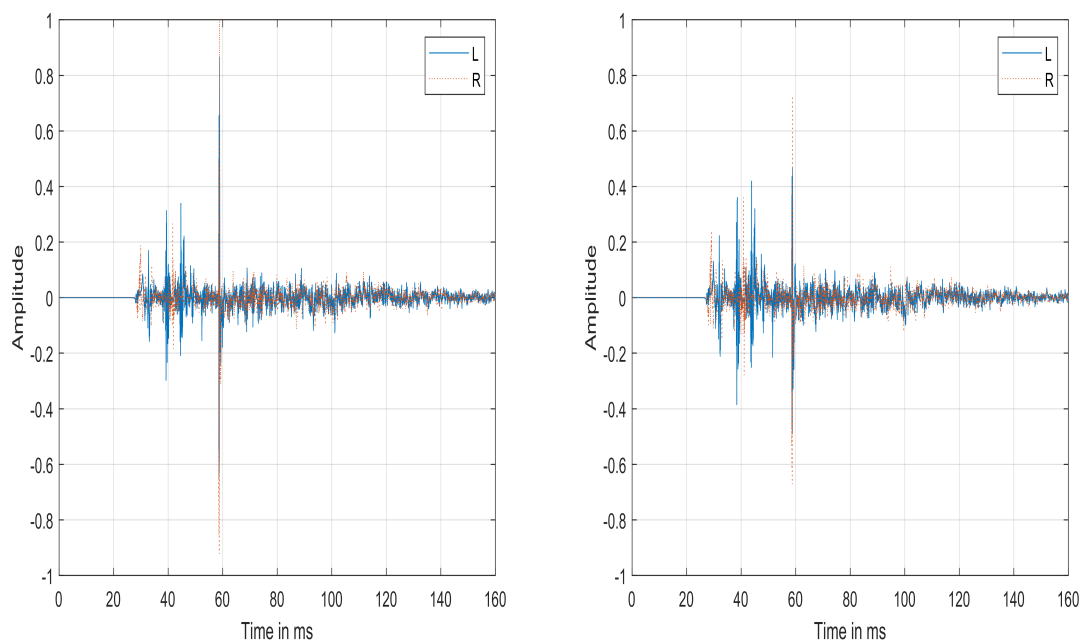
The IC of Maes howe when the listener has moved 1 m respective 3.5 m in front of the center is presented in Figure 4.17, where it can be seen that the left and right signals becomes more coherent the further from the center the listener moves.

The opposite happens when the listener instead moves backwards from the center. The direct sound reaches the listener after the reflections do and the further from the center the listener is, the higher the reflections become in amplitude while the direct sound is attenuated. This is illustrated in Figure 4.18, where it can be seen that at 2 meters from the center, the direct sound is already attenuated a lot. Furthermore, amplitude of the reflections increases for listening positions up to and including 3.5 meters from the center before they starts to attenuates. This is due to that the distance between the listener and loudspeaker 1 and 2 starts to increase again for listening positions further than 3.5 meters backwards from the center. The reflections contribute to a richer sound compared to how it sounded when the listener moved forward, but the direct sound do not sound as clear as it did when the listener was positioned at the center and this affects the sound quality negative.

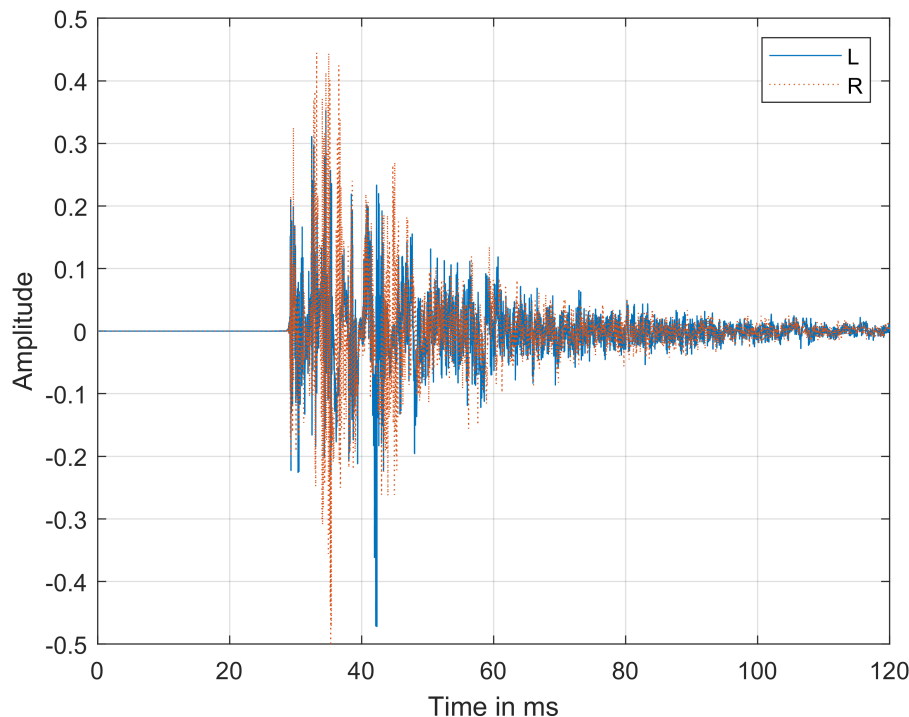
For the Genesis 6 studio, six early reflections are distributed over loudspeaker 2 and 3, where only one of them are played-back from loudspeaker 2. The other 4 reflections and the direct sound are assigned to loudspeaker 1. This creates a loudspeaker array synthesis which is dominated at the right side, as seen in Figure 4.19. By moving straight aside to the left, loudspeaker 2 will dominate the total



**Figure 4.17:** Interaural coherence of Maes howe for a 3-loudspeaker array in the horizontal plane at the listener position 1 meter in front of the center (left) and 3.5 meters in front of the center (right).



**Figure 4.18:** Loudspeaker array synthesis of Maes howe for a 3-loudspeaker array in the horizontal plane at the listener position 1 meter behind the center (left) and 2 meters behind the center (right).

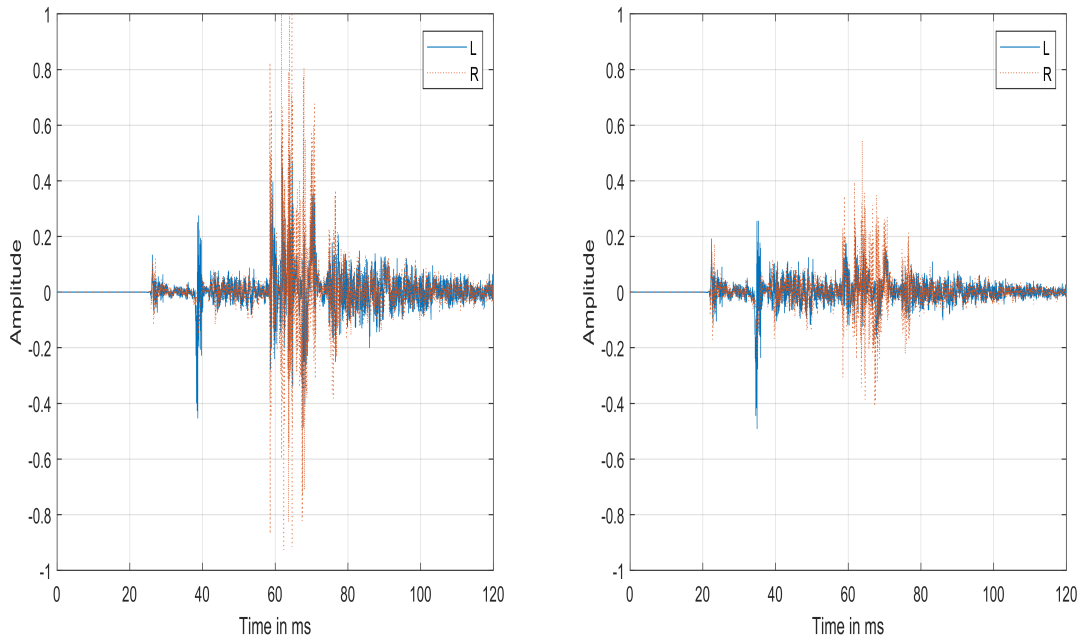


**Figure 4.19:** Loudspeaker array synthesis of the Genesis 6 studio for a 3-loudspeaker array in the horizontal plane at the sweet spot.

output from all loudspeakers and as this loudspeaker are closer to the left ear, the signal will be amplified at the left ear. But both loudspeaker 1 and 3 will dominate the right ear because the listener’s head will shadow these loudspeaker signals at the left ear. This will in turn amplify the sound on this side of the listener. This can be seen in Figure 4.20, where the one reflection played-back from loudspeaker 2 is only heard by the left ear and appears earlier in time than the direct sound and rest of the reflections. Moreover, the whole loudspeaker array synthesis is dominated at the right ear as most reflections reaches the listener from this side. The amplitude of the loudspeaker signals of loudspeaker 1 and 3 attenuates more and more the further to the left the listener is, while it increases for loudspeaker 2. The sound quality is good as far as 1.5 meters from the center then it drastically deteriorates the further from the center the listener moves. The separation in time of the reflections contributes to a more flat sound and it sounds like the previous wide sound color has tapered off, but the spaciousness, though, can still be heard.

The opposite applies when moving to the right, where loudspeaker 3 will be the dominate loudspeaker. This is illustrated in Figure 4.21. It sounds just as when moving to the left, but the listener can move further to the right compared to the left, before the quality goes down. The HRIR from loudspeaker 3 to the listener at the new listener positions 1.5 m as well as 3.5 m aside to the right is presented in Figure 4.22, where it can be seen how the listeners head, upper torso and pinna filters the loudspeaker signal at the two ears.

When the listener moves diagonally forward, both to the left and right, loud-



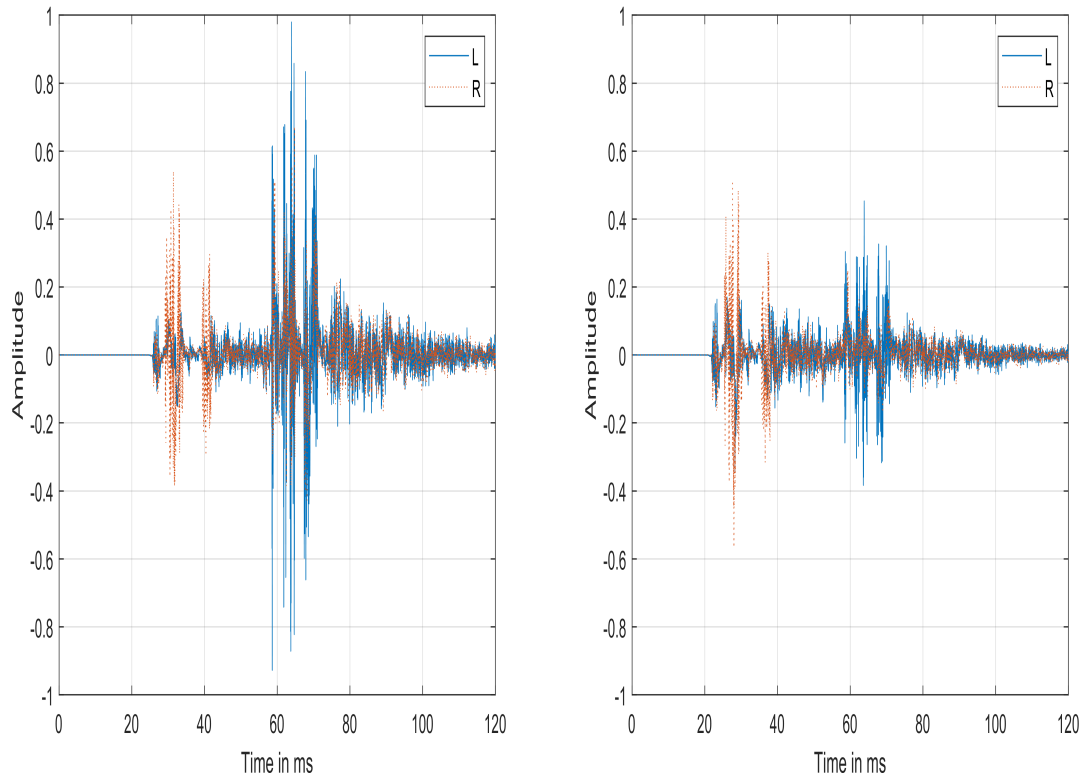
**Figure 4.20:** Loudspeaker array synthesis of the Genesis 6 studio for a 3-loudspeaker array in the horizontal plane at the listener position 1.5 meters aside to the left of the center (left) and 3.5 meters aside to the left (right).

speaker 1 dominates the total output from all loudspeakers and reaches the listener before the other loudspeaker signals do. The listener can move further from the center in the loudspeaker array of Maes howe, in comparison to the Promenadikeskus concert hall, which only has two and one reflection assigned to loudspeaker 2 and 3, respectively, where the quality deteriorates already by moving 7 cm diagonally forward. The quality deteriorates because the sound sounds flat and the spaciousness decreases. This is also the case for Maes howe, with the exception that it sounds just as at the sweet spot up to 3.5 meters from the center.

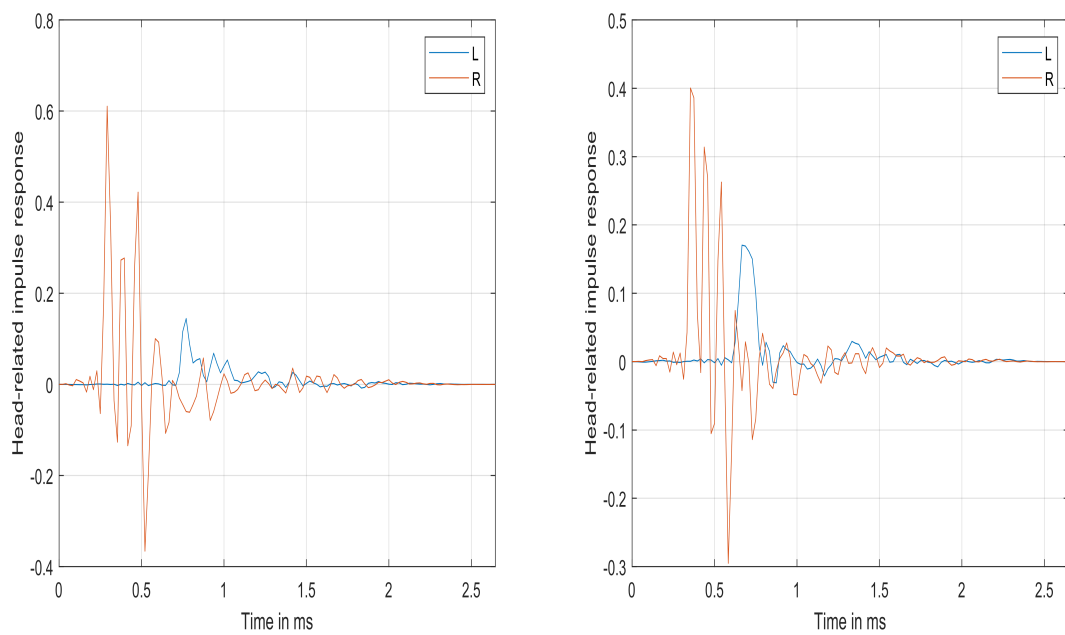
Due to the symmetry of the loudspeaker array on each side of the listener, the same occurs when the listener moves diagonally forward to the left as to the right, but the results are reversed. The further away from the center the listener is, the less the loudspeakers are angled towards the listener, which contributes to the sound being reduced in amplitude on both ears. For listening positions up to 2.8 m diagonally in front of the center of the array, the loudspeaker signals behind the listener are attenuated, not only due to how the loudspeakers are angled relative the listener but also due to the increase in distance between the loudspeakers and the listener. This applies especially to the loudspeaker positioned on the same side to the one the listener moves towards. This can be seen in Figure 4.23, where the syntheses of Maes howe at listening position 2 m diagonally to the left and to the right is presented.

## 4. Results

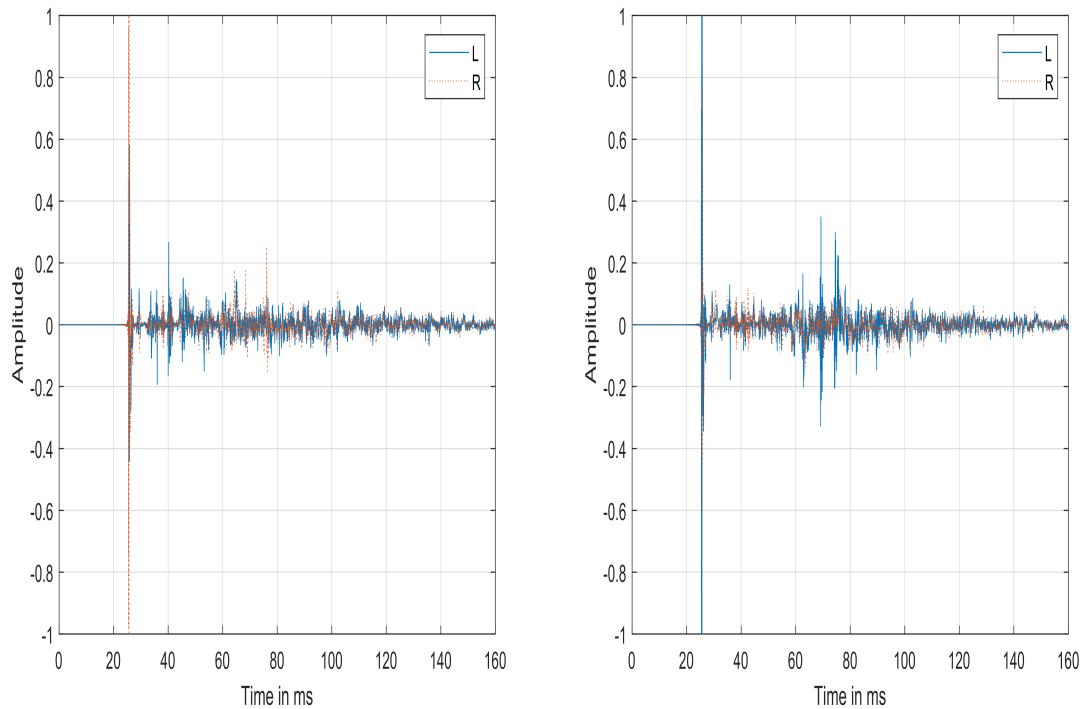
---



**Figure 4.21:** Loudspeaker array synthesis of the Genesis 6 studio for a 3-loudspeaker array in the horizontal plane at the listener position 1.5 meters aside to the right of the center (left) and 3.5 meters aside to the right (right).



**Figure 4.22:** HRIR from loudspeaker 3 to the listener at the new listener positions 1.5 m aside to the right (left) and 3.5 m aside to the right (right).



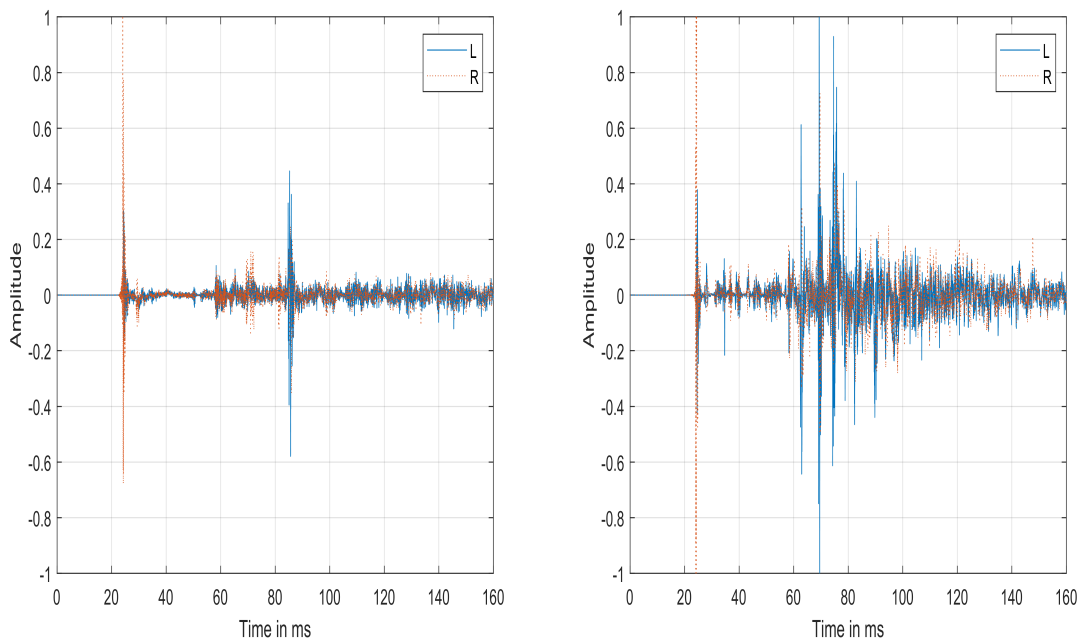
**Figure 4.23:** Loudspeaker array synthesis of Maes howe at the listening position 2 m diagonally in front of the center to the left (left) and at the listening position 2 m diagonally in front of the center to the right (right).

As seen in Figure 4.23, the direct sound is amplified equally in both cases, but is dominated at different ears. Moreover, as the listener moves to the left, loudspeaker 2 becomes closer to the listener than loudspeaker 3. The opposite happens when the listener moves to the right and this can be seen in Figure 4.23 by looking at the reflections. The reflections assigned to loudspeaker 2 are amplified when the listener moves to the left, while the reflections distributed to loudspeaker 3 are attenuated. When the listener instead moves to the right, the reflections played-back from loudspeaker 3 will be amplified while the loudspeaker signal of loudspeaker 2 is attenuated. Moreover, the reflections of loudspeaker 3 are higher in amplitude when the listener moves to the right in comparison to the reflections of loudspeaker 2 when the listener moves to the left.

At the position 3.5 m diagonally in front of the center of the array to the left, the minimum distance between listener and loudspeaker 2 occurs, which highly amplifies this loudspeaker signal. The distance to loudspeaker 3, on the other hand, increases and the reflections played from this loudspeakers are highly attenuated. This can be seen in Figure 4.24, where the reflections, in the syntheses of the Promenadikeskus concert hall as well as Maes howe at this listener position, played-back from loudspeaker 2 are drastically amplified, while the loudspeaker signal of loudspeaker 3 is attenuated to the extent that the reflections of loudspeaker 3 are lower in amplitude than the diffuse sound coming from the other speakers.

## 4. Results

---



**Figure 4.24:** Loudspeaker array synthesis for the listener position 3.5 m diagonally forward to the left of the Promenadikeskus concert hall (left) and of Maes howe (right).

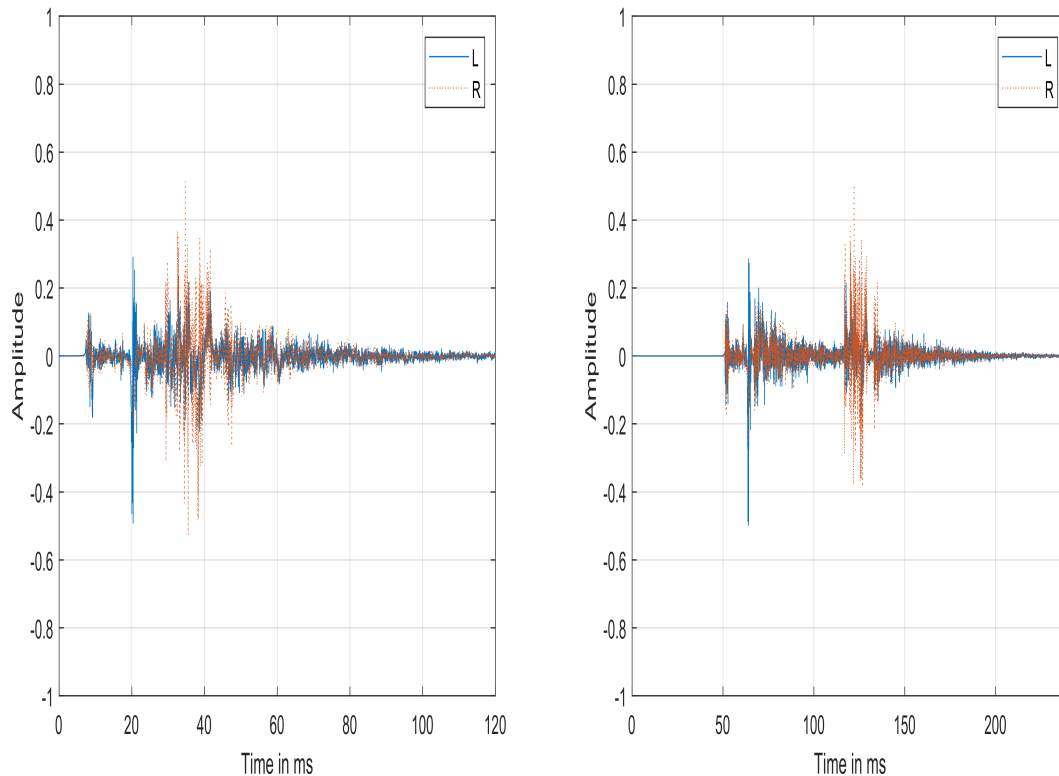
The same but opposite applies to loudspeaker 2 and 3 when the listener do the same move, but to the right. As the listener continuous to move further away from the center of the array, the loudspeaker signals of loudspeaker 2 and 3 gets more and more attenuated as the distance between the listener and these loudspeakers increases.

When the listener moves diagonally forward to the left, the signals from loudspeaker 1 and 3 will mostly reach the listener right ear, just as when moving straight aside to the left. The signal coming from loudspeaker 2, on the other hand, will dominate the left ear of the listener. The opposite applies when moving diagonally forward to the right.

By moving diagonally backwards, the listener approaches either loudspeaker 2 or 3, depending on whether the listener goes to the left or right. This results in that the reflections arrives to the listener before the direct sound do. Despite this, it sounds better when the listener moves diagonally backwards to the left in Maes howe, in comparison to when the listener do the same move but to the right. For the Promenadikeskus concert hall, on the other hand, the quality deteriorates when moving diagonally backwards on both sides. The soundscape retains its richness, but it sounds like the synthesis is made from a different, more reverberant room that contributes to more reflections that it does in real life. This applies to both spaces but especially the Promenadikeskus concert hall which now sounds more like a church than a concert hall. The sound can also be experienced to some extent as scattered.

The influence of different parameters of the loudspeaker array are studied by changing the radius and adding more loudspeakers. The radius is set to half and

twice the original radius. The sweet spot widens when the radius is longer and if the listener takes small steps, in comparison to the radius, from the center of the array then it is advantageous to have a longer radius. However, as the listener moves further from the center it is more advantageous to have a smaller radius. The loudspeaker array synthesis of the Genesis 6 studio for the listener position 3.5 m aside to the left when the radius is decreased and increased, respectively, is presented in Figure 4.25 to illustrate the why.



**Figure 4.25:** Loudspeaker array synthesis for the listener position 3.5 m aside to the left of the Genesis 6 studio when the radius is decreased to 5 m (left) and increased to 20 m (right).

The most obvious difference is that the sound hits the listener earlier when the radius is shorter. Moreover, since the listener moves to the left, the loudspeaker signal containing the reflections that reaches the left ear of the listener are louder for shorter radius because the listener is moving closer to this loudspeaker which will dominate the overall loudspeaker output to a greater degree than the same loudspeaker does for a loudspeaker array of longer radius. When the listener moves in a particular direction, one loudspeakers will get closer but at the same time another will get further away. The difference in distance between the loudspeakers becomes greater when the radius is longer for the same step the listener takes inside the two loudspeaker arrays and this results in the directional components assigned to the different loudspeakers being separated to a higher degree for the long radius loudspeaker array, while the BRIR is more coherent as the radius is small. If the

reflections are separated enough, as they are in Figure 4.25, the listener perceives the time difference. This effect becomes stronger as the first reflections are dominated on one ear while the later ones are dominated on the other, which is the case for genesis 6 studio. Overall, the sound for the same listening distance was improved with a shorter radius than with a longer one, compared to the original radius.

By adding more loudspeakers the listener can move more freely within the loudspeaker array. The directional components are no longer grouped in the same way but are distributed more evenly between the loudspeakers, resulting in that various reflections will be amplified or attenuated as well as shifted differently which sounds more natural, like it does in reality.

# 5

## Discussion

### 5.1 Methods

For calculating the mixing time using the echo density profile proposed by Abel and Huang [5], it is recommended to use a window length of 20-30 ms. If the window is too short then small changes between the windows can become noticeable and in some cases the echo density profile can contain jumps rather than is a steadily growing curve when the RIR contains sudden reflections that are high in amplitude compared to nearby energy. In addition, short windows contribute to a high variance about its local mean. Abel and Huang also writes that too long windows will not maintain sufficient time resolution for purposes due to psychoacoustics. The impulse responses that uses a shorter window length than 20 ms has a steadily growing echo density profile and the mixing time arrives after all early reflections and therefore the window length can be seen as an exception. On the contrary, some of the spaces with longer window lengths than 30 ms still contains jumps in the echo density profile but provides a good estimated mixing time. For these RIRs the rectangular window, that is used for all tested RIRs, can be replaced with a Hanning window that smooths out the jumps. In this way, even longer window lengths can be used while maintaining the mixing time since the Hanning window has an effective window length of half the window length of a rectangular window. For example, this can be applied to the Falkland palace royal tennis court whose energy density profile both contains jumps and has a high variance around 1, as seen in Figure 3.1.

The room dimensions set in the method is not always exact and in some cases the approximation differs greatly from reality. For example, it is hard to describe a stair in means of a rectangular-shaped room, and especially the sloping ceilings due to the steps that are in wraparound. This has contributed to the DOA pattern of the Stairway being approximated to have greater variation in azimuth than it has in reality, since the source and receiver positions are approximated to be in the middle of a room whose height is much longer than the length and width. The lack of roof in the Falkland palace royal tennis court is also approximated by a long height, i.e. to 300 m which in comparison to the other room dimensions of 30 m and 10 m is quite a high ceiling. But the reflections from the image source simulation arises only from elevation angles that corresponds to the ground and side walls of the court and therefore this type of approximation is considered acceptable. Moreover, in the image source simulation the rooms are empty which not may be the case during the measurements of the monaural RIR. There may therefore be reflections

that are assigned incorrectly azimuth and elevation angles the in DOA patterns. For example, both Central hall and the Arthur sykes rymer auditorium are auditoriums with many spectator seats. It is common in these types of spaces to place reflectors close to the stage to direct early reflections towards the audience and this may be the reason why these spaces do not vary in azimuth at all.

The parameterized amplitudes of the direct sound and early reflections differs from the measured amplitudes in the monaural RIR. This is due to how they are calculated in the method, where the RMS amplitude is taken in a 1.5 ms long window around the TOAs, and if the energy is high in the window, the parameterized amplitude will be slightly higher than the measured one, and vice versa. This is nothing that affects the synthesis negatively. However, this is nothing that is audible in the synthesis later.

## 5.2 At the center of the loudspeaker array

The method performs well in finding a parametric description of the monaural RIR, and by comparing it with the corresponding synthesis of the loudspeaker array at the sweet spot, they agree in structure and shape, and it is quite audible that both types describe the same room. This applies to all the tested environments. Some reverberant spaces, for example the Helsington church, contributes to a lot of reflections in the monaural RIR which are not included in the synthesis, as seen in Figure 4.7. Since the method still succeeded in recreating this environment, parameterization of up to the 10 loudest early reflections are actually enough, just as the authors of the Paraspax proposed in [1].

The loudspeaker array depends highly on the detected early reflections. The results shows that when at least three early reflections are distributed over the loudspeakers positioned diagonally behind the listener in a 3-loudspeaker array in the horizontal plane, binaural sound is recreated. Although three reflections are sufficient, the result is unclear compared to when four reflections or more are assigned to loudspeaker 2 and 3. From this it can be concluded that if one or two reflections are played-back from loudspeaker 2 and 3 altogether, the result will be even harder to interpret. Moreover, at least one reflection each should be distributed over loudspeaker 2 and 3 so that not only diffuse sound is played from any of these, otherwise the synthesis will sound more mono than binaural since the signal from loudspeaker 1 equally reaches both ears of the listener.

For those environments where binaural sound could not be reproduced using the loudspeaker array, neither for a loudspeaker array in the horizontal or vertical plane, is due to a few different factors. As seen in the DOAs plots in Figure 3.7 as well as 4.11-4.13, the reflections only varies in elevation but not in azimuth and thus only a loudspeaker array in the vertical plane can be used. However, the loudspeaker array setup in the vertical plane will not contribute to recreated binaural sound. The DOAs of the reflections in Arthur sykes rymer auditorium, presented in Figure 3.7, have the azimuth angle at  $0^\circ$  as well as  $300^\circ$ , but this is not varied enough as  $300^\circ$  corresponds to  $60^\circ$  to the right relative the listener positioned at the sweet spot in the loudspeaker array, which means that all reflections will still hit the listener from the front. Regarding the two RIRs that correspond to outdoor environments

and thus contain only one reflection, binaural sound can never be recreated by the loudspeaker array and it is due to the fact that the two loudspeakers positioned behind the listener needs at least one reflection assigned each.

Just as for the loudspeaker array that only varied in elevation, a loudspeaker array in the horizontal plane containing only two loudspeakers will produce a monotonous sound as well. This is due to that one of the loudspeakers are positioned right in front of the listener at  $0^\circ$ , while the other loudspeaker is positioned behind the listener at  $180^\circ$ . The loudspeaker signals will therefore reach the listener equally on both ears.

For some of the RIRs, the azimuth angles of the DOAs are more in line with the azimuth angles of the loudspeaker positions for a 4-loudspeaker array compared to a 3-loudspeaker array. For example, in Hamilton Mausoleum, the method found one reflection at  $180^\circ$  in azimuth and four reflections at  $279^\circ$ . The rest of the reflections comes from the azimuth angle of  $0^\circ$ . The azimuth of these reflections matches the loudspeaker positions of a 4-loudspeaker array rather than a 3-loudspeaker array, where the reflection at  $180^\circ$  is approximated to the loudspeaker positioned at  $120^\circ$  and the reflections at  $279^\circ$  are approximated to  $240^\circ$ . However, the difference in the results from these two types of loudspeaker arrays in the horizontal plane is negligible. As the importance of how the loudspeakers should be placed in order to reproduce binaural sound, this is primarily more important than letting the reflections reach the listener from the exact angle detected by the method when estimating the DOAs.

The interaural coherence differs from room to room, but they all have in common that the left and right signals are coherent for the very low and high frequencies. It also shows that when approximating the room dimensions to form a shoebox-shaped room, the directional components in the loudspeaker array synthesis are coherent to a higher degree compared to rectangular-shaped rooms. Also, the DOA pattern found in the image source simulation will also only be an approximation when not the real room dimensions are used. However, when comparing the monaural sound, created by convolving audio signals with the monaural RIR, with the binaural sound resulting from convolution with the synthesis, it can still clearly be heard that it is the same room that has been recreated. Therefore, it can be concluded that it is possible to use non-rectangular rooms to reproduce a binaural sound from a measured monaural RIR using the Paraspax method applied in a circular loudspeaker array that surrounds the listener. In addition, the authors of the Paraspax method [1] have compared the ICs of the synthesized and measured BRIR at different positions in the room of test and have come to the conclusion that the IC of the synthesized BRIR is similar to the IC of the measured BRIR, although the IC of the synthesized BRIR shows coherent directional components to a higher degree compared to the measured one.

In the case of Terrys factory warehouse a 4-loudspeaker array is used instead of a 3-loudspeaker array. The reason for this is the azimuth angle of the DOAs of the early reflections. When a 3-loudspeaker array is used, the DOAs in azimuth of the early reflections matches only the loudspeaker positions of loudspeaker 1 and 2, such that the loudspeaker signal of loudspeaker 3 will only contain diffuse sound. The same happens for a 4-loudspeaker array, where the early reflections are assigned

to all loudspeakers, except the one at  $90^\circ$ . Thus the method has not found any reflections that reach the listener from the left side. This means that even though four loudspeakers are required for the loudspeaker array, only three of them will play-back the reflections and thus only three are actually used at the end.

Why is it so that a 3-loudspeaker array in the horizontal plane work in most cases, as long as the DOAs of the early reflections are more or less evenly distributed in azimuth? To recreate binaural sound from the loudspeaker array, the sound should reach the listener from both the front and the back as well as from the sides. By placing one of the loudspeakers right in front of the listener, the direct sound will always hit the listener first which is an important characteristic of a RIR. The two loudspeakers positioned diagonally behind the listener acts for the sound to reach the listener from both the back and the sides. Since they are positioned at the same distance and are directed towards the listener at the same angle, but on either side, their respective HRIRs describing how the sound reaches the ears of the listener from these loudspeakers are the same but opposite for the different ears.

### 5.3 Investigation of the sweet spot

When the listener changes position inside the loudspeaker array there is not a specific position or a certain radius from the center where the quality of the sound starts to deteriorate generally. It rather depends on the impulse response used and how its reflections are distributed over the different loudspeakers. When the DOAs are simulated, the method finds the majority of the reflections on one side of the room and this creates loudspeaker array syntheses that are dominated either on left or right ear. For a left-dominated synthesis, the listener can move to the left to a higher degree compared to when moving to the right.

By moving straight forward towards loudspeaker 1 in the loudspeaker array of Maes howe, the reflections are separated from the direct sound. For the positions up to 3 m from the center, the reflections are still relatively high in amplitude in comparison to the diffuse sound field. But when the listener has moved 3.5 m and further from the center the quality improves again. The loudspeaker array synthesis plot in Figure 4.16 shows that at these distances from the center, the amplitudes of the reflections has attenuated so much that they are now in line with the diffuse sound, and it no longer matters as much that they were separated from the direct sound. However, as the synthesis begins to consist more and more of only direct sound and diffuse sound, in combination with that the output of the rear loudspeaker signals decreases, the synthesis will sound more and more mono. This is due to the dominating loudspeaker that reaches both ears of the listener the same and this fact is applied to all tested RIRs too since Maes howe is the space where most of the reflections are distributed to the loudspeakers at the back. If fewer reflections are assigned to these loudspeakers, this effect will be even greater.

In a room impulse response, the early reflections and reverberation tail are echoes arisen from the direct sound. As the listener moves backwards in the loudspeaker array, the direct sound reaches the listener after some of the reflections and diffuse sound which does not sound intuitive correct. This can be described with the following simile. When two people standing in front of each other with a close distance,

one will perceive mostly the direct sound from the other when he or she speaks. If the distance between the two persons increases then the direct sound from the speech is mixed with the reflected sound. The direct sound is usually the sound with the highest amplitude, but since it is only a single event that only lasts for a few milliseconds, it contributes more to a loud sound than a fully one, for which the reflections are responsible. If the direct sound reaches the listener after all the reflections then the impulse response will contribute to a rich sound due to the reflections but it will also be perceived as lower in sound level. Also, due to the short time of the direct sound, it does not contribute very much to the sound as it arises after the reflections. This effect holds for all tested RIRs when the listener moves backwards, both straight and diagonally, but for diagonal steps backwards towards one of the loudspeakers at the back, the total output of the loudspeaker array is dominated by the loudspeaker towards which the listener walks. As the direct sound reaches the listener after the reflections, the quality will deteriorate, but not as much as when moving straight backwards if the dominating loudspeaker are assigned a lot of reflections. The dominating loudspeaker will then play a louder sound due to the decreasing distance to this loudspeaker and the more reflections played from this, the richer the sound from this loudspeaker will be, resulting in that listener can move further from the center towards this loudspeaker. What happens here is that this loudspeaker now will contribute to richer sound and since it will also dominate the total output from all loudspeakers, it will also mask the output from the other loudspeakers that will contribute to a reduced sound quality. Therefore, the RIRs where the reflections are evenly distributed over the different loudspeakers in the array are the spaces that the listener can move more freely within the loudspeaker array. This claim is backed up by the fact that moving to the right loudspeaker sounded much worse than moving the left for Maes howe, where the reflections are mostly played-back from the loudspeaker at the left. For the Promenadikeskus concert hall, on the other hand, where only two and one reflections is assigned to loudspeaker 2 and 3, respectively, the quality started to deteriorate for movements both to the left and right.

For new listener positions diagonally in front of the sweet spot, the direct sound will hit the listener before the reflections as the loudspeaker at the front will be the one closest to the listener. However, it is still true that if one of the loudspeakers in the rear is assigned enough reflections, the listener can move further from the center towards this loudspeaker before the quality goes down. This was observed for Maes howe when moving diagonally forward, but it was also noted for the Genesis 6 studio when moving straight to the left or the right. Then the listener could move up to 1.5 meters to the right before the quality went down, while longer distances were possible if the listener went to the left instead.

When comparing the two plots in Figure 4.21, illustrating new listener positions 1.5 m respective 3.5 m aside to the left, with the HRIRs of loudspeaker 3 at these positions, showed in Figure 4.22, it can be seen that the amplitude of the reflections not only depends on the distances to the loudspeakers but also how the loudspeakers are angled towards the listener. The reflections played-back from loudspeaker 3 are the first sound that will reach the listener and mainly heard by the right ear. These reflections are almost the same in amplitude at 1.5 m and 3.5 m from the center,

even though the HRIR plots shows that the loudspeaker signal at the right ear is more attenuated at 3.5 m from the center in comparison to when the listener is 1.5 m from the center. It means that even though the amplitude has been amplified due to decreasing distance, it is taken out by the speaker not being angled towards the listener anymore. For the reflections played-back from loudspeaker 2, which appears later in time and mostly at the left ear, the distance attenuation is higher and this is due to the fact that this loudspeaker is angled towards the listener at these new listening positions. Furthermore, the angle dependency decreases when the loudspeakers are positioned further away from the listener. There are new listening positions inside the loudspeaker array where both the distance and the angle to the loudspeakers in the rear are maximized and it is about 3.5 m diagonally slanted forward in the same direction as the loudspeaker with the most reflections is positioned.

### 5.4 Influence of different parameters

The sweet spot for a longer versus shorter radius can be described using the following simile. Imagine that you have a flashlight that shines with a red light and forms a circle. If the flashlight is far away, the red light is smeared out so that it covers a larger circumference but is lower in strength. The closer the flashlight gets, the brighter the light becomes, but only on a limited surface. This can be applied to a loudspeaker array by replacing the light by sound. If the loudspeakers are placed with a longer distance from the center of the array, the sweet spot gets wider, resulting in that the listener can move further from the center without deteriorate the quality. However, if the listener moves far enough from the center so that the difference in TOAs of the directional components from the different speakers becomes large enough, the sound quality will deteriorate significantly and in some cases the TOA differences are experienced and the sound between the different loudspeakers ends up out of sync. This phenomenon does not occur at a radius of less than 10 meters. Moreover, the difference in results between the loudspeaker arrays of different radius implies that the distance attenuation/amplification has a greater influence on how the sound changes when a listener moves.

By increasing the number of loudspeakers in the simplified loudspeaker array in the horizontal plane, the listener can move more freely without degrading the sound quality to the same extent as when only three loudspeakers are used. The reason is that not only one loudspeaker signal will dominate the total output from all loudspeakers when the listener moves in a certain direction. Moreover, the distance attenuation between these loudspeakers will be less perceived. Fewer reflections will also be assigned to the various loudspeakers as they are spread out more and that will result in the various reflections being amplified or attenuated differently. They will also be spread out more in time and not reach the listener as a group, just as reflections in real life behaves. When each loudspeaker is not assigned the same amount of reflections, as they are when they are fewer, individual loudspeakers do not have the same impact on the total output either. However, it does not matter how many loudspeakers are added to the simplified loudspeaker array in the vertical plane because the sound reproduced will always be mono.

# 6

## Future research

As the loudspeaker array is highly dependent on variation in the azimuth angle of the DOAs of the early reflections, so future research could have focused on a way to produce non-empty shoebox-shaped rooms used for the image source model. By placing single interior details in the rooms such as reflectors that are used in, for example, concert halls to create early reflections, the method could have been used even in rooms where the variation was currently very small. It would also be interesting to find out if it is possible to reduce the number of early reflections to reduce the computational load.

The method of the thesis is built upon a statically state where the sound image created by the loudspeaker array is synthesized for new positions that are predefined before the method is executed. Just as for the Paraspax method, which can either be used for BRIR synthesis or real-time rendering, this extended version for loudspeaker arrays could as future work be used to work in real-time too. Either the method can be used together with a head-tracker or the listeners movement can be analysed using a camera, such as a kinect or webcam. For example, Song et. al. presents a method for a 3D audio system of 6 DoF for using a web camera [14] which could be worth reading for inspiration of other types of headtracking systems than the one used in the Paraspax method.



# 7

## Conclusion

The Paraspax method successfully parameterized all tested RIRs as long as the following are available: measured monaural RIR, source distance and direction, room dimensions as well as source and receiver positions. The image source model approximates a space to an empty shoebox-shaped room which can give incorrect DOAs of the direct sound and selected early reflections if this approximation deviates too much from reality. However, this does not affect the reproduction of the room. DOAs with little or no variation in azimuth angle cannot be applied to a simplified loudspeaker array, where all loudspeakers are placed around the listener with varied azimuth angles.

For synthesis of the simplified loudspeaker array, it must consist of at least three loudspeakers, where one is positioned just in front of the listener and is responsible for the direct sound that should hit the listener first. The other two loudspeakers should be placed on the sides and behind the listener so that the sound reaches the listener from all directions. The best possible loudspeaker positions form a triangle with the listener at the midpoint of the triangle and this is the case for 10 out of the 15 tested RIRs. To possibly synthesize a sound field created in such a loudspeaker array, at least one reflection should be assigned to each of the loudspeakers at the sides/behind the listener.

The listener has more freedom to move within a loudspeaker array if it has most of the reflections distributed over the loudspeakers behind the listener, while the direct sound and a few early reflections should be distributed to the loudspeaker in front of the listener. When only a few reflections reach the listener from behind or from the sides, the quality deteriorates considerably in comparison, both when standing at the sweet spot or changing positions. The sound behaves differently depending on the direction in which the listener chooses to move, but in general, the sound is least affected when the listener moves straight forward. The synthesis of the loudspeaker array lacks direct sound as the listener moves backwards, or the total output from all loudspeakers will dominate one ear when the listener moves to the sides. Moreover, by moving to the sides, both straight and diagonally, deteriorates the quality the most. There are also certain positions in the loudspeaker array where some loudspeaker signals are highly amplified and attenuated.

By increasing the radius or the number of loudspeakers in the array, the listener can move more freely. However, this only applies to positions near the sweet spot.



# Bibliography

- [1] JO. M.. Arend, SE. AM. Garí, C. Schissler, F. Klein, and PH. W.. Robinson, "Six-Degrees-of-Freedom Parametric Spatial Audio Based on One Monaural Room Impulse Response," *J. Audio Eng. Soc.*, vol. 69, no. 7/8, pp. 557-575, (2021 July). doi: <https://doi.org/10.17743/jaes.2021.0009>
- [2] Damian Murphy and Joe Rees-Jones, "Open Acoustic Impulse Response (Open AIR) Library, IR Data" Audiolab, Department of Electronics, University of York [https://www.openair.hosted.york.ac.uk/?page\\_id=36](https://www.openair.hosted.york.ac.uk/?page_id=36)
- [3] Juha Merimaa, Timo Peltonen and Tapio Lokki, "Concert Hall Impulse Responses," TKK, Akukon Oy Consulting Engineers <http://legacy.spa.aalto.fi/projects/poririrs/>
- [4] C. Pörschmann, P. Stade and J. M. Arend, "Binauralization of Omnidirectional Room Impulse Responses - Algorithm and Technical Evaluation," in *Proceedings of the 20th International Conference on Digital Audio Effects (DAFx-17)*, pp. 345-352 (Edinburgh, UK) (2017 Sep.).
- [5] J.S. Abel and P. Huang, "A Simple, Robust Measure of Reverberation Echo Density," presented at the *121st Convention of the Audio Engineering Society*, pp. 1-10 (2006 Oct.), paper 6985.
- [6] H. Kuttruff, *Room Acoustics, Fifth Edition*. CRC Press, 2009.
- [7] J. Thiemann and S. van de Par, "A multiple model high-resolution head-related impulse response database for aided and unaided ears," in *EURASIP Journal on Advances in Signal Processing*, Article number: 9 (2019).
- [8] ISO/TC 43/SC 2 Building Acoustics, "ISO 3382-2:2008: Acoustics - Measurement of room acoustic parameters - Part 2: Reverberation time in ordinary rooms," ICS: 91.120.20, Acoustics in building. Sound insulation, (2008 June). ISBN: 978 0 580 67763 2
- [9] Purves D, Augustine GJ, Fitzpatrick D, et al., "The audible spectrum," in *Neuroscience, 2nd edition*, Sinauer Associates (2001).
- [10] Fabian Brinkmann, Hannes Gamper, Nikunj Raghuvanshi and Ivan Tashev, "Towards encoding perceptually salient early reflections for parametric spatial audio rendering," presented at the *148th Convention of the Audio Engineering Society*, pp. 1-11, Paper 10380, (2020 May).

- [11] Allen, J. B. and Berkley, D. A., "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, 65(4), p. 943, 1979.
- [12] Philipp Stade, Johannes M. Arend and Christoph Pörschmann, "Perceptual Evaluation of Synthetic Early Binaural Room Impulse Responses Based on a Parametric Model," presented at the *142nd Convention of the Audio Engineering Society*, Paper 9688, (2017 May).
- [13] Leon Müller and Jens Ahrens, "Perceptual Differences for Modifications of the Elevation of Early Room Reflections," presented at the *2022 International Conference on Audio for Virtual and Augmented Reality*, Chalmers University of Technology, Gothenburg, Sweden, (2022 August).
- [14] M. -S. Song, C. Zhang, D. Florencio and H. -G. Kang, "Personal 3D audio system with loudspeakers," 2010 IEEE International Conference on Multimedia and Expo, 2010, pp. 1600-1605, doi: 10.1109/ICME.2010.5583184.

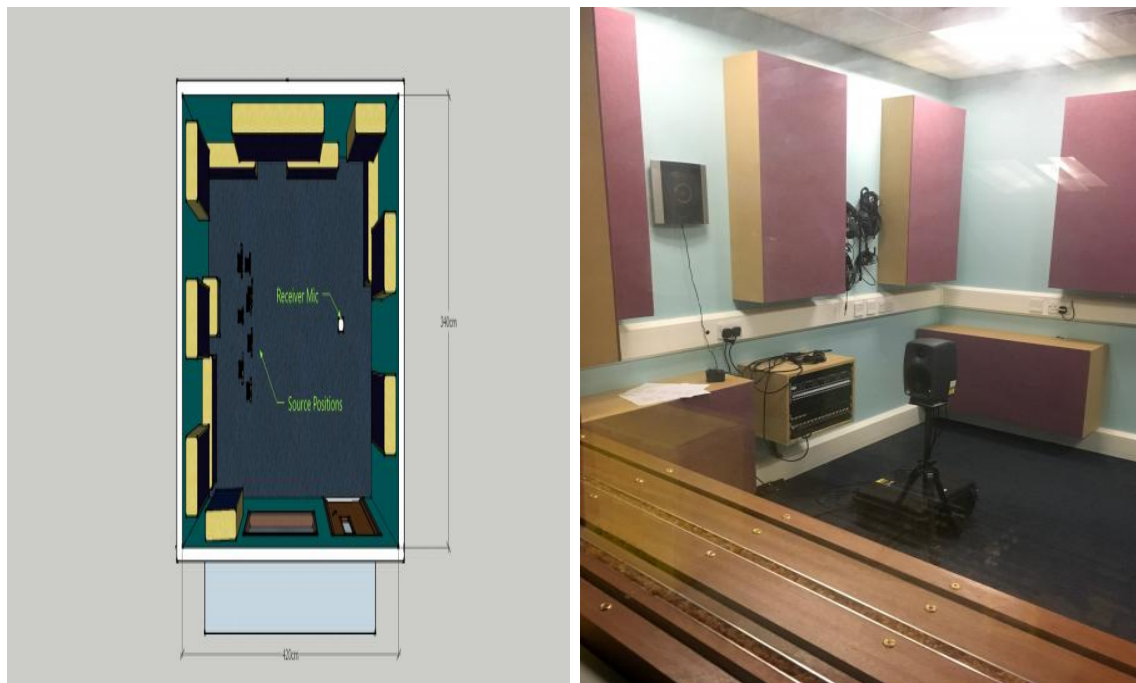
# A

## Tested room impulse responses

Within this work, the Paraspax method was tested for 15 different monaural RIRs of various environments. Subsequent sections contains a description of each of them including measurement positions and pictures of the spaces.

### A.1 Genesis 6 studio

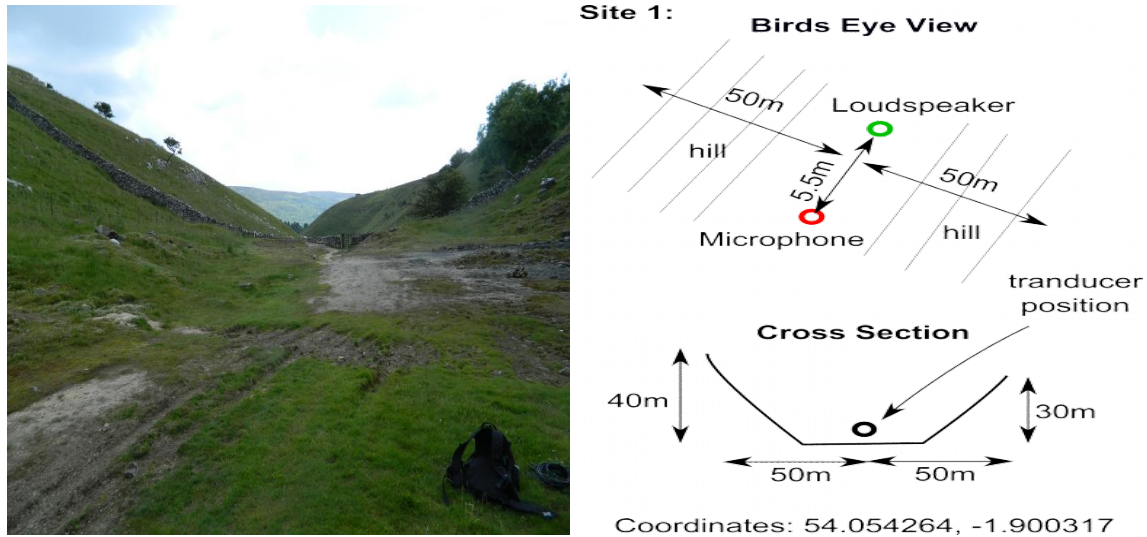
*Genesis 6 studio* is a recording studio consisting of a control room and a live recording studio. It is situated in the York University Campus in England. The measurements are performed in the recording studio which have a low reverberation time due to acoustically treated panels placed on the walls with the task to absorb sound and prevent reflections. Each panel is filled with Rockwool acoustic insulation and has a frame consisting of medium-density fiberboard wood. The sound source and receiver are placed just in front of each other with the height of the receiver corresponding to a human's ears.



**Figure A.1:** Floor plan with source and receiver positions (left) and photo taken from the control room (right) of the Genesis 6 studio.

## A.2 Trollers gill

In the limestone gorge *Trollers gill* the impulse response is measured in a valley of V-shape and hills on each side of the microphone and loudspeaker positions.



**Figure A.2:** The measured valley of Trollers gill (left) and floor plan with source and receiver positions (right).

## A.3 Maes howe

At the Orkney islands in Scotland is one of the most excellent chambered cairns in Europe. It belongs to an ancient site, dated to 3000 BC in the 30th century BC. The shape of the chamber is almost cubic with resonances, see Figure A.7, which it is said was used to affect rituals and interaction with the space. There is a hypothesis that they are within the male voice frequencies. Strong modal frequencies are present here. The walls of the interior that can be seen in Figure A.4 is made of large, flat slabs of stone.



**Figure A.3:** The outside location of Maes Howe (left) with floor plan (right).



**Figure A.4:** The interior of Maes Howe.

During the measurements the source was placed at the centre of the chamber and the microphone at the centre of the wall opposite to the entrance.

## A.4 Arthur sykes rymer auditorium

At the University of York in England is the *Arthur sykes rymer auditorium* located which is said to reproduce sound of high quality thanks to its unique acoustics. Its preferred noise criterion (PNC) is better than the PNC 15 standard, which means it shuts out outside noise. Source and receiver positions during the measurements are the ones called "S1" and "R2" in Figure A.5.

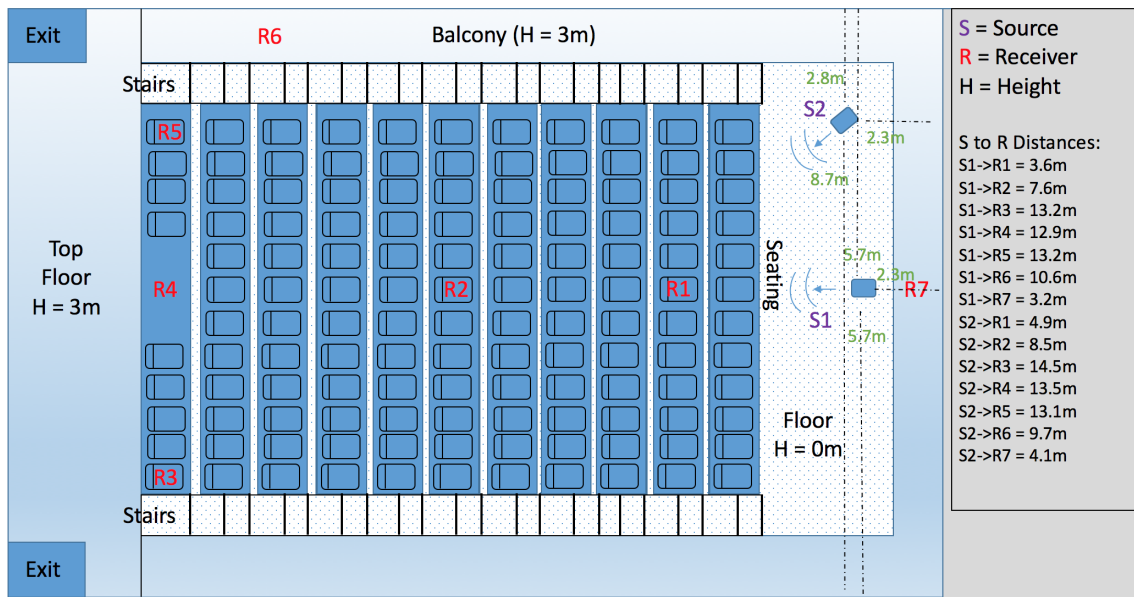
## A.5 Koli national park

*Koli national park* is a forest in Finland. The measurements was performed around mid-summer. The source and receiver are positioned opposite to each other at the same height.

## A.6 Stairway

At the University of York in England, an impulse response was measured in a stairway. The measurements was performed at some mid floor, seen in Figure A.8, with the loudspeaker placed at 3 m distance from the and microphone and both having the same height.

## A. Tested room impulse responses



**Figure A.5:** Floor plan with measurement positions of source and receiver at the Arthur sykes rymer auditorium.



**Figure A.6:** The interior of the Arthur sykes rymer auditorium.



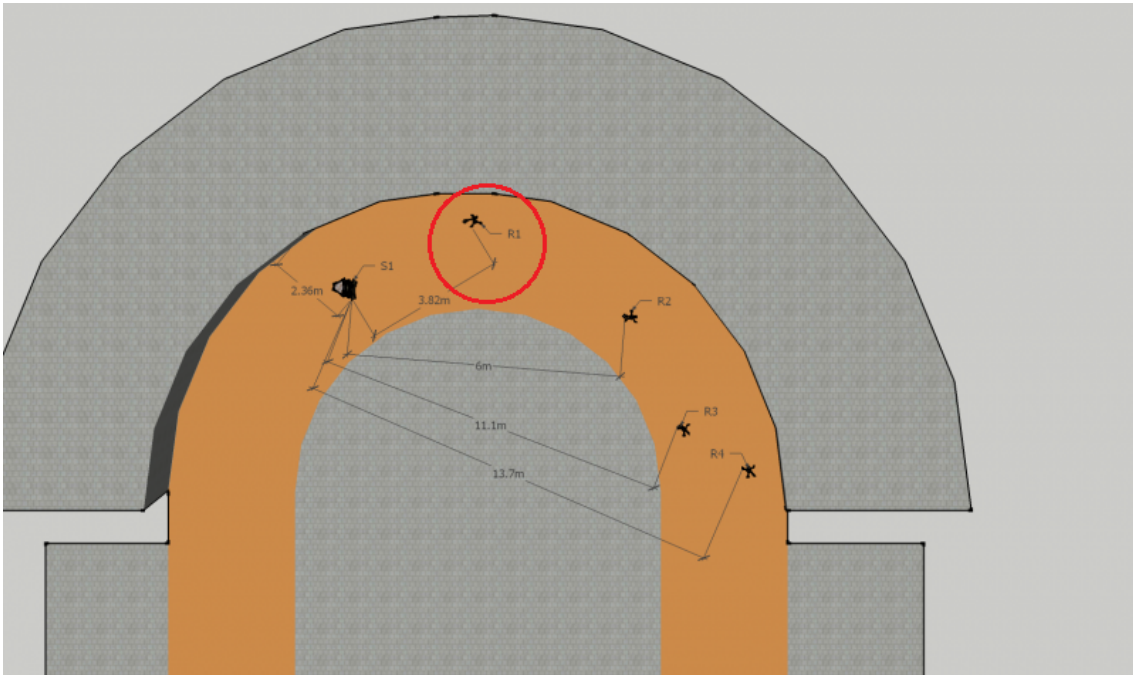
Figure A.7: The Koli national park at summer.



Figure A.8: The floor at which measurements were made at the Stairway (left). The floors below the measurement floor (right).

## A.7 Hoffman lime kiln

In the Yorkshire dales national park in the city of Langcliffe in England you can find the *Hoffman lime kiln* that was built in 1873, used on an industrial scale and was in operational until 1931. This is a tunnel-like chamber with open entrances along the side walls that are made of large stones. The source and receiver positions are marked "S1" and "R1" in Figure A.9. Due to the many receiver positions in Figure A.9, the "R1" position is marked by a red circle.



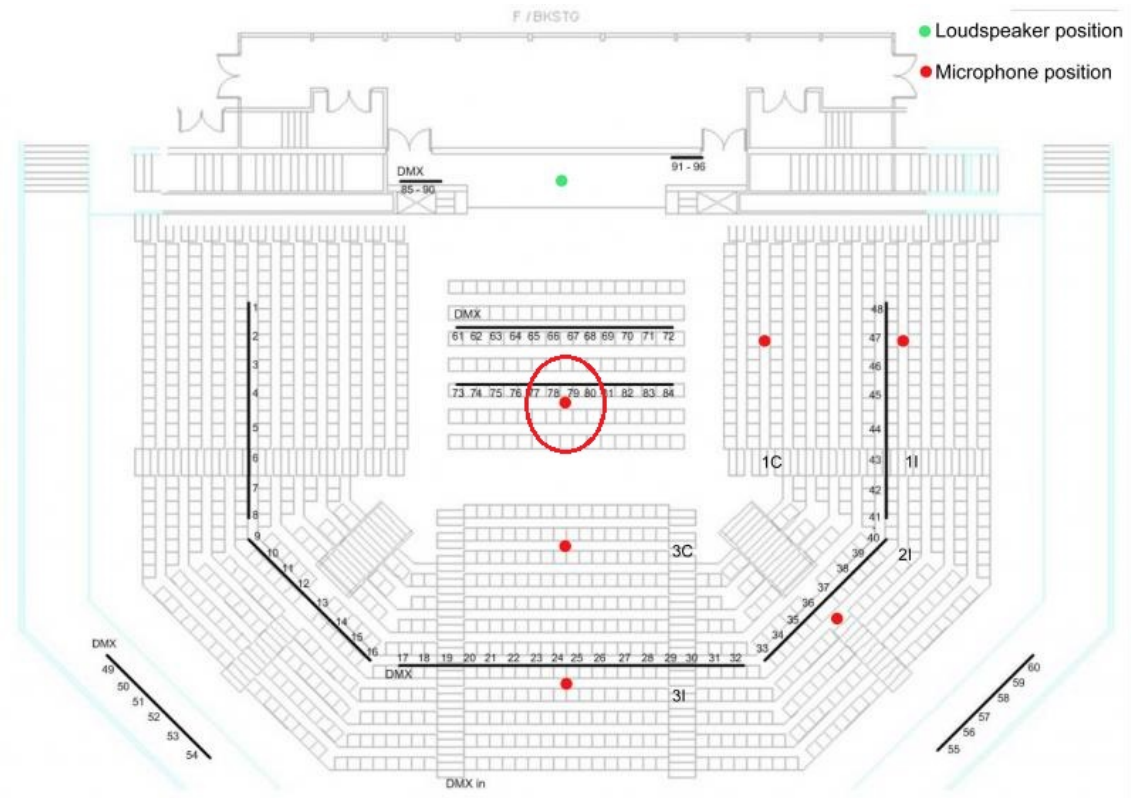
**Figure A.9:** Floor plan of Hoffman lime kiln with source and receiver positions.



**Figure A.10:** The exterior (left) and interior (right) of the Hoffman lime kiln.

## A.8 Central hall

*Central hall* is an auditorium at the University of York in England. The main area of use of the hall is classes and conferences. The impulse response was measured in front of the stage at the centre of the chairs placed on the floor (the position between chair 78 and 79), marked with a red dot inside a red circle in Figure A.11.



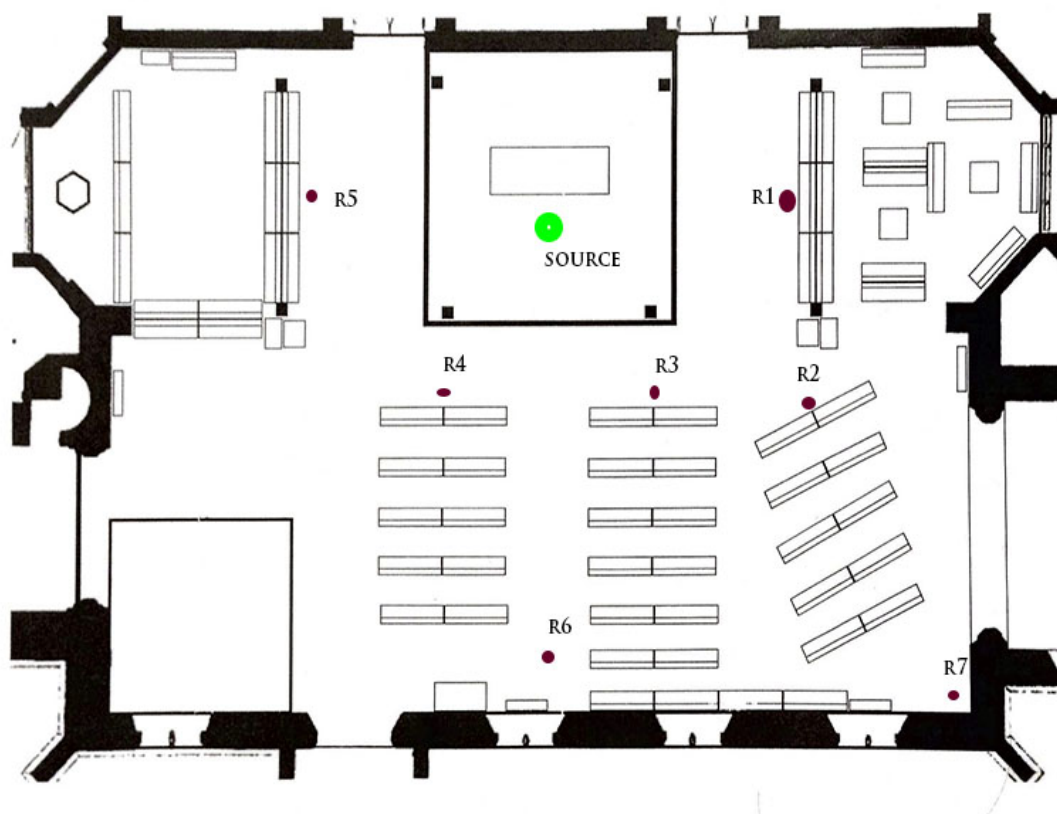
**Figure A.11:** Floor plan of the Central hall with source and receiver positions.



**Figure A.12:** The interior of Central hall. The speaker used in the measurements is visible on stage (right). The hall is equipped with bleachers at the back and a bunch of chairs at the front (left).

## A.9 Helsington church

*Helsington church* is situated in York, a city of England. It is a church of Methodism, a group associated with historically related denominations within Protestant Christianity. The foundation of the building has existed ever since the 11th century but today's construction was renovated and expanded on two occasions; in the middle of the 19th century and in 1973. The most remarkable thing about these renovations is the new orientation where the altar is placed on the long side so that the assembly is facing the direction it would not normally be. This could cause interesting acoustical results as the original building was designed traditionally from an acoustic perspective. Figure A.13 of the floor plan shows the loudspeaker used positioned at the altar and microphone position at "R6".



**Figure A.13:** Floor plan of the Helsington church with source and receiver positions.

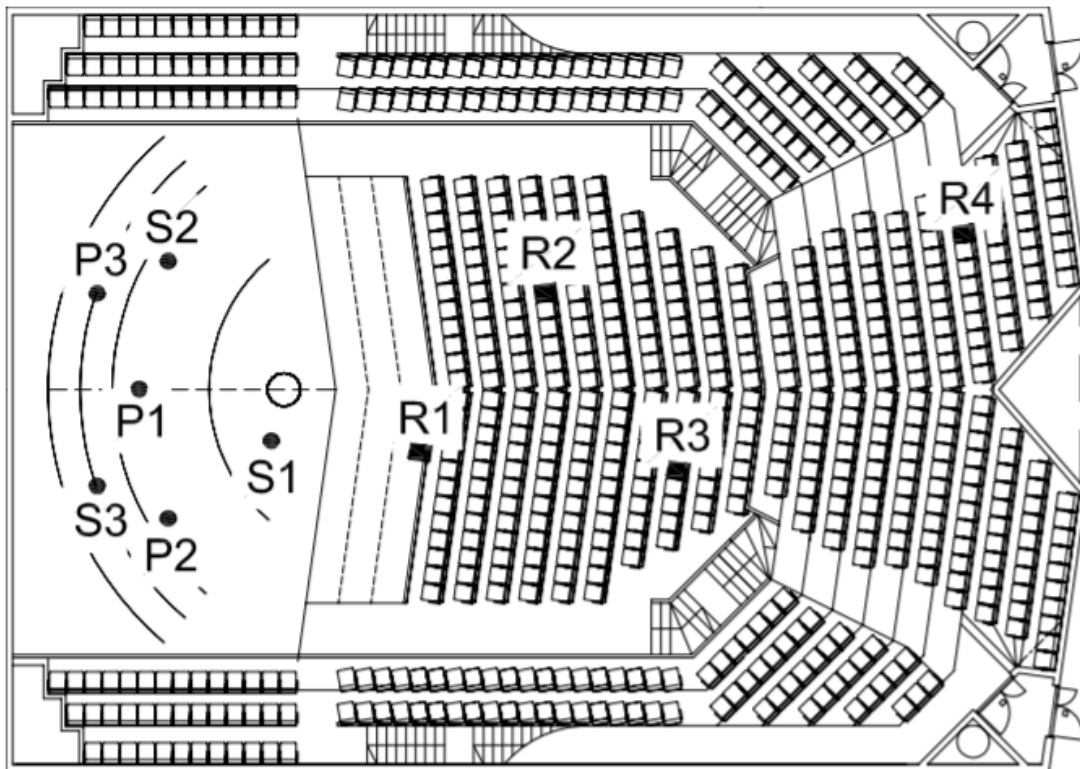
## A.10 Promenadikeskus concert hall

The Promenadikeskus is a concert hall in the city called Pori in Finland. The hall was built in 1999 and contains 700 seats for audience. The monaural RIR is measured when the sound source is placed on the stage and the microphone on the first row, at the fourth seat on the left hand side of the mid line of the hall. The



**Figure A.14:** The interior of Helsington church. The loudspeaker at the altar (right) and the microphone at position "R6" (left).

source and receiver positions are marked "S1" and "R1" in the floor plan in Figure A.15.



**Figure A.15:** Floorplan of the Promenadikeskus concert hall with source and receiver positions.

Moreover, the sound source consists of two loudspeakers, one omnidirectional loudspeaker for the high frequencies and a subwoofer to extend the frequency range for the low frequencies. The source source is recorded with a dummy head, having

two DPA 4006 microphones in each ear.

## A.11 Innocent railway tunnel

Innocent railway tunnel is also called St Leonards tunnel and it is a tunnel situated in south east of the centre of Edinburgh. It is 517 m long and is semicircular in shape with the dimensions 6 m  $\times$  4.5 m (width  $\times$  height). When it was built in 1831 it was used as a railway until year 1968. It was then unused for a couple of years and reopened again 1981, not as a railway tunnel but for pedestrians and now days it is also used as a cycle path. Figure A.16 shows how the tunnel looks like today.



**Figure A.16:** The interior of the Innocent railway tunnel.

The source and receiver positions are found in Figure A.17.

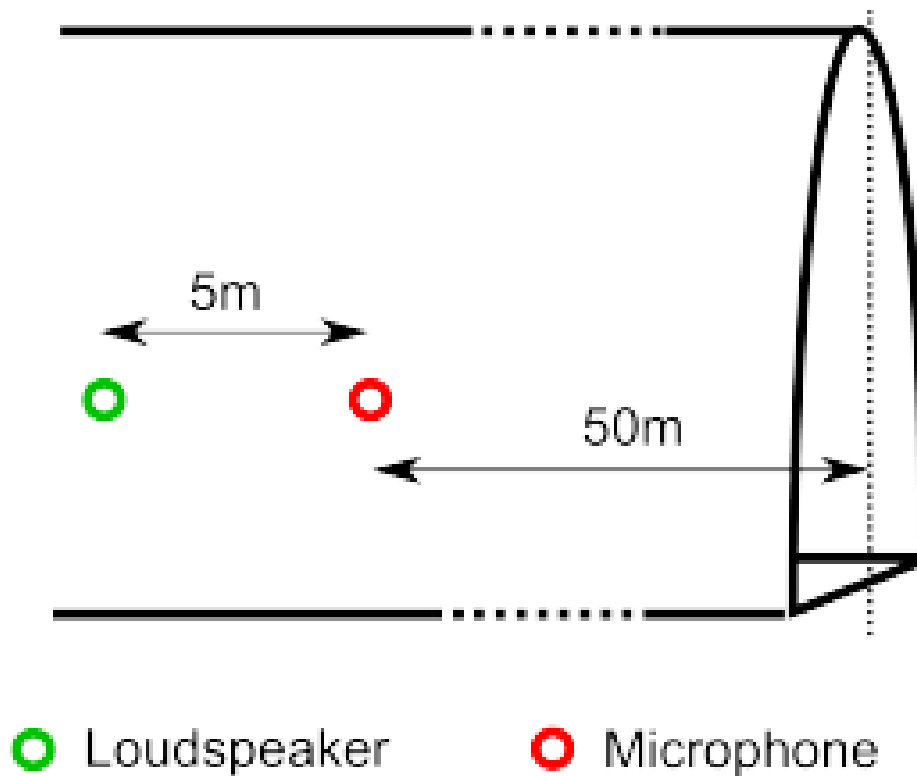


Figure A.17: Source and receiver positions of the Innocent railway tunnel.

## A.12 Falkland palace royal tennis court

The semi outside environment called the Falkland palace royal tennis court is a tennis court consisting of ground, four walls but no roof. The court was built in 1539 and is situated in Fife in Scotland. The measurements were performed with the source and receiver distance of 10.76 m, both at the same height of 1.5 m. The microphone used can be seen in Figure A.18.

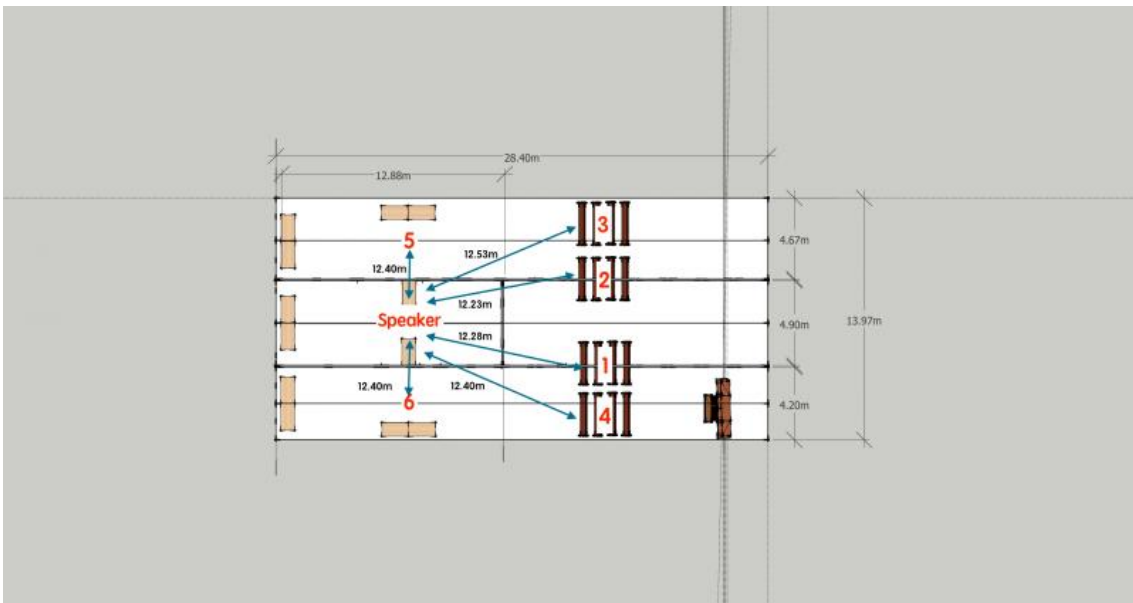
## A.13 Shrine and parish church of all saints

The church called the Shrine and parish church of all saints is Anglican church style designed and includes three high roofs aisles. The most interesting interior detail is the wooden altar made of panels of carving wood from the 17th-century. It is located close to the singing section and creates a characteristic sound field. The church is 28.4 m in height and 14 m in width. During the measurements the source was placed at the altar and the receiver position can be found in the floor plan in Figure A.19, marked by "5". Moreover, the interior of the space is presented in Figure A.20.

## A. Tested room impulse responses



**Figure A.18:** The interior of the Falkland palace royal tennis court.



**Figure A.19:** Floor plan of the Shrine and parish church of all saints together with source and receiver positions.



**Figure A.20:** The interior of the Shrine and parish church of all saints.

## A.14 Hamilton mausoleum

In the town called Hamilton in Scotland can you find the *Hamilton mausoleum*, which took 16 years to build and was finished in 1858. The building is made of marble and sandstone and it has the shape of a octagon in plan with a diameter of 18 meters. Its cylindrical roof stretches up to 36 meters. The space has two parts, a chapel and a crypt. Its unique and long reverberation contributes to acoustic properties that musicians uses for recording.

During the measurements, the microphone was placed at the center of the space and the loudspeaker close to on of the walls, creating distance of 4.8 meters between them, which can be seen in Figure A.21.



**Figure A.21:** The exterior (left) and the interior, showing the microphone and loudspeaker position (right) of Hamilton mausoleum.

## A.15 Terrys factory warehouse

*Terrys factory warehouse* is an large, shoebox-shaped industrial premises situated in York in England. It was previously used by the company "Terry's chocolate and confectionery factory", but is now closed and empty since 2005. Its large volume contributes to the longest reverberation time of all tested room impulse responses. The source and receiver are positioned at 28.35 meters from each other and at same height.



**Figure A.22:** The interior of Terrys factory warehouse including microphone position (left) and loudspeaker position (right).

## A.16 Table

Additional information about the spaces described earlier in this chapter can be found in Table A.1. The table presents, among others, the reverberation times, volumes as well as type of sound and receiver sources.

A. Tested room impulse responses

Name	Type	$RT_{60}$ in s	$V$ in $m^3$	Sound source	Receiver source	Type of IR
Genesis 6 studio	recording studio	0.63	35.2	Soundfield Genelec 8040	ST350 Kit: Kick	omni
Trollers gill	open air	0.95	-	Genelec 8130A	Soundfield ST450 kit	omni
Maes howe	chamber	1.13	100	Genelec S30D	Soundfield SPS422B	omni
Arthur sykes rymer auditorium	auditorium	1.34	1 560	Genelec 8130A	Soundfield ST 350 Kit	omni
Koli national park	open air	1.46	-	Genelec 8130A	Soundfield ST450 kit	omni
Stairway	indoor corridor	2.05	-	Genelec S30D	Neumann KM 140 Cardioid	cardioid
Hoffman lime kiln	cave	2.13	-	Genelec 8130A	Soundfield ST450 MKII	omni
Central hall	auditorium	2.67	8 000	Genelec	Soundfield	-
Helsington church	church	3.1	2 000	Genelec 8030	Soundfield ST450	omni
Promenadikeskus concert hall	concert hall	4.29	9 300	4 Vifa MG10SD09 4" & EAW SB48e	DPA 4006	omni
Innocent railway tunnel	outdoor tunnel	6.31	13 000	Genelec 8130A	Soundfield ST450 kit	omni
Falkland palace royal tennis court	semi outdoor (no roof)	8.28	2 300	Genelec S30D	Soundfield SPS422B	omni
Shrine and parish church of all saints	church	20.31	2 398.8	Genelec 8030	Soundfield SPS422B	omni
Hamilton mausoleum	mausoleum	20.93	9 000	Genelec S30D	Soundfield SPS422B	omni
Terrys factory warehouse	warehouse	56.16	4 500	Genelec S30D	Soundfield SPS422B	omni

Table A.1: Information about the tested room impulse responses.

# B

## Results

The window length used in the echo density profile, the estimated mixing time and reverberation level for all tested spaces is presented in Table B.1.

Name	Window length (edp) in ms	Mixing time in ms	Reverberation level in dB
Genesis 6 studio	16.67	42	-6.7
Trollers gill	21.3	45	-37.4
Maes howe	43.75	62	-19.6
Arthur sykes rymer auditorium	16.67	64	-29.6
Koli national park	9.38	50	-43.4
Stairway	21.3	40	-20.6
Hoffman lime kiln	58.3	97	-1
Central hall	21.3	78	-15.5
Helsington church	89.6	105	-10.2
Promenadiskeskus concert hall	21.3	115	-25
Innocent railway tunnel	42.67	115	-14.7
Falkland palace royal tennis court	25	169	-13.3
Shrine and parish church of all saints	42.67	100	-9.4
Hamilton mausoleum	31.2	131	-24.7
Terrys factory warehouse	21.3	176	-12.7

**Table B.1:** Various results for all tested room impulse responses.