Iceberg: A loudspeaker-based room auralization method for auditory research

Sergio Luiz Aguirre

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Hearing Sciences – Scottish Section - School of Medicine
University of Nottingham

Supervised by William M. Whitmer, Lars Bramsløw, & Graham Naylor

2022

There is no “nonspatial hearing” — Jens Blauert

Abstract

Depending on the acoustic scenario, people with hearing loss are challenged on a different scale than normal hearing people to comprehend sound, especially speech. That happen especially during social interactions within a group, which often occurs in environments with low signal-to-noise ratios. This communication disruption can create a barrier for people to acquire and develop communication skills as a child or to interact with society as an adult. Hearing loss compensation aims to provide an opportunity to restore the auditory part of socialization. Technology and academic efforts progressed to a better understanding of the human hearing system. Through constant efforts to present new algorithms, miniaturization, and new materials, constantly-improving hardware with high-end software is being developed with new features and solutions to broad and specific auditory challenges. The effort to deliver innovative solutions to the complex phenomena of hearing loss encompasses tests, verifications, and validation in various forms. As the newer devices achieve their purpose, the tests need to increase the sensitivity, requiring conditions that effectively assess their improvements.

Regarding realism, many levels are required in hearing research, from pure tone assessment in small soundproof booths to hundreds of loudspeakers combined with visual stimuli through projectors or head-mounted displays, light, and movement control. Hearing aids research commonly relies on loudspeaker setups to reproduce sound sources. In addition, auditory research can use well-known auralization techniques to generate sound signals. These signals can be encoded to carry more than sound pressure level information, adding spatial information about the environment where that sound event happened or was simulated. This work reviews physical acoustics, virtualization, and auralization concepts and their uses in listening effort research. This knowledge, combined with the experiments executed during the studies, aimed to provide a hybrid auralization method to be virtualized in four-loudspeaker setups. Auralization methods are techniques used to encode spatial information into sounds. The main methods were discussed and derived, observing their spatial sound characteristics and trade-offs to be used in auditory tests with one or two participants. Two well-known auralization techniques (Ambisonics and Vector-Based Amplitude Panning) were selected and compared through a calibrated virtualization setup regarding spatial distortions in the binaural cues. The choice of techniques was based on the need for loudspeakers, although a small number of them. Furthermore, the spatial cues were examined by adding a second listener to the virtualized sound field. The outcome reinforced the literature around spatial localization and these techniques driving Ambisonics to be less spatially accurate but with greater immersion than Vector-Based Amplitude Panning.

A combination study to observe changes in listening effort due to different signal-to-noise ratios and reverberation in a virtualized setup was defined. This experiment aimed to produce the correct sound field via a virtualized setup and assess listening effort via subjective impression with a questionnaire, an objective physiological outcome from EEG, and behavioral performance on word recognition. Nine levels of degradation were imposed on speech signals over speech maskers separated in the virtualized space through Ambisonics' first-order technique in a setup with 24 loudspeakers. A high correlation between participants' performance and their responses on the questionnaire was observed. The results showed that the increased virtualized reverberation time negatively impacts speech intelligibility and listening effort.

A new hybrid auralization method was proposed merging the investigated techniques that presented complementary spatial sound features. The method was derived through room acoustics concepts and a specific objective parameter derived from the room impulse response called Center Time. The verification around the binaural cues was driven with three different rooms (simulated). As the validation with test subjects was not possible due to the COVID-19 pandemic situation, a psychoacoustic model was implemented to estimate the spatial accuracy of the method within a four-loudspeaker setup. Also, an investigation ran the same verification, and the model estimation was performed with the introduction of hearing aids. The results showed that it is possible to consider the hybrid method with four loudspeakers for audiological tests while considering some limitations. The setup can provide binaural cues to a maximum ambiguity angle of 30 degrees in the horizontal plane for a centered listener.

Acknowledgements

Thank you,

Obrigado, Gracias, Grazie, Tak skal du have, Dank u zeer & Danke sehr

Firstly, I would like to express my gratitude to my supervisors, Drs. Bill Whitmer, Lars Bramsløw, and Graham Naylor, for their guidance, expertise, and remarkable patience throughout this process. Your support and mentorship have been invaluable in helping me to fill my knowledge gaps, tirelessly encouraging me to ask the right questions, and guiding me to produce high-quality scientific research. I also thank Dr. Thomas Lunner for his initial guidance, insightful questions, and comments.

Thank you to the special people at Eriksholm Research Centre in Denmark and the people at Hearing Sciences – Scottish Section. Working with such fantastic top teams has been a joy and a privilege. A special thank you to Jette, Michael, Bo, Niels, Claus, Dorothea, Sergi, Jeppe, James, Lorenz, Johannes, and Hamish. Thanks to all the people involved in HEAR-ECO for their hard work, especially Hidde, Beth, Patrycja, Tirdad, and Defne.

I am deeply grateful to my sweet wife, Lilian, for the love, encouragement, and support she has given me throughout this journey. Her unwavering support has been an enduring seed of resilience and inspiration. I cannot thank her enough for being such an integral part of my life.

Thanks to my friends Math and Gil, for always being there. Special thanks to my former professors, Drs.: Arcanjo Lenzi, William D'Andrea Fonseca, Eric Brandão, Paulo Mareze, Stephan Paul, and Bruno Sanches Masiero for stimulating critical thinking and for all the support, knowledge, and encouragement. Thank you also to my former Oticon Medical colleagues Simon Ziska Krogholt, Patrick Maas, Brian Skov, and Jens T. Balslev for the and support.

I would like to thank you, my Professor, *Professora* Dra. Dinara Xavier Paixão. All of this is possible because of you and your determination to create an official undergraduate course in Acoustical Engineering in Brazil. This course is praised for forming remarkable professionals who are recognized worldwide. It is not just my dream that you have made possible, but the countless people for whom this course has been a life-changing experience. We know that this was a collective effort, but your role was vital. Your way of showing that politics is a part of everything, and that we need to be gentle but correct, made all the difference. Thank you. *Muito Obrigado.*

I sincerely thank my friends and colleagues from my undergraduate studies in acoustical engineering (USFM/EAC) and the master's (UFSC/LVA). Your support, happiness, patience and encouragement have been invaluable throughout this journey. Thank you for helping me to develop my skills and knowledge and for being such a positive influence on my academic career. I am deeply grateful for all you have done for me and look forward to continuing our professional relationship. I want to thank the oldest friends, Fabrício and Juliano, and the Panteon. I value the bond and history that we share. Thank you for being such wonderful friends.

I want to express my heartfelt thanks to the Brazilian CNPq and the government (Lula/Dilma) policies that support students with low income and from public schools. With their financial assistance, I could pursue my studies and achieve my goals. I am deeply grateful for their support and the opportunity to receive a quality education. I would also like to express my gratitude to Marie Skłodowska-Curie Actions for their support of my doctoral education. Their reference programme for doctoral education has provided me with invaluable resources and opportunities, and I am extremely grateful for their support. Thank you for helping me achieve my goals and being a valuable part of my academic journey.

I want to express my gratitude to all those who will read this thesis in the future. Your time and attention are greatly appreciated. I wish you a good reading experience and hope that you will find the ideas and research presented in this work to be both thought-provoking and beneficial. Thank you again for considering this work.

Author's Declaration

This thesis is the result of the author's original research. Chapter 4 is a collaboration work with Tirdad Seifi-Ala. The author has composed it and has not been previously submitted for any other academic qualification.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie-Skłodowska-Curie grant agreement No 765329; The funder had no role in study design.

Sergio Luiz Aguirre

Contents

List of Tables

List of Figures

Nomenclature

General Symbols

C_{50} Clarity: the ratio between the first 50 ms of the RIR and from 50 ms to the end, Eq. (2.12), page 36.

C_{80} Clarity: the ratio between first 80 ms of the RIR and RIR from 80 ms to the end, Eq. (2.12), page 36.

D_{50} Clarity: ratio between first RIR 50 ms of a RIR and the complete RIR, Eq. (2.14), page 37.

D_{80} Clarity: ratio between first 80 ms of a RIR and the the complete RIR, Eq. (2.14), page 37.

g Gain matrix, page 25.

$h(t)$ Impulse response energy in time domain, Eq. (2.10), page 35.

$h_{\mathrm{b}}(t)$ RIR measured with a pressure gradient microphone, Eq. (2.16), page 38.

$h_L(t)$ Impulse responses collected from the left ear, page 39.

$h_R(t)$ Impulse responses collected from the right ear, page 39.

$\boldsymbol{l}_1$ Vector from center point to channel 1, Eq. (2.3), page 25.

$\boldsymbol{l}_2$ Vector from center point to channel 2, Eq. (2.3), page 25.

L_{12} Speaker Position Matrix (Channels), page 25.

m Ambisonics components order, Eq. (2.9), page 30.

N number of necessary sources to Ambisonics reproduction, Eq. (2.9), page 30.

$\boldsymbol{p}$ Vector from center point to virtual font, Eq. (2.3), page 25.

p p-value for the t-statistic of the hypothesis test that the corresponding coefficient is equal to zero or not., page 114.

$p_L^n(t)$ Bandpassed left impulse response, Eq. (3.7), page 71.

$p_R^n(t)$ Bandpassed right impulse response, Eq. (3.7), page 71.

$p_L(t)$ Impulse response at the entrance of the left ear canal, Eq. (3.5), page 70.

$p_R(t)$ Impulse response at the entrance of the right ear canal, Eq. (3.5), page 70.

RT Reverberation time, page 34.

RT_{60} Reverberation time, page 34.

$s(t)$ Arbitrary sound source signal, page 29.

$\mathrm{S}_l(t)$ Time signal recorded with the set microphone and the loudspeaker l, page 66.

SE Standard error of the coefficients, page 114.

T_{20} Reverberation Time (T_{60}) extrapolated from 25 dB of energy decay, Eq. (2.10), page 35.

T_{30} Reverberation Time (T_{60}) extrapolated from 35 dB of energy decay, Eq. (2.10), page 35.

t Time, page 12.

t_s Center Time, Eq. (2.15), page 37.

T_{60} Reverberation Time, Eq. (2.10), page 34.

$v(t)_{1\mathrm{kHz}}$ Sinusoidal 1 k Hz signal recorded from the calibrator in VFS, Eq. (3.2), page 66.

V Volume of the room, Eq. (2.10), page 34.

$v_l(t)$ Calibrator signal recorded in the left ear, Eq. (3.1), page 65.

$v_r(t)$ Calibrator signal recorded in the right ear, Eq. (3.1), page 65.

Greek Symbols

$\alpha_{l,\mathrm{rms}}$ Calibration factor for the left ear, Eq. (3.1), page 65.

$\alpha_{r,\mathrm{rms}}$ Calibration factor for the right ear, Eq. (3.1), page 65.

$\bar{\alpha}$ Averaged Absorption Coefficient, Eq. (2.10), page 34.

Γ_l Level factor to the loudspeaker l, Eq. (3.3), page 66.

ω Angular frequency, page 12.

ϕ Elevation angle related to the ears axis of the listener, page 29.

θ Azimuthal angle related to the ears axis of the listener, page 29.

Mathematical Operators and Conventions

β Fixed-effects regression coefficient, page 114.

e Exponential function, where $\mathrm{e}^{(1)} \approx 2,7182$, page 12.

$\int$ Integral, page 12.

j $\sqrt{-1}$, imaginary operator, page 12.

τ Time delay, Eq. (2.18), page 39.

t_x t-statistic for each coefficient to test the null hypothesis, page 114.

$Y_n^m(\theta, \phi)$ Spherical harmonics function of order n and degree m, Eq. (2.7), page 29.

$\mathrm{L_{eq}}$ Equivalent continuous sound level, page 103.

max() Function that returns the element with the maximum value for a sequence of numbers, or for a vector, Eq. (2.18), page 39.

RMS() Root mean square, Eq. (3.2), page 66.

Acronyms and Abbreviations

2D Two-dimensions in space, page 24.

3D Three-dimensions in space, page 24.

vs. From Latin *Versus* is the past participle of *vertare.* which means “against” and “as opposed or compared to., page 81.

AD/DA Analog-to-Digital Digital-to-Analog converter, page 59.

AR Augmented reality, page 27.

ASW	Apparent Source Width, page 38.
BRIR	Binaural room impulse response, page 15.
CTC	Cross-talk cancellation, page 23.
dB HL	Hearing Loss in decibels, page 102.
DBAP	Distance-Based Amplitude Panning, page 47.
DS	Direct sound, page 15.
EcoEG	Combination study number 3: Eco (Reverberation/Ecological) and EEG, page 97.
EEG	Electroencephalogram, page 50.
FFRs	Brainstem frequency responses, page 50.
FFT	Fast Fourier transform, page 70.
FIR	Finite Impulse Response, page 71.
HATS	Head and torso simulator, page 60.
HC	B& K 4128 HATS at center position, page 79.
HC K+X	B& K 4128 HATS at center position and KEMAR at X cm to the left, page 80.
HC K-X	B& K 4128 HATS at center position and KEMAR at X cm to the right, page 79.
HEAR-ECO	Innovative Hearing Aid Research – Ecological Conditions and Outcome Measures, page 97.
HINT	Hearing in Noise Test, page 103.
HOA	Higher Order Ambisonics, page 31.
HRTF	Head-Related Transfer Function, page 12.
IACC	Interaural Cross-Correlation Coefficient, page 39.
IACF	Interaural cross-correlation function, Eq. (3.5), page 70.
ILD	Interaural Level Difference, page 10.
IPD	Interaural Phase Difference, page 10.

ITD	Interaural Time Difference, page 10.
ITF	Interaural Transfer Function, page 14.
JND	Just noticeable difference, page 93.
KEMAR	Knowles Electronics Manikin for Acoustic Research, page 60.
LEF	Lateral Energy Fraction, Eq. (2.16), page 38.
LEV	Listener Envelopment , page 39.
LG	Lateral Strength, Eq. (2.17), page 39.
LMM	Linear Mixed-effect Model, page 113.
LPF	Low-pass filter, page 73.
LTI	Linear and Time-Invariant System, page 33.
MDAP	Multiple-Direction Amplitude Panning, page 47.
MOA	Mixed Order Ambisonics, page 52.
MTF	Monaural Transfer Function, page 13.
NSP	Nearest Speaker, page 52.
PLE	Perceptual Localization Error, page 41.
PTA4	Four bands pure tone audiometry, page 102.
RIRs	Room impulse response, page 15.
SH	Spherical Harmonics, page 28.
SPL	Sound pressure level, page 35.
SRT	Speech Reception Threshold, page 52.
VBAP	Vector-Based Amplitude Panning, page 23.
VBIP	Vector-Based Intensity Panning, page 41.
VFS	Volts full scale, page 66.
VSE	Virtual Sound Environment, page 18.
W	Omnidirectional channel, Eq. (2.9), page 29.
WFS	Wave Field Synthesis, page 31.

X Bi-directional pattern channel towards the source, Eq. (2.9), page 29.

Y Bi-directional pattern channel perpendicular to the source in azimuth, Eq. (2.9), page 29.

Z Bi-directional pattern channel perpendicular to the source in elevation, Eq. (2.9), page 29.

Chapter 1

Introduction

Individuals with normal hearing often can effortlessly comprehend complex listening scenarios involving multiple sound sources, background noise, and echoes [226]. However, those with hearing loss may find these situations particularly challenging [273, 289, 304, 317]. These environments are commonly encountered in daily life, particularly during social events. They can negatively impact the communication abilities of individuals with hearing loss [137, 260]. The difficulties associated with understanding complex listening scenarios can be a significant barrier for individuals with hearing loss, leading to reduced participation in social activities [16, 63, 119].

1.1 Motivations

Several hearing research laboratories worldwide are developing systems to realistically simulate challenging scenarios through virtualization to better understand and help with these everyday challenges in people's lives [41, 79, 102, 116, 118, 160, 161, 188, 195, 218–220, 259, 272, 298] The virtualization of sound sources is a powerful tool for auditory research capable of achieving a high level of detail, but current methods use expensive, expansive technology [293]. In this work, a new auralization method has been developed to achieve sound spatialization with a reduction in the technological hardware requirement, making virtualization at the clinic level possible.

1.2 Aims and Scope

Overall the objective of the research was to investigate parameters of sound virtualization methods related to its localization accuracy, especially the perceptually based ones [39], in their optimal but also in challenging conditions. Furthermore, an auralization method oriented to a smaller setup to reduce the hardware requirements is proposed.

The specific objectives were:

- To investigate spatial distortions through binaural cue differences in two well-known virtualization setups (Vector-Based Amplitude Panning and Ambisonics (VBAP)).
- To investigate the influence of a second listener inside the sound field (VBAP and Ambisonics).
- To evaluate the feasibility of a speech-in-noise test within Ambisonics virtualized reverberant rooms.
- To study the relation between reverberation, signal-to-noise ratio (SNR), and listening effort in environments virtualized in first-order Ambisonics.
- To investigate the binaural cues, objective level and reverberation time for a new auralization method utilizing four loudspeakers.

- To investigate the influence of hearing aids on binaural cues and objective parameters within virtualized scenes utilizing the new auralization method with an appropriate setup.

The main objective of this research was to examine various parameters of sound virtualization methods related to their localization accuracy, with a focus on perceptually-based methods [39], in optimal and challenging conditions. Additionally, a new auralization method was proposed for a smaller setup to reduce hardware requirements. The specific goals of the research included:

- Examining spatial distortions through differences in binaural cues in two well-known virtualization setups (Vector-Based Amplitude Panning and Ambisonics (VBAP)).
- Evaluating a second listener's impact within the sound field (VBAP and Ambisonics).
- Assessing the feasibility of a speech-in-noise test within Ambisonics virtualized reverberant rooms.
- Investigating the relationship between reverberation, signal-to-noise ratio (SNR), and listening effort in environments virtualized using first-order Ambisonics.
- Using four loudspeakers, propose an auralization method, measure it, analyze objective parameters against existent methods.
- Test and analyze the influence of acquiring signals with hearing aids microphones on virtualized scenes using the new auralization method with a four-loudspeaker virtualization setup.

1.3 Contributions

The main contribution of this research to the scientific field of auditory perception is the development of a new auralization method that addresses the current gap in the virtualization of sound sources using a small number of loudspeakers. Specifically, this method aims to achieve both good localization accuracy and a high level of immersion simultaneously, which has been a challenge in previous approaches. Furthermore, the proposed method combines existing techniques. It can be implemented using readily available hardware, requiring a minimum of four loudspeakers. This technology makes it more accessible for audiologists and researchers to create realistic listening scenarios for patients and participants while reducing the technical resources required for implementation. Overall, this work represents a valuable contribution to the field of auditory perception and has the potential to advance the understanding of spatial hearing and the development of effective hearing solutions.

1.4 Organization of the Thesis

In Chapter 2, a review examines previous work carried out in several different areas concerning virtualization and the auralization of sound sources. The chapter starts with an overview of the basic concepts of human sound perception. Next, virtual acoustics are explored, reviewing the generation of virtual acoustic environments using different rendering paradigms and methods. In addition, relevant room acoustics concepts and objective parameters, and their relation to hearing perception, are described. Finally, the review considers auralization and virtualization as applied to auditory research. This review stresses the importance of virtual sound sources for greater realism and ecological validity in auditory research and the challenges of adequately creating a virtual environment focused on auditory research.

Chapter 3 presents an investigation of binaural cue distortions in imperfect setups. First, the methods are described, including the complete auralization of signals using two different methods and the system's calibration. The investigation first compares both auralization methods through the same calibrated virtualization setup in terms of spatial distortions. Then the spatial cues are examined with the addition of a second listener to the virtualized sound field. Both investigations are performed with the primary listener on and off-center.

In Chapter 4, a behavioral study examines subjective effort within virtualized sound scenarios. As the study was part of a collaborative project, only one auralization method was selected, first-order Ambisonics. The aim was to examine how SNR and reverberation combine to affect effort in a speech-in-noise task. Also, the feasibility of using first-order Ambisonics was examined. However, the sound sources were well separated in space, and localization accuracy was not a factor. An important aspect of the study was an auralization issue involving head movement observed during pilot data collection. This issue led to a solution that allowed the study to continue. The results verified the relationships between subjective effort and acoustic demand. Furthermore, this issue led to the further investigation of the effect of off-center listening, considered in both Chapter 3 and Chapter 5.

In Chapter 5, a hybrid method of auralization is proposed combining the methods examined and used in previous chapters: VBAP and Ambisonics. This method was designed to allow auralized signals to be virtualized in a small reproduction system, thus providing better accessibility to research within the virtualized sound field in clinics and research centers that do not have a sizeable acoustic apparatus. The hybrid auralization method aims to unite the strengths of both techniques: localization by VBAP and immersion by Ambisonics. Both of these psychoacoustic strengths are related to the room's impulse response. The hybrid method convolves the desired signal with distinct parts of an Ambisonics-format impulse response that characterizes the desired environment. The potential for generating auralizations for a reproduction system with at least four loudspeakers is demonstrated. The virtualization system was tested with three different scenarios. Parameters relevant to the perception of a scene, such as reverberation time, sound pressure level, and binaural cues, were evaluated in different positions within the speaker arrangement. The effects of a second participant inside the ring were also investigated. The evaluated parameters were as expected, with the listener in the system's center (sweet spot). However, deviations and issues at specific presentation angles were identified that could be improved in future implementations. Such errors also need to be further investigated as to their influence on the subjective perception of the scenario, which was not performed due to the COVID-19 pandemic. An alternative robustness assessment was performed offline, examining the localization accuracy with a model proposed by May *et al.* [182] The method also proved effective for tests with hearing aids for listeners positioned in the center of the speaker arrangement. However, the method performance considering hearing instruments with compression algorithms and advanced signal processing still needs to be verified.

Chapter 6 presents a general discussion of the feasibility of applying tests using the proposed method and an overview of the processes. In addition, the relevant contributions of the work are presented, as are the limitations and the suggestions for further improvements.

Chapter 2

Literature Review

2.1 Introduction

The field of audiology is concerned with the study of hearing and hearing disorders, as well as the assessment and rehabilitation of individuals with hearing loss [110]. In this review chapter, we will explore various topics related to human binaural hearing, spatial sound, and virtual acoustics to provide a comprehensive overview of the current state of knowledge in these fields and highlight their important contributions to our understanding of hearing and auditory perception. First, we will delve into the intricacies of human binaural hearing. Next, we will examine the concepts of spatial hearing, including the various binaural and monoaural cues that contribute to our ability to localize sound in space. We will also explore the head-related transfer function, which describes the way that sounds are filtered as they travel from their source to the ear drum, as well as the subjective aspects of audible reflections. Next, we will turn our attention to spatial sound and virtual acoustics. We will discuss the virtualization of sound, including the various methods used to achieve this, such as auralization and virtual sound reproduction. We will also examine the different auralization paradigms used in auditory research, including binaural, panorama, vector-based amplitude panning, ambisonics, and sound field synthesis. We will then examine the role of room acoustics in virtualization, and auditory research, including the various parameters, used to describe room acoustics, such as reverberation time, clarity and definition, center time, and parameters related to spatiality. Finally, we will explore the use of loudspeaker-based virtualization in auditory research, including hybrid methods and sound source localization, as well as the assessment of listening effort.

2.2 Human Binaural Hearing

The engineering side of the listening process can be simplified modeled through two input blocks separated in space [92]. These inputs, frequency, and level are limited and are followed by a signal processing chain that relates the medium transformations for the wave propagation from air to fluid and electrical pulses [315].

Although this block modeling can be reasonably accurate for educational purposes, it falls short of capturing the true effect and importance of listening on our essence as human beings. The ability to feel and interpret the world through the sense of hearing, and to attribute meaning to sound events, enables humans to enrich their tangible world [56, 244]. For instance, a characteristic sound can evoke memories or trigger an alert [128]. A piece of music can bring tears to one's eyes or persuade someone to purchase more cereal [13, 114]. A person's voice can activate certain facial nerves, turning hidden teeth into a smile. These are some of the reasons why researchers and clinicians dedicate their lives to understanding the transformation of sound events into auditory events, with a scientific dedication focused on creating solutions and opening opportunities for more people to experience the sound they love and deserve - **a dedication focused on people and their needs**.

As the auditory system comprises two sensors, normal-hearing listeners can experience the benefits of comparing sounds autonomously, relating them to the space around them [21]. This constant signal comparison is the main principle of binaural hearing, where the differences between these sounds allow for the identification of the direction of a sound event, as well as the sensation of sound spatiality [9, 40]. Usually, these signals are assumed to be part of a linear and time-invariant system, which helps to study how humans interpret the information present in the different signals across the time and frequency domains. However, this assumption of linearity can fail when analyzing fast sound sources, reflective surfaces, or sound propagating through disturbed air [200, 255]. Nonetheless, the advantages of quantifying and capturing the effect have led to significant progress in hearing sciences.

2.2.1 Spatial Hearing Concepts

Identifying the direction of incidence of a sound source based on the audible waves received by the listener is defined as an act or process of human sound localization [285]. For research in acoustics, it is relevant to acknowledge that the receiver is, in general, a human being. The human hearing mechanism's main anatomical characteristic is the binaural system. There are two signal reception points (external ears positioned on opposite sides of the head). Albeit, the whole set (torso, head, hearing pavilions) can also modify the signal that reaches the two tympanic membranes at some extent [153, 216]. Human binaural hearing and associated effects have been extensively reported by Blauert [38].

In addition to analyzing sound sources' spatial location, the central auditory system extracts real-time information from the sound signals related to the acoustic environment, such as geometry and physical properties [153]. Another benefit is the possibility of separating and interpreting combined sounds, especially from sources in different directions [170, 242].

2.2.2 Binaural cues

The sound propagation speed in the air can be assumed to be finite and approximately constant, considering it as an approximately non-dispersive medium [18]. Thus, when the incidence is not directly frontal or rear, the wavefront travels through different paths to the ears, reaching them at different times. The time interval between that a sound takes to arrive on both ears is commonly expressed in the literature as Interaural Time Difference (ITD) [39]. It is crucial cue for sound source localization in low-frequency sounds [39, 153, 242]. Moreover, it is considered the primary localization cue [306]. For continuous pure tone signals and other periodic signals, the ITD can be expressed as the time Interaural Phase Difference (IPD) [285]. On the other hand, most mammals' high-frequency sound source localization is based on a comparative analysis of sound energy in each ear's frequency bands, the Interaural Level Difference (ILD). The named duplex theory surmises ITD cues as the basis to sound localization of low-frequency and ILD cues to high-frequency. The authorship of this is assigned to Lord Rayleigh at the beginning of the last century [246]. These binaural cues are related to the azimuthal position. However, they do not present the same success explaining the localization on elevated positions [37, 250]. An ambiguity in binaural cues caused by head symmetry and referred to as the cone of confusion [296] can create difficulties to a correct sound source localization. The cone of confusion is the imaginary cone extended sideways from each ear where sound source locations will create the same interaural differences (see Figure 2.1).

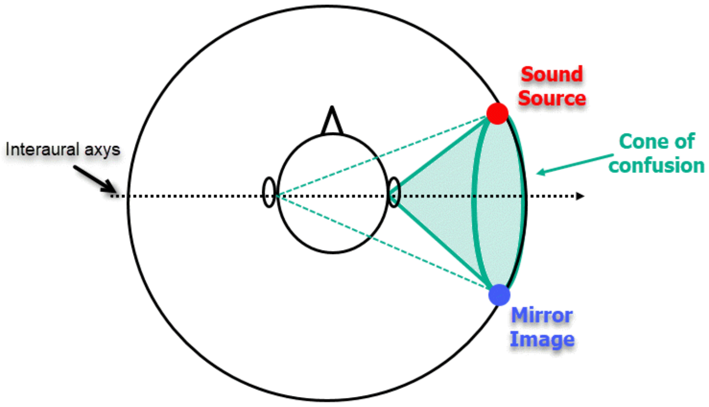


Figure 2.1: *Two-dimensional representation of the cone of confusion.*

Head movements are essential for resolving the ambiguous cues from sound sources located on the cone of confusion. As the person moves their head, they change the reference and the incidence angle helping them to solve the duality. This change is reflected in the cues associated with directional sound filtering caused by the human body's reflection, absorption, and diffraction.

2.2.3 Monaural cues

Monoaural cues are related to spatial impression, especially in the localization of elevated sound sources. These cues give, to some extent, some limited but crucial localization abilities to people with unilateral hearing loss [72, 307]. This type of cue is centered on instant level comparison and frequency changes. As the level of a continuous enough sound source changes, the approximation or distancing of that source can be estimated. Furthermore, when there are head movements that shape the frequency content, the disturbance, mainly the pinnae provide, can benefit the listener to learn the position of a sound source [129, 292]. In addition, the importance of the previous knowledge of the sound to the deconvolution process is also investigated, revealing mixed results [307].

2.2.4 Head-related transfer function

The Head-Related Transfer Function (HRTF) describes the directional filtering of incoming sound due to human body parts such as the head and *pinnae* [189]. The free-field HRTF can be expressed as the division of the impulse responses in the frequency domain measured at the entrance to the ear canal and the center of the head but with the head absent [108] (see Figure 2.2).

HRTFs depend on the direction of incidence of the sound and are generally measured for some discrete incidence directions. Mathematical models can also generate individualized HRTFs based on anthropometric measures [52] or through geometric generalization [70].

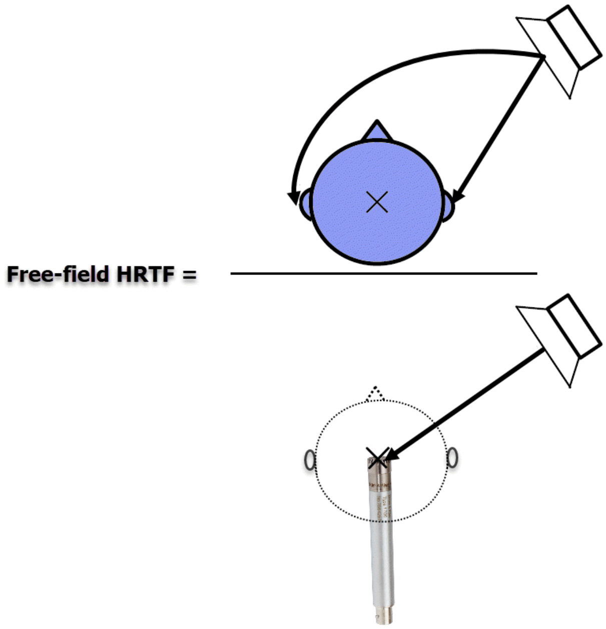


Figure 2.2: *A descriptive definition of the measured free-field HRTF for a given angle.*

The referential system related to the head can be seen in Figure 2.3, where β is the elevation angle in the midplane, and ϕ is the angle defined in the horizontal plane.

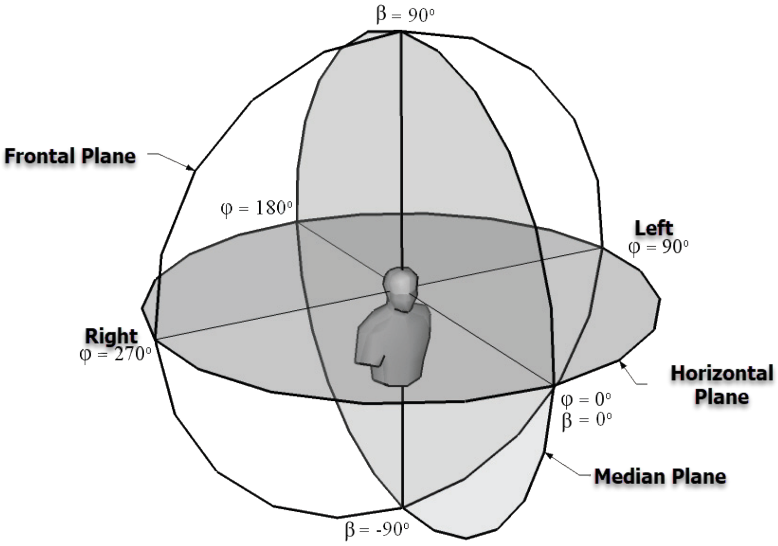


Figure 2.3: *Polar coordinate system related to head incidence angles, adapted from Portela [240].*

Suppose the distance to the sound source exceeds 3 meters. In that case, it can be considered approximately a plane wave, thus making the previous HRTFs almost independent of the distance to the sound source [38]. Blauert [39] also explain two other types of HRTF, namely:

- Monaural Transfer Function (MTF): relates the sound pressure, at a measurement point in the ear canal, from a sound source at any position to a sound pressure measured at the same point, with a sound source at a reference position ($\phi = 0$ and $\beta = 0$). MTF is given by

$$\mathrm{MTF} = \frac{\left(\frac{P_i}{P_1}\right)_{r,\phi,\beta,f}}{\left(\frac{P_i}{P_1}\right)_{\phi=0^\circ,\beta=0^\circ,f}}, \tag{2.1}$$

  where p_i it can be p_1, p_2, p_3 or p_4.

  - p_1 sound pressure in the center of the head position with the listener absent;

  - p_2 sound pressure at the entrance of the occluded ear canal;
  - p_3 sound pressure at the entrance to the ear canal;
  - p_4: eardrum sound pressure.

- Interaural Transfer Function (ITF): relates the sound pressures at corresponding measurement points in the two auditory canals. The reference pressure will then be the ear that is directed towards the sound source. The ITF can be obtained through

$$\mathrm{ITF} = \frac{P_{i \text{ Opposite side of the source}}}{P_{i \text{ Side facing the source}}}. \tag{2.2}$$

More considerable variations are seen above 200 Hz in HRTFs [293] because the head, torso, and shoulders begin to significantly interfere in frequencies up to approximately 1.5 kHz (mid frequencies). In addition, the pinna and the cavum conchae (space inside the most inferior part of the helix cross; it forms the vestibule that leads into the external acoustic meatus [270]) distort frequencies greater than 2 kHz.

HRTF measurements vary from person to person, as seen in Figure 2.4, where TS 1, TS 2, TS 3, and TS 4. represent HRTFs of different people. When recording using mannequins or different people's ear canals (non-individualized HRTFs), the reproduction precision in terms of spatial location and realism tends to be diminished [51, 178]. This poorer precision is because the transfer function will differ for each individual, especially at high frequencies [155]. This dependence is related to the wavelength and the singular irregularity of the ear canal of each human being [38].

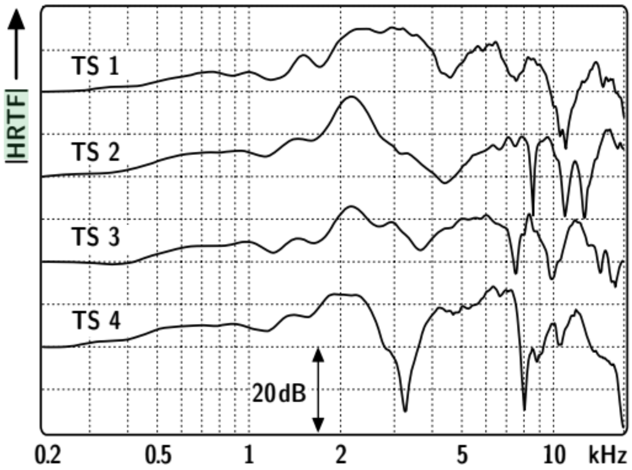


Figure 2.4: *Head-related transfer functions of four human test participants, frontal incidence, from Vorländer [293].*

Binaural Impulse Response A Binaural Room Impulse Response (BRIR) results from a measurement of the response of a room to excitation from an (ideally) impulsive sound [183]. The BRIRs are composed of a sequence of sounds. Parameters like the magnitude and the decay rate, the phase, and time distribution are the key to understanding how a BRIR can audibly characterize a room to a human perception [167]. Albeit the air contains a small portion of Co2 that is dispersive, sound propagation velocity can be considered homogeneous in the air (non-dispersive medium) [312] for the Room Impulse Responses (RIRs). The first sound from a source that reaches a receptor inside the room travels a smaller distance, and it is called direct sound (DS). Usually, the following sounds result from reflections that travel a longer path, losing energy on each interaction and resulting in an exponential decay of magnitude. The BRIR is proposed to collect the room information as a regular Impulse Response, although having two sensors separated as the typical human head. Nowadays, BRIR can be recorded with small microphones placed in the ear canal of a person or utilizing microphones placed in mannequins [197].

A BRIR is the auditory time representation of a set source-receptor defined by its position, orientation, acoustic properties as directionality of the sound source, as well as from the physical elements within the environment [38, 108]. The convolution of BRIR with audio signals is a feasible task for modern computation, which allows the creation and manipulation of sounds even in real-time applications [62, 217]. Thus, it is possible to impose spatial and reverberant characteristics of different spaces to a given sound [109].

2.2.5 Subjective aspects of an audible reflection

The impulse response is composed of the direct sound followed by a series of reflections (initial and later reflections) [45, 165]. Essential knowledge on how the human auditory system processes the spectral and spatial information contained in the impulse response has been obtained through studies with simulated acoustic fields. [6, 17, 93, 125, 141, 174, 176, 188, 193, 257, 305, 321]. The results of Barron's experiments, depicted in Figure 2.5, involved the reproduction of both a direct sound and a lateral reflection. These two auditory stimuli were manipulated in terms of their time delay and relative amplitude, with the goal of eliciting subjective impressions correlated to these factors. By varying the time between the direct sound and the reflection, as well as the relative amplitude of these stimuli, it was possible to understand better how these characteristics impact the overall auditory experience.

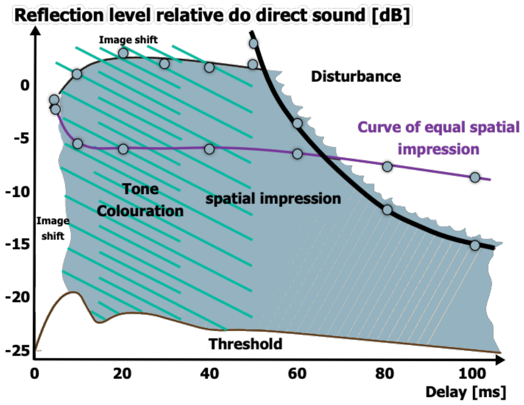


Figure 2.5: *Audible effects of a single reflection arriving from the side (adapt from Rossing [254]).*

The audibility threshold curve indicates that the reflection will be inaudible if the delay or the relative level is minimal. The reflection's subjective effect also depends on the direction of incidence of the sound source in the horizontal and vertical plane. It is possible to note that for delays of up to 10 milliseconds, the relative difference in level must be at least -20 dB for the reflection to be noticeable.

The echo effect is typically observed in delays of more than 50 milliseconds, being an acoustic repetition with a high relative level—approximately the same energy as the direct sound. The coloring effect is associated with the significant change in the spectrum caused by the constructive and destructive interference of the superposition of sound waves.

The image change happens when there are reflections with relative levels higher than direct sound or minimal delays. In this case, the subjective perception is that the sound source has a different position in space than the visual system perceives.

2.3 Spatial Sound & Virtual Acoustics

The sound perceived by humans is identified and classified based on physical properties, such as intensity and frequency [242]. Human beings are equipped with two ears (two highly efficient sound sensors), enabling a real-time comparison of these properties between the captured sound signals [9]. The sounds and the dynamic interaction between sound sources, their positions, movements, and the physical interaction of the generated sound waves with the environment can be perceived by normal-hearing people, providing what is called spatial awareness [153]. That auditory spatial awareness includes the localization of the sound source, estimation of distance, and estimation of the size of the surrounding space [38, 305]. A person with hearing loss may lose this ability partially or entirely; the spatial awareness is also tied to the listener's experience with the sound and the environment, motivation, or fatigue level [54, 304].

In the field of virtual acoustics, the ultimate goal is to generate a sound event that elicits a desired auditory sensation, creating a Virtual Sound Environment (VSE) [293]. In order to achieve this, it is necessary to synthesize or record the acoustic properties of the target scene and subsequently reproduce them in a manner that accurately reflects the original acoustic conditions [97]. This involves a careful consideration of the various factors that contribute to the overall auditory experience, including the spectral and spatial characteristics of the sound. By accurately recreating these properties, it is possible to create a highly immersive and realistic VSE that effectively conveys the intended auditory experience to the listener [196, 213, 293].

2.3.1 Virtualization

Nowadays, it is possible to create audio files containing information about sound properties related to a specific space [293]. For example, it is possible to encode information about the source and receptor position, the transmission path, reflections on surfaces, and the amount of energy absorbed and scattered (*e.g.,* Odeon [59], a commercially available acoustical software). The sound field properties can be simulated, synthesized, or recorded *in-situ* [113, 293]. These signals can be encoded and reproduced correctly in various reproduction systems [122, 161]. The creation of files that can be reproduced containing such information is called auralization. As different interpretations of the terms occur in literature, in this thesis, the virtualization process is considered to encompass the auralization and the reproduction of a sound (recorded, simulated, or synthesized) that includes spatial properties.

2.3.1.1 Auralization

Auralization is a relatively recent procedure. The first studies were conducted in 1929; Spandöck and colleagues tried to process signals measured in a scale-created room. After that, in 1934, Spandöck [280] succeeded in the first auralization, in the analogical way, using ultrasonic signals of scale models recorded in magnetic tapes. In 1962 Schroeder [263] incorporated the computing process into the auralization. In 1968 Krokstad [146] developed the first acoustic room simulation software. The term auralization was introduced in the literature by Kleiner in 1993: "Auralization is the process of rendering audible, by physical or mathematical modeling, the soundfield of a source in a space, in such a way as to simulate the binaural listening experience at a given position in the modeled space." (Kleiner [138])

In his book titled Auralization, published in 2008, Vorländer defined: "Auralization is the technique of creating audible sound files from numerical (simulated, measured, or synthesized) data." (Vorländer [293])

In this work, auralization is understood as a technique to create files that can be executed as perceivable sounds. An auralization method describes the technique; it can involve one or more auralization techniques. These sounds can then be virtualized (reproduced) via loudspeakers or headphones and provide audible information about a specific acoustical scene in a defined space, following Vorlander's definition. That was chosen to encourage the separation of the process as an auralized sound file can contain information that allows it to be decoded in different reproduction systems [320].

Auralization is consolidated in architectural acoustics [45, 148, 165, 254], and it is also emerging in environmental acoustics [19, 68, 69, 139, 162, 231, 232]. This technique allows a piece of audible information to be easily accessed and understood. It is also an integral part of the entertainment industry in games, movies, and virtual or mixed reality [320]. Knowing an environment's acoustic properties allows it to manipulate or add synthesized or recorded elements, leading the receiver to the desired auditory impression, including the sound's spatial distribution [62]. This process is also used in hearing research, allowing researchers to introduce more ecologically valid sound scenarios to their study (See Section 2.3.4).

Sound spatiality, or the perception of sound waves arriving from various directions and the ability to locate them in space, is a crucial aspect of the auditory experience [40]. Auralization, which is analogous to visualization, involves the representation of sound fields and sources, the simulation of sound propagation, and the strategy to decode in the spatial reproduction setup [293]. That is typically achieved through tri-dimensional computer models and digital signal processing techniques, which are applied to generate auralizations that can be reproduced via acoustic transducers [293].

The modeling paradigm used to create the spatial sensation can be perceptually or physically based [39, 106, 164, 276]. Multiple dimensions influence sound perception; the type of generation of the sound, the wind direction, the temperature, the source and the receptor movement, space (size, shape, and content), receptor's spatial sensitivity, and source directivity are some examples. That implies the importance of physical effects as Doppler shifts [96, 284, 293]. Furthermore, the review of room acoustic and psychoacoustics elements (See Section 2.3.3) corroborates the auralization modeling procedure's understanding.

2.3.1.2 Reproduction

Sound signals containing acoustic characteristics of a space can be reproduced either with binaural techniques (headphones or loudspeakers) or with multiple loudspeakers (multichannel techniques) [293]. Moreover, an acoustic model for a space can be analytically or numerically implemented, having a series of competent algorithms and commercial software and tools available [49]. With that, it is also possible to measure micro and macro acoustic properties for materials in a laboratory or *in-situ* [206] and access databases of various coefficients and indexes to an extended catalog of materials [50, 71, 158, 266].

On the reproduction end of the virtualization process, factors such as frequency and level calibration, signal processing, and the frequency response of the hardware can significantly impact the accuracy of the final sound (e.g., the orientation/correction of the microphone when calibrating the system [274]). Depending on the chosen paradigm, a lack of attention to these details may disrupt an accurate description of the sound field, sound event, or sound sensation [214, 282, 283, 320]. Additionally, the quality of the stimuli may be compromised depending on the chosen reproduction technique, which is often tied to the hardware available [77, 166, 275, 276]. That can lead to undesired effects on the level of immersion and problems with the accuracy of sound localization and identification (*e.g.*, source width, source separation, sound pressure level, and coloration and spatial confusion effects [97]). The process of building a VSE is called sound virtualization, which involves both the auralization and reproduction stages to create audible sound from a file. The main technical approaches or paradigms for reproducing auralized sound are Binaural, Panorama, and Sound Field Synthesis (Section 2.3.2). These paradigms can be distinguished by their output, which can be physically or perceptually motivated. For example, while binaural methods are treated apart, they can be intrinsically classified in a physically-motivated paradigm since its success relies on reproducing the correct physical signal at a specific point in the listener's auditory system, typically the entrance of the ear canal [106].

2.3.2 Auralization Paradigms

2.3.2.1 Binaural

Binaural hearing, which refers to the ability to perceive sound in a three-dimensional auditory space, is a fundamental concept in auditory research and has been extensively studied by researchers such as Blauert [40]. In the context of auralization, the term "binaural" refers to the specific paradigm that aims to reproduce the exact sound pressure level of a sound event at the listener's eardrums. That can be achieved through the use of headphones or a pair of loudspeakers (known as transaural reproduction) [314]. However, when using distant loudspeakers, it is necessary to consider the interference that can occur between the sounds coming from each speaker. To mitigate this issue, techniques such as cross-talk cancellation (CTC) [60, 262] can be employed, which involve manipulating a set of filters to cancel out the distortions caused by the sound from one speaker reaching the other ear. Another form of binaural reproduction involves the use of closer loudspeakers that are nearfield compensated.

Binaural methods over headphones is commonly applied. It requires no extensive hardware (in simple setups that do not track the listener's head), providing a valid acoustic representation and spatial awareness [293]. A Disadvantage of this method can include the accuracy dependence of individualized HRTF (as each human being have his own slightly different anatomic "filter set") [314]. Over headphones also, the movement of the listener's head can be disruptive to the immersion [179]. It may require tracking the head's movement [11, 115, 252], *e.g.*, when movements are required or allowed in an experiment. Furthermore, a listener wearing a pair of headphones may not represent a realistic situation. For example, an experiment with a virtual auditory environment that represents a regular daily conversation with aged participants may lose the task's ecological validity. Also, usually, headphones prevent the listener from wearing hearing devices. Figure 2.6 illustrates the main idea behind different binaural reproduction setups.

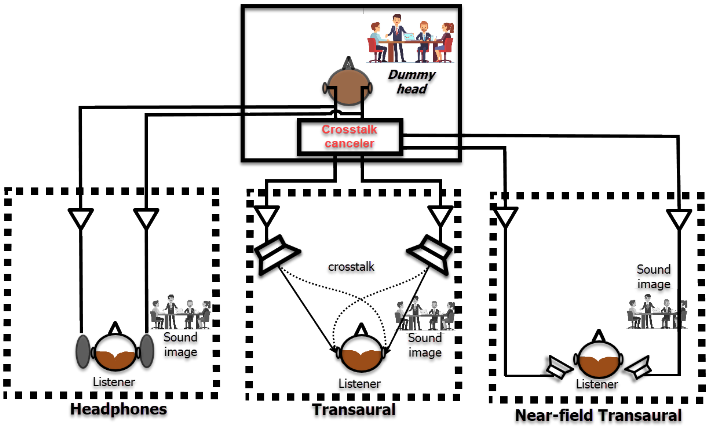


Figure 2.6: *Binaural reproduction setups: Headphones, transaural and near-field transaural (Adapted from Kang and Kim [131]).*

2.3.2.2 Panorama

The Panorama paradigm encompasses auralization methods focused on delivering accurate ITDs and ILDs at the listener's position, also known as stereophonic techniques [106, 276]. The most well-known methods are based on amplitude panning [180], including Low-order Ambisonics [91] and Vector-Based Amplitude Panning (VBAP) [241]. High Order Ambisonics is an extension of the Ambisonics method, which is not typically considered a panning method but rather a sound field synthesis method (See Section 2.3.2.3). VBAP employs local panning by rendering sound using pairs or triplets of loudspeakers. In contrast, Ambisonics uses global panning to produce a single virtual source using all available loudspeakers [282].

Vector Based Amplitude Panning: The Vector-Based Amplitude Panning (VBAP) is a first-order approximation of the composition of emitted signals that creates virtual sources [241]. The virtualization process using VBAP is based on amplitude panning in two dimensions (variation in amplitude between the speakers), which is derived from the Law of Sines and Law of Tangents (see Benesty et al. [23] for a derivation of these laws). The original hypothesis of VBAP assumes that the speakers are arranged symmetrically, equidistant from the listener, and in the same horizontal plane. VBAP does not limit the number of usable speakers but uses a maximum of 3 simultaneously. The speakers are arranged in a reference circle (2D case) or sphere (3D case), and a limitation of the technique is that virtual sources cannot be created outside of this region. VBAP is mainly used for the reproduction of synthetic sounds [180].

The formulation of the VBAP method (from Pulkki [241]) for two dimensions starts from the stereophonic configuration of two channels (see Figure 2.7). Reformulated to a vector base, formed by unit length vectors $l_1 = [l_{11} l_{12}]^{\mathrm{T}}$ and $l_2 = [l_{21} l_{22}]^{\mathrm{T}}$ that point to the speakers and the unit length vector $p = [p_1 p_2]^{\mathrm{T}}$ which points to the virtual source and presents itself as a linear combination of the vectors l_1 and l_2. The notation $^{\mathrm{T}}$ is used here to identify the matrix transposition.

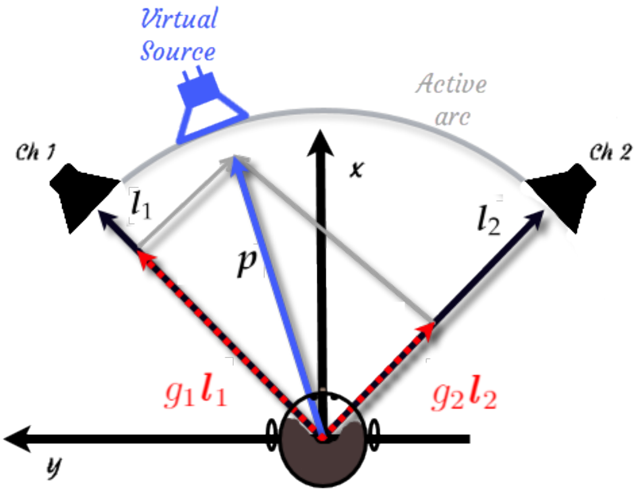


Figure 2.7: *Vector-based amplitude panning: 2D display of sound sources positions and weights.*

Consider the vetor $\boldsymbol{p}$:

$$\boldsymbol{p} = g_1 \boldsymbol{l}_1 + g_2 \boldsymbol{l}_2 \,, \tag{2.3}$$

where g_1 and g_2 (scalar) are the gain factors to be calculated for positioning the vector relative to the virtual source. In matrix form, there is

$$\boldsymbol{p}^{\mathrm{T}} = g \boldsymbol{L}_{12} \,, \tag{2.4}$$

where $g = [g_1 \ g_2]$, and $L_{12} = [\boldsymbol{l}_1 \ \boldsymbol{l}_2]^{\mathrm{T}}$. The gains can be calculated by

$$g = \boldsymbol{p}^{\mathrm{T}} L_{12}^{-1} = [p_1 p_2] \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix}^{-1} . \tag{2.5}$$

The formulation is also expanded to 3 dimensions:

$$\boldsymbol{p} = g_1\boldsymbol{l}_1 + g_2\boldsymbol{l}_2 + g_3\boldsymbol{l}_3\,, \tag{2.6}$$

and

$$\boldsymbol{p}^{\mathrm{T}} = g\boldsymbol{L}_{123}\,, \tag{2.7}$$

where g_1, g_2, and g_3 are gain factors, $g = [g_1 g_2 g_3]$, and $L_{123} = [l_1 l_2 l_3]^{\mathrm{T}}$.

The detailed derivation can be found at [241]. The derivation can use triangles and the three-dimensional system. Figure 2.8 presents an example of the sound source distribution for virtualization of a virtual source P using VBAP in three dimensions.

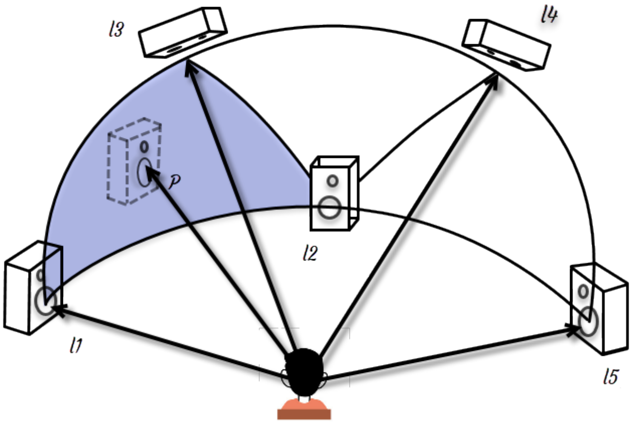


Figure 2.8: *Diagram representing the placement of speakers in the VBAP technique Adapted from [241].*

Some factors collaborate so that methods based on Amplitude Panorama are widely used in virtual audio applications, such as the low computational cost and flexibility in the speakers' placement.

Ambisonics: The original Ambisonics auralization method is an amplitude panning method that differs from the Vector Base Amplitude Panning (VBAP) method in several ways. While VBAP only uses positive weights to pan sound across speakers, Ambisonics uses a combination of positive and negative weights to create a shift in frequency and amplitude. This results in a more homogeneous sound field, albeit with a broader virtual source. Additionally, Ambisonics has all loudspeakers active for any source position. At the same time, VBAP only activates specific speakers based on the desired source position [199].

One of the benefits of Ambisonics is its scalability for reproduction on different loudspeaker arrays and the ability to encode and decode the sound field during the recording and reproduction process [161]. This versatility is possible because Ambisonics signals can be directly recorded using an appropriate microphone array or simulated through numerical acoustic algorithms that model the directional sensitivity of the microphone array [5, 46, 59]. The signal can then be decoded and rendered in real time to different arrays with various numbers of loudspeakers. Hence, an Ambisonics decoder is a tool for converting an Ambisonics representation of a sound field into a multichannel audio format that can be reproduced over a given speaker setup [130, 235, 238]. In order to reproduce an Ambisonics signal, it must first be transformed, or ”decoded,” into a format compatible with a specific speaker configuration. Simple decoders consist of a frequency-independent weighting matrix [282]. It is also possible to reproduce the signal via headphones, which can be considered a specific speaker setup, by scaling it down to binaural signals [320]. Additionally, Ambisonics can enhance realism by tracking head movements and correcting binaural signals utilizing HRTFs as filters [277]. This feature is particularly relevant in the recording and broadcasting industry, particularly with emerging technologies such as augmented reality (AR) [320]. In summary, an Ambisonics decoder is a tool used to transform an Ambisonics representation of a sound field into a multichannel audio format that can be reproduced over a given loudspeaker setup, enabling the creation of immersive sound experiences.

According to Schröder [261], decomposition in spherical harmonics (SH) is a recent analysis and widely used in the modeling of directivity paths. Analogous to a Fourier transform in the frequency domain, SH in the spatial domain decomposes the signal into spherical functions (in the Fourier transform, the decomposition is in sine or cosine functions) weighted by the coefficients of the corresponding harmonic spheres. According to Pollow [239], it is commonly applied in multi-dimensional domain problems. However, the analytical requirements for cases with few dimensions (two in the case of the sound field) are considerably simplified. Manipulating the wave equation by separating variables is an essential tool here.

Appendix C shows the derivation of SH through the separation of variables of the wave equation in spherical coordinates (Equation C.1).

The solutions to the linear wave equation in spherical coordinates expressed in the frequency domain (Helmholtz equation) are orthogonal basis functions $Y_n^m(\theta, \phi)$ where n is the degree and m is the order. These angle-dependent functions are called spherical harmonics and can represent, for example, a sound field [309]. That is the core assumption to Ambisonics recording and reproduction. Figure 2.9 depicts SHs up to order N = 2.

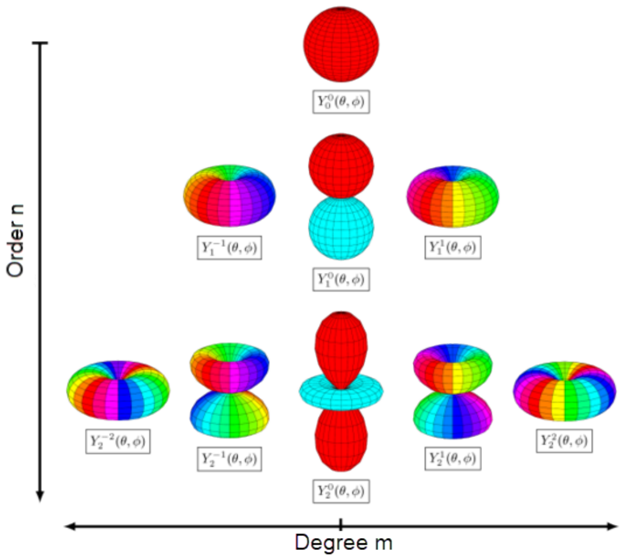


Figure 2.9: *Spherical Harmonics $Y_n^m(\theta, \phi)$. Rows correspond to orders $0 \leq n \leq 2$, columns to degrees $-n \leq m \leq n$ (adapted from Pollow [239]).*

The four SH weights $Y_n^m(\theta, \phi)$ to encode all the spatial audio information into a First-Order Ambisonics file is given by:

$$B_n^m(t) = s(t) Y_n^m(\theta_s, \phi_s) \tag{2.8}$$

where the $s(t)$ is the source signal in the time domain and $Y_n^m(\theta_s, \phi_s)$ the encoding coefficients to the source $s(t)$. Computed as first order in the B-format, the normalized components can be described as [172]:

$$\begin{cases} W = B_{00} = SY_{00}\left(\theta_S, \phi_S\right) = S(0.707) \\ X = B_{11} = SY_{11}\left(\theta_S, \phi_S\right) = S\cos\theta_S\cos\phi_S \\ Y = B_{1-1} = SY_{1-1}\left(\theta_S, \phi_S\right) = S\sin\theta_S\cos\phi_S \\ Z = B_{10} = SY_{10}\left(\theta_S, \phi_S\right) = S\sin\phi_S \end{cases} \tag{2.9}$$

The resulting four-channel signals are the equivalent to an omnidirectional (W) and three orthogonal bi-directional (commonly called figure-of-eight) microphones (X, Y, and Z). The channels can represent the pressure and the particle velocity of a given sound (See Figure 2.10. It is possible to transcode and manipulate the generated signal to change its orientation with a matrix multiplication in signal processing. Also, it is possible to decode the same encoded signal to a single sound source, headphones, or a multichannel array.

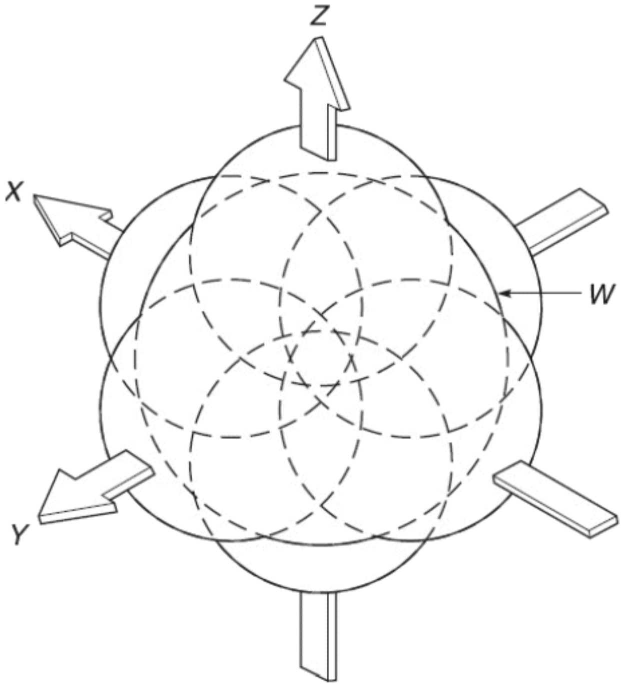


Figure 2.10: *B-format components: omnidirectional pressure component W, and the three velocity components X, Y, Z. Extracted from Rumsey [256].*

The limitation of first-order Ambisonics is spatial precision since it is only effective at a point centered within a defined area. This limitation can be overcome with higher-order components. Adding a set of higher-order components improves the directionality. However, increasing the number of components will also increase the number of loudspeakers required to play higher-order Ambisonics. That means a more accurate sound field representation if the order is increased. The number of channels N for a periphonic Ambisonics of order m order is $N = (m + 1)^2$ for 3D reproduction and $N = (2m + 1)$ for 2D [65].

2.3.2.3 Sound Field Synthesis

The objective of techniques from Sound Field Synthesis remains the same as in the techniques from the perceptually-motivated paradigm: a spatial sound field reproduction. The perceptually motivated are centered on the psychoacoustic effects of summing binaural cues that lead the listener to perceive a virtual source. On the other side, the Sound Field Synthesis techniques rely on the physical reconstruction of the original/simulated sound field to a specific area. The main techniques are the extension of Ambisonics reproduction to higher orders called Higher Order Ambisonics (HOA) and the Wave Field Synthesis (WFS) [24, 25].

The HOA extends the order of the classical Ambisonics and, therefore, the number of sound sources arranged in a spherical array. As the Ambisonics order increases, the perceived sound source direction accuracy also increases, although requiring more loudspeakers [97]. An important distinction can be made between the Ambisonics and HOA. Given the truncation possibility in Ambisonics, the method is treated as a soft transition from a perceptually based method to a physically based one. Although the HOA utilizes the same principle as Ambisonics, it is classified as a sound field synthesis paradigm (physically based) along with WFS. The HOA limitations are reported in the literature by several studies [26, 27, 64, 73, 236, 299], especially the aliasing in frequency that leads to pressure errors and the sweet spot size [253].

The WFS formulation relies on Huygen's principle: a propagating wavefront at any instant is shaped to the envelope of spherical waves emanating from every point on the wavefront at the prior instant [281], the principle is illustrated in Figure 2.11).

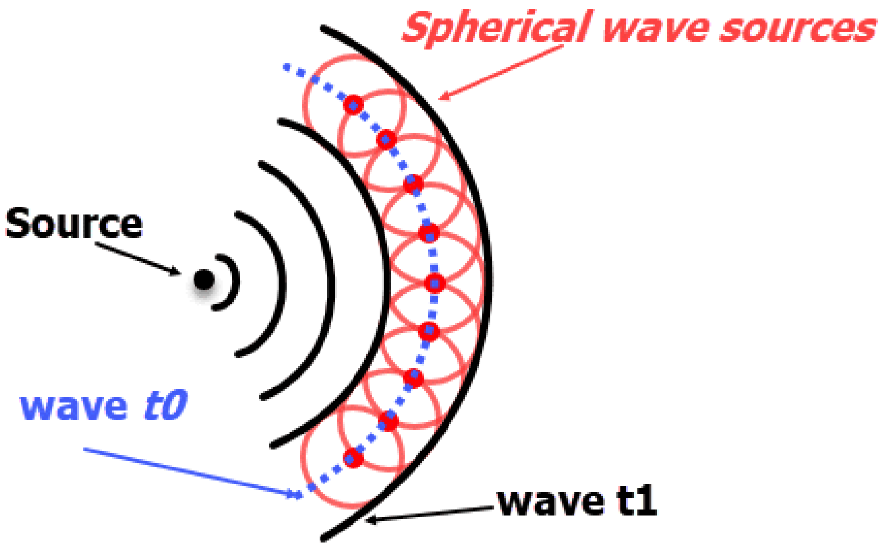


Figure 2.11: *Illustration of Huygen's Principle of a propagating wave front.*

A conceptual difference between WFS and HOA is that for HOA, the sound characteristics are described in a point (or small area) inside the array, while in WFS, the sound pressure that must be known is on the border of the reproduction area. A review and a comparison of both methods and their compromises in terms of spatial aliasing errors and noise amplification is presented in [65]. Their findings indicated similar constraints to both methods. However, they are both characterized by the requirement of large arrays of loudspeakers. The HOA has been found to have a higher limit of the size of the center area. In contrast, the WFS has limitations on the distortion of higher frequencies (aliasing), depending on the number of loudspeakers. Regardless, as the scope of the thesis aims to work with a small number of loudspeakers, they will not be thoroughly discussed.

2.3.3 Room acoustics

Different aspects are taken into account when describing the hearing experience of a human being in a space, for example, the individuality of auditory training, familiarity with space, personal preferences, humor, fatigue, culture, and the spoken language [10, 32, 123, 169, 184, 190, 250]. However, there are similarities in the expressions used between sample groups. Such an effect is attributed to the similarity of the auditory-cognitive mechanism of human beings [40].

In architectural acoustics and room acoustics, studies of sound properties assume that a sound source and a receiver in a given space is a linear and time-invariant system (LTI) [45]. Thus, a complete LTI characterization to each source-receiver can be expressed by its impulse response in the time domain or the transfer function in the frequency domain [45, 293].

2.3.3.1 Room acoustics parameters

Objective parameters are essential in acoustic projects and in compositions of statistical models that aim to predict the human interpretation of acoustic phenomena [254]. Objective parameters derived from the LTI's impulse response aim to create metrics that quantify subjective descriptors from numerous experiments [254]. The calculation and measurement of many objective parameters are described in an appendix to the International Organization for Standardization (ISO) standard 3382 [127].

2.3.3.2 Reverberation Time

The reverberation time (RT) measures the time it takes for the impulse response's sound pressure level to decrease to one-millionth of its maximum value, equivalent to a decline of 60 dB; it is also often referred to as RT_{60} or T_{60}. Note that the reverberation time measures how fast the decay of sound energy occurs and not how long the reverberation lasts in the environment, depending on the sound source power and the background noise. The RT was the first parameter studied, modeled, and understood, related to several subjective aspects of the human hearing experience in a room. Today, this is considered the most critical parameter, although it is not enough to describe human perception completely. Wallace C. Sabine [258] initially described it through mathematical relations obtained by an empirical method, followed later by developing the theoretical bases together with W. S. Franklin [86].

The expression gives the analytical form of the reverberation time obtained by Sabine:

$$T_{60} = \frac{0.161V}{S\bar{\alpha}} \text{ [s]} \tag{2.10}$$

where V is the volume of the room and $S\bar{\alpha}$ represents the amount of absorption present in the environment, the unit is named [Sabins] in honor of Sabine. Subsequent models improved the calculation of the reverberation time by considering the evolution of the energy density, and the sound absorption carried out by the air [74], the specular reflection of each sound wave [187, 271], the propagation path [148], the triaxial arrangement of the different absorption coefficients [14, 82], among others.

In addition to statistical theory, T_{60} can be obtained from the measurement of the impulse response. In measurements, the T_{60} is obtained considering limitations regarding the background noise level and the sound source's maximum sound pressure level. Thus, according to the ISO 3382 standard [127], the measurement's dynamic range must present the end of the decay at least 15 dB above the background noise and start 5 dB below the maximum. For example, the sound pressure level required to measure the T_{60} in a room with a background noise of 30 dB is 110 dB (30 + 15 + 60 + 5).

Linear behavior is noted by observing the square of the energy $h^2(t)$ in the decay curve plotted in dB (See Figure 2.12). Thus, to reduce the dynamic range required for measurement, it is possible to estimate the T_{60} through other limits. The T_{60} is commonly mistaken for double the T30, which is not true. The T_{20} and T_{30} also correspond to the time the sound pressure level (SPL) inside the room takes to drop 60 dB but estimated from measurement restricted to ranges of -5 dB to -25 dB, and -5 dB to -35 dB, respectively. Therefore, the linear energy decay produces the relation $T_{60} = T_{30} = T_{20}$.

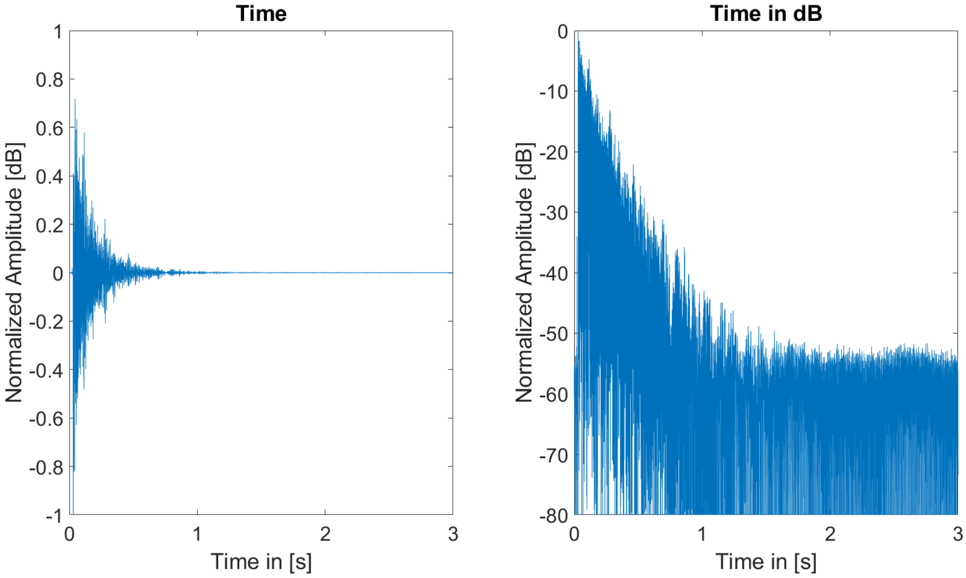


Figure 2.12: *Normalized Room Impulse Response: example from a real room in the time domain (left), and in the time domain in dB (right).*

The T_{20} is obtained as the decay rate by the linear least-squares regression of the measured decay curve, also called the Schroeder curve, in the range -5 dB to -25 dB. In comparison, the T_{30} is obtained when the curves' adjustment is carried out in the range between 5 dB and -35 dB [127].

2.3.3.3 Clarity and Definition

The clarity and definition parameters express a balance between the energy that arrives earlier and later in the impulsive response, which is related to human beings' particular ability to distinguish sounds in sequence [44, 45, 57, 247, 254]. With the first reflections arriving within the limits of 50 or 80 milliseconds, the tendency is to be integrated by the auditory system into the direct sound. Thus, if the first reflections contain relatively greater energy than the reverberating tail, the sound will be experienced as amplified. On the other hand, if the reverberating tail has more energy and is long enough, it will be perceived and mask the next direct sound. The limits of 50 and 80 milliseconds are defined in the literature as appropriate in optimizing speech and music, respectively [245, 247].

The Clarity defined in the ISO 3382 standard measures the ratio between the energy in the first reflections and the energy in the rest of the impulse response. Clarity's positive values, which are given in dB, mean more energy in the first reflections. Negative values indicate more energy in the reverberating tail. A null value indicates the balance between the parts of the impulse response. The "Clarity" is given by:

$$C_{80} = 10\log\left(\frac{\int_0^{80\mathrm{ms}} h^2(t)\mathrm{d}t}{\int_{80\mathrm{ms}}^{\infty} h^2(t)\mathrm{d}t}\right) \tag{2.11}$$

and

$$C_{50} = 10\log\left(\frac{\int_0^{50\mathrm{ms}} h^2(t)\mathrm{d}t}{\int_{50\mathrm{ms}}^{\infty} h^2(t)\mathrm{d}t}\right) \tag{2.12}$$

The "Definition" parameter, in turn, is presented on a linear scale and computes the ratio between the energy contained in the first reflections by the total energy of the impulse response. Values greater than 0.5 indicate that most of the impulse response's energy is contained in the first reflections. The "Definition" is given by:

$$D_{80} = \frac{\int_0^{80\mathrm{ms}} h^2(t)\mathrm{d}t}{\int_0^{\infty} h^2(t)\mathrm{d}t} \tag{2.13}$$

and

$$D_{50} = \frac{\int_0^{50\mathrm{ms}} h^2(t)\mathrm{d}t}{\int_0^{\infty} h^2(t)\mathrm{d}t} \tag{2.14}$$

2.3.3.4 Center Time

The central time is a parameter analogous to the previous ones, measuring the balance between the energy contained in the early reflections and the reverberating tail's energy. However, the central time is particularly interesting in pointing out what can be seen as the center of gravity of the squared impulse response. Moreover, the central time does not previously delimit the transition barrier between first reflections and a reverberating tail. Thus, the definition of the central time for an impulse response is given by:

$$t_s = \frac{\int_0^{\infty} t h^2(t)\mathrm{d}t}{\int_0^{\infty} h^2(t)\mathrm{d}t} \tag{2.15}$$

2.3.3.5 Parameters related to spatiality

The relation to the human auditory spatiality sensation and the objective parameters derived from measurements are studied in detail in the literature [20, 21, 88]. They observe how the sound energy distribution is arranged from the directions and timing aspect. The principal sensations and their related parameters are presented for better understanding.

Apparent Source Width The Apparent Source Width (ASW) is related to the impression of the sound source's size or how the source is distributed in the space.

An objective metric associated with ASW is the Lateral Energy Fraction (LEF). The Equation 2.16 gives the LEF:

$$\mathrm{LEF} = \frac{\int_{5\mathrm{ms}}^{80\mathrm{ms}} h_{\mathrm{b}}^2(t)\mathrm{d}t}{\int_{0}^{80\mathrm{ms}} h^2(t)\mathrm{d}t} \tag{2.16}$$

Where $h(t)$ is the impulse response measured with a microphone that has an omnidirectional sensitivity pattern and $h_{\mathrm{b}}(t)$ is the impulse response measured with a microphone that has bidirectional sensitivity (pressure gradient) at the same position as the omnidirectional.

Thus, this objective parameter represents the ratio between the lateral energy that reaches the receptor between 5 and 80 milliseconds (*i.e.*, the energy contained in the early reflections, excluding the direct sound) and the total energy arriving from all directions between 0 and 80 milliseconds [21]. As low and mid frequencies make the dominant contributions to the LEF, this parameter is usually represented by the arithmetic mean of the octave bands' values obtained between 125 Hz and 1000 Hz [45, 254].

Listener Envelopment The Listener Envelopment (LEV) is related to the impression of being immersed in the room's reverberant field.

From Bradley and Soulodre's experiments [44] with test participants inside an anechoic room, the sense of involvement was assessed with loudspeakers. The authors find the LEV associated with the ratio between the lateral energy and the total energy reaching the receptor. The lateral energy is contained in the impulse response measured with a bidirectional microphone after the first 80 milliseconds. The total energy is defined as the impulse response measured with an omnidirectional microphone, in free field condition, and 10 meters away from the sound source utilizing the same sound source at the same power. The ratio is called "Lateral Strength" (LG) and is given by:

$$\mathrm{LG} = \frac{\int_{80\mathrm{ms}}^{\infty} h_{\mathrm{b}}^2(t)\mathrm{d}t}{\int_0^{\infty} h_{10}^2(t)\mathrm{d}t} \tag{2.17}$$

Interaural Cross-Correlation Coefficient In his work, Keet [133] proposed an auditory-cognitive process relating the spatial impression to comparing the signals received by both ears.

The cross-correlation function measures the degree of similarity of the signals. Therefore, the Inter-Aural Cross-Correlation Coefficient (IACC) was incorporated as a third parameter related to the spatial impression. The IACC is defined as the absolute maximum value of the ratio between the cross-correlation function of the impulse responses collected from the left ear ($h_L(t)$) and the right ear ($h_R(t)$) by the total energy contained in each of them.

$$\mathrm{IACC} = \max\left|\frac{\int_{t1}^{t2} h_L(t)h_R(t+\tau)\mathrm{d}t}{\sqrt{\left(\int_{t1}^{t2} h_L^2(t)\mathrm{d}t\right)\left(\int_{t1}^{t2} h_R^2(t)\mathrm{d}t\right)}}\right| \tag{2.18}$$

where $\int_{t1}^{t2} h_L^2(t)\mathrm{d}t$ and $\int_{t1}^{t2} h_R^2(t)\mathrm{d}t$ are the energy between the instant t1 and the instant t2 in the impulse response from the left and right ears; the expression $\int_{t1}^{t2} h_L(t)h_R(t+\tau)\mathrm{d}t$ is the cross correlation function between the impulse response; τ is given between 0 and 1 ms.

2.3.4 Loudspeaker-based Virtualization in Auditory Research

Virtualization of sounds through auralization of simulated environments have been used in architectural design to preview the sound behavior in rooms when changing space design or even to preview a completely new space that is not built yet before the building process [293]. As the room acoustics simulation and the auralization process evolves as a sound equivalent to the visual preview rendered in 3D models, it has found applications in research also outside the architectural field [282, 320]. Lately, the virtualization of sound sources has been applied to extend the ecological validity of sound scenarios in auditory research [161].

Research that utilizes binaural virtualization with headphones are common in auditory research literature [4, 38, 142, 248, 305]. A series of advantages may include, but are not limited to, the individual control of the stimuli reproduced in each ear, the smaller setup, and easier calibration [251]. Although binaural reproduction is a suitable method for some research questions, others may require a more complex test environment, especially research encompassing hearing aids.

In that regard, the use of loudspeakers can be associated with single loudspeaker presentations where a loudspeaker reproduce a single sound source positioned in space (*e.g.*, [176, 230, 268, 321]) or virtualization methods that manage auralized files to be perceived as single sources or complex environments [89, 177, 295].

For the virtualization of sound sources, the loudspeaker number depends on the selected method of encoding and decoding the spatial information [293]. For example, a quadrophonic loudspeaker arrangement was found to be sufficient to reproduce a diffuse sound field to a perceptual spatial impression when constraining listener movements [117]. However, utilizing Directional audio coding, Laitinen and Pullkki [150] found that to have an adequate reproduction of diffuse sound, it would be necessary from 12 to 20 loudspeakers.

VBAP and HOA techniques were evaluated with different numbers of loudspeakers in simulations by Grimm *et al.* [97]. Perceptual localization error (PLE) was computed for the arrays utilizing these techniques. Eight loudspeakers were estimated to be sufficient in terms of sound source localization. In the same work, Grimm textitet al. showed that the effects of virtualization with VBAP and HOA on hearing-aid beam patterns are present with less than 18 loudspeakers in a bandwidth of 4 kHz (spatial aliasing higher than 5.7 dB criterion). However, the spectral distances, a weighted sum of the absolute differences in ripple and spectral slope between virtual and reproduced sound sources, were all very low, indicating high naturalness when compared to subjective data from Moore and Tan [191].

Aguirre [1] evaluated VBAP and its variation Vector-Based Intensity Panning (VBIP) in terms of spatial accuracy with 30 normal-hearing participants within an array of eight loudspeakers. There was no significant difference among stimuli (speech, intermittent white noise, and continuous white noise) on both techniques. It was found that an average PLE around 4°, consistent with the values simulated by Grimm *et al.* [97].

Evaluating SNR benefits on HA beamformer algorithms within a spherical array with 41 loudspeakers, Oreinos and Buchholz [212] found similar results between the real environment and the auralized one. Reproduction errors in HOA reproduction to hearing aids were studied in [213]. The reverberation was found to reduce the time-averaged errors introduced by HOA, implying that the frequency limit of usable renderings with HOA can be extended in those environments.

Loudspeaker-based virtualization have been used in a hearing research context evaluating normal hearing, hearing impaired and hearing aid users through different methods [6–8, 30, 31, 55, 61, 80, 93, 102, 136, 168, 174, 188, 220, 303, 322]. Furthermore, some studies explored the ecological validity of the techniques with subjective responses based on psycho-linguistic measure comparing *in-situ* and those virtualized in laboratory [66, 103, 286].

The process of virtualizing sound sources using loudspeakers is complex [282] and requires a thorough understanding of physical acoustics, psychoacoustics, signal processing fundamentals, and proper calibration of software and hardware [126, 165, 293]. As a result, research centers have developed systems to establish reliable procedures for virtualizing sound sources for auditory testing. Examples of such systems include the transaural CTC system developed by Aspöck *et al.* [17], the system with a spherical array of 42 loudspeakers capable of rendering scenarios using HOA up to fifth order and VBAP presented by Parsehian *et al.* [215], and the Loudspeaker Based Room Auralization (LoRa) system developed by Favrot [79], which is capable of rendering auditory scenes using pure HOA and a hybrid version with Nearest Speaker (NSP) and HOA. In addition, Grimm [100, 101] introduced the Toolbox for Acoustic Scene Creation and Rendering (TASCAR), which is capable of rendering perceptually plausible scenes in real-time using VBAP and HOA 2D implementation. A recent study by Hamdan and Fletcher [107] proposed a method using only two loudspeakers in 2022 based on the transaural method with cross-talk cancellation. While this list is not exhaustive, these studies provide recommendations and guidelines for the field and highlight the importance of implementing reliable systems and verifying their sound fields objectively and subjectively to increase the ecological validity of auditory research and hearing aid development.

2.3.4.1 Hybrid Methods

Hybrid methods that combine the reproduction of direct sound and reverberation are not new, having been developed since at least the 1980s with the Ambiophonics group [42, 95]. They proposed the Ambiophonics method to reproduce concerts to one or two home listeners as if they were in the hall where the recording was performed. This method combined crosstalk canceled stereo-dipole and convolved signals with the IR from the recorded spaces [76, 94]. The system aims to enhance the reproduction of recordings from existing systems (e.g., stereo and 5.1). The group also developed a new recording methodology called Ambiophone. This method is a microphone arrangement composed of two head-spaced omnidirectional microphones covered by a baffle in the rear to favor room reflections from frontal directions.

In 2010, the Loudspeaker based Room Auralization (LoRa) method developed by Favrot [79] applied the hybrid concept using HOA and the nearest speaker (NSP) for the direct sound and early reflections. The method uses the envelope from simulated rooms to reduce the computational cost by multiplying it with uncorrelated noise. The scheme was originally conceptualized for a large spherical 69 loudspeaker array. Figure 2.13 depicts its system schematic.

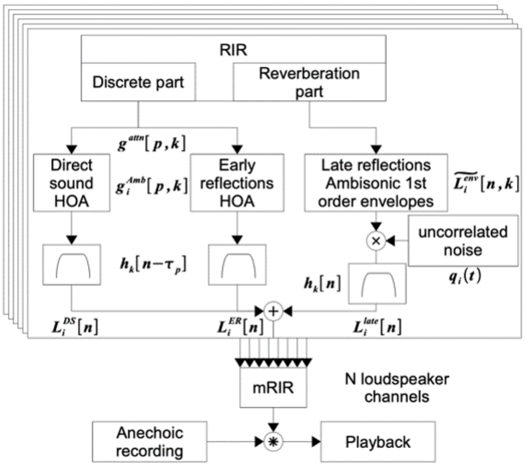


Figure 2.13: *LoRa implementation processing diagram. The multichannel RIR is derived in eight frequency bands and for each part of the input RIR (Figure from Favrot [79]).*

Pelzer *et al.* [221] presented a comparison between transaural or cross-talk cancellation (CTC), VBAP, and 4th-order Ambisonics among two new hybrid proposals: (1) direct sound and early reflections through CTC and late reflections with 4th-order Ambisonics and (2) direct sound and early reflections through VBAP and late reflections with 4th-order Ambisonics. The hybrid methods were implemented in a single case without generalization to different simulations. These methods were tested within a 24 loudspeaker array with no statistically significant change in human localization performance by any of the methods. Pausch *et al.* [217] presented a method designed for investigations with subjects with hearing loss. The method mixes binaural techniques to process components in complex simulated environments and CTC to present them over loudspeakers. At the same time, the head position can be tracked, allowing user interaction.

In 2017 Pulkki *et al.* [243] presented the first-order directional audio coding (DirAC) method is a technique for reproducing spatial sound over a standard stereo audio system. It is based on using first-order ambisonic channels, which encode the sound pressure and particle velocity at a listener's location to represent the sound field. These channels are transformed into a stereo audio signal using a frequency-dependent matrix, which preserves the spatial cues that are important for localizing sound sources. The method implies the direction of arrival of the sound source to be able to virtualize it through amplitude panning. It uses real-world recordings.

The DirAC method is effective for various types of audio content, including music, speech, and sound effects. It can potentially improve the spatial realism of audio experiences over traditional stereo systems and has applications in myriad fields, including entertainment, gaming, and virtual reality.

Table 2.1 presents an overview of the listed methods and the techniques involved, their purpose, and their parameters.

Table 2.1: Non-exhaustive overview list of hybrid auralization methods proposed in the literature. The A-B order of the techniques does not represent any order of significance.

Year	Method	Authors	Technique A	Technique B	Proposed Loudspeaker Number	Proposed to
1986	Ambiophonics	Farina *et al.* [76]	Crosstalk Cancelation	Binaural	2	Music Reproduction
2005	DirAC	Pulkki *et al.* [243]	Ambisonics	VBAP	2+	multiple applications
2010	LoRa	Favrot [79]	HOA	NSP	64	
2014	-	Pelzer *et al.* [221]	Crosstalk Cancelation	HOA	24	
2014	-	Pelzer *et al.* [221]	VBAP	HOA	24	
2018	Extended Binaural Real-Time	Pausch *et al.* [217]	Binaural	CTC	2	Hearing Loss Investigations

2.3.4.2 Sound Source Localization

A comparison among VBAP and Ambisonics conducted by Frank [84] demonstrated a median deviation in experimental results from the ideal localization curve of 2.35º± 2.93º on VBAP and 1.05º± 4.07º to third order Ambisonics using max-r_E decoder. The setup was placed in a typical non-anechoic studio and a regular array of 8 loudspeakers, listening in a 2.5m radius circle at the central position. The subjective results from 14 participants were listening to pink noise. These experimental results were compared to a localization model (Lindemann [157]) based on ITD and ILD from impulse responses. The results showed a deviation close to the standard deviation in subjective listening, 2.35º on VBAP and 3.37º to third order Ambisonics using max-r_E. Off-center measurements were pointed out as necessary for future investigation by the author.

Ambisonics in first, third, and fifth order was examined in another study by Frank and Zotter [85], with 15 normal hearing, 12 loudspeakers setup, and listening to pink noise with interval attenuation. This study investigated the effect of the position (centered and off-center) and the order. The results showed, to the first order rendered, a localization error of around 5º to the centered listener and 30º for the off-center position.

Also, Ambisonics in the first order with four loudspeakers and the third order with eight loudspeakers was investigated by Stitt *et al.*, [283]. This study was conducted in a non-reverberant environment to verify the off-center position and the Ambisonics order. The setup was a circular array with 2.2 meters of radius and an RT of 0.095 s. Eighteen test participants listened to white noise bursts of 0.2 s. At this acoustically dry condition, the centered first-order median absolute error was around 10º, while in the off-center positions tested was close to 30º. As expected, the error was lower in the third order achieving a median of absolute error around 8º in the center and 11º off-center.

A study by Laurent *et al.,* [275] investigated the effect of 3D audio reproduction artifacts on hearing devices assessing ITD, ILD, and DI on HOA (third and fifth orders), VBAP, distance-based amplitude panning (DBAP), and multiple-direction amplitude panning (MDAP). The study was conducted in a non-anechoic room with 32 loudspeakers in a spherical configuration. The loudspeaker distance from the center was 1.5 m, except for the four loudspeakers at the top, which were distant only 98 cm. This study investigated centered and off-center positions (10 and 20 cm). The results presented an expected Ambisonics limitation of reproducing ITD because of the spatial aliasing at high frequencies, accordingly to the authors. In addition, they investigated MVDR monoaural beamformer, which did not reproduce the correct ITD, especially off-center. At the centered position, only DBAP could not correct reproduced ITD. Ambisonics ITDs deteriorate more than VBAP on off-center positions. ILD errors in virtualized sound sources can make the system unreliable for testing HI with processing based on ILDs. In the experiment, the ILDs were less affected by beamforming processing in VBAP, and Ambisonics benefited from the max$_{RE}$ decoding that maximizes the energy vector. However, the authors expect a better ILD representation from VBAP as HOA has an aliasing frequency limitation.

Hamdan and Fletcher [107] present the development of a compact two-loudspeaker virtual sound reproduction system for clinical testing of spatial hearing with hearing-assistive devices. The system is based on the transaural method with cross-talk cancellation and is suitable for use in small, reverberant spaces, such as clinics and small research labs. The authors evaluated the system's performance regarding the accuracy of sound pressure reproduction in the frontal hemisphere. They found that it could produce virtual sound fields up to 8kHz. They suggest that tracking the listener's position could improve the system's performance. Overall, the authors believe this system is a promising tool for the clinical testing of spatial hearing with hearing-assistive devices.

Finally, a study by Bates *et al.*, [22], evaluated second-order Ambisonics and VBAP localization errors in subjective listening tests and ITD and IAFC comparisons. They presented the stimuli to a simultaneous set of nine listeners in different positions inside a concert room (around 1 s of RT). With 16 loudspeakers, they selected 1 second of speech (male and female), white noise, and music. The results indicate that VBAP and Ambisonics techniques cannot consistently create spatially accurate virtual sources for a distributed audience in a reverberant environment. The off-center positions are compromised by technique and stimulus. Depending on the stimuli, centered positions resulted in localization errors between 10º and 20º degrees. In the spatial distribution inside the ring, a bias from the target image position and towards the nearer contributing loudspeaker is more present in the Ambisonics than the VBAP. The authors mentioned that the room acoustics could also impact localization accuracy.

The number of variables across these previous studies and their contributions is massive, e.g., objective measures, technique variations, number of loudspeakers, loudspeaker distance, number of simultaneous listeners, reverberation time, and form of the array.

Table 2.2 presents an overview of methods and estimated or measured localization error.

Table 2.2: Overview of Localization Error Estimates or Measurements from Loudspeaker-Based Virtualization Systems Using Various Auralization Methods.

	Method	Error at center position	Error at off-center position	LS number
Present Study	Iceberg VBAP/Ambisonics	30º (Max Estimated) 7º (Average Estimated)	30º (Max Estimated) 7º (Average Estimated)	4
Frank [84]	VBAP	2.35º (Average)	N/A	8
Frank [84]	HOA (3rd order)	3.37º (Average)	N/A	8
Zotter [85]	Ambisonics	5º (Median)	30º (Median)	12
Zotter [85]	HOA (3rd order)	2º (Median)	15º (Median)	12
Zotter [85]	HOA (5thorder)	1º (Median)	10º (Median)	12
Stitt *et al.*, [283]	Ambisonics	10º (Median)	30º (Median)	8
Stitt *et al.*, [283]	HOA (3rd order)	8º (Median)	11º (Median)	8
Bates *et al.*, [22]	Ambisonics Second Order	10º (mean)	20º (mean)	16
Bates *et al.*, [22]	VBAP	10º (mean)	20º (mean)	16
Grimm *et al.* [97]	HOA (3rd order)	2º (Estimated)	6º (Estimated)	8
Grimm *et al.* [97]	VBAP	4º (Estimated)	6º (Estimated)	8
Aguirre [1]	VBAP	4º (Median)	N/A	8
Hamdan and Fletcher [107]	CTC	2º (Max Head Displacement)	N/A	**2**
Huisman *et al.* [125]	Ambisonics	30º (Median)	N/A	4
Huisman *et al.* [125]	HOA (3rd order)	≈ 15º (Median)	N/A	8
Huisman *et al.* [125]	HOA (5th order)	≈ 8º (Median)	N/A	12
Huisman *et al.* [125]	HOA (11th order)	≈ 5º (Median)	N/A	24

2.4 Listening Effort Assessment

The regular task of following a conversation, listening to a person's speech, or interacting with someone in a conversation may require additional effort in an unfavorable or challenging sound environment [227]. The listening effort is defined as ”the deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a [listening] task” [224]. Studying aspects of the listening effort related to different acoustic situations through reliable methods can lead to the development of solutions to reduce it, improving the quality of life [304]. However, there is no consensus in the literature on the best method to measure listening effort.

Attempts to measure how much energy a person takes in a specific acoustic situation may rely on different paradigms. There are objective measurements of physiological parameters in literature associated with changes in effort, such as pupil dilation [151, 209, 211, 301, 302, 319], responses to brainstem frequency (FFRs) and cortical electroencephalogram (EEG) activity from event-related potentials [28, 33], or alpha band oscillations [186, 223]. In addition, the behavioral perspective studies changes in response time in single [204] or dual-task paradigm tests, also assuming that they are related to changes in cognitive load in auditory tests [87, 228] [225]. In turn, subjective assessments of listening effort are performed through questionnaires [323] or effort scales [147, 149, 249, 260] and their results generally agree with performance metrics [192].

Although subjective measurements are intuitive and valid, they tend to be less accepted as an indication of the amount of listening effort because of differences between objective and subjective outcomes [151, 225]. For instance, Zekveld and Kramer [318] present evidence of disagreement between the physiological and the subjective measure where the young normal-hearing participants attributed high subjective effort to the most challenging conditions despite their smaller pupil dilation. The authors assumed that the methodological aspects and the participant's tendency to drop out were also related to pupil dilation at low levels of intelligibility. In a study on syntactic complexity and noise level in the auditory effort, Wendt *et al.* [300] evaluated it through self-rated effort and pupil dilation. They found both background noise and syntactic complexity reflected in their measurements. However, at high levels of intelligibility, the methods show different results. According to the authors, the explanation is that each measure represents a different aspect of the effort. In its turn, Picou *et al.* [226]; and Picou and Ricketts [229] found subjective ratings of listening effort were correlated with performance instead of the listening effort utilizing the response time as a behavioral measure in a dual-task. Interestingly though, in this study, a question about control was correlated to their response time results. The varied outcomes from subjective and objective paradigms proposed to achieve a proxy to listening effort can indicate that these methods are quantifying separated aspects of a complex global process [12, 224].

Another explanation suggests a bias in the subjective method due to the heuristic strategies adopted by the participants to minimize the effort [192]. The mentioned strategy would consist of replacing the question about the amount of effort spent with a more straightforward question related to how they performed in the task. Concomitantly, studies based on objective measurement paradigms also have divergent results. For example, even among physiological measures sensitive to the spectral content of stimuli, such as pupil dilation and alpha power, they are not always related, and can be sensitive to different aspects of listening effort [186]. Even within the same paradigm, a different task may indicate that different aspects are being observed. For example, Brown and Strand [53] analyzed the role of the working memory as a weighting factor on listening effort. Although increasing background noise indeed increases listening effort measured by the dual-task paradigm, the memory load was not affected. They also suggested that the working memory and listening effort are related in the recall-based single-task, unlike in the dual-task. In Lau *et al.* [151] significant differences between sentence recognition and word recognition were found on pupil dilation measurements and in subjective ratings, although with no correlation between objective and subjective measures.

The demand for mental resources can also be affected by personal factors, such as fatigue and motivation [224]. At the same time, several physical-acoustical artifacts can degrade a sound, creating or leading to difficulties in everyday communication (increasing listening effort), especially in social situations. The masking noise, the spectral content of the noise, the Signal-to-Noise Ratio (SNR), and the environment reverberation are examples of artifacts capable of smearing the temporal envelope cues [163].

Also, speech intelligibility was assessed in a virtual environment that consists in a large spherical array of 64 loudspeakers reproducing Mixed Order Ambisonics (MOA) [6] presented comparable results of Speech Reception Threshold (SRT) compared to real room in a co-located situation of masker and target. With spatial separation of 30 degrees the virtual environment led to an SRT benefit of 3 dB, it was argued that benefit was not present in more reverberant or complex scenes suggesting the masking effect of more challenge scenes.

SRTs for normal hearing and hearing-impaired using hearing aids were also investigated by [31]. A complex scenario (reverberant cafeteria) and an anechoic situation were evaluated in a spherical array of 41 loudspeakers. The virtualization was provided convolving the direct sound and the early reflections parts of the RIR with the anechoic sentence and presenting the sound through the Nearest Speaker (NSP) and the late reflections part of the RIR are created through the directional envelope of each loudspeaker with uncorrelated noise.

The reviewed studies were conducted in laboratories mainly taking advantage of spatial sound and virtual acoustics via loudspeaker or headphones reproduction. Thus, the complex nature of human auditory phenomena and the importance of reproducibility in hearing research highlight the need for innovative tools such as spatial sound [134]. Virtualized sounds allows for realistic and controllable sound environments, enabling control over selected parameters and consistent reproduction of experiments [61, 161, 282, 293]. This technology can help hearing investigations become more true-to-life and reliable [134, 161, 251]. For example, it can be used to study listening effort and speech intelligibility using virtual sound sources to create ecologically valid and controlled environments [7, 177]. It also can enable the integration of virtual sound scenarios with ecological tasks involving multiple people, providing an ecologically valid assessment of the performance of hearing solutions more accessible than large-field studies (*e.g.,* in Bates *et al.,* [22]).

Additionally, spatial audio enables the accessible investigation of spatial separation's effects on binaural cues considering different environments, the role of binaural hearing in spatial perception, and new hearing aid hardware and algorithms [61, 97, 213]. Overall, spatial sound & virtual acoustics in hearing research offers numerous benefits and represents a valuable tool for advancing our understanding of hearing and developing effective hearing solutions.

2.5 Concluding Remarks

The literature review suggests a contrast between localization and immersion in auralization methods that virtualize sound using a low number of loudspeakers. Thus, there is a need for a method that can achieve useful performance on both localization and immersion with a small number of loudspeakers and that is reliable in rendering sound for listeners in the presence of another listener within the virtualized sound field. Previous methods, including hybrid approaches, have been developed using a larger number of loudspeakers and different techniques for balancing energy. A recent study proposed a method using only two loudspeakers in 2022. However, it implemented a different auralization method and had its limitations. The proposed method in this study is innovative, using a room acoustic parameter called center time to calculate the energy balance of room impulse responses and combining it with two known auralization methods.

Chapter 3

Binaural cue distortions in virtualized Ambisonics and VBAP

3.1 Introduction

In acoustics, the complex communication scenarios can involve, simultaneous sound sources, distracting background noise, moving sound sources, sources without large spatial separation, and low signal to noise ratio. Although people with normal hearing can deal with most of these conditions in a relatively efficient way, people with hearing loss perform poorly [273, 289, 317]. Since social events are often a real example of complex communication, the interaction barriers make people avoid this and sometimes ostracizing themselves [16, 63]. That can be a factor in decreasing the quality of life of people with hearing problems.

In hearing research, innovative signal processing techniques, new devices, more powerful hardware, and updated parameter settings are continuously developed and evaluated. These technological improvements aspire to resolve communication problems in everyday situations for hearing aid users [227], increasing their socialization and quality of life [119]. Tests as speech recognition in noise are developed and tailored to evaluate the human auditory response on everyday acoustics situations better than clinical based in pure tones stimulation [145]. Even though the tasks are moving towards a more realistic representation, they still need to improve the ecological validity [134].

Auralization methods are designed to create files meant to be reproduced to a specific listener or a group of listeners; these files contain particular characteristics that try to mimic a recorded or digitally created sound scene according to the method. The mathematical formulations that produce these characteristics for the psychoacoustically based methods focus on delivering accurate binaural cues. The listener position, physical obstacles as the listener movement will impact differently on distinct methods and cues.

A VSE is an auralized sound field that can contain realistic elements. Currently, it is possible to create VSE employing loudspeaker arrays or headphones for the listener, such as high background noise, high reverberation, and concomitant sound events from different directions [61, 79, 294]. Furthermore, through a VSE, it is also possible to enable a participant to wear, for example, a hearing aid during the test. Thus, the researcher can maintain control of the stimuli, the incidence direction, signal-to-noise ratio (SNR), among other settings, while examining the hearing device performance in a more ecological situation [98, 161, 269].

Although novel technologies emerge and contribute to emulating sound sources and even entire complex sound scenes with humans' social interaction [267], these opportunities are often overlooked in auditory evaluations. Typically, tests are performed by observing only one individual within the laboratory [81, 89, 104, 152, 169, 175]. Furthermore, the systems are designed to acquire responses from a single individual at a time [41, 79, 102, 118, 195, 218–220, 259]. A reasonable explanation for this is the low cost and complexity of auralization through headphones. More complex techniques, like Wave Field Synthesis, do not limit the listener to a restricted spot [207], reproducing a complete sound field, although at the cost of a large number of sound sources in a specifically treated room.

Social situations can have effect on people's listening effort [230, 234] and their motivation to listen [181, 224]. In this context, social interactions have been simulated through avatars or audiovisual recordings in virtual environments, gaining space in auditory research [116, 160, 161, 272, 298]. Although it can be considered a significant asset, it also focuses on a single individual's responses to simulated social stimuli.

The scenario creates a ground for this study to investigate controlled acoustical changes on the VSE. This study assesses two main situations within a ring of loudspeakers virtualizing sound sources on Ambisonics and VBAP: (1) the displacement of the listener from the center (sweet spot), and (2) the effect including a second simultaneous listener inside the ring. These topics can help understand the perception of sound in these specific virtualization methods, increasing the fundamental scientific basis for future hearing research applications. The changes to the sound field were observed in three major spatial cues: ITD, ILD, and IACC. That was explored by changing the listener's position and including a second listener inside the ring of loudspeakers to measure BRIRs.

These metrics can describe the spatial perception of an auralized sound signal [47, 48], being ITD and ILD responsible by localization and IACC perceived spaciousness and the listener envelopment [44]. Therefore, these measurements can indicate the possibility of a simultaneous second participant in any hearing test with virtualized spatially distributed sound sources. Two different auralization techniques were used to virtualized sound sources, vector-based amplitude panning (VBAP) [241] and Ambisonics [91]. Both techniques rely on the same receptor-dependent psychoacoustic paradigm to provide an auditory sense of immersion for those with normal hearing [161, 180]. These techniques aim delivering the correct binaural cues to a point or area to create a realistic spatial sound impression, albeit through different mathematical formulations. The work investigates if the techniques can provide an appropriate spatial impression for young normal-hearing listeners.

Hypothesis The main research question is how auralized scenarios with VBAP and Ambisonics are affected when displaced from the center and with another listener inside the ring. The hypothesis is that localization cues can be better provided by VBAP, especially in off-center positions. In contrast, Ambisonics can provide a better sense of immersiveness. Also, the second listener would impact Ambisonics more than VBAP virtualized sound sources.

3.2 Methods

The experiment was conducted in two different locations; The first one is a sound treated test room at the Hearing Sciences - Scottish Section in Glasgow (See Figure 3.1), the second is an anechoic test room at Eriksholm Research Centre (See Figure 3.2). This section presents the rooms' acoustic characterizations and the methods used in this experiment.

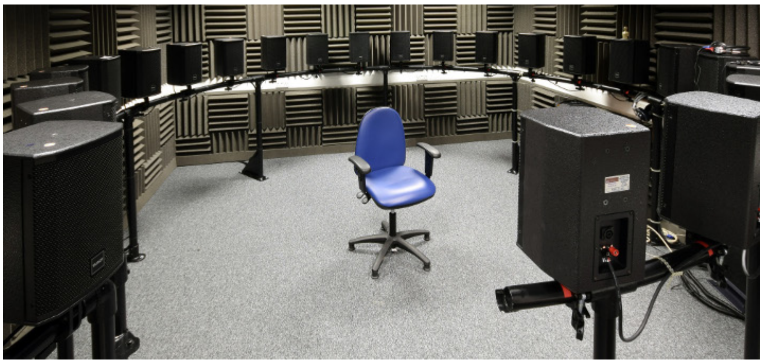


Figure 3.1: *Hearing Sciences - Scottish Section Test Room.*

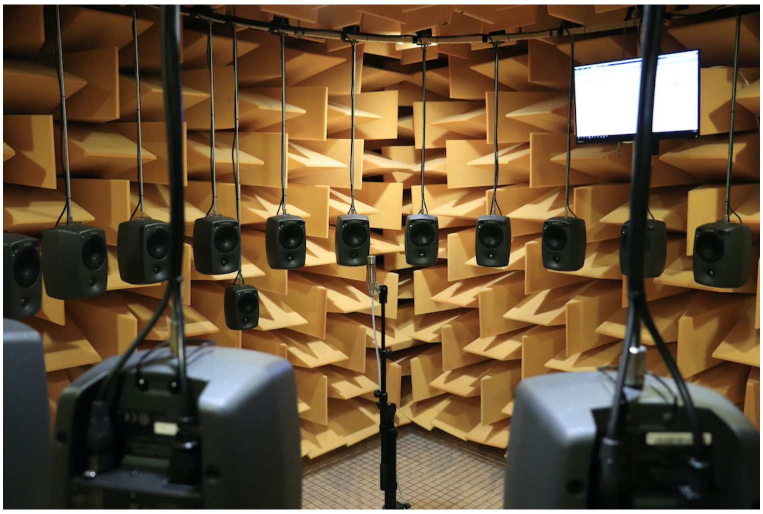


Figure 3.2: *Eriksholm Test Room.*

3.2.1 Setups and system characterization

The experiment conducted in Glasgow was in a large sound-proof audiometric booth (4.3 × 4.7 × 2.9 m; IAC Acoustics). An azimuthal circular array configuration of 24 loudspeakers (3.5-m diameter; 15° of separation; Tannoy VX6) was used. The ceiling and walls were covered with 100-mm deep acoustic foam wedges to reduce reflections; the floor was carpeted with a foam underlay. The AD/DA audio interface that was used was a Ferrofish Model A32. The loudspeakers received signals that were amplified by ART SLA4 amplifiers. The reference microphone used to characterize the Glasgow Test Room was a 1/2" G.R.A.S 40AD pressure-field microphone set with e GRAS 26CA preamplifier. It was oriented 90 degrees vertically from the sound source. At Eriksholm, an equivalent setup was fitted. This time in a full anechoic room from IAC

Acoustics. The room's outer dimensions (6.7 × 5.8 × 4.9 m;) and inner dimensions, from the tip of the foam edges (4.3 × 3.4 × 2.7 m). An azimuthal circular array configuration of 24 active loudspeakers (16 Genelec 8030A and 8 Genelec 8030C; 2.4-m diameter; 15° of separation) was used. The AD/DA was a MOTU PCI-e 424 combined with a firewire 24-channel audio extension. The reference microphone used to characterize the Eriksholm test room was a 1/2" B&K 4192 pressure-field and a preamplifier type 2669, supplied by power module 5935. It was oriented 90 degrees vertically from the sound source.

The signal acquisition and processing were entirely through Matlab 2020a software using the ITA-Toolbox v.9 [29].

The technical setup was equivalent in both rooms, a B&K head and torso simulator (HATS) model 4128-C mannequin for measurements, and a **K**nowles **E**lectronics **M**annequin for **A**coustic **R**esearch (KEMAR) was used as a physical obstacle. Although technically, both devices are head and torso simulators, in this thesis, HATS will refer to the B&K 4128-C for simplicity. The sampling rate of the recordings was fixed at 48 kHz, resulting in an uncertainty of ±20 µs, therefore not compromising the final analysis.

3.2.1.1 Reverberation time

The reverberation time is one of the most critical objective parameters of a room [154]. The decay of sound energy to 60 dB below its peak after the cessation of a sound source characterizes the RT. The parameter is frequency-dependent; it is associated with speech understanding speech, sound quality, and the subjective perception of the size of the room. For controlled environments, the values are fractions of seconds. The T_{60} for both rooms in the third octave is presented in Figure 3.3.

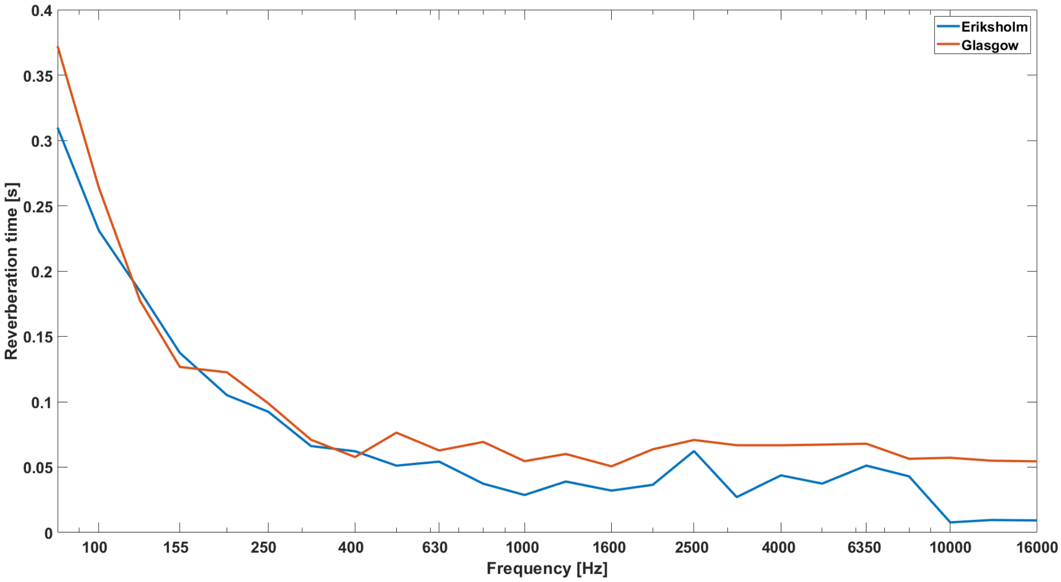


Figure 3.3: *Reverberation time in third of octave bands up to 16 kHz.*

The room's reverberation time T_{20} was measured using a loudspeaker, arbitrarily chosen, and microphone setup as in Section 3.2.1. The measurement and analysis were performed in Matlab through the ITA-Toolbox software.

3.2.1.2 Early-reflections

To ensure that there is no influence of the environment, Recommendation ITU-R 1116-3:2015 [126], determines that the magnitude of the first reflections should be at least 10 dB below the magnitude of the direct sound $\Delta\text{SPL} \geq 10$ dB. The differences in the SPL that are determined in the environments of this work met this requirement. Table 3.1 shows the difference in sound pressure level between the direct sound and early reflections. Higher differences in the Erkisholm environment are consistent with its anechoic setup compared to the sound treated booth in Glasgow, where the floor provide some energy to the reflections.

Table 3.1: Sound pressure level difference between direct sound and early reflections Δ SPL [dB]

	Δ SPL [dB]	
Angle	Eriksholm	Glasgow
0	-20.99	-14.94
15	-23.40	-15.31
30	-22.66	-14.61
45	-21.97	-15.45
60	-20.39	-13.28
75	-21.22	-15.19
90	-17.71	-15.33
105	-21.49	-15.22
120	-17.83	-15.68
135	-20.12	-15.23
150	-19.70	-14.62
165	-19.13	-16.11
180	-24.57	-15.03
195	-23.56	-13.52
210	-22.62	-14.81
225	-21.04	-15.39
240	-22.29	-14.25
255	-23.73	-14.37
270	-20.90	-14.01
285	-24.06	-12.56
300	-19.61	-15.95
315	-17.68	-15.03
330	-21.46	-15.66
345	-23.08	-15.95

3.2.2 Procedure

The experiment studied how the presence of a second listener within a loudspeaker ring affects the spatial cues of the reproduced sound field. The data were collected through the HATS, and the second listener being simultaneously inside the virtualized sound area was simulated through another mannequin (KEMAR), as shown in Figures 3.4 and 3.5.

Using the results for the reverberation time as presented in Section 3.2.1, the appropriate length of a logarithmic sweep signal was calculated as approximately four times larger than the higher value of T_{60} (1.49 seconds). Also, a stop margin of 0.1 seconds was set to ensure the quality of the room impulse

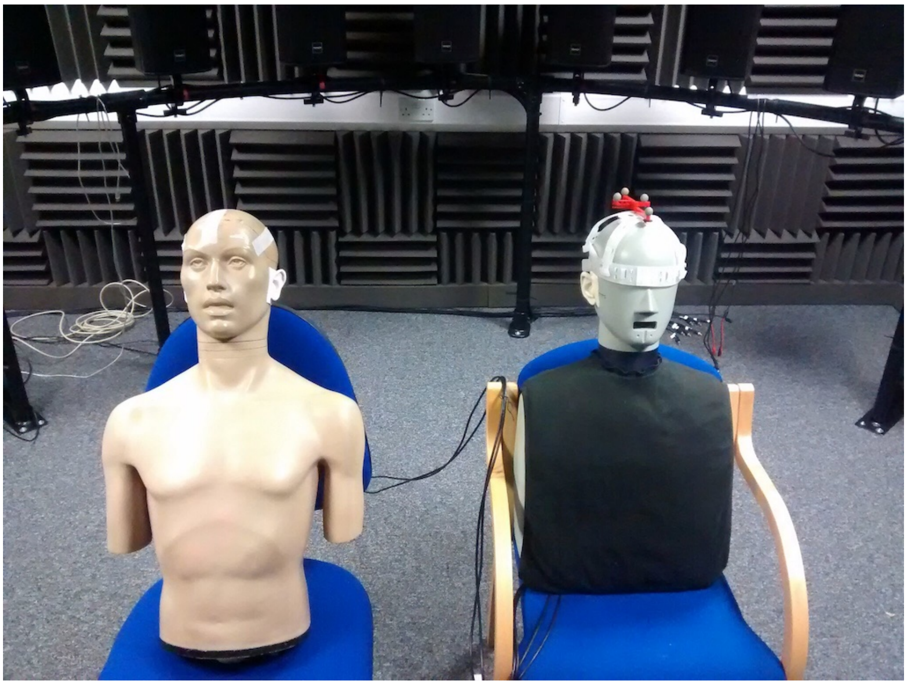


Figure 3.4: *HATS (with motion-tracking crown) and KEMAR inside test room in Glasgow.*

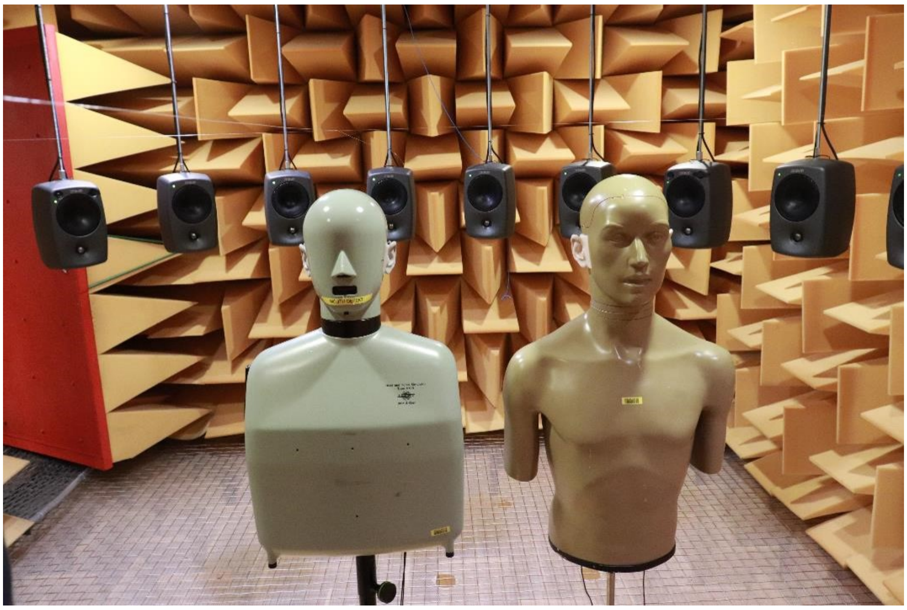


Figure 3.5: *HATS and KEMAR inside anechoic test room at Eriksholm.*

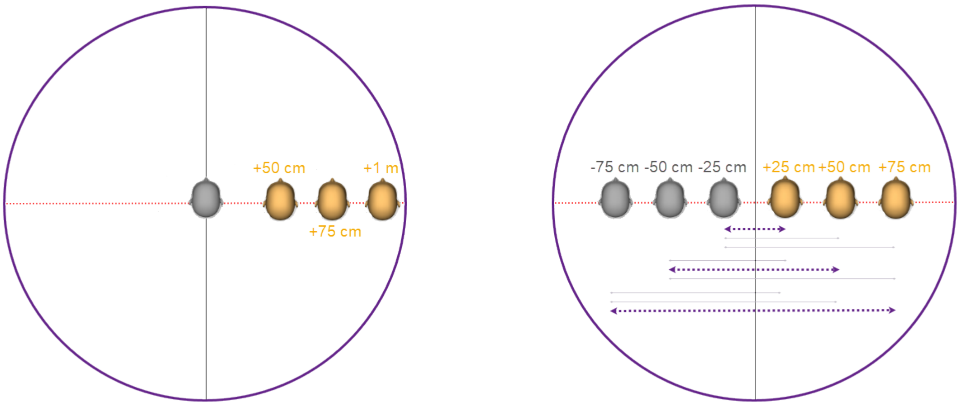
responses (RIRs) that were obtained [75, 194]. The frequency of the sweep was from 50 Hz to 20 kHz.

The position of the head has a significant effect on the signals that are measured. To have a reliable assessment of the absolute tri-dimensional position of the HATS, its position was measured with a the Vicon infra-red tracking system with an accuracy of 0.5 mm in Glasgow. At Eriksholm a laser tape measure was used to ensure the correct positions. The microphones' height position in both experiments was set to match the geometrical center of the loudspeakers enclosure in all measurements. The first position measured used the HATS in the center, without interference from another obstacle inside the ring, to provide a baseline.

Figure 3.6a illustrates a set of positions to study the influence of a second listener inside the ring while keeping the test subject in the center (the sweet spot). Three different positions for the KEMAR (50, 75 and 100 cm of separation) were measured with the HATS fixed at the center of the loudspeaker array. The data collected are from microphones in the HATS ears; the KEMAR was only a physical obstacle to simulate a listener inside the ring. Figure 3.6b, illustrates a different set of positions, maintaining a minimum separation of 50 cm between the center of the heads, were measured. The purpose of these positions with the HATS off-center was to identify the presence of distortions caused by the decentralization of the subject and the effect of the addition of a listener within the circle of loudspeakers as a physical obstacle to sound waves. The positioning was standardized so that the movement along the x-axis to the left and right directions of the dummies were annotated as negative and positive, respectively.

3.2.3 Calibration

To calibrate the HATS recordings, the adapter B&K UA-1546 was connected to the B&K 4231 calibrator. That provided a 97.1 dB SPL signal, which corresponds to 1.43 Pa, instead of 94 dB without an adapter. The recorded signal from each ear was used to calibrate the levels of all measurements. The calibration factor was calculated as:



(a) *Centered position* **(b)** *Off-center position*

Figure 3.6: *HATS in gray, KEMAR in yellow. a) Measured positions with the HATS centered and the KEMAR present in the room in different positions (three combinations). b) Measured positions with the HATS in different positions and the KEMAR present in the room in different positions (nine combinations).*

$$\alpha_{l,\mathrm{rms}} = \frac{1.43}{\mathrm{rms}(v_l(t)_{1\mathrm{kHz}})} \left[\frac{\mathrm{Pa}}{\mathrm{VFS}}\right], \tag{3.1a}$$

$$\alpha_{r,\mathrm{rms}} = \frac{1.43}{\mathrm{rms}(v_r(t)_{1\mathrm{kHz}})} \left[\frac{\mathrm{Pa}}{\mathrm{VFS}}\right], \tag{3.1b}$$

where

$\alpha_{l,\mathrm{rms}}$ is the calibration factor for the left ear;

$\alpha_{r,\mathrm{rms}}$ is that for the right ear;

$v_l(t)$ is the calibrator signal recorded in the left ear;

$v_r(t)$ is that for the right ear;

The individual loudspeakers' sound pressure level to the same file can differ depending on several factors (*e.g.*, the amplification system's level). To balance that, a factor was then measured for a GRAS 1/2" pressure-field microphone recording a pistonphone's calibrated sound signal from 1 kHz. The calibration factor α_{rms} was calculated from the root mean square (RMS) using:

$$\alpha_{\text{rms}} = \frac{10}{\text{RMS}(v(t)_{\text{1kHz}})} \left[\frac{\text{Pa}}{\text{VFS}}\right], \tag{3.2}$$

where

$v(t)_{\text{1kHz}}$ is the sinusoidal signal at 10 Pa recorded from the calibrator in volts full scale (VFS). The loudspeaker correction factor is calculated through the iterative process that starts reproducing a RMS scaled version of a pink noise signal at 70 dB SPL.

$$\text{pink noise}(t) = \left(\frac{\text{pink noise}(t)}{\text{rms}(\text{pink noise}(t))} 10^{\frac{70-\text{dBperV}}{20\mu}}\right) \Gamma_l \tag{3.3}$$

where Γ_l is the level factor to the loudspeaker l with initial value = 1; $\text{dBperV} = 20\log_{10}\left(\frac{\alpha_{\text{rms}}}{20\mu}\right)$.

The signal pink noise(t) is played through a loudspeaker l and simultaneously recorded with the microphone $\text{S}_l(t)$; the SPL of the recorded signal is calculated as follows

$$\text{SPL}_l[\text{dB}] = 20\log_{10}\left(\frac{\text{S}_l(t)[\text{VFS}]\alpha_{\text{rms}}\left[\frac{\text{Pa}}{\text{VFS}}\right]}{20[\mu\text{Pa}]}\right), \tag{3.4}$$

Ten measurements are sequentially performed, making intervals of 1 second; the next iteration happens if the SPL obtained exceeds the tolerance of 0.5 [dB] on any of the measurements. A step of ± 0.1 [VFS] is set to update Γ_l in its next iteration accordingly to the SPL obtained.

3.2.4 VBAP Auralization

In the first measurement, VBAP was the technique used to auralize the files. The first step in signal processing was recording the 24 (RIRs) one from each loudspeaker. Knowing the RT of the room, a sweep (50-20000 Hz) was created, fulfilling the length requirement; in this case, a logarithmic sweep of 1.49 seconds. After that, an inverse filter (minimum-phased) was created to compensate for the frequency responses from the different loudspeakers. This signal is then processed through the VBAP technique to the specified array of 24 loudspeakers. The output is a file with 24 channels containing the sweep signal appropriately weighted to the specific angle. The signal can be processed through a single channel (when the angle to be played is at the loudspeaker position) or up to two combined channels when it is a virtual loudspeaker's position. Each channel was also convolved with the designed filter. The final (auralized) signal was used as an excitation in the transfer function where the receptors were a pair of microphones in the B&K HATS.

3.2.5 Ambisonics Auralization

In the second measurement at the Eriksholm test room, the files were auralized with first-order Ambisonics in a similarlly to VBAP. To be able to process the excitation signal, to acquire the impulse responses, some adaptations were required.

In this case, the Ambisonics auralization process requires an encoded impulse response that contains the magnitude and the direction of incidence information for each instance of time. This RIR can be attained via computer simulation or recorded with a specific array of microphones. The ODEON software version 12.15 was used to simulate the sound behavior in an anechoic environment and encode the impulse responses in Ambisonics first-order format around the listener.

Odeon software is based on a hybrid numeric method [59]. In general, the Image-Source, a deterministic method, is favored in the region of the first reflections up to an order predetermined by the user. Then, reflections from subsequent orders than the predetermined transition order are calculated using ray tracing, a stochastic method [148, 201]. Therefore, it is possible to simulate the sound behavior from a 3D model description of the space and details of its acoustic properties. From that simulation result, any music or sound can be exported as recorded inside that space from the given positions of source and receptor [288]. Another option is to export the room impulse response, which represents the sound behavior of the given source receptor positions. The RIR can also be exported as BRIR and Ambisonics in first and second order in the version 12 of the Odeon software.

The selected materials used to compose the simulation, and their correspondent absorption coefficients used in the ODEON simulation are listed in the Appendix E. In total, 72 different RIRs (5 degrees separation) were simulated for different positions of source-receptor. The simulated source positions were at the same distance of 1.35 meters from the center as the loudspeakers in the anechoic room. These RIRs were convolved with the appropriate sweep signal, producing a four-channel first-order Ambisonics sweep signal. These signals were then processed by a decoder to the loudspeakers array's specific positions, generating the auralized 24 channel files. The inverse filter procedure to each loudspeaker was applied as well as the calibration of the sound pressure level across loudspeakers. The alpha factor was calculated as $\alpha_{\text{rms}} = \frac{1}{\text{rms}(v(t)_{1\text{kHz}})} \left[\frac{\text{Pa}}{\text{VFS}}\right]$, since the recorded input was from a sound calibrator type 4231 by B&K delivering 1 [Pa] SPL. The equalized, convolved, decoded, and filtered sweep signals contain the simulated source-receptor sound distributions in magnitude, time, and space as if recorded inside the simulated room. In this experiment, the simulated anechoic room has an absorption coefficient equal to one on all surfaces, simulating the anechoic condition.

The setup in first-order Ambisonics was chosen given the possibility of exploring a reduction in the number of loudspeakers in future experiments and the possibility of generating it through validated software such as Odeon.

3.3 Results

In this study, the performance of the system was evaluated by collecting and analyzing results based on the positions of a mannequin within the virtual sound field (*i.e.*, center and off-center) and the conditions under which the system was tested (*i.e.*, with and without the presence of a second head-and-torso simulator). The results were presented in terms of angles referenced counter-clockwise, which allowed for a detailed analysis of the system's performance under various conditions. Through this analysis, it was possible to gain a comprehensive understanding of the system's capabilities and identify potential areas for improvement.

3.3.1 Analysis

The signals were played and simultaneously recorded; the recorded result carried the auditory spatial effects from auralization and also the physical limitations given by the virtualization setup (*e.g.,* loudspeakers' frequency response, and presence of loudspeakers inside the room). As the recorded sweep has a greater length than the original one, zero-padding was performed. In that process, zeroes are appended to the end of the time domain signal, obtaining the equivalent convolution nonetheless [242]. After that, it was possible to calculate the virtual environment's impulse response by dividing the recorded signal by the zero-padded version of the initial sweep, both in the frequency domain.

For both measurements, the interaural time difference is calculated by comparing the sound's arrival time in the impulse response between the two channels of a binaural room impulse response (BRIR). There are different methods for ITD calculation [132, 314]. In this work, ITDs were estimated as the delay that corresponds to the maximum of the normalized interaural cross-correlation function (IACF). According to the ISO-3382-1:2009 [127], the IACF is calculated as:

$$\mathrm{IACF}_{t1,t2}(\tau) = \frac{\int_{t1}^{t2} p_L(t) p_R(t+\tau)\mathrm{d}t}{\left(\int_{t1}^{t2} p_L^2(t)\mathrm{d}t\right)\left(\int_{t1}^{t2} p_R^2(t)\mathrm{d}t\right)} \tag{3.5}$$

where

$p_L(t)$ is the impulse response at the entrance of the left ear canal;

$p_R(t)$ is that for the right canal;

The interaural cross correlation coefficients, IACC [127], are given by:

$$\mathrm{IACC}_{t1,t2} = \max|\mathrm{IACF}(\tau)|,\ \text{for } -1\text{ms} < \tau < 1\text{ms}\,. \tag{3.6}$$

Similarly, to calculate the interaural level difference (ILD), a fast Fourier transform (FFT) is applied to the time domain's impulse responses, the spectrum is divided into averaged octave bands, and the ratio in dB between the frequency magnitudes are calculated as the ILD:

$$\mathrm{ILD}(n) = 20\log_{10}\left(\frac{\sqrt{\int p_R^n(t)^2}}{\sqrt{\int p_L^n(t)^2}}\right), \tag{3.7}$$

where:

n is the given frequency band;

$p_R^n(t)$ is the bandpassed right impulse response;

$p_L^n(t)$ that to the left channel.

3.3.2 Centered position

In the centered-position configuration, (Figure 3.6a), the listener remains at the ideal VSE position (center) to focus on the effect of an added listener inside the loudspeaker ring. This framework can be valuable to auditory research as it can be used to analyze group responses to interviews, arguments, collaborative work, social stress or disputes between individuals in listening tasks.

The IACC to the frontal angle (0°) across frequencies is shown in Figure 3.7. High values indicate that the system delivers the same signal to both ears. Conversely, the drop in IACC values at high frequencies can indicate that the Ambisonics may fail to render specific frequencies affecting the octave bands analysis. The IACC values measured across all angles for VBAP and Ambisonics can be found in Figure 3.8. They indicate that Ambisonics tend to provide less lateralization in lower frequencies (constant and higher IACC values) and lower but constant values in high frequencies, possibly translating to blurred sound localization.

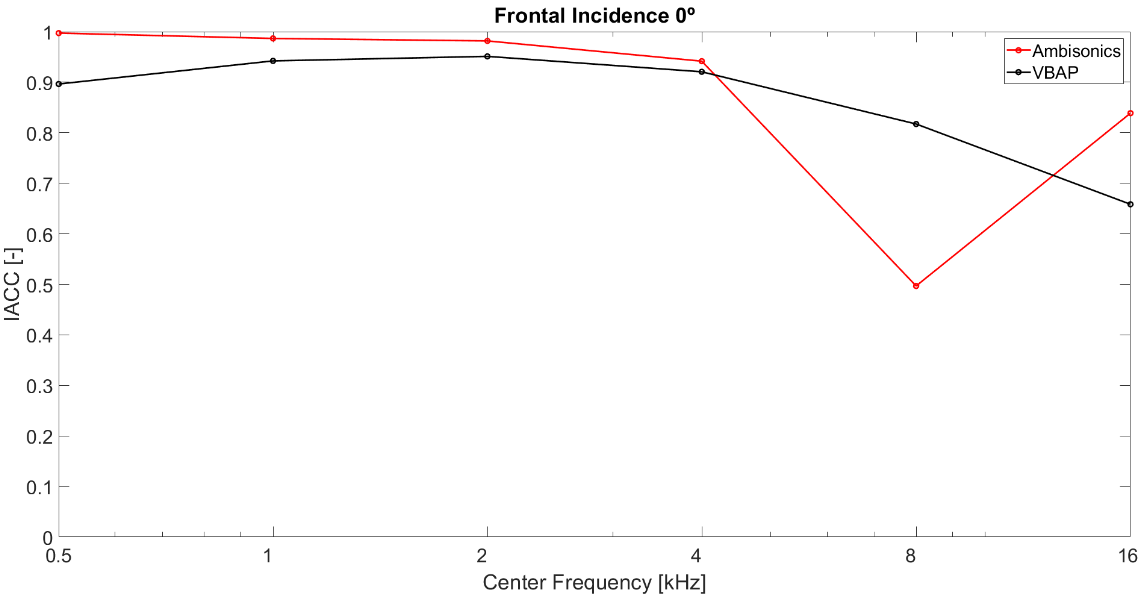


Figure 3.7: *Interaural cross correlation as a function of frequency in octave bands - Frontal angle 0º.*

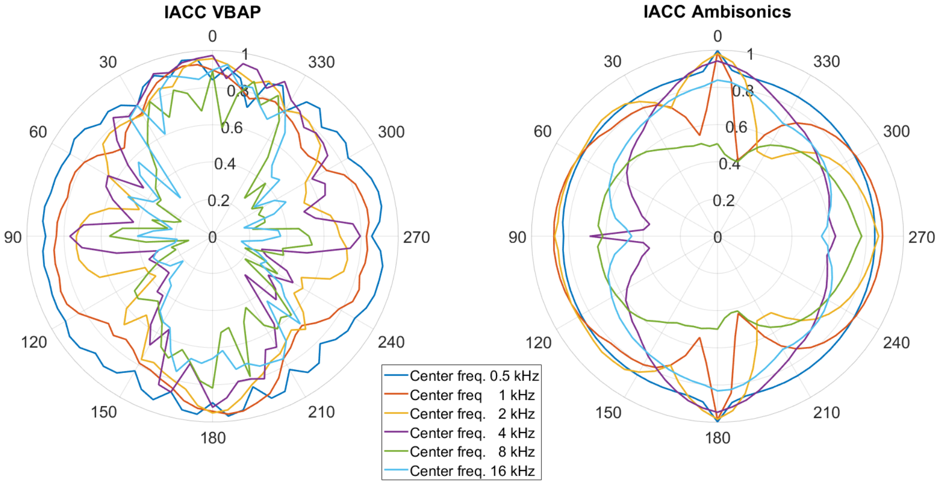


Figure 3.8: *Interaural cross correlation for averaged octave bands in Ambisonics and VBAP techniques represented in polar coordinates.*

That can happen due to a tilt in positioning the hats or imprecision from the virtualization system. For example, a high-frequency sound wave at 8 kHz has a wavelength of approximately 4 cm and 2 cm at 16 kHz, which means that even a slight tilt can influence high-frequency IACC. Furthermore, the inverse FIR filter applied was not the inverse broadband signal, but the filtered in third of octave bands. That decision was a signal processing compromise, as a broadband filter would only partially compensate for loudspeakers' geometry or phase differences in high frequencies. This point can be further investigated as a way to improve Ambisonics reproduction.

There is a relative increase of variations with frequency in VBAP results, which are present to a lesser extent in the Ambisonics IACC results. That reveals a difficulty from Ambisonics to drive a good sense of localization as a high coherence level indicates the sound coming from front or back [58]. At the same time, due to the Ambisonics activation of all available loudspeakers to render the sound in the sweet spot area, the sense of immersion is higher.

3.3.2.1 Centered ITD

The ITD results presented were obtained after a tenth-order low-pass Butterworth filter (LPF) was applied. The filter's cutoff frequency was 1,000 Hz to approximate the low frequency dominance in ITD [38, 124, 197, 242].

Vector Based Amplitude Panning The light blue line in Figure 3.9 shows the results for the ITD from the initial setup (HATS alone centered). The system presented a magnitude peak in response time of approximately 650 µs, which corresponds to approximately 22 cm for a wave traveling at the velocity of sound propagation in the air. This distance is comparable to the distance between HATS microphones (19 cm). It is appropriate to note that the symmetry of HATS is also presented in the HATS alone results (triangles in Figure 3.9) providing reassurance in the quality of the data collected.

The HATS was kept in the center of the loudspeaker ring for the next set of measurements. A second listener's influence was then simulated by introduc-

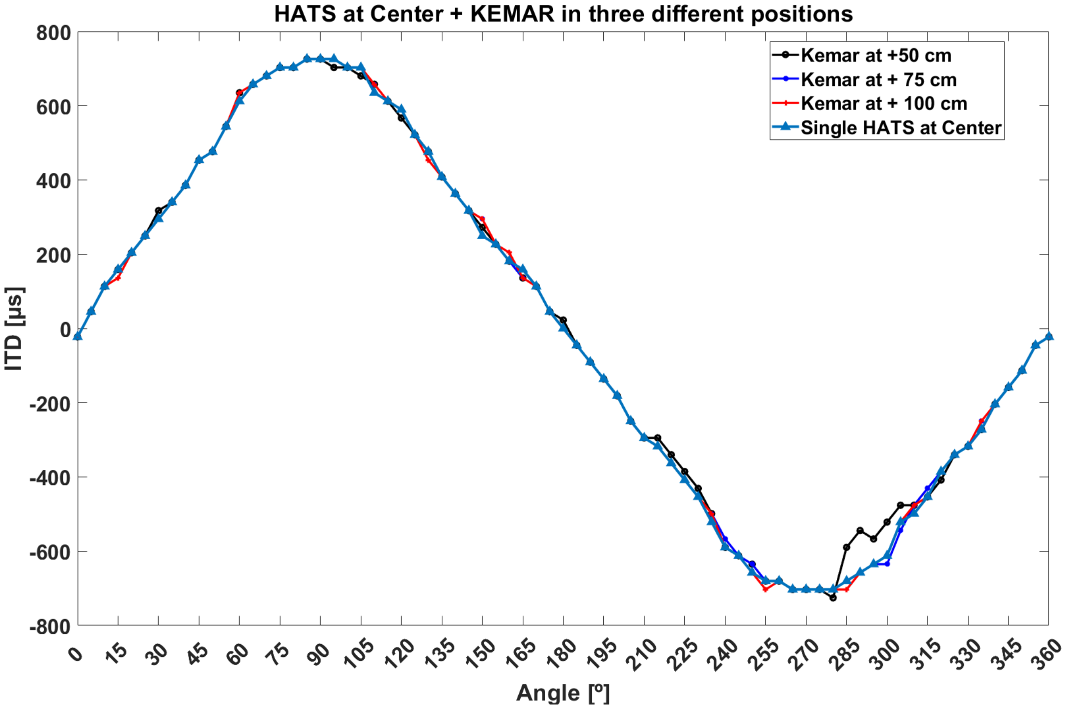


Figure 3.9: *a) HATS alone at center. b) Light blue line: HATS alone at center. Black line: HATS centered and KEMAR at 0.5 m to the right. Blue line: HATS centered and KEMAR at 0.75 m to the right. Red line: HATS centered and KEMAR at 1 m to the right.*

ing a KEMAR and laterally varying its position along the lateral axis (x-axis). The results are presented in Figure 3.9. The ITD data obtained from this experiment make it possible to comprehend that the second mannequin (KEMAR) has an impact as an obstacle on the interaural time difference in the HATS at the center of the loudspeaker ring.

In the closest position of the second listener (50 cm from the center), there is a reduction of ITD values (angles between 285 and 305 degrees). Thus, the maximum difference is 50 us. That effect is related to the insertion of the physical obstacle represented by the second listener. As the sound wave diffracts, different paths to the listener's ears are imposed, reducing the sound's arrival time between ears. Therefore, the effect should be centered at 270 degrees. However, the second listener was not perfectly aligned to the lateral of the centered listener. That was a limitation of the experiment as the KEMAR was placed in an ordinary chair, and its bottom is not flat.

Ambisonics The ITD results for the initial setup (HATS alone centered) virtualized from Ambisonics auralization are presented in Figure 3.11. The system showed a magnitude peak in response time, roughly 600 μ, 50 μ lower than the VBAP method. Another characteristic of Ambisonics ITDs is the flat behavior around the lateral angles, which is generated mainly by the chosen order of the Ambisonics auralization. In first-order, the horizontal directivity is determined by the to an intersection of three bi-directional (figure-eight) sensitivity patterns, circumvented by a omnidirectional one, as illustrated in Figure 3.10. That can also limit the localization performance when utilizing first-order Ambisonics, even when reproduced through a higher number of loudspeakers.

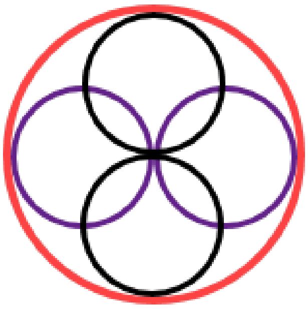


Figure 3.10: *Horizontal 2D Ambisonics directional sensitivity crop representation. The red line represents an omnidirectional pattern, the black line represents a bidirectional pattern, y-axis oriented (null points at the sides), and the purple line is a bidirectional pattern representation x-axis oriented (null points in front and the back).*

The HATS was kept in the center of the loudspeakers ring and simulated a second listener's influence on the sound field by introducing a KEMAR to three different positions along the x-axis 50, 75, and 100 cm to the left of HATS (*i.e.* at 270°). The results are presented in Figure 3.11 by the black, blue, and red lines. The data clearly demonstrated that as an obstacle, the second listener (KEMAR) does not influence the interaural time difference when using Ambisonics, and HATS is at the center of the loudspeaker ring.

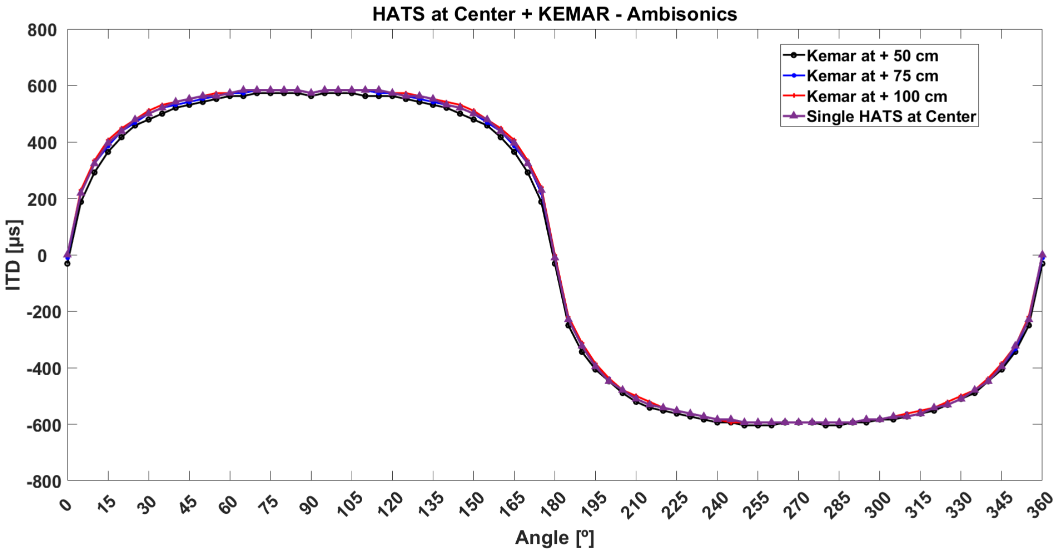


Figure 3.11: *a) HATS alone at center. b) Light blue line: HATS alone at center. Black line: HATS centered and KEMAR at 0.5 m to the right. Blue line: HATS centered and KEMAR at 0.75 m to the right. purple line: HATS centered and KEMAR at 1 m to the right.*

3.3.2.2 Centered ILD

The effects in higher frequencies due to a second listener require an analysis of a different parameter. Instead of studying the difference in the arrival time of the sound between the ears, the representative metric is the level difference between the ears.

There are effects as absorption, reflection, and diffraction before the sound pressure signal reaches the eardrums. The torso, shoulders, outer ear, and pinna mechanically affect an incoming sound wave. These effects are angle and frequency-dependent, as different frequency waves have different wavelengths [39, 40, 90].

The effects on ILD caused by the virtualization process were calculated as the differences between the reference ILDs measures with HATS alone and centered and the ILDs measured with HATS and a second mannequin (KEMAR). As a reference, Figure 3.12 presents the ILDs by each method from twelve different angles (30 degrees separation) around the listener.

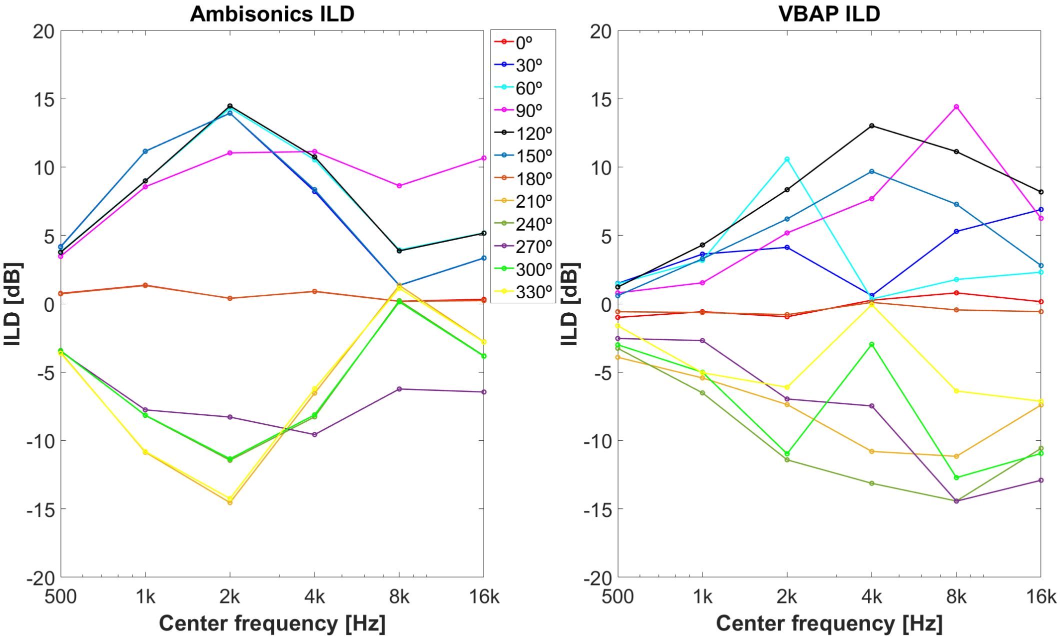


Figure 3.12: *Interaural Level Differences as a function of octave-band center frequencies in twelve different angles around the central point.*

There are differences between ILDs calculated from measurement with both techniques on the energy in the averaged octave bands. However, the ILDs from VBAP present a significant effect based on incidence angle (more natural) than the Ambisonics [222]. Furthermore, the ILD peak for the Ambisonics is observed around 2000 Hertz, which can be interpreted as the limit in frequency reproduction of level difference between ears when decoding through 24 loudspeakers [299]. A more comprehensive comparison between techniques with the HATS centered alone can be observed in the heatmap representation from Figure 3.13 including all 72 angles (5 degrees separation) measured. The homogeneity across angles from Ambisonics measurements indicates that its ILD lacks precision as a binaural spatial cue. Localization accuracy in Ambisonics reproduction, especially to lateral angles, is highly dependent on its order (acquisition and reproduction) [27].

Figure 3.14 shows the energy difference across the octave bands for eight different incidence angles on both techniques with and without the presence of the second mannequin. On both techniques, the strongest influence happens

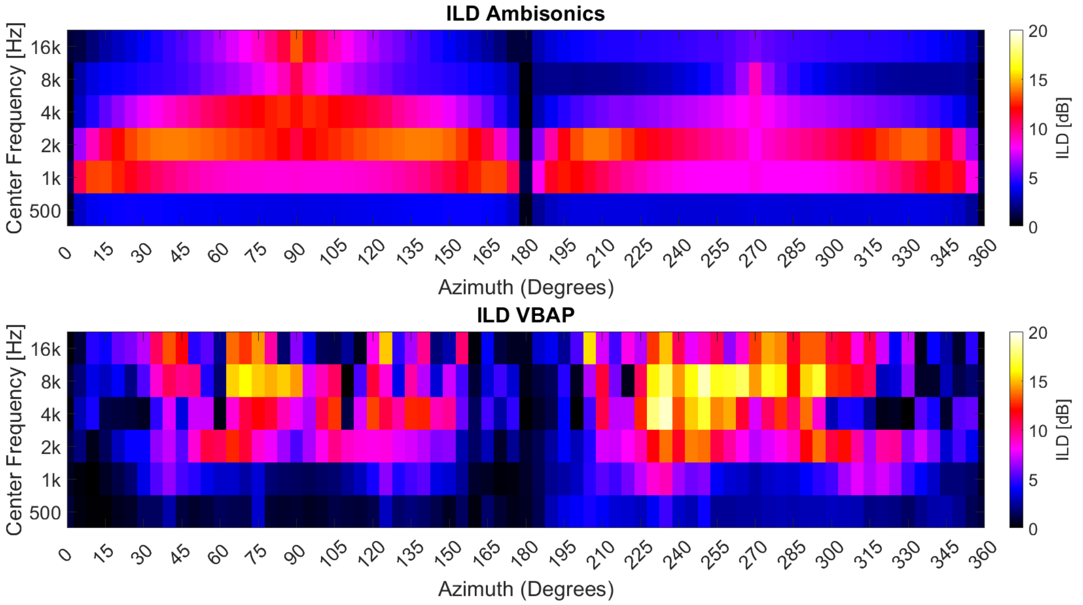


Figure 3.13: *Interaural level differences averaged octave bands as a function of azimuth angle for a HATS Brüel and Kjær TYPE 4128-C in the horizontal plane.*

when the second mannequin is closest to the center. The second listener is to the right angle in VBAP (270°), while in the Ambisonics is positioned to the left (90°).

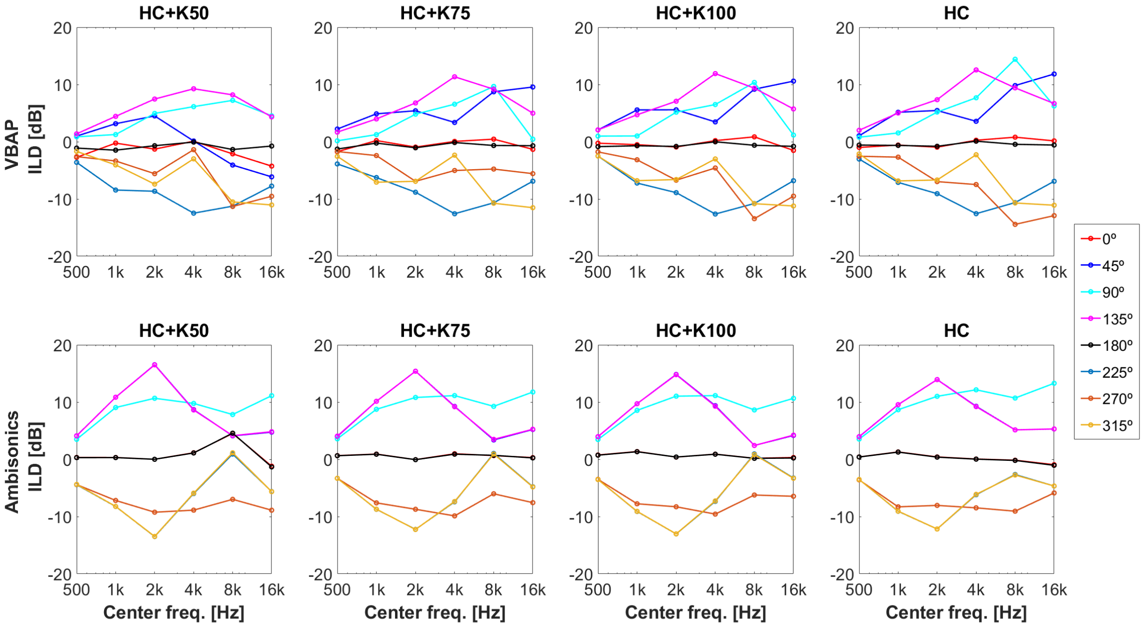


Figure 3.14: *Interaural Level Differences (octave band) angles around the central point considering different displacement of the second listener.*

The ILDs calculated from measurements with the second mannequin present are not extensively different compared to the reference ILD. The difference is proposed to be observed as a distortion parameter. These differences were calculated by subtracting the ILDs with the second mannequin from the specified center alone reference ILD. Ideally all graphs should be black for a full match (no difference between different setups/positions), meaning no measured distortion.

Vector Based Amplitude Panning Figure 3.15 presents the differences between ILDs calculated from HATS centered (HC) and the configuration that combines the HATS centered plus the KEMAR in one of the three positions (*e.g.*, HC K-50 is the defined notation to HATS centered and KEMAR at 50 cm to the right). The sounds were auralized via VBAP for all 72 angles (5° spacing). The angles that correspond to loudspeaker locations (15° spacing) is placed were reproduced directly by the physical loudspeaker at that angle.

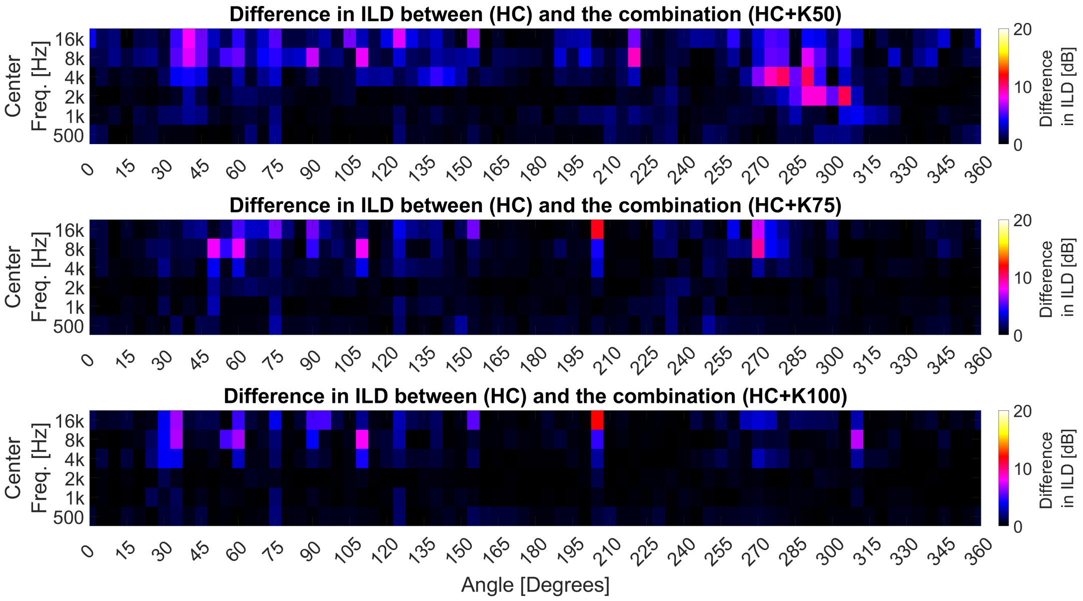


Figure 3.15: *VBAP discrepancies in ILD between HATS at the center and: (Top) HATS at the center plus KEMAR at 50 cm to the right, (Middle) HATS at the center plus KEMAR at 75 cm to the right, (Bottom) HATS at the center plus KEMAR at 100 cm to the right.*

The differences in frequencies over 1 kHz are pronounced for angles to the right side of centered HATS, 270-305° azimuth. Smaller effects can also be noted on other angles that correspond to virtual sound sources (where there is no loudspeaker, and the sound source is produced via the auralization technique). These effects are diminished as the second mannequin position increases away from the centered receptor, indicating a smaller acoustic shadow.

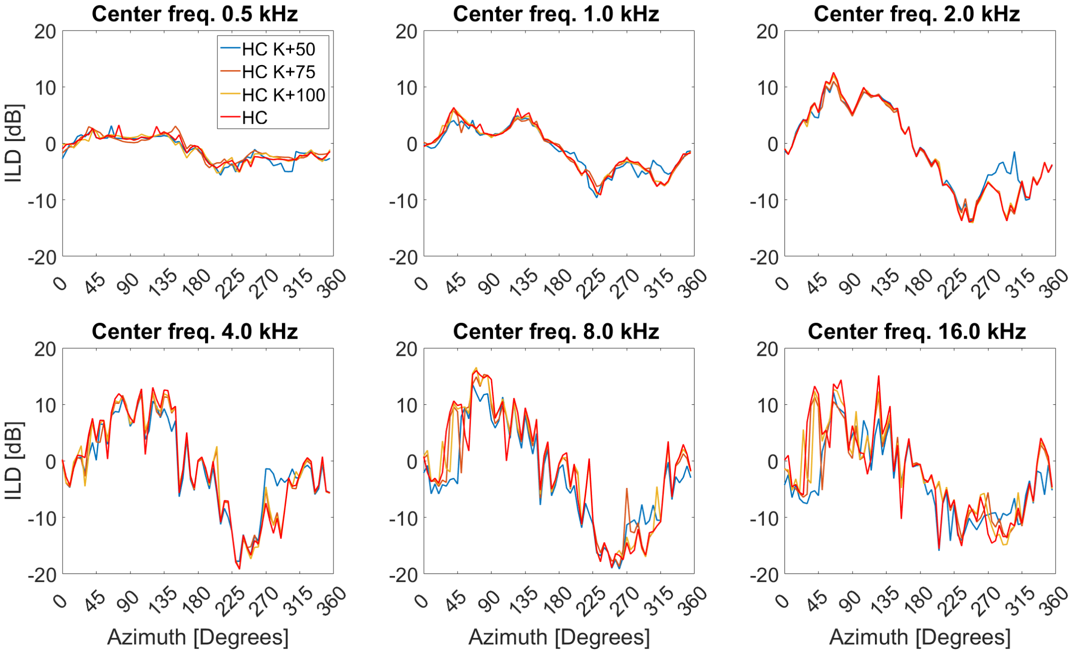


Figure 3.16: *VBAP Interaural Level Differences as function of azimuth angle around the centered listener.*

Figure 3.16 shows the ILD in six octave bands from impulse responses recorded with files auralized using VBAP. The HATS centered (HC) position refers to HATS alone and it is compared to the configurations adding the second listener (KEMAR) in three different positions 50, 75 and 100 cm displaced from the center (K+50, K+75, and K+100, respectively).

The mismatch is pronounced when KEMAR is closer (blue line), especially in the angles blocked by KEMAR. As the second listener blocks the sound wave, an acoustic shadow is created, which reduces the sound energy to the ear facing the sound source, decreasing the level difference between ears. There is also a reduction in ILD for angles from 35 to 50. That can be related to the opposite effect where the mannequin reflects part of the sound, increasing the level to HATS-centered counter ear. The finding supports interpreting that a substantial effect occurs on ILDs to the KEMAR's closer position.

Ambisonics Figure 3.17 presents the calculated differences between ILDs from the Ambisonics auralization with the same configurations (*i.e.*, HC *vs.* HC K-50, HC *vs.* HC K-75, and HC *vs.* HC K-100).For convenience, the second mannequin was positioned to the left of the center (90°). The switch from right to left does not affect the comparison as both HATS and the Eriksholm test room are symmetric. Figure 3.18 shows the ILD in octave bands, to highlight the stronger effect being at the 8 kHz.

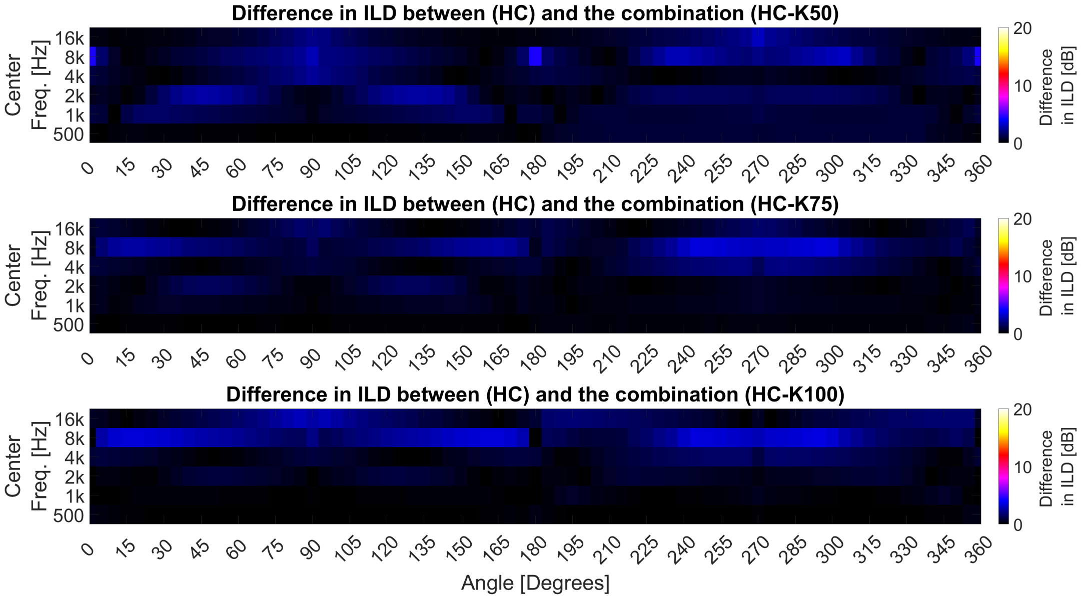


Figure 3.17: *Ambisonics discrepancies in ILD between HATS at the center and: (Top) HATS at the center plus KEMAR at 50 cm to the left, (Middle) HATS at the center plus KEMAR at 75 cm to the left, (Bottom) HATS at the center plus KEMAR at 100 cm to the left.*

The results demonstrate that including a second listener has a negligible effect on Ambisonics first-order ILDs. However, on the reference measurement (HATS alone), the ILDs did not adequately reproduce this spatial cue throughout the angles around the listener, given the observable minor ILD differences across angles, especially over 2 kHz.

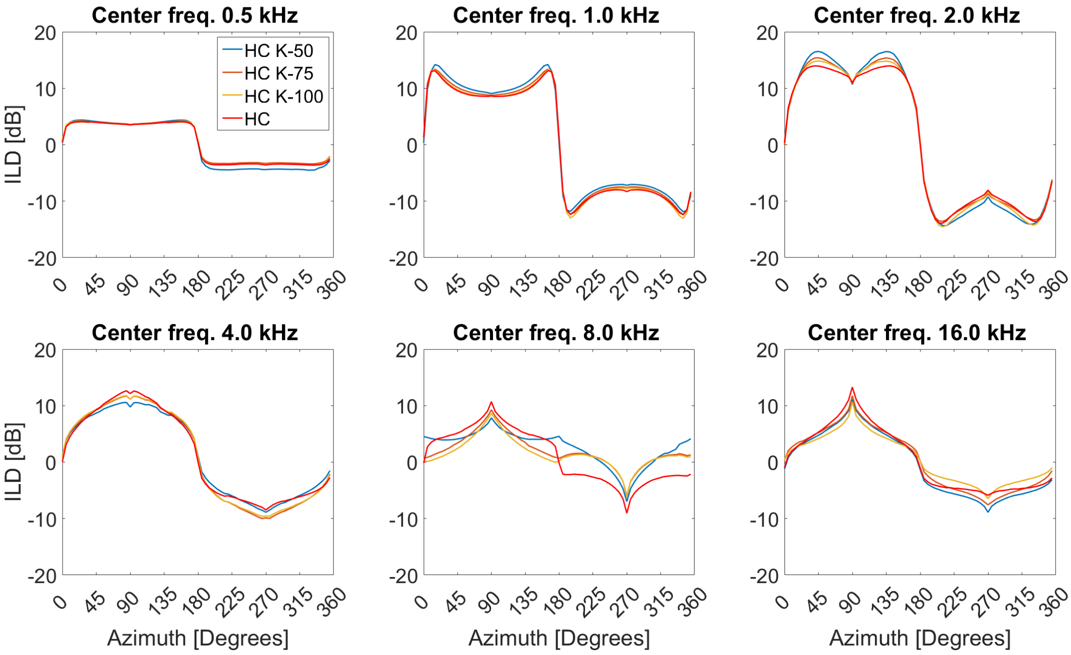


Figure 3.18: *Ambisonics Interaural Level Differences as function of azimuth angle around the centered listener.*

3.3.3 Off-centered position

Being able to have the participant away from the center of the loudspeakers ring can be valuable for testing simultaneous participants or a particular physical apparatus' influence (*e.g.* Listening effort evaluated under presence of another individual [230]). Auditory research that aims to test the influence of a particular noise, SNR, or the direction of the noise on the interaction in participants' conversation can benefit from a setup that would make it possible to virtualize a sound scene and present it without spatial distortions. Measurements aiming to study the influence of off-center HATS displacement were performed in nine different configurations: with HATS and KEMAR independently displaced 25, 50, and 75 cm from the center, resulting in separations of 50, 75, 100, 125, and 150 centimeters (See Figure 3.6b).

The listening position is critical to the auralization process techniques presented in this work as they are derived and programmed to render the sound in the center of a loudspeaker array. Adding computer power to real-time processing could handle participant movements; although that can be considered, it was not in this part of the experiment scope. Such processing focuses on dynamics (head motion). The focus here is the effects of sub-optimal positions and the influence of a second listener as an obstacle to the sound field.

3.3.3.1 Off-center ITD

The effects of off-center positioning on sound's arrival time can affect the subjective perception of the sound incidence direction.

Vector Based Amplitude Panning Observing the ITD results shown in Figures 3.19, 3.20, and 3.21, almost no influence of the second mannequin (KEMAR) can be noted even with the HATS off-center. The ITD at off-center positions deviates from the ITD from HATS centered at the same proportion regardless the second listener (KEMAR) position.

Nonetheless, Figure 3.22 shows that a pronounced effect appears by shifting out the HATS off center. When exceeding 25 cm, the spikes represent a difficulty of the vector-based amplitude panning process to generate the virtual sound sources. This behavior is expected as the VBAP mathematical formulation is derived by a unitary vector pointing to the center.

In figures 3.20 and 3.21 it is possible to observe more considerable distortions (sharp peaks crossing the reference line in addition to being offset from the reference line) in the ITD for the virtual sound sources. Such distortions increase as HATS is moved away from the central position. Sound sources reproduced using VBAP in this loudspeaker ring to these receptor positions would not be correctly interpreted in terms of direction by the listener.

The ITDs difference is greater when the sound sources are at angles close to the

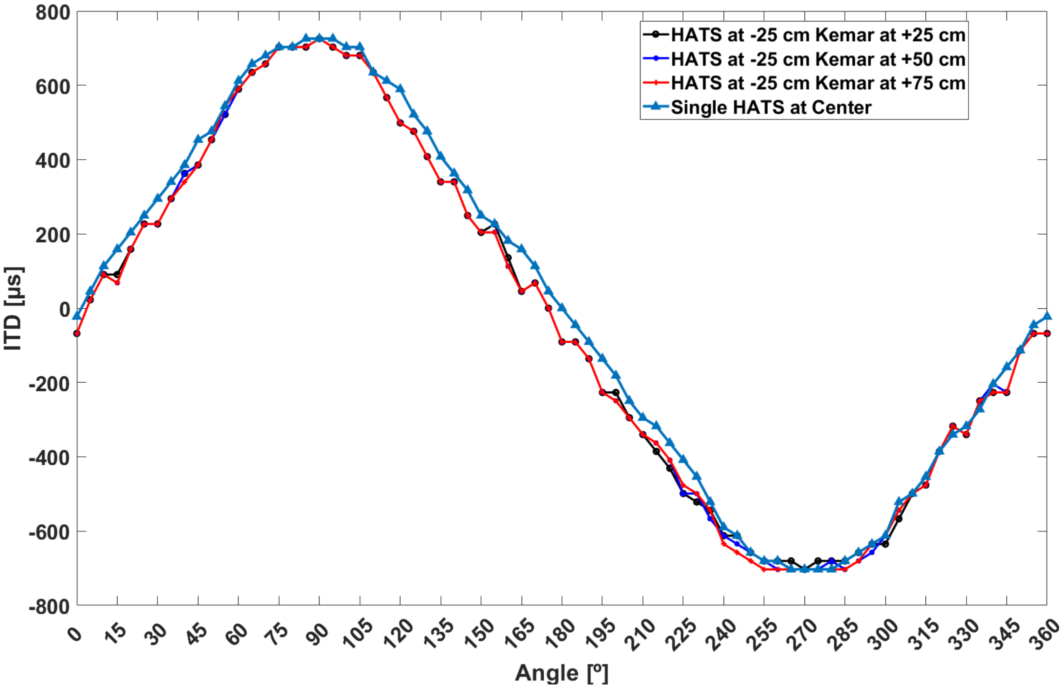


Figure 3.19: *ITD as a function of source angle Light blue line: HATS alone at the center. Black line: HATS at -25, KEMAR at +25. Blue line: HATS at -25, KEMAR at +50. Red line: HATS at -50, KEMAR at +75.*

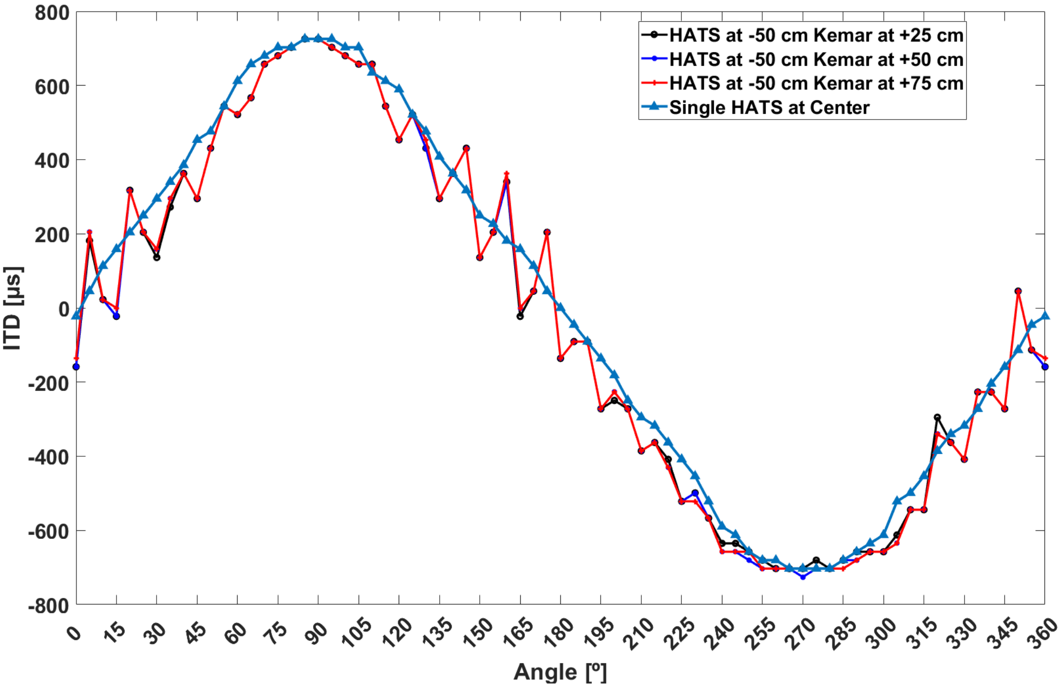


Figure 3.20: *ITD as a function of source angle Light blue Line: HATS alone at the center. Black Line: HATS at -50, KEMAR at +25. Blue line: HATS at -50, KEMAR at +50. Red line: HATS at -50, KEMAR at +75.*

front or rear (0° and 180°) directions. This effect is related to HATS physical displacement. The ITD results at lateral angles are representing a larger lobe to the HATS right ear (270°), and a sharpened lobe at the HATS left ear (90°)

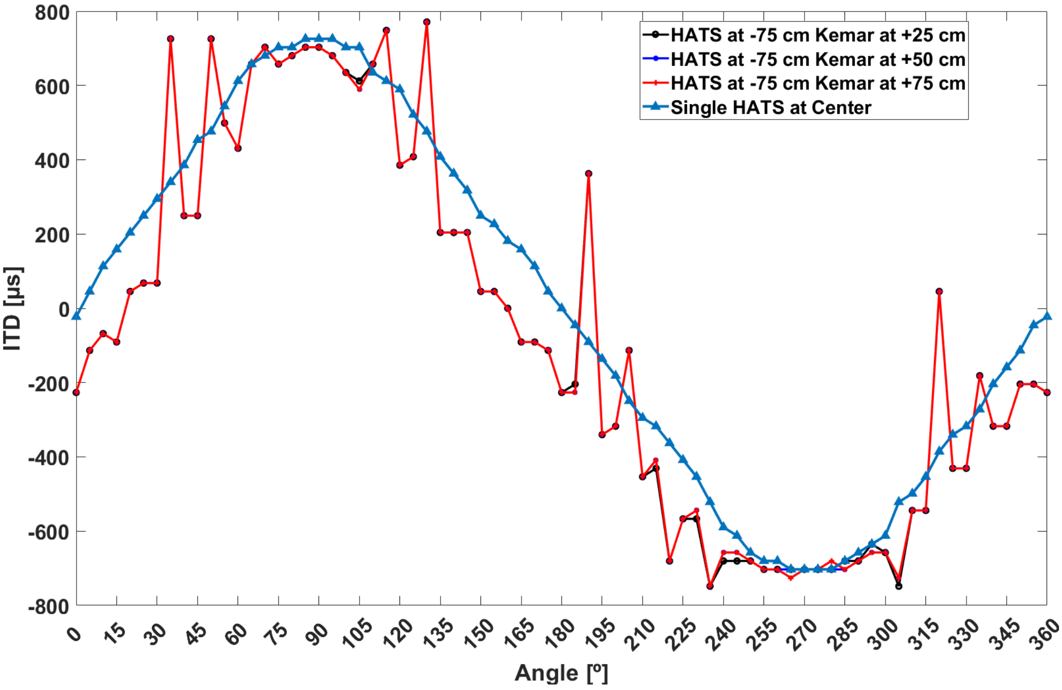


Figure 3.21: *ITD as a function of source angle. Light blue line: HATS alone, centered. Black line: HATS at -75, KEMAR at +25. Blue line: HATS at -75, KEMAR at +50. Red line: HATS at -50, KEMAR at +75.*

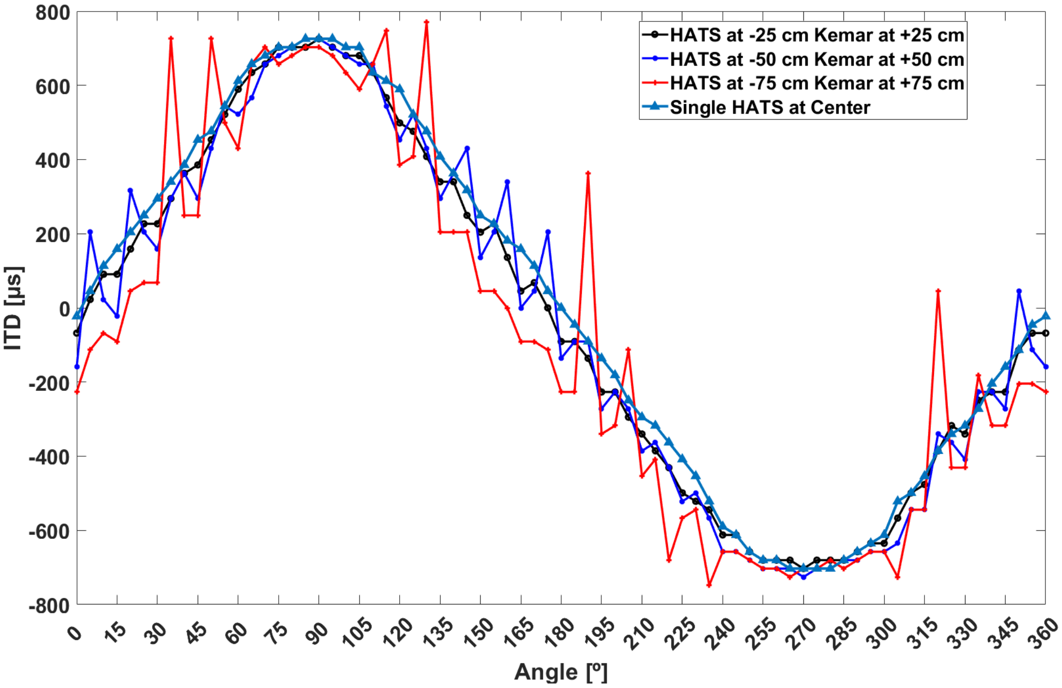
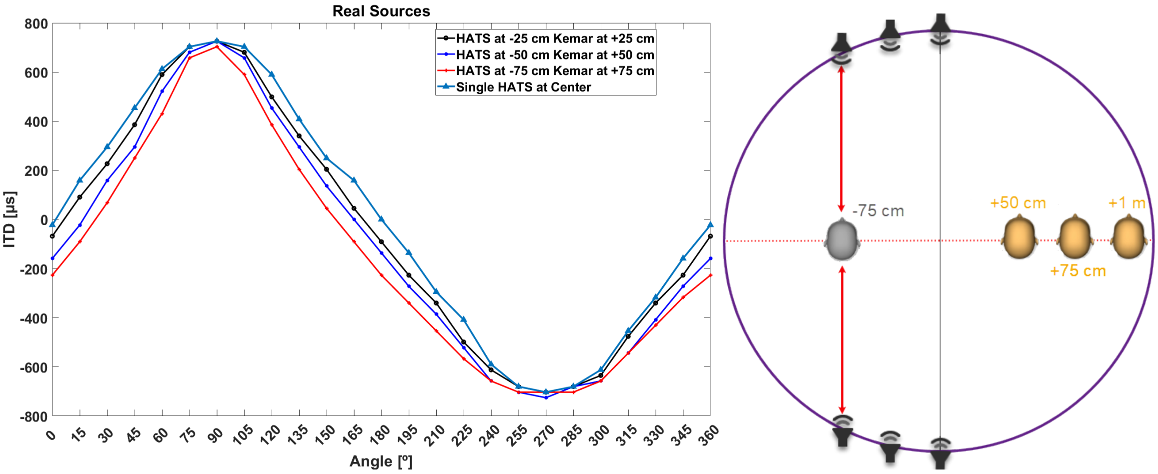


Figure 3.22: *ITD as a function of source angle. Light blue line: HATS alone, centered. Black line: HATS at -25, KEMAR at +25. Blue line: HATS at -50, KEMAR at +50. Red line: HATS at -75, KEMAR at +75.*

shows the off-center displacement. This effect occurs because HATS is not at the center of the ring (See Figure 3.23b), and the angles and separations between the loudspeakers are modified. The effect is even more apparent when looking only at the real sound source (angles correspondent to loudspeaker locations) ITDs, without the distortions created by VBAP auralization, (See Figure 3.23a).



(a)

(b)

Figure 3.23: *a) ITD for real sound sources. Light blue line: HATS alone, centered. Black line: HATS at -25, KEMAR at +25. Blue line: HATS at -50, KEMAR at +50. Red line: HATS at -75, KEMAR at +75. b) HATS off-center position -75 cm scheme facing the third loudspeaker.*

Ambisonics The VBAP method constructs the auditory spatial cues through one to three loudspeakers in this setup, usually in the same quadrant. Ambisonics, in contrast, uses all the available loudspeakers in the rendering process. Hence, the sound localization is benefited on VBAP auralization compared to Ambisonics due to the nature of the methods [104, 105, 175, 180, 221]. Furthermore, the ITDs results observed from the first-order Ambisonics reflect the method's limitation on sweet spot size.

Figure 3.24 shows the calculated ITD in three different configurations H+25 K-25, H+50 K-50, H+75 K-75, and the center configuration for comparison. To improve readability, the ITD results for the remaining spatial configuration (which were similar across conditions) can be found in Appendix A.

The expected size of a listening area is 20 cm when combining 24 speakers to reproduce Ambisonics in a 2D horizontal matrix [299]. The displacement

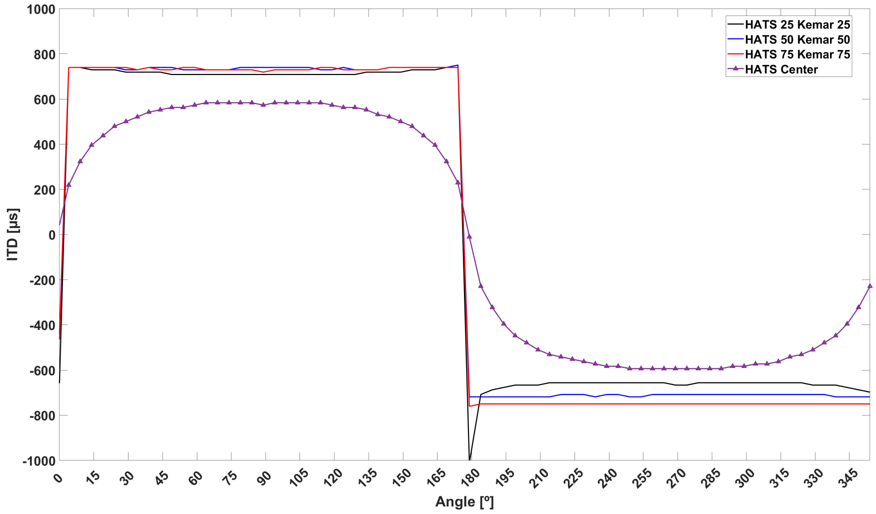


Figure 3.24: *ITD as a function of source angle in Ambisonics setup. Light blue line: HATS alone, centered. Black line: HATS at -25, KEMAR at +25. Blue line: HATS at -50, KEMAR at +50. Red line: HATS at -75, KEMAR at +75.*

of 25 cm and greater puts the receptor outside the sweet spot. Therefore, it is possible to observe in Figure 3.24 that Ambisonics does not virtualize this acoustic track correctly outside the center position, as the values remain mostly constant for the side being played.

3.3.3.2 Off-center ILD

ILDs can be highly sensitive to the listener's position on a virtualized sound field, given the considered smaller wavelengths. The composition of a virtualized sound wave is be performed by simultaneously combining sounds from several sound sources, which requires a highly precise combination.

This section investigates the ILD changes due to having the listener away from the optimal position while having another listener present. ILD's influence when the HATS and a second participant are away from the center.

A comparison of the ILD results across the positions is shown in Figure 3.25; it presents for both techniques the calculated ILDs over frequency on eight incidence directions spaced over 45 degrees in the azimuth on three different positions plus the centered position as a reference. The pattern deviation as the receptor is moved from the center is not the same across the techniques. As expected, the physical construction of the summed sound wave from Ambisonics that relies on all loudspeaker has a higher impact on ILDs than the VBAP which only combines few sound sources from the same quadrant.

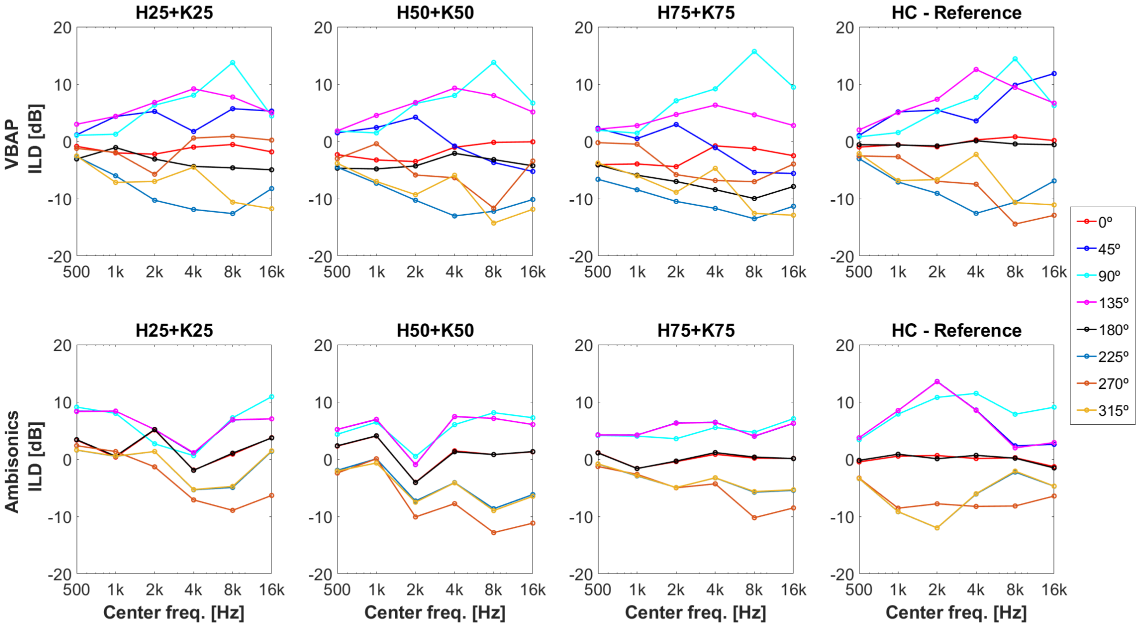


Figure 3.25: *ILD as a function of frequency at different angles (line color) for VBAP (top row) and Ambisonics (bottom row) for symmetrical displacement in off-center setups.*

On files auralized through VBAP, the discrepancies between ILD measured having HATS in the center (optimal position) and the other positions can be interpreted as acoustic artifacts capable of conveying the wrong localization of the sound source. Although the second listener did not have a primary influence, the observed displacement from the center affects the ILD pattern, especially the higher frequencies. For Ambisonics, the listener position is critical. The ILD differences from center to off-center positions create artifacts that compromise ILD used as a cue to the sound localization on all tested positions.

Vector Based Amplitude Panning The top row of Figure 3.25 shows the ILD screening in some of the incidence angles. Comprehensive visualization of ILDs across angles is presented in Figure 3.26 for the reference-centered (top) and off-centered positions. There is an effect on ILDs when moving the receptor from the center position and adding a second listener inside the loudspeaker's ring. Although noticeable, the effect still preserves the pattern, allowing the difference to be interpreted as artifacts. The vertical zeroes ILDs indicated the frontal and rear angles (0° and 180°) where the sound should arrive at the ears with the same level. These vertical black lines are shifted as the listener is displaced from the center. At 75 cm displacement the lowest value vertical line on Figure 3.26 appears is 35° (frontal) and 145° (rear)

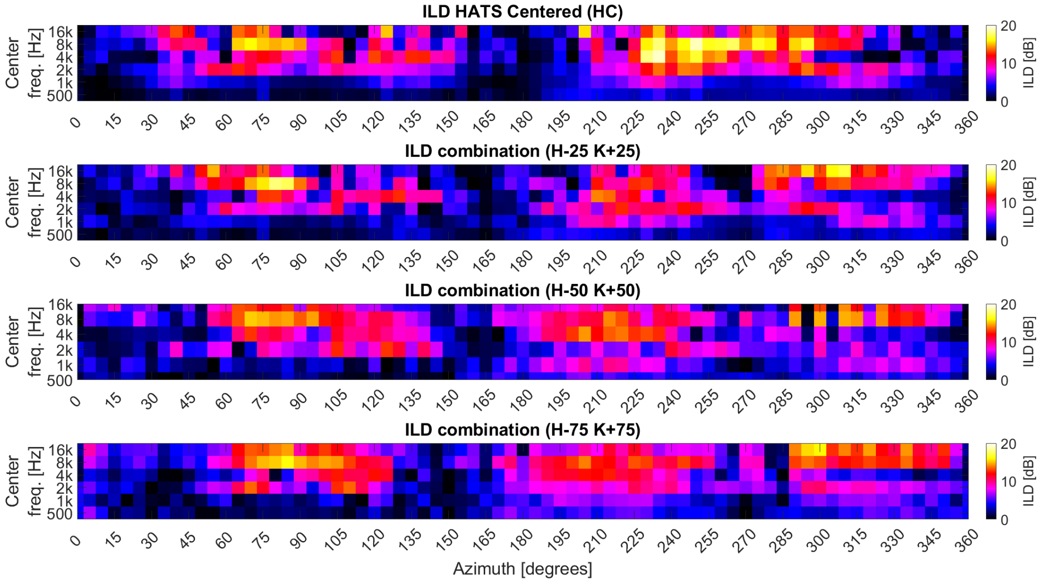


Figure 3.26: *VBAP setups: ILD on centered position (top); ILD on off-center setups: HATS at 25 cm to the left with KEMAR at 25 cm to the right (middle top); HATS at 50 cm to the left with KEMAR at 50 cm to the right (middle bottom); HATS at 75 cm to the left with KEMAR 75 cm to the right (bottom).*

The difference between ILD with the HATS in the reference position (alone and in the center) and the configurations with HATS outside the center simultaneously with KEMAR are shown in Figures 3.27, 3.28 and 3.29.

The acoustic field behavior outside the center of the ring at frequencies above

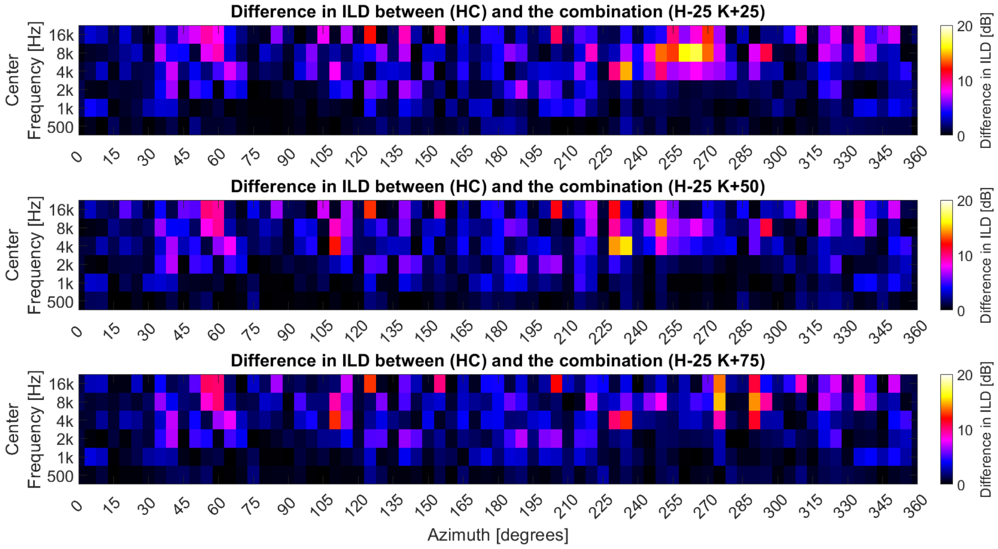


Figure 3.27: *VBAP differences in the ILD between centered Alone and off-center with KEMAR setups: HATS at 25 cm to the left with: KEMAR at 25 cm to the right (top); KEMAR at 50 cm to the right (middle); KEMAR 75 cm to the right*

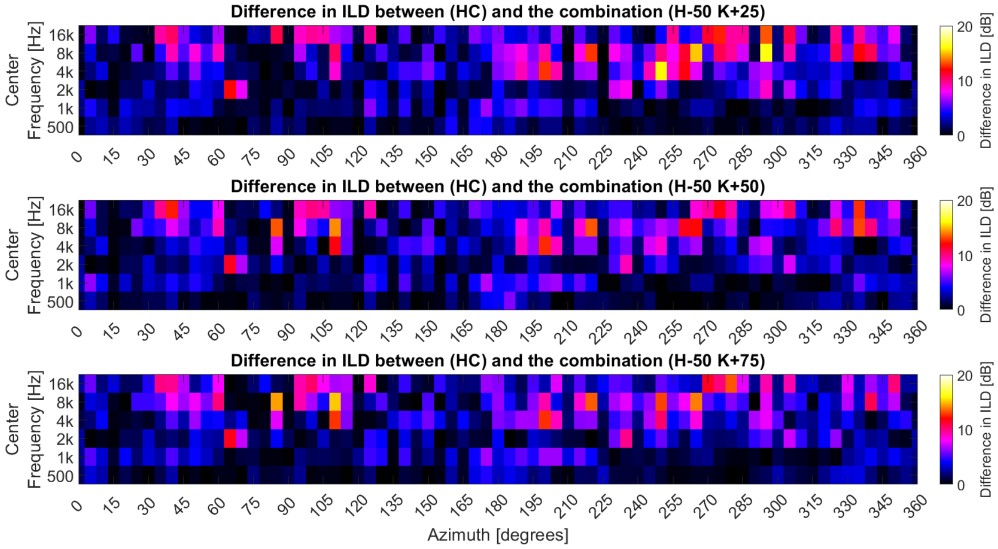


Figure 3.28: *VBAP differences in the ILD between centered setup and 25 cm off-center VBAP setups: HATS at 50 cm to the left with: KEMAR at 25 cm to the right (top); KEMAR at 50 cm to the right (middle); KEMAR 75 cm to the right (bottom).*

1 kHz presents significant ILD differences for the measured configurations, especially on angles that are virtual sound sources. The ILD difference reaches up to 15 dB.

As in the ITD, the ILD data from HATS in the off-center position shows the acoustic shadowing effect caused by KEMAR. It is possible to note that as

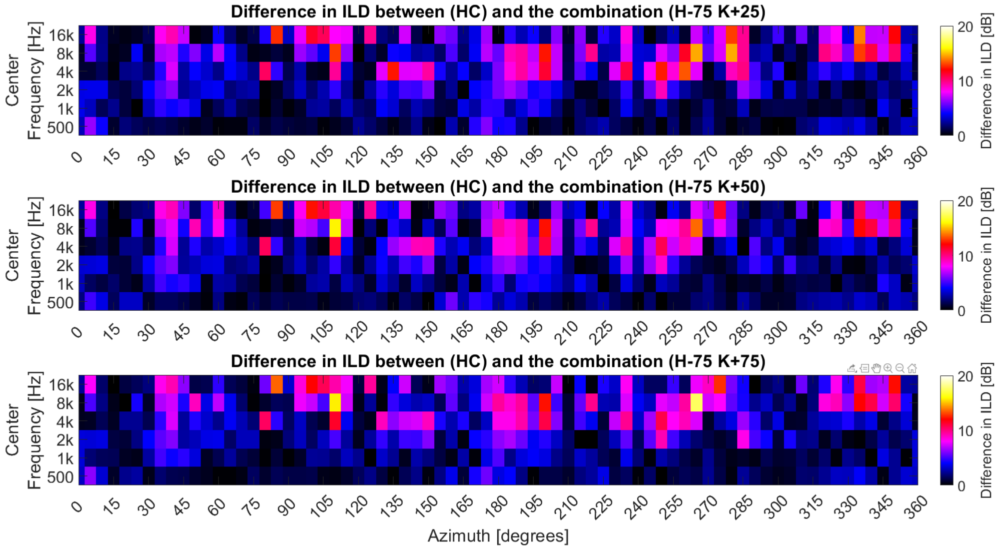


Figure 3.29: *VBAP differences in the ILD between centered setup and off-center setups: HATS at 75 cm to the left with: KEMAR at 25 cm to the right (top); KEMAR at 50 cm to the right (middle); KEMAR 75 cm to the right (bottom).*

close as KEMAR is positioned to HATS, greater discrepancies in ILD around positions near 270 degrees occur. This effect is due to the diffraction and absorption of the sound on the second listener (KEMAR), and happens for both real (loudspeaker) and virtual sound sources locations.

Ambisonics Ambisonics presents a more considerable limitation regarding movement outside the center of the ring due to its nature. The sound composition requires a combination of amplitude and phase from all available loudspeakers being the correct representation achieved only for an area at the center and without obstructions. The ILD in octave bands is shown in Figure 3.30.

The low amplitude and homogeneity across frequencies demonstrate that Ambisonics is limited to render the binaural cue proposed, not appropriately delivering the level differences outside the center. The ILD differences from the off-center positions to the HATS centered are presented in Appendix B.

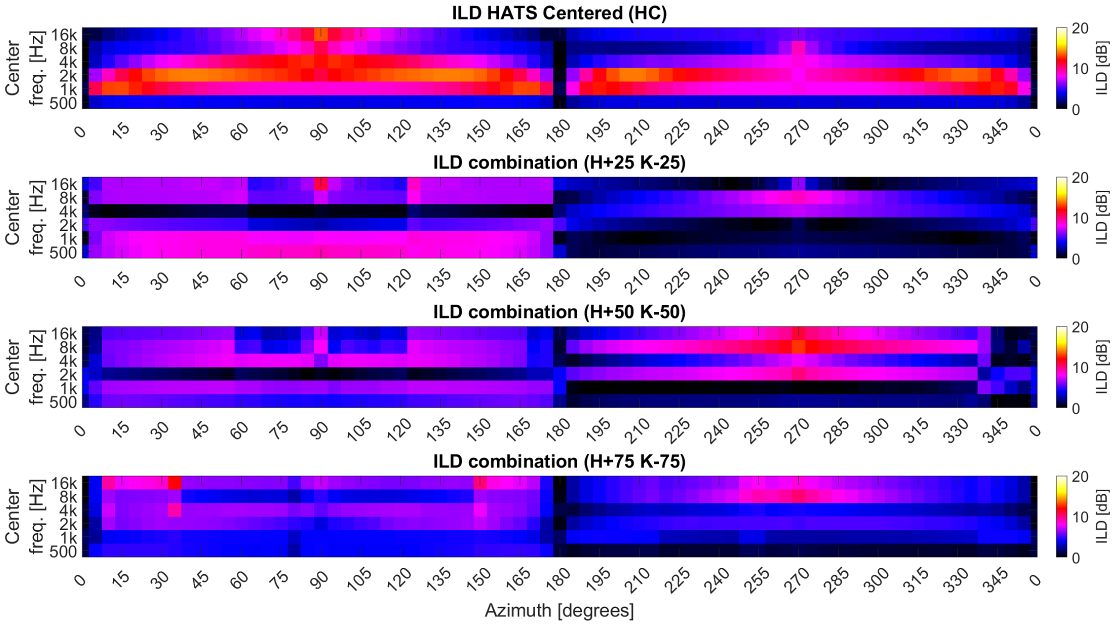


Figure 3.30: *Ambisonics setups: ILD on centered position (top); ILD on off-center setups: HATS at 25 cm to the left with KEMAR at 25 cm to the right (middle top); HATS at 50 cm to the left with KEMAR at 50 cm to the right (middle bottom); HATS at 75 cm to the left with KEMAR 75 cm to the right (bottom).*

3.4 Discussion

Once the listener is centered in the loudspeaker array, the second listener did not affect the auralization other than the angles physically shadowed by the second listener. Thus, a second listener does not deteriorate the spatial cues on both auralization techniques analyzed in this work.

For VBAP, the discrepancies in ITD only occur as the second listener was positioned 50 cm away, the closest measured position in this experiment. Also, differences in ILD for VBAP are more notable at the second listener's closest position. Concurrently, Ambisonics has not presented an apparent difference in ITD to a centered listener by placing a second listener inside the ring. The difference in Ambisonics ILDs from the centered reference indicates an acoustic shadow (this time at the left angle of 90 degrees) and an additional slight difference across other angles.

There is an apparent effect on ITD as the listener is moved out of the center. For VBAP, the peak of magnitude remains practically the same, approximately 650 microseconds, as the ITD 0 value (sound reaching simultaneously in both ears) is shifted. At 75 cm off-center to its left side, the difference in arrival time corresponds to a shift of 30 degrees approximately. That is in line with the setup, as the mannequin was placed in front of another loudspeaker. However, the Ambisonics ITDs demonstrate that the composition of magnitude and phase is not completed in off-centered positions. The Ambisonics weights are calculated to the sound waves from the loudspeakers to interact in the center position and then form a sound field representing a sound wave from a defined incidence angle. Moving the primary listener to the right makes the interaction between the loudspeakers inaccurate. In this case, the time difference turns wrong due to the Ambisonics truncation order to be low increasing the aliasing effect as it can be similarly observed in third and fifth order in Laurent *et al.,* [275].

The sound from the right mainly reaches the right ear and travels to the left ear before the sound from the left side can travel the extra distance. That is an expected effect since even the minimum displacement (25 cm) is larger than the expected reproducible area (around 20 cm) for this setup. There was no difference observed as the second listener (KEMAR) positions were changed (25, 50, and 75 cm to the right of the center) in all VBAP measurements with HATS positioned to the left of the center. Considering the ITD just noticeable difference (JND) in an anechoic condition is of the order of the 10 to 20 microseconds [38, 140, 241], the ITD results when the off-center HATS position was 25 cm to the left were a good approximation of the reference-centered measurement.

That means that a listener would not be able to discern the difference concerning the direction of incidence if placed in these positions relying only on the ITD cue. It is also worth considering that the JND to reverberant conditions is even higher [140] and the artifact can be masked by reverberation [97] which would benefit the auralization process. The HATS measurements positioned on 50- and 75-cm presents peaks and crossover values across the line that corresponds to centered ITD, which indicate distortion problems at low frequencies regarding the spatial cue. A similar analysis of the KEMAR impact on ITDs from Ambisonics virtualization can not be achieved since the ITD is not accurately rendered outside of the sweet spot.

Each result of the interaural level difference position combination (HATS and KEMAR) was subtracted from the HATS results alone to perform the ILD analysis off-center. In the VBAP method, a shadow effect generated by a second listener is present as expected, mainly when the first listener is 25 or 50 cm left of center. However, the differences in high frequencies are essentially on virtual sources, which indicates the difficulty of creating the virtual sound source impression outside the center position, independently of the second listener presence [2].

Off-center positions did not allow the accurate synthesis of the ILDs from the loudspeakers using Ambisonics. The method did not reproduce time or level differences accurately in these conditions, which could lead to not achieving the correct spatial impression. That is in line with literature, although generally investigating higher Ambisonics orders, the complexity of accurately render high frequency cues is present [279, 290], and also the off-center increase on accuracy by increasing the Ambisonics order with a proper number of loudspeakers [275].

It should be noted that the current study did not measure changes in ITD and ILD for off-center listener positions without the presence of a second listener. Based on the effects of having the first listener off-center with a second listener present, coupled with the smaller changes with a second listener when the first listener is centered, it can be deduced from the current results that the off-center position has a degradation effect on the ITD and ILD. Considering that many simulations are limited by a "sweet spot" for the listener(s), the off-center position, as opposed to the presence of a second listener, is probably the greatest liability for multi-listener methods in hearing research.

3.5 Concluding Remarks

The more demanding the test requirement in terms of localization of the sound source (out from left, right, front and back), the more the researcher should drive towards VBAP. In case of fixed positions and requirement of more sense of immersion, Ambisonics should be able to build more convincing sound scenarios.

The techniques do not affect the ILD and ITD acoustic cues in the central position for one test participant. The addition of a second listener within the ring also does not significantly affect these parameters at the three distances tested, except for the angles usually hidden by the shadow second listener. Thus, it is suitable to move towards subjective tests with a center participant and an actor on the side. Although the second listener has not deteriorated the techniques, they present different performances in terms of spatial representation and notably present a different sense of immersion. Thus, the test's purpose to be designed must be taken into account when defining the auralization method.

There is a clear degradation when two test subjects are simultaneously present both in off-center positions, regardless of the distance of a second listener. The VBAP measurements showed an increase in differences for ITD increasing the distance from the center and significant differences in ILD. These differences indicate the creation of acoustic artifacts, possibly generated by the method's difficulty in correctly virtualizing high frequencies outside the sweet spot. For the ITD parameter, the displaced position of 25 cm of the center has little difference or evidence of artifacts generated by virtualization errors. At the same time, the other distances present significant differences and artifacts. The binaural cues analysis suggests that VBAP is less sensitive to the participant positions than the Ambisonics setup.

However, it is relevant to note that although the differences in the binaural cues denote differences in audio spatialization, reflecting on the perceived angle of incidence of the sound, both techniques can be calibrated to reproduce the stimuli at a desired level of sound pressure. That means that an auralized sound can be reproduced with the correct sound pressure level although its direction may not be correctly interpreted by the listener as their binaural cues are not being delivered appropriately.

Chapter 4

Subjective Effort within Virtualized Sound Scenarios

This experiment was a collaborative study (EcoEG [3]) with fellow HEAR-ECO PhD student Tirdad Seifi-Ala, also from the University of Nottingham, that combined the virtualization of sound sources and electroencephalography (EEG) to assess listening effort in ecologically valid conditions. Both students contributed equally to the study design, preparation, data collection and interpretation. TSA additionally performed the data analysis; SA additionally performed the room simulations, stimuli preparation, software interface and sound calibration.

As definitions can vary, this chapter uses the following terms:

- Simulation: Numerical acoustic simulation of spatial behavior of a sound in a defined space.
- Auralization: Creation of a file that can be converted to a perceivable sound and contains spatial information.

- Sound Virtualization: Reproduction of an auralized sound file through loudspeakers or headphones.

4.1 Introduction

The interest from researchers and clinicians in the listening effort measures has grown recently [83, 135, 210], the importance of studying listening effort in an ecologically valid sound environment follows the same trend [134].

The previous chapter discussed the feasibility and constraints of the virtualized sound field through binaural cues and foreseeable effects on spatial impression and localization. This chapter investigates whether reverberation and the signal-to-noise ratio (SNR) are modeled in behavioral data, being a proxy of subjective listening effort in a virtualized sound environment.

The reverberation is the accumulation of energy reflections (sound) in an enclosed space that creates diffusion in its sound field [256]. Reverberation Time, in turn, is an objective parameter that represents the amount of time required to dissipate the energy of a sound source by one-millionth of its value (60 dB) after the sound source has ceased [254]. This parameter was reviewed in Section 2.3.3.2. The remaining sound energy can blur the auditory cues, rapid transitions between phonemes, and decrease the low-frequency modulation of a signal; it may compromise speech intelligibility [39, 112].

Since reverberation is a complex phenomenon, depending on space and frequency [111, 185], a wide range of physical-acoustical factors may limit some comparisons. For example, the reproduction method, the masker type, the position and number of sources, the SNR, the sound pressure level of the presentation, the reverberation time interval studied, and the simulated position being in a free or in a diffuse sound field. Like the methodologies, the findings in terms of reverberation influence on listening effort across experiments can also vary.

Previous studies investigated the effect of reverberation on speech intelligibility and listening effort. Variations across reverberation time, level, and population groups were observed. For example, a correlation between age and reverberation was traced in work by Neuman *et al.* [203]. This study found that reverberation negatively impacts the necessary SNR to reach 50% of speech recognition. This impact varies across ages, with the effect decreasing as age increases. The sensitivity of subjective measures and electrodermal activities were evaluated by Holube *et al.* [121]. The effect of reverberation was found statistically significant to subjective measures but not to the electrodermal activity. A study from Picou *et al.* [225] presents a response time in a dual-task paradigm as a behavioral measure of listening effort. In their study, there was no significant effect in response time neither in the same SNR conditions nor comparing the response time of equal performance scores. The impact on listening effort was studied by Kwak *et al.* [149] through subjective ratings resulting in a significant effect of reverberation on ratings of listening effort and the sentence recognition performance. In Nicola and Chiara's study [204], the negative influence of reverberation on response time was considered indicative of an increase in listening effort. The study assessed the influence of reverberation and noise fluctuation on response time. The different methodologies applied in the studies and their groups of participants must be carefully analyzed, as they can explain the different results.

Ambisonics arrangements (Mixed Order Ambisonics (MOA) [78, 177] and HOA) are already used in audiological studies [7, 77, 173, 303]. This study proposed a low-order (first-order) Ambisonics implementation. The low-order technique is more sensitive to the listener position [64, 65], which was also verified in this study. That can be seen as a counter-intuitive and non-conventional choice, although it was meant to assess low-order Ambisonics' feasibility in audiological studies and its constraints. This decision was a step towards confirming the feasibility of a listener in a centralized position found in Chapter 3, observing its constrains, and further developing an auralization method with lower hardware requirements in Chapter 5.

Hypothesis The main research question is how the auralized acoustic scenario, specifically the room and the SNR, increases auditory effort when virtualized. The hypothesis for the experiment is that a longer RT provided through sound virtualization and a lower SNR both lead to a more significant listening effort. Reverberation time can influence normal hearing and hearing-impaired people in different ways. For example, on average, hearing-impaired listeners experience more significant difficulties with understanding speech in a reverberant condition than normal hearing listeners, so they can suffer more from the strain of listening. As reverberation's effects on hearing-impaired listeners vary (see Chapter 2), this study employed only normal-hearing participants to investigate the effects of audio degradation. To subjectively assess changes in hearing effort, a questionnaire was provided to participants, asking how much effort they found for each condition (described in Section 4.2). This investigation is the first step to understanding the feasibility of including the simplified virtualization of sound sources in the expanding field of listening effort research.

4.2 Methods

This experiment was designed to gather data for two parallel analyses: the first was to evaluate differences in behavioral performance (speech recognition) and subjective impressions of listening effort driven by different scenarios, manipulating the room type and the signal-to-noise ratio (SNR). The second study compared physiological responses of the brain as measures of listening effort to the same behavioral performance. This chapter focuses on the experiment's first study (behavioral data *vs.* subjective impressions). Three rooms were chosen for this study: a classroom, a restaurant dining area, and an anechoic room.

For this experiment, a setup was developed to investigate the influence of listening effort caused in nine different situations: three room simulations characterized by their reverberation time and three SNRs. The setup was composed of four recorded talkers acting as maskers and one talker acting as the target. The talkers' positions were all spatially separated.

The test paradigm involved the auditory presentation of Danish hearing in noise test (HINT) sentences [205] on top of four speech maskers and recalling the words they could keep in memory after 2 seconds. The sound sources are spatially distributed and the participant is informed that the target speech is always from the front. The participants responses were word scored (*i.e.*, word-based speech intelligibility) by Danish-speaking clinicians.

The method in this study follows a similar setup with a four-talker babble setup as in [209, 302], which investigated SNR and masker types using pupilometry as a proxy for listening effort. Also, a study from Wendt *et al.*, [301] investigated the impact of noise and noise reduction through an equivalent setup. This method's innovation relies on using first-order Ambisonics to generate the reverberation based on Odeon simulated rooms.

4.2.1 Participants

For the data collection, 18 normal-hearing native Danish-speaking adults (eight females) with an average age of 36.9 ± 11.2 years first gave written consent form and initially participated in the test. One participant was placed outside the sound field sweet spot, so his data were discarded, and the data for the other 17 participants were used for further analysis. Ethical approval for the study was obtained from the Research Ethics Committees of the Capital Region of Denmark. For each participant, the pure-tone average of air conduction thresholds at 0.5, 1, 2 and 4 kHz pure tone audiometry (PTA4) were tested and confirmed below 25 dB HL.

4.2.2 Stimuli

The target stimulus consisted of simple Danish sentences spoken by a male speaker. The sentences were from the HINT in Danish [205] and were 1.3-1.8 s in duration. The masking signal consisted of four different speakers, two female and two male, reading a Danish-language newspaper [302]. The total duration of each of the masker recordings was approximately 90 seconds. The maskers' onset was 3 s before and offset was 2 s after the target, resulting in a masker duration of 6.3-6.8 s. In each trial, the time segment used of each masker was randomized. In addition, the spatial position for each masker was also randomized in each trial, but always interspersing male and female talkers. The overall maskers' equivalent continuous sound level L_{eq} was set at 70 dB (64 dB each masker), and the target L_{eq} were set at 62 dB, 67 dB and 72 dB to generate three different SNR conditions: -8, -3 and +2 dB. In this study, SNR was defined as the equivalent continuous sound level of the target signal compared to the competing masking L_{eq}. The chosen reverberation conditions aimed to represent common everyday situations. The RT of the anechoic and reverberant conditions studied were defined as the overall reverberation time obtained through the output of the simulation software (ODEON Software$^{©}$ v.12). The chosen reverberation time values aim to represent common everyday situations. The absorption coefficients and relative area utilized to obtain the mentioned conditions are presented on Appendix E.

Five source positions (one target and four maskers) were created around a receptor in each simulated room. All positions were 1.35 m from the center of each room where the receptor is located. The approach of creating two different rooms instead of changing the parameters of a single room was chosen to achieve a more natural sound field. That way the absorption coefficient applied to the room's materials was kept close to real.

The virtualization of the proposed acoustic scenarios follows the path indicated in Figure 4.1. An acoustic simulation is performed to create the appropriate characteristics of the sound according to the room. The software calculates the amplitude and the incidence directions of sound and its reflections arriving from specific sources to a receptor position inside the room. For each combination of source-receptor, the software generates a room impulse response that is encoded in Ambisonics first-order in AmbiX [198] format (which is a channel order specification for Ambisonics auralization first 4 channels are WYZX compared to WXYZ in the FuMa specification). The generated file was convolved with anechoic audio and decoded to the specific array of 24 loudspeakers.

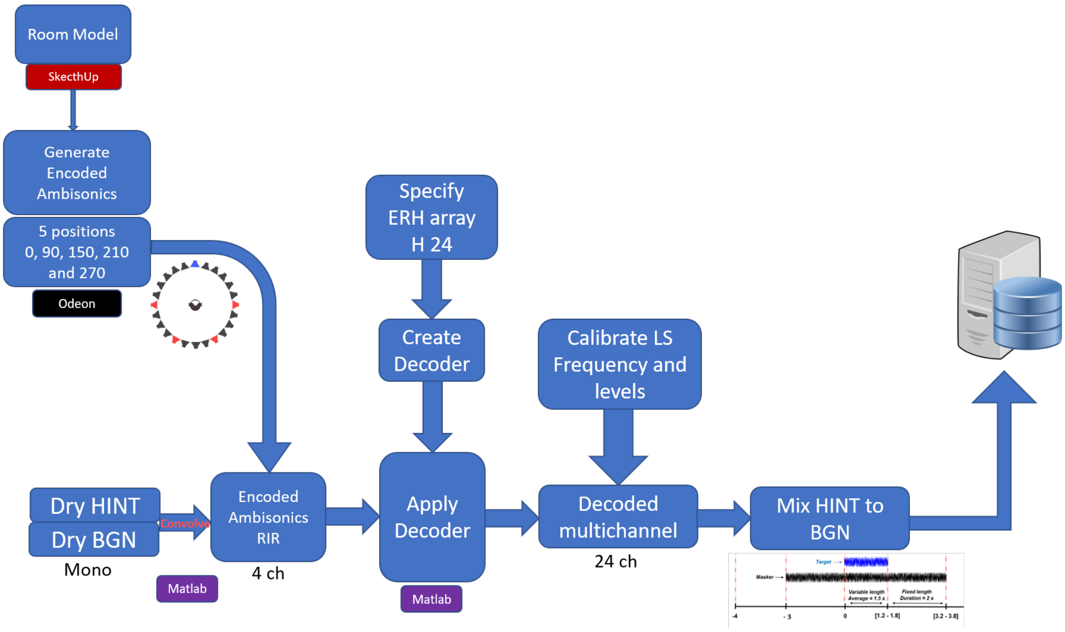


Figure 4.1: *Auralization procedure implemented to create mixed audible HINT sentences with 4 spatially separated talkers at the sides and back (maskers) and one target in front.*

4.2.3 Apparatus

The experiment was set up in an anechoic room (IAC Acoustics) with 4.3 m × 3.4 m × 2.7 m (inner dimensions). The experimental setup consisted of a circular array of 24 loudspeakers positioned on 15° interval on the azimuth and 1.35 meters distance from the center. The target sound was reproduced at 0° (participant's front), the maskers were auralized at ± 90° and ± 150° (Figure 4.2). The position of the participant during all the test was monitored through a laser line and a camera ensuring they remained in the sweet spot. Stimuli were routed through a sound card (MOTU PCIe-424) with Firewire 440 connection to the MOTU Audio 24 I/O interface) and were played via 16 loudspeakers Genelec 8030A and 8 loudspeakers Genelec 8030C (Genelec Oy, Iisalmi, Finland) aligned in frequency and level. The Biosemi EEG device was used to collect the physiological data, which helped to restrain participants' movement; the EEG data were not analyzed in this study.

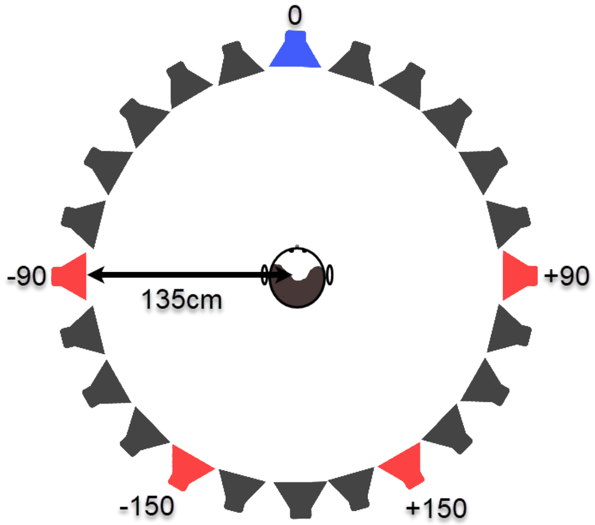


Figure 4.2: *Spatial setup of the experiment: Test subjects attended to target (in blue) stimuli from a 0° angle in front.The masking talkers (in red) are presented at lateral ±90 and rear ±150 positions.*

All enclosed spaces have a certain degree of reverberation due to acoustically reflective surfaces and background noise due to equipment, including controlled audiological environments. The levels of reverberation and background noise meet the criteria from Recommendation ITU-R BS.1116-3 [126] and are respectively shown in Figures 4.4 and 4.3.

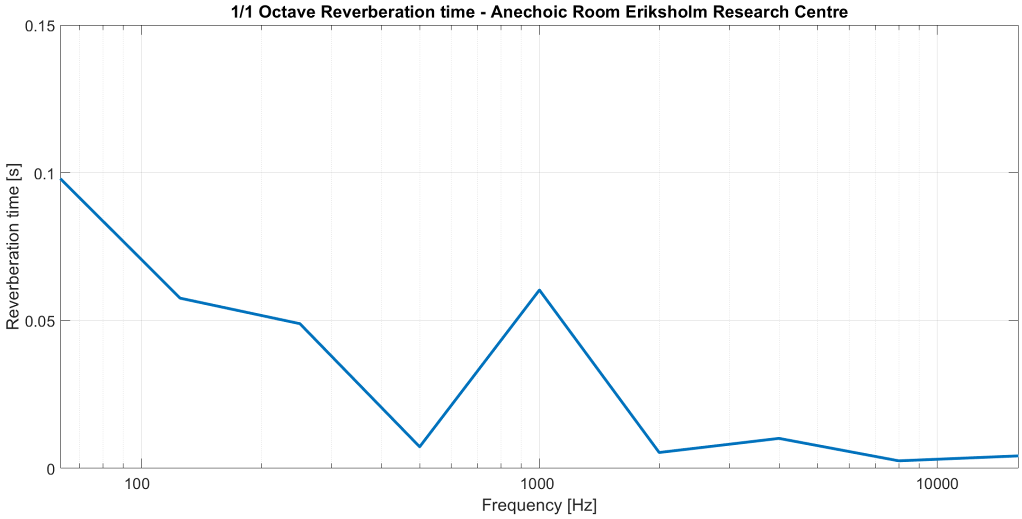


Figure 4.3: *Reverberation Time inside anechoic room at Eriksholm Research Centre with setup in place.*

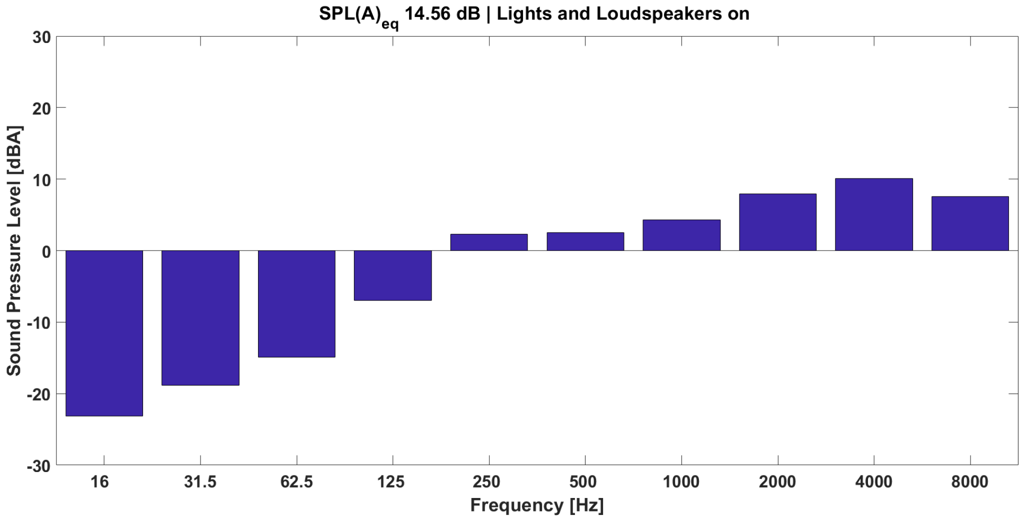


Figure 4.4: *Eriksholm Anechoic Room: Background noise A-weighted. Loudspeakers and lights on, motorized chair off.*

The parameters were measured with the setup (loudspeakers, motorized chair and BioSemi eeg equipment) inside the room and positioned as in the experiment. Figure 4.5 shows the setup placed inside the anechoic room.

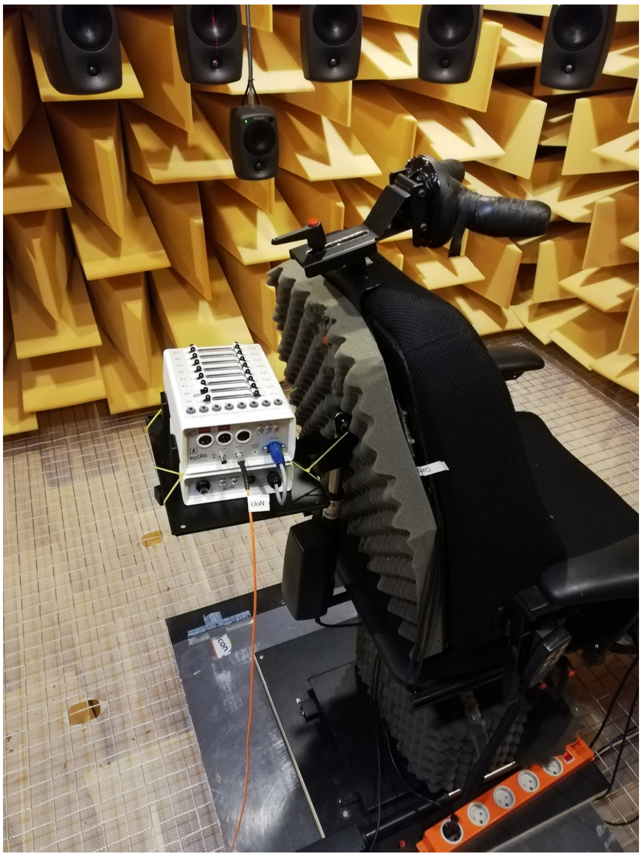


Figure 4.5: *Setup inside anechoic room (Motorized chair, adjustable neck support and EEG equipment).*

4.2.4 Auralization

Acoustic Scene Generation and Room Acoustic Simulation

To simulate the acoustics characteristics of the chosen scenarios, geometric models were created in the room acoustics software ODEON. Next, the Ambisonics Room Impulse Responses were simulated using ODEON software, version 12 [59]. The absorption coefficients of the room surfaces are listed in Annex E. All sentences were auralized in Ambisonics [15], truncated by 1st order and encoded to 24 channels. The analysis utilized the Institute of Technical Acoustics (ITA)Toolbox [29, 67]. Rooms were chosen as representative of realistic and not extreme acoustic conditions. The spaces simulated were a classroom (9.46 m × 6.69 m × 3.00 m) with an overall RT of 0.5 seconds, and a restaurant's dining area (12.19 m × 7.71 m × 2.80 m) with an overall RT of 1.1 seconds. The distance between source and receptor was kept the same, 1.35 m, across rooms. Target and masker positions were simulated by selecting the appropriate simulated RIR to convolve *i.e.,* the simulated source-receptor RIR that corresponds to the desired reproduction angle.

Ambisonics Sweet Spot In this study, two different metrics were used to compare the off-center performance of virtual sources auralized with first-order Ambisonics: the RT and the Sound Pressure Level (SPL). That is, the presented virtualized soundfield was delivering the correct amount of reverberation and also the correct sound pressure level of each source resulting in the appropriate signal-to-noise ratio when was not perfectly centered. To estimate each position's metrics, a logarithimc sweep signal (50-20000 Hz, 2.73 s (FFT Degree 18, Sample Frequency 96 kHz)) was generated and convolved with the Ambisonics first-order RIR calculated by ray-tracing in ODEON for each modeled room. The simulated rooms presented an overall theoretical reverberation time of 0, 0.5, and 1.1 s. These auralized files were encoded to 24 channels distributed in the horizontal axis. In the following, the files were played inside the anechoic room and simultaneously recorded. From the division in the frequency domain of the recorded signal and the zero-padded initial signal (deconvolution), the calculated impulse response (or binaural RIR (BRIR) when recorded with HATS) represents the virtualized system, including the physical effects of the array and all calibration.

Reverberation Time The RT was calculated with ITA-Toolbox from initial 20-dB decrease from peak level (T_{20}) in the virtualized IRs. Figure 4.6 shows the overall RT results at the center position and by moving the receptor (manikin) towards the front.

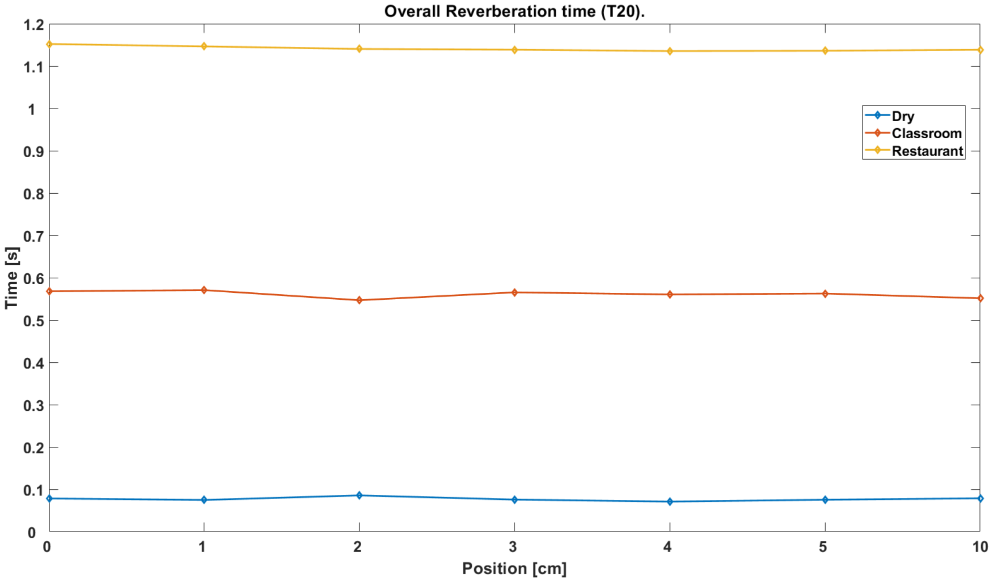


Figure 4.6: *Overall reverberation time (RT) as a function of receptor (head) position in the mid-saggital plane re center (0 cm)*

The results showed slightly greater RTs (0.58 and 1.16 s) than what was simulated in the ODEON software (0.5 and 1.1 s). However, this was expected since there is equipment inside the anechoic room (e.g., a large chair and loudspeakers) that can be considered reflective surfaces that were not present in the simulation. The results showed that there is no major effect on the energy decay for small head movements.

Sound Pressure level The sound pressure level was determined by convolving the target and masker sounds with the impulse responses collected across twelve positions with horizontal displacements of 2.5, 5 and 10 cm and forward (mid-saggital) displacements of 2.5 and 5 cm. The results are shown in Figure 4.7. Four speech talkers are individually convoluted. The equivalent sound pressure level is determined using the calibration factor. The measure is the average of 20 different sentences.

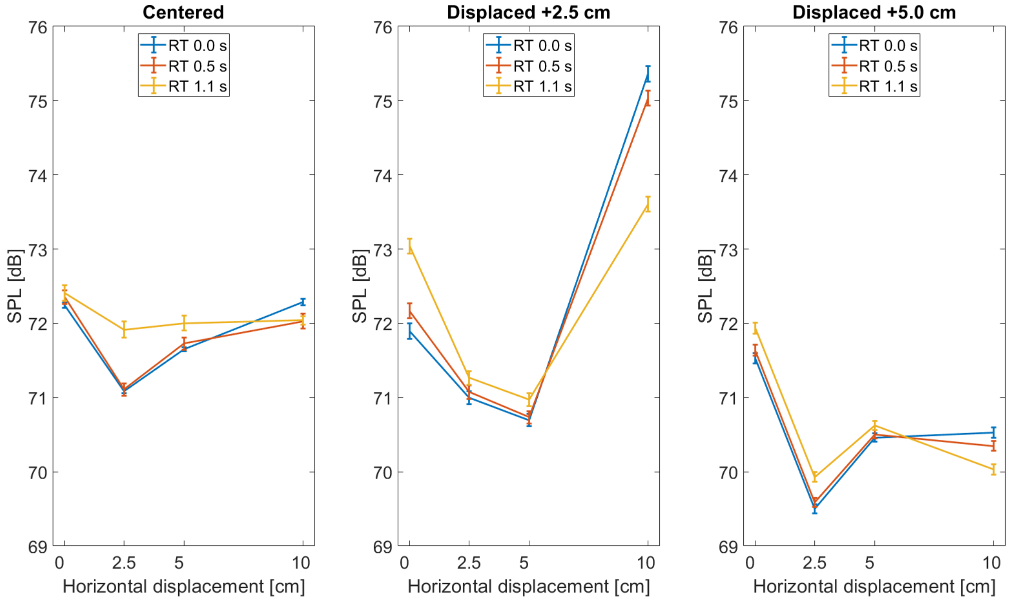


Figure 4.7: *Sound pressure level virtualized through Ambisonics at different listener positions.*

The changes in SPL as a function of off-centre position do not follow a consistent pattern. The SPL changes were, however, mostly similar across the three simulated rooms, with the exception of three positions where the restaurant (1.1 s RT) was 1-1.5 dB different (x = 2.5, y = 0; x = 0, y = 2.5; x = 10; y = 2.5). The center position is the optimal position for sound pressure level accuracy. To help get reliable, appropriate data from the experiment, a neck rest as well as a video feed and laser line were added to the setup after the first pilot test. The participants were asked to be in contact with the neck rest all the time. The clinician was able to see the laser line at the patient's head throughout the test. They could ask the participant to quickly correct posture at the start of each block or at any point of the session after the participant needed a break. Figure 4.8 shows a participant positioned with all sensors connected. Another important find was, after adjusting the participant position, the motorized chair should be unplugged, otherwise the EEG data would be compromised.

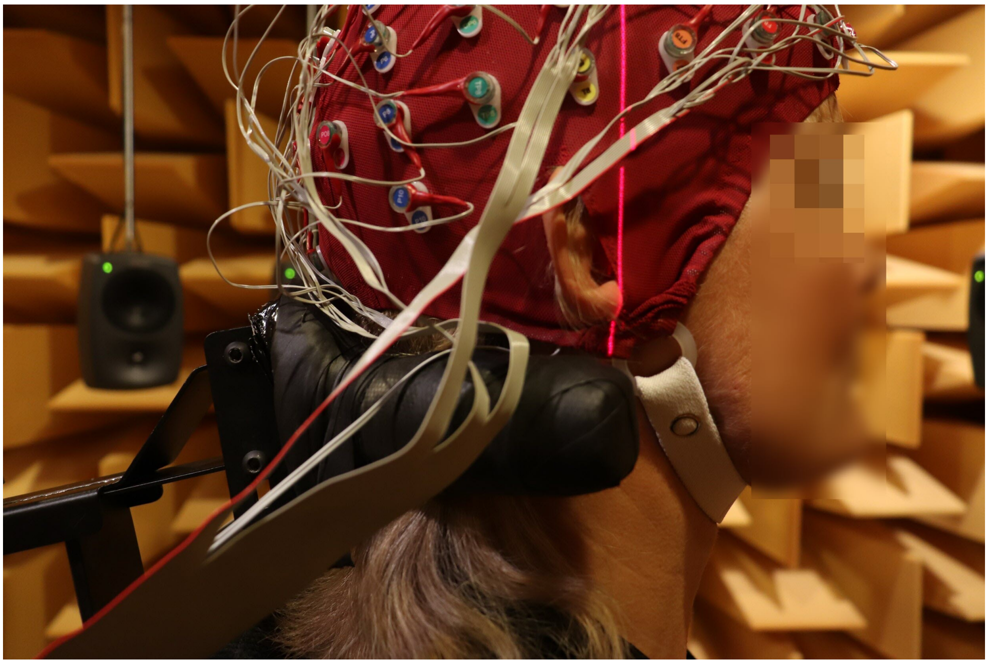


Figure 4.8: *Participant positioned to the test.*

4.2.5 Procedure

There were 9 different conditions based on SNR (+2 dB, -3 dB, -8 dB) and reverberation time (0 s, 0.5 s, 1.1 s) of the sound. Each condition was presented in separate blocks, and each block consists of 20 sentences, so in total there were 9 blocks and 180 sentences presented to the participants in the main test. In addition to that, each participant went through a training round in the beginning, consisting of 20 sentences with different conditions. The procedure for each trial is illustrated in Figure 4.9. Each trial started with 2 s of silence (preparation), then 3 s of background noise which served primarily as a baseline period for the separate EEG analysis. Then a HINT sentence was played as the background noise continued for 1.5 s on average. After the target sentence finished, the background noise continued for another 2 seconds during which participants needed to maintain the words they just listened to (maintenance), also serving primarily for the companion analysis of EEG responses re baseline. When the background noise was stopped, the participants were instructed to repeat all the words within the sentence (recall). The listening effort reflected in alpha power changes in the maintenance phase have been investigated by [208, 310, 311, 313]

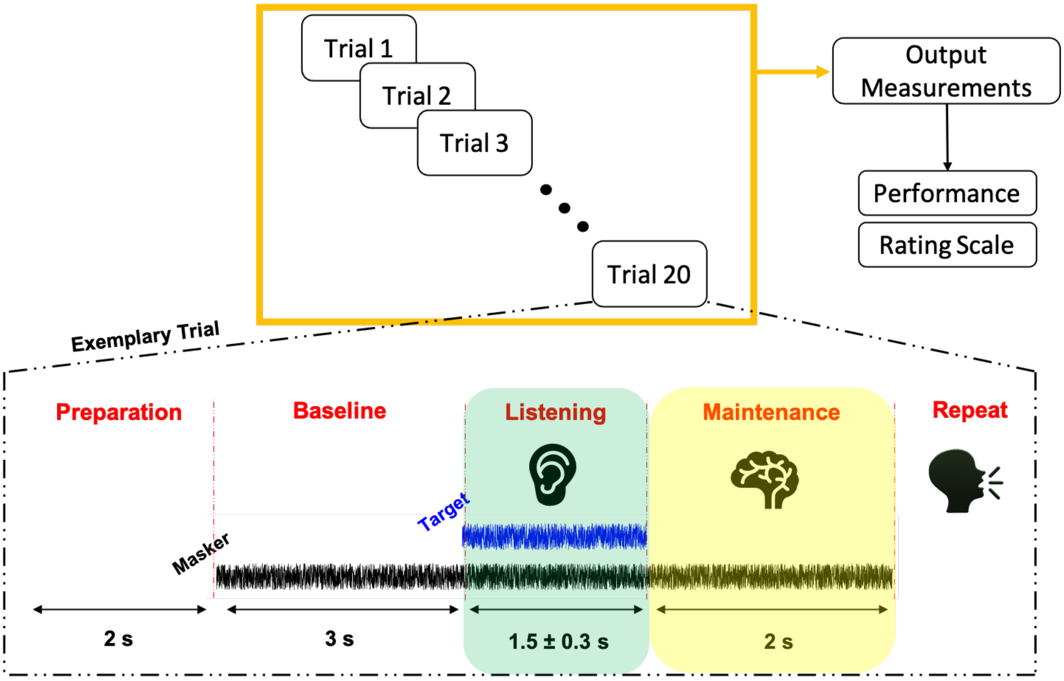
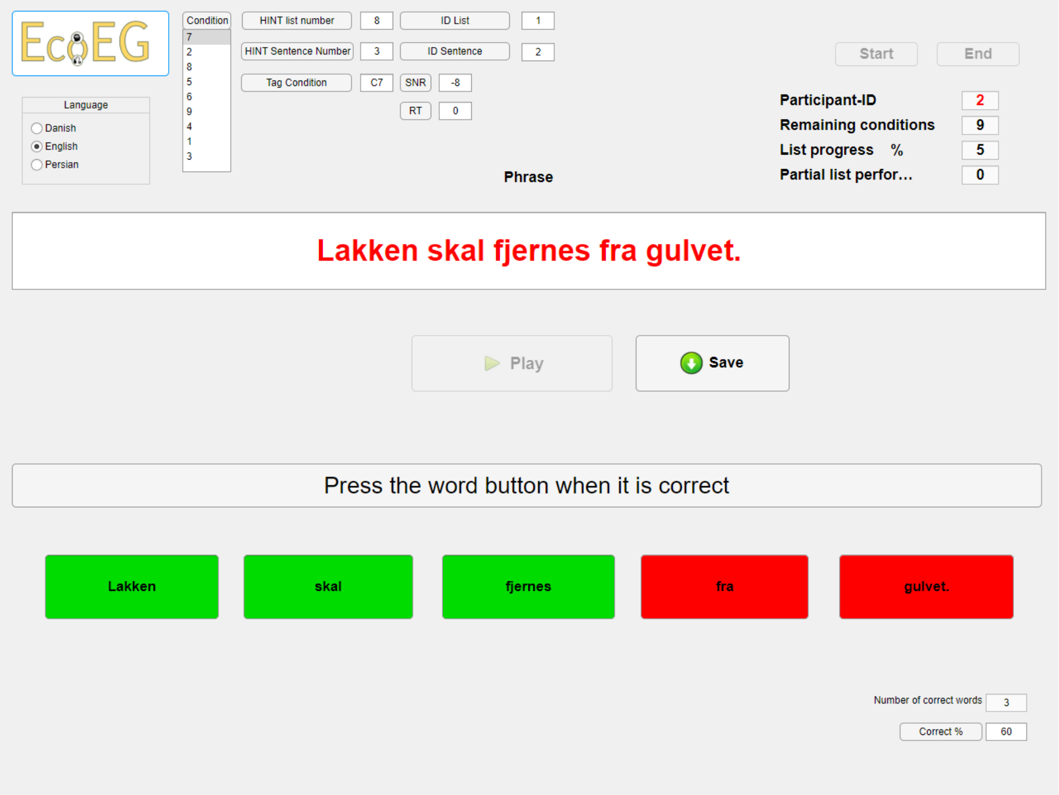


Figure 4.9: *Trial design. For each trial, 20 in each block, there was 2 s of silence, then 3 s of masker (4 spatially separated talkers), then a Danish HINT sentence as target stimuli in the presence of continuing masker, then 2 additional s of masker, followed by silence when the participant repeated as many target words as they could understand and keep in memory.*

Figure 4.10 shows the user graphical user interface designed and implemented for this experiment. The 24-channel audio files were produced beforehand (off-line), being calibrated to the specific setup. Along with audio presentation, the software also sent a series of triggers in synch with presentation timings to the EEG software (Actiview, BioSemi) to mark the EEG measurement appropriately for the companion analysis.



Language
○ Danish
● English
○ Persian

Condition: 7, 2, 8, 5, 6, 9, 4, 1, 3

HINT list number: 8
ID List: 1
HINT Sentence Number: 3
ID Sentence: 2
Tag Condition: C7
SNR: -8
RT: 0

Participant-ID: 2
Remaining conditions: 9
List progress %: 5
Partial list perfor...: 0

Phrase

Lakken skal fjernes fra gulvet.

Press the word button when it is correct

Number of correct words: 3
Correct %: 60

Figure 4.10: *Graphic User Interface used to acquire the data from participants. Words are state buttons that alternates between green and red being saved as 1 or 0 respectively.*

4.2.6 Questionnaire

At the end of each block (SNR × room condition) a three-item questionnaire was presented to the participants; the English translation is shown in Table 4.1. The questionnaire was translated from Zekveld and Kramer [318] to Danish. The response to each question had a scale of 0 to 100 in integer units Appendix F. The first question was aimed to measure participants' estimation of their performance, referred to as "Subjective intelligibility" for the rest of the text. The second question was to measure participants' perception of effort, referred to as "Subjective effort". The third question provided to measure how often participants gave up during the test, referred to as "Subjective disengagement".

Table 4.1: The questionnaire for subjective ratings of performance, effort and engagement (English translation from Danish)

Question 1	How many words do you think that you understood correctly?
Question 2	How much effort did you spend when listening to the sentences?
Question 3	How often did you give up trying to perceive the sentences?

4.2.7 Statistics

A linear mixed model [171, 233] (LMM) was used to investigate SNR and RT effects on performance and questionnaire. The effects on different alpha bands through EEG power by SNR and RT were also explored through LMM in the collaborative analysis performed by Seifi Ala. SNR and RT were fixed factors, while participants were random factors in the model. Implemented in MATLAB, the syntax for LMM was Dependent ∼ 1+SNR*RT+(1—Subject ID), with Dependent being either performance or questionnaire. Both the SNR (-5, 0, 5) and RT (-0.53, -0.03, 0.56) levels were re-centered around zero for the model.

4.3 Results

This section highlights the findings concerning the study's questions, the feasibility of having a hearing in noise test virtualized in first-order Ambisonics, and the influence of degradation through SNR and Reverberation in the Speech Intelligibility.

The participant's behavioral performance (*i.e.*, speech recognition accuracy) demonstrated significant effects of SNR ($\beta = 5.98, SE = 0.30, t_{158} = 19.67, p < 0.001$), and RT ($\beta = -31.17, = 1.78, t_{158} = -17.49, p < 0.001$) and a significant interaction between the two ($\beta = 1.76, SE = 0.43, t_{158} = 4.04, p < 0.001$). Figure 4.11 presents the mean performance (percent correctly recalled words) as a function of SNR for each room. Less signal degradation, whether higher

SNR or lower RT led to higher performance accuracy.

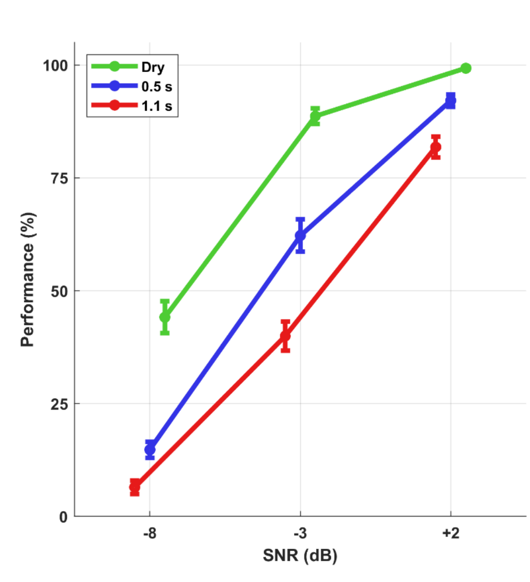


Figure 4.11: *Performance accuracy based on percentage of correctly recalled words as a function of SNR and RT (line color/shading). Error bars represent the standard error of the mean. Lines/symbols are staggered for legibility and do not indicate variation in SNR.*

The statistical analysis of the results for subjective intelligibility (Figure 4.12), subjective effort (Figure 4.13), and subjective disengagement (Figure 4.14) are shown in Table 4.2. All the measures show a significant interaction between SNR and RT. Lower signal degradation (higher SNR and lower RT) led to higher subjective estimation of intelligibility performance accuracy, decreased reported effort and disengagement.

Table 4.2: Results of linear mixed model based on SNR and RT predictors estimates of the questionnaire.

DF = 158	Self-report scales		
Question Predictor	**Subjective intelligibility**	**Subjective effort**	**Subjective disengagement**
SNR	β = 5.71 SE = 0.42 t = 13.48 p <0.001	β = -5.60 SE = 0.41 t = -13.57 p <0.001	β = -5.78 SE = 0.48 t = -11.85 p <0.001
RT	β = -33.74 SE = 2.47 t = -13.61 p <0.001	β = 23.58 SE = 2.41 t = 9.76 p <0.001	β = 33.39 SE = 2.85 t = 11.68 p <0.001
SNR x RT	β = 1.56 SE = 0.60 t = 2.57 p = 0.010	β = 1.50 SE = 0.59 t = 2.54 p = 0.012	β = -2.06 SE = 0.69 t = -2.94 p = 0.003

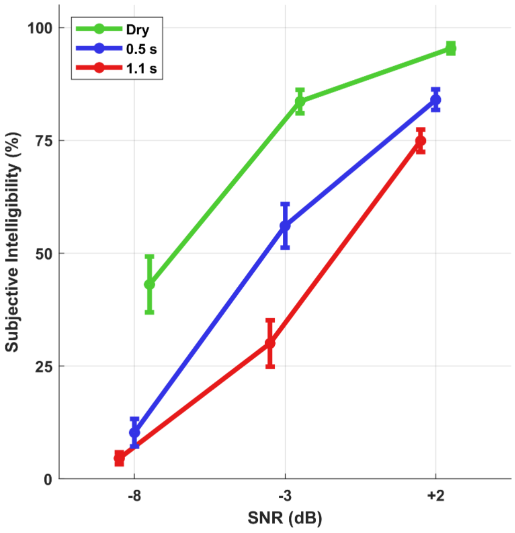


Figure 4.12: *Subjective intelligibility as a function of SNR and RT (line color/shading). Error bars represent the standard error of the mean. Lines/symbols are staggered for legibility and do not indicate variation in SNR.*

The subjective impression of how much effort was required and how willing they were to give up in each situation are presented in Figures 4.13 and 4.14, respectively.

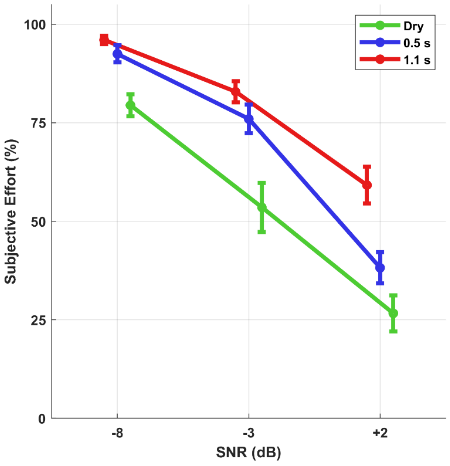


Figure 4.13: *Subjective effort as a function of SNR and RT (line color/shading). Error bars represent the standard error of the mean. Lines/symbols are staggered for legibility and do not indicate variation in SNR.*

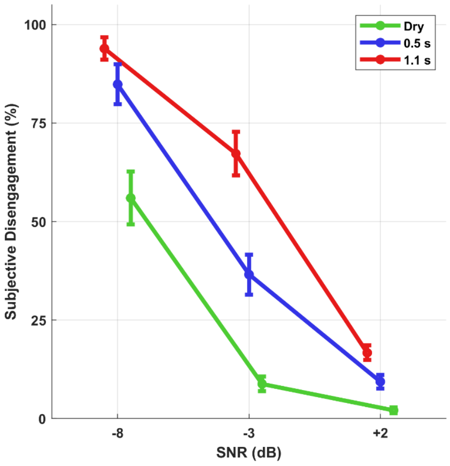


Figure 4.14: *Subjective disengagement as a function of SNR and RT (line color/shading). Error bars represent the standard error of the mean. Lines/symbols are staggered for legibility and do not indicate variation in SNR.*

The results show the statistically significant contributions of reverberation and SNR to perceived performance, effort and disengagement. From Figures 4.12, 4.13, and 4.14, the self-report scales varied near-linearly with the signal degradations across conditions, agreeing generally with the behavioral data (See Figure 4.11).

The subjective effort is related to the inverse of the reverberation time; the more time the energy needs to dissipate in the environment, the greater the perceived effort. The results from all the self-report scale questions were highly correlated to performance. Pearson skipped correlations [308] revealed a significant ρ coefficient (See Table 4.3):

Table 4.3: Pearson skipped correlations between performance and self-reported questions.

	performance vs subjective intelligibility	performance vs subjective effort	performance vs subjective disengagement
r	0.95	-0.79	-0.94
CI	0.93, 0.96	-0.84, -0.74	-0.96, -0.92

4.4 Discussion

This study presented an interesting challenge to the researchers. The pilot data pointed to the direction of the virtualization not rendering the correct sound, especially not the correct sound pressure level. The setup was retested and investigated in different positions, and the problem was identified. The first-order Ambisonics rendering has a relatively small sweet spot. Thus participants were monitored to be in the correct position during the testing. The sweet spot capabilities in terms of the correct overall SPL reproduction presented limitations of plus-minus 1 dB up to 5 cm and plus-minus 3 dB to the target SPL up to 10 cm off center. Although not testing the exact Ambisonics implementation through a different performance measure, the findings agree with literature observing contrasts caused by the reproduction method in similar distances out of the center. As a reference, Grimm *et al.* [97] analyzed simulated Ambisonics environments with different numbers of loudspeakers, studying its influence on a representative hearing aid algorithm. It showed a decrease in SNR errors when increasing loudspeakers and decreasing frequency. A bandwidth of 2 kHz in the central listening position, 12 loudspeakers would be required for HOA. If 24 loudspeakers are available, the bandwidth in the central listening position would be 6 kHz. Laurent *et al.* [276] analyzed the reconstruction error to assess the rendering system's frequency capabilities. A KEMAR was fitted with a hearing aid, without processing, to collect the impulse responses. Regarding range, a third-order implementation with 29 loudspeakers decreased from 3,150 Hz in the center to 2,500 Hz when positioned 10 cm from the center.

Tests that involve separated sound sources and are auralized and virtualized by loudspeaker setups need to be verified in terms of sweet spot size to the specific sound parameters (e.g., RT and SPL). An off-centered or moving head can, in an Ambisonics 1st order auralization, easily encounter a spot in space where, for example, the wave field combination may partially cancels one or more maskers increasing the SNR, even if the intended SNR is low (See Figure 4.7). In other off-center spot it could also be possible to partially cancel the target. These distortions could profoundly impact the results, and not represent what would be achieved in the real scenario that was being simulated.

For normal hearing participants, a more psychologically oriented psychoacoustic auralization method such as lower order Ambisonics can provide the desired acoustic impression insofar as objective and subjective performance when the calibration is performed, and the setup limitations (e.g., very restricted sweet spot) are respected. An investigation of performance in off-center positions using hearing impaired participants would be an important next step towards understanding a broad clinical application of this method.

Participants were tested in three different SNRs (-8, -3, +2 dB) and three virtual rooms (with RTs of 0, 0.5, and 1.1 s). The more the manipulated signal was degraded (lower SNR and higher RT), the more demanding the listening conditions became, which led to lower the participant's speech intelligibility. A questionnaire was used as a subjective measure of effort. Comprehensively, participants reported increased speech intelligibility, less cognitive effort, and less tendency to disengagement when diminish the signal degradation. That denotes that if they could recall the speech well, they perceived that they performed well and also spent less effort. The results from all three questions within the questionnaire were strongly correlated (either positively or negatively) to the speech intelligibility of the participants. They significantly changed with both SNR and RT and the interaction between them. When asked about subjective impressions of each block, the participants demonstrated to have perceived the proposed signal degradation both in SNR and RT. That is in line with the studies from Zekveld *et al.* [319], Holube *et al.* [121], Neuman *et al.* [203], Kwak *et al.* [149], Nicola & Chiara's study [204] and Picou & Ricketts [229]. Furthermore, studies that cross objective measurements of physiological parameters in the literature associated with changes in effort can have divergent outcomes, as discussed in Chapter 2. From that discussion, it is speculated that these different methods, proposed to achieve a proxy to listening effort, are sensitive separated aspects of a complex global process [12, 224]. Another explanation would be the minimization of effort utilized by the participant through the heuristic strategies in the subjective method [192], and lastly, the effect of working memory being related differently to different methods [53, 186]. A separated study by Tirdad Seifi-Ala from this combined experiment examined the correlation between objective (physiological responses of the brain) and subjective paradigms.

4.5 Concluding Remarks

In this study, nine levels of degradation were imposed on speech signals over speech maskers separated in space and virtualized. Three different SNRs (-8, -3, +2 dB) and three different simulated rooms (with RTs of 0, 0.5, 1.1 s) were used to manipulate task demand. The speech intelligibility was assessed through a word-scored speech-in-noise test performed in a 24 loudspeaker setup utilizing Ambisonics first-order. The results showed a high correlation between participants' performance and responses to questions about subjective intelligibility, effort and disengagement. The main effects and interaction of SNR and RT were demonstrated on all questions. Furthermore, it was observed that the reverberation time inside a room impacts both speech intelligibility and listening effort. This study demonstrated the possibility of virtualizing a combination of sound sources in low order Ambisonics and extracting quality behavioral data.

Chapter 5

Iceberg: A Hybrid Auralization Method Focused on Compact Setups.

5.1 Introduction

People usually wear their hearing devices in spaces very different from the laboratories' soundproof booths in everyday life. Additionally, the everyday sounds are more complex and different from the pure tones, words, and phrases without context utilized in many hearing tests. Therefore, hearing research has increasingly aimed to include acoustic verisimilitude on auditory tests to make them more realistic and/or ecologically valid [61, 79, 101, 177, 212, 217]. Thus, they can evaluate new features and algorithms implemented on hearing devices and experiment with different fittings and treatments while maintaining their repeatability and control.

One can utilize a particular auralization technique to create reproducible sound files in a listening area. These sounds attempt to mimic the acoustical characteristics of environments (from actual recordings or acoustic simulations). It can then be played through a set of loudspeakers or pair of headphones, creating both the subjective impression and objective representation of listening to the intended sound environment [293].

Through an auralization method, it is possible to create a sound file containing spatial information about the scene and a series of details about the configuration of the reproduction system [293]. The reproduction system includes, for example, the number of loudspeakers and their physical position, the number of audio channels available, and the distance from the loudspeakers to the listening position. The size of the effective listening reproduction area - where the auditory spatial cues of the scene are most accurate - is usually called the ”sweet spot” [253]. The spatialization accuracy is differently affected by different systems as well as auralization methods [65, 97, 166, 275, 276].

The auralization method can be decisive in the reproduction system choice; for example, certain methods require certain numbers of loudspeakers [62, 217]. Consequently, the auralization method can be a limiting factor depending on the tests or experiments. A dedicated setup capable of handling different auralization methods with a large listening area [188] may require an excessive amount of funding and physical space. These requirements can be a limiting factor to conducting research and developing innovative treatments.

This chapter proposes a compact setup with a hybrid auralization method. It is characterized in some conditions (RTs, presence of a second listener, and listener position) by considering the intended use in auditory evaluations as in the previous chapter. The setup aims to reproduce sound scenes maintaining spatial localization and creating an immersive sound environment from either a scenario in an actual room or virtual rooms created in acoustic software.

5.2 Iceberg an Hybrid Auralization Method

The Iceberg auralization method combines two well-known methods: VBAP and Ambisonics. In Chapter 3, VBAP and Ambisonics binaural cues were objectively evaluated. The VBAP method was found to render accurate cues in the center position, even with a second listener inside the array. That corroborates the use of VBAP to increase tests' ecological validity in auditory tests [134]. On the other hand, Ambisonics delivered less precise localization cues, imposing more restrictions on the listener's position. The results are in line with literature presenting poor localization but high immersiveness from low-order Ambisonics [104, 105] and, conversely, lesser immersiveness and greater localization accuracy from VBAP [89, 104]. Therefore, the idea here is to provide an auralization that contains temporal and spectral features of the sounds encoded through VBAP while the spaciousness provided through the reverberation envelope is encoded through Ambisonics.

This specific combination of auralization methods has also been considered to decrease the number of necessary loudspeakers for a setup that requires regular hearing devices. At the same time, the setup may allow some degree of head movement without the need for tracking equipment. That is a countermeasure to overcome common limitations in ordinary auditory test spaces [316].

5.2.1 Motivation

The primary motivation for creating this auralization method was to test hearing aid users in typical situations while wearing hearing devices in a small setup. Therefore, the method is loudspeaker-based, but at the same time, the number of loudspeakers and the system complexity were also constraints. The theoretical support for combining these auralization methods and proposing the smaller virtualization setup is gathered from room acoustic parameters and psychoacoustics principles presented in the review and during this chapter. These parameters and principles led to a system able to use RIRs from simulated environments (spaces that may only exist in a computer) and recorded RIRs from real ones. The initial Iceberg focus is on tests that manipulate sound scenarios to evaluate speech intelligibility masked by noise from static positions as tested with low-order Ambisonics in Chapter 4.

5.2.2 Method

The Iceberg method is a relatively easy-to-use algorithm that can be introduced to test environments with a simple calibration process. The virtualization system presented auralized files in a quadraphonic array with loudspeakers positioned at 0, 90, 180 & 270 ° (see Figure 5.1). Other horizontal setup arrangements can be implemented depending on the need, considering the system's angle rotation, frequency response, and the potential variation in localization accuracy. Although there is a minimum number of necessary loudspeakers (four), the method can be used to auralize files to setups with an extended loudspeaker number. The presented algorithm was implemented in MATLAB (Mathworks).

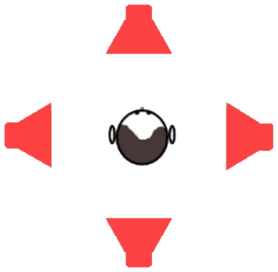


Figure 5.1: *Top view. Loudspeakers position on horizontal plane to virtualization with proposed Iceberg method.*

The proposed loudspeaker setup had a radius of 1.35 m. Other distances need to be evaluated regarding the system frequency response.

The proposed Iceberg's implementation derives an appropriate multi-channel audio signal with specific information from a sound and its reflections (incidence angle, sound energy, spatial and temporal distribution). These parameters can be encoded into a sound file with the reproduction setup's specific calibration values and positioning orientation. Finally, the auralized file can be reproduced (virtualized) as spatial sound.

5.2.2.1 Components

The Iceberg method proposed is a hybrid auralization method, a combination of VBAP and first-order Ambisonics; section 2.3.2.2 reviews the derivation of both methods. Both techniques are based on the panorama of amplitude. The main difference is in the mathematical formulation of the gains applied to the amplitude of each sound source. VBAP treats the reproduced sound as a unitary vector in a two- or three-dimensional plane (Equations 2.4 and 2.7, respectively). The weights applied to the amplitude of the signal at each loudspeaker are derived from the tangent law. It is traced as a vector from the nearest available sources between the listening position and the desired source position (Equation 2.3). On the other hand, Ambisonics utilizes all loudspeakers available to compose the sound field. The method combines the amplitude of the sources, calculating their weights according to the sum of spherical harmonics (Equation 2.9) that represents the pressure field formed by the sound wave (Equation 2.8). While VBAP concentrates the energy between two loudspeakers in its 2D implementation, Ambisonics spreads it through all available loudspeakers. That leads to a more immersive experience on Ambisonics, while the VBAP can better represent the sound source direction.

5.2.2.2 Energy Balance

The energy balance between the methods is calculated based on the Ambisonics first-order impulse response (See example in Figure 5.2); on the left is the impulse response (or decay curve) and is not the decay of the squared value of the sound pressure signal. On the right, the 10log curves ($h^2(t)$) for the different channels. Note that in these curves, the maximum level is 0 [dB], as the interest is in the time it takes for the power to drop by 60 [dB]. Also, note that there is a small time gap between time 0 [s] and when the energy value of $h(t)$ is maximum. This interval corresponds to the time it takes the sound wave to travel between the source and the receiver and allows an estimate of the distance between them. From recorded IRs, the gap also includes the system delay, which should be compensated. That was the choice since the impulse response of an environment can easily be acquired utilizing an Ambisonics first-order microphone array. Furthermore, it is possible to find commercially available acoustic software tools to simulate sound environments capable of exporting impulse responses in Ambisonics format.

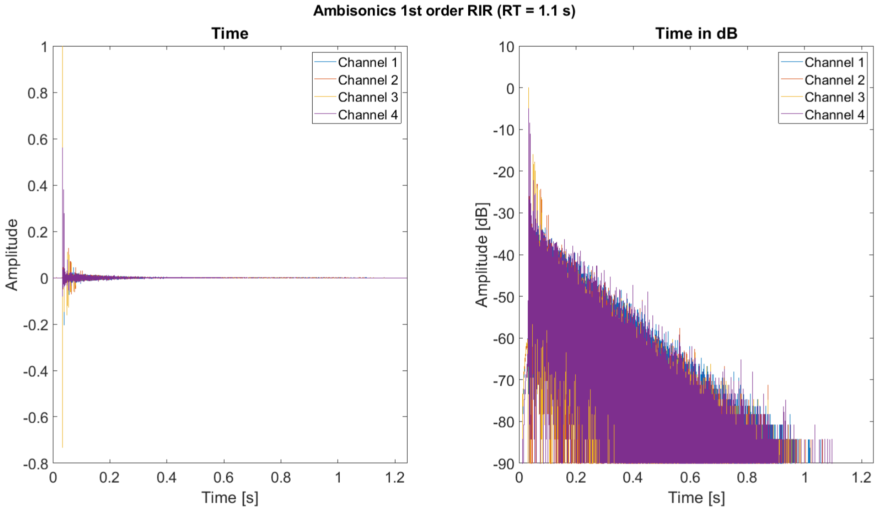


Figure 5.2: *Normalized Ambisonics first-order RIR generated via ODEON software. Left panel depicts the waveform; right panel depicts the waveform in dB.*

The system's design requires an RIR to be split into two parts. The first part contains the amount of energy to be delivered through VBAP. The second part will be computed through Ambisonics. From the reflectogram, the time representation of the latency and attenuation of the direct sound (DS), early reflections (ER), and late reflections (LR; see Figure 5.3), it is possible to find the point in time representing the direct sound (the first peak) and then separate it correctly from the rest of the RIR. Although splitting the RIR into DS and remainder may be the most straightforward method, the achieved results were initially perceived in personal experience as unnatural highlighted “dry” (not reverberant) sound from a defined position followed by really distant/disconnected reverberation, counter to the aims of a more ecologically valid sound reproduction. Thus, in the proposed method the ER part was included with the DS part.

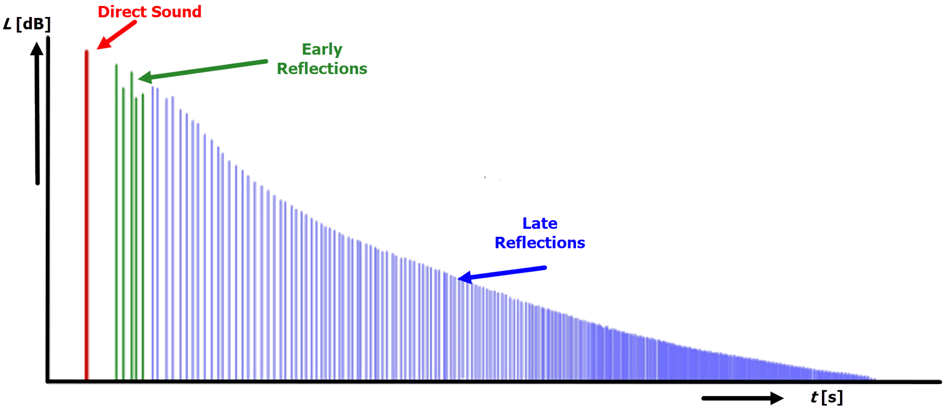


Figure 5.3: *Reflectogram split into Direct Sound Early and Late Reflections.*

The late reflections of an RIR refer to the signal wavefronts reflected and scattered several times across the different possible paths. These reflections overlap each other, and as time progresses, successive wavefronts interact with any surface, increasing reflection order, changing direction and decreasing remaining sound energy.

The literature indicates that a psychoacoustical approximation of the time point in a specific RIR when the human auditory system can no longer distinguish single reflections due to reflection density [38]. Lindau [156] proposed a transition point in time (transition time (t_m)) based on mean free path length for the wavefront (Equation 5.1).

$$t_m = \frac{20V}{S} + 12\ [\mathrm{ms}]\,, \tag{5.1}$$

where, V is the volume of the room in m^3, and S is the surface area inside the room in m^2.

The minimum necessary order of reflections to represent a uniform and isotropic sound field that leads to diffuse reverberation from an Image Source (IS) model is 3. That agrees with observations from Kuttruf [148] on the specular reflections' contribution to diffuse energy in an RIR.

This approach was implemented in a similar hybrid method by Pelzer *et al.* [221]. Another method developed by Favrot [79] also uses the IS order information from simulated RIR computed with ODEON software. Its IS reflection order information provides a point to obtain a segment of the file with the late reflections envelope used by the system to deliver a hybrid multichannel RIR. These methods consider RIR and mix specific stimuli to the output, as does the proposed method. Other hybrid auralization methods such as DirAC [243] consider the recording of a sound event (in Ambisonics) and drive the reproduction based on energy analysis spanning all sound source directions. Thus, DirAC is intended to primarily work with recorded scenes instead of convolutions with RIR.

5.2.2.3 Iceberg proposition

The proposed Iceberg method, however, uses neither the t_m method, which is dependent on the volume of the rooms and the IS simulated reflection order, nor the LR envelope time, derived from an IS simulation. Instead, a different parameter is proposed that allows generalizing to both recorded and simulated Ambisonics RIRs.

Parameters of clarity and definition are metrics to determine early/late energy balance [43]. However, the fixed time of 50 or 80 milliseconds is not appropriate to represent the transition point (from early to late reflections) on every RIR, as the slope will differ and depend on many factors [45]. The transition point changes as the amount of energy and the decay distribution change from RIR to RIR. A similar parameter that is not time fixed is the T) given by Equation 2.15 (see Section 2.3.3). This parameter is also derived from the squared RIR by calculating the transition point from early to late reflections represented as the RIR's center of gravity. Therefore, the method's name is given because of the singularity of the RIRs. They present a center of gravity on its power decay representation, which is similar to the physical blocks of frozen water called icebergs. The center of gravity is the equilibrium point between the gravity force and the water buoyancy for icebergs [34]. This representation is translated to the Iceberg method as the transition point between early and late reflections from an RIR.

This process entails an RIR applied through multiplication in the frequency domain, equivalent to a convolution in the time domain, to a sound that can be virtualized through the system. The first action of the method's algorithm is the identification the center time Ts in the channel relative to the omnidirectional channel of the Ambisonics RIR. A schematic overview of the method is presented in Figure 5.4.

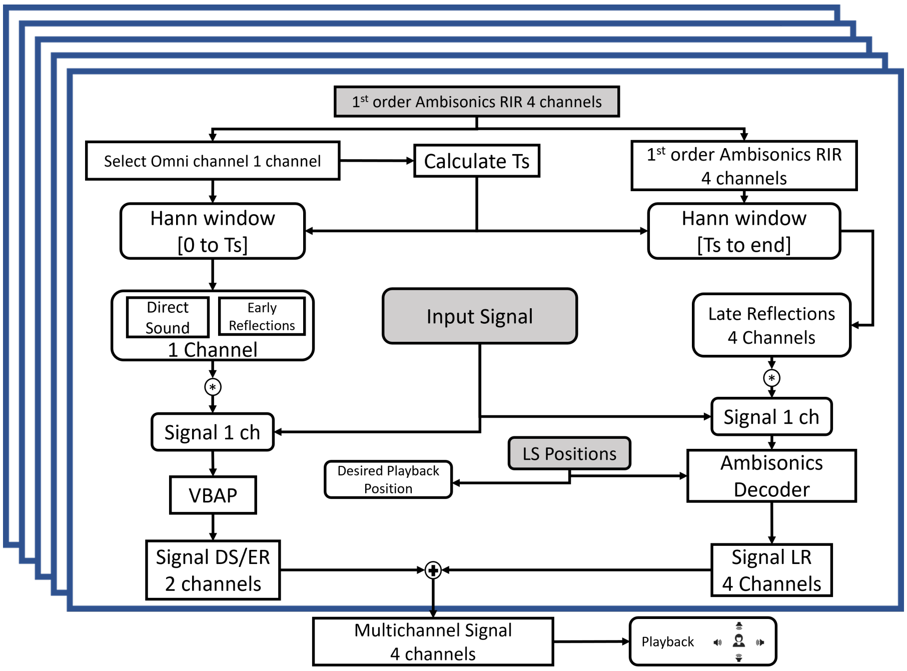


Figure 5.4: *Iceberg's processing Block diagram. The Ambisonics RIR is treated, split, and convolved to an input signal. A virtual auditory scene can be created by playing the multi-channel output signal with the appropriate setup..*

Figure 5.5 shows an example of the RIR relative to the omnidirectional input channel simulated through ODEON V.12 [59] relative to the simulated restaurant dining room used in Chapter 4 with 1.1 seconds of reverberation time.

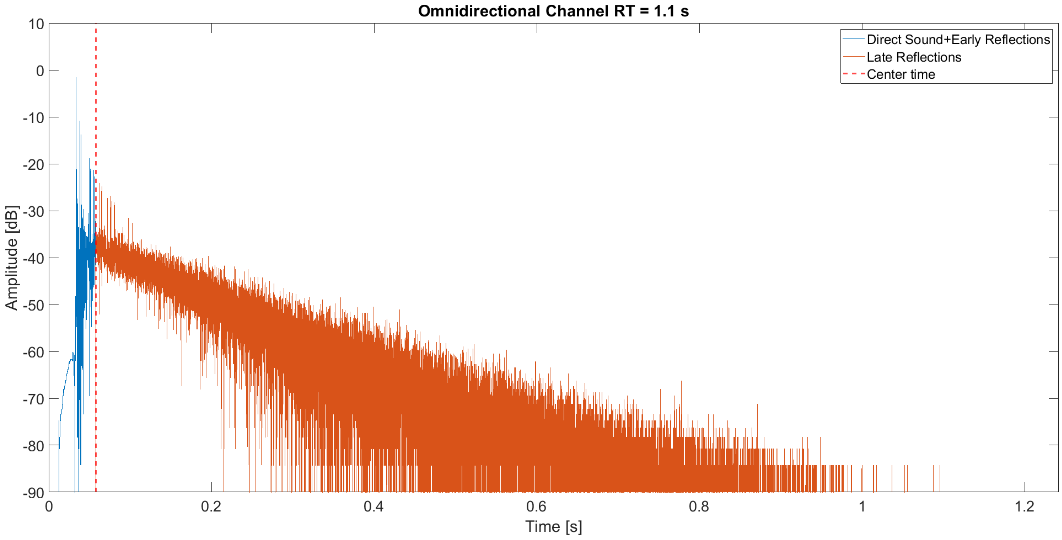


Figure 5.5: *Omnidirectional channel of Ambisonics RIR for a simulated room. The blue line indicates the part that previously selected the calculated Center Time, hence indicated as the direct sound plus the early reflections. The orange line indicates the late reverberation part of the RIR.*

Figure 5.6 presents an example of the Ambisonics RIR in the left column, the omnidirectional channel relative to the DS+ER in the middle column. The right column graphs represent the four channels late reflections part of the Ambisonics RIR.

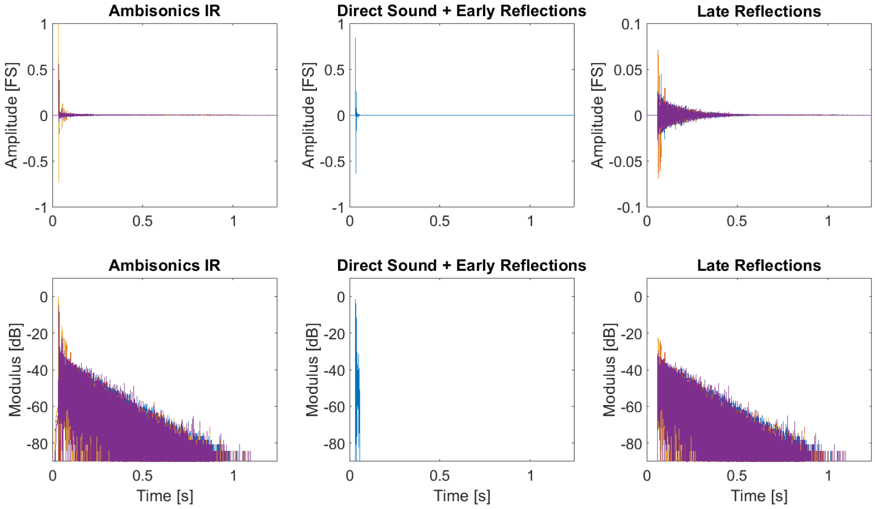


Figure 5.6: *First column: four channels Ambisonics RIR. Middle column: omnidirectional channel (DS+ER part). Right column: four channels Ambisonics RIR (LR part).*

In the sequence, the method first splits the RIR based on the TS. Then the direct sound and the early reflections are convolved with the signal to be reproduced. In this step, only the omnidirectional channel is used. Finally, the signal is processed using VBAP to provide its directional properties. The VBAP method utilized was implemented in [237]. The VBAP output is two-channel panned audio that is sent to channels of the corresponding loudspeakers. The output signal corresponds to the relative full scale of the panned signal if the provided Ambisonics RIR is normalized, or the absolute value in case of an un-normalized RIR. With the normalized RIRs, calibration of a sound pressure level is required, and the reproduction level can be set accordingly to the application needs. Assuming a coherent sum between two loudspeakers that are set to reproduce the scaled signal to a predefined level, a proportion is computed as follow:

$$LS1 = 20\log_{10}(10^{\text{level}/20} * \sin^2\theta), \tag{5.2a}$$

$$LS2 = 20\log_{10}(10^{\text{level}/20} * \cos^2\theta), \tag{5.2b}$$

where the user sets the level in dB SPL and θ is the incidence angle. A similar level calibration recording a pure tone from a calibrator with a microphone to find a system's α coefficient (as explained in 3.2.3) will allow playing the signal over each loudspeaker with the intended level. A frequency filter for each loudspeaker is also possible if the loudspeakers' FRF needs to be individually adjusted to achieve a flat(ter) response.

The second part of the impulse response is then convolved with the signal, all four channels in the proposed quadraphonic system being used. First, an Ambisonics decoder matrix observing the loudspeakers' position is created. Thus, the convolved signal is decoded from its bFormat to aFormat. The implementation utilized in the algorithm to create the decoder matrix and to decode the signal uses the functions from the Politis [237] work.

The separated signals are then merged being ready to be reproduced.

Figure 5.7 show an example for an auralization of five seconds of the International Speech Test Signal (ISTS) [120]. The top graph is the original signal, and the mid-top is the signal convolved with the DS and ER of the omnidirectional channel of the Ambisonics RIR. The envelope is minimally affected by the ER. The mid-bottom shows the signal convolved to the LR part of four channels and decoded from Ambisonics bFormat. The diffuse nature of the Ambisonics-generate LR evident in the smoother overall envelope. The bottom graph shows the result of the Iceberg method, the merged signal.

This process provides an auralized file that should be reproduced through an

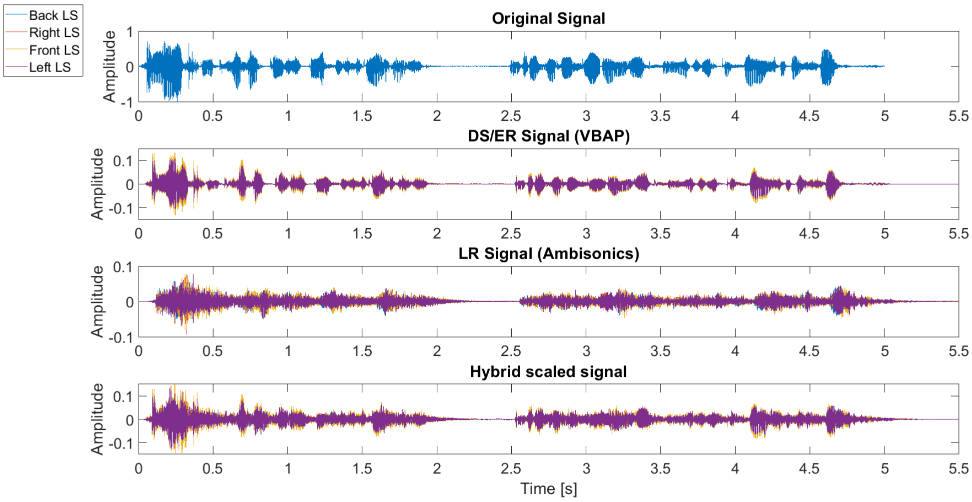


Figure 5.7: *Iceberg method example. Top graph: original signal. Mid top graph: DS+ER part (VBAP). Mid bottom graph: LR part (Ambisonics). Bottom part (Iceberg).*

equalized and calibrated setup. An equalization and calibration proposal is described in the Section 5.2.3 and can be applied to similar setups with equivalent hardware. However, the results may vary depending on hardware quality, loudspeaker amplification, and frequency response. In this work, the electroacoustical requirements (7.2.2) and Reference listening room (8.2) from the Recommendation ITU-R 1116-3 [126]: Methods for the subjective assessment of small impairments in audio systems were observed. The frequency-specific reverberation times were lower than the Recommendation: 0.04 s from 0.2-4 kHz (0.08 s at 0.125 kHz) and 0.18 s in the Recommendation. The anechoic characteristic of the room was intentionally chosen in this case to evaluate reverberation in the virtualization setup. A setup within a different space will have different room acoustics characteristics. The experimenter can compensate for the need for greater reverberation by controlling the RIRs input. The electroacoustical requirements for the loudspeakers are also relevant as they aim to guarantee the correct frequency reproduction or the possibility of compensating the frequency response with the appropriate hardware. The room proportions are also essential when setting a test environment, especially if the reproduction will include low frequencies affected by the room's

Eigentones (standing waves).

The address https://github.com/aguirreSL/HybridAuralization contains an example and the necessary resources to auralize files according to this Iceberg method. This study utilizedAmbisonics first-order impulse responses generated with the ODEON V.12 software. The choice was made by convenience, and it can be extended to any equivalent Ambisonics RIR, simulated or recorded.

The resulting RIR from ODEON is normalized. With that, the user can play a sound on a different level (from the simulated one) without rerunning the simulation using the normalized version. As an option, the method can denormalize it (dividing the RIR by its corresponding factor provided in ODEON grid [159].). The denormalized result sound will be auralized to the level simulated in ODEON (or equivalent software).

5.2.3 Setup Equalization & Calibration:

The setup can include a calibration and equalization procedure that is included in the MATLAB scripts to ensure a correct sound level reproduction and also a flatter frequency response from the system's loudspeakers, avoiding additional undesired coloration artifacts. First it was calculated a factor to transform the acquired signals from full scale to dB SPL. This step consists in recording a pure tone at a specific frequency (1 kHz) with a known input level of 1 Pa, and calculating a factor to convert the input from Full-Scale to Pa. The term indirect refers to the fact that this calculated factor is applied to all frequencies, under the assumption that the setup (microphone, pre-amplifier, power supply, and AD/DA converter) has a flat frequency response in the audible frequency range. To calculate the conversion factor a sound pressure calibrator device (in this case the B&K 4231) was connected to the microphone (1/2" B&K

4192 pressure-field and a pre-amplifier type 2669, supplied by power module B&K 5935). That provided a 93.98 dB SPL signal, which corresponds to 1 Pa. The calibration factor (α_{rms}) was calculated as in Equation 5.3. Although this step was not needed to the frequency equalization, it was convenient as once measured all the following measurements were performed without the need of entering the room.

$$\alpha_{\text{rms}} = \frac{1}{\text{RMS}(v(t)_{1\text{kHz}})} \left[\frac{\text{Pa}}{\text{FS}}\right], \tag{5.3}$$

The next step consists in equalize the frequency response of each loudspeaker. A RIR from each loudspeaker was measured and based on that an inverted FIR filter was individually created to be applied to the signals that will be reproduced. The frequency response was converted to its third-octave version, normalized, and inverted to create a vector with 27 values from 50 Hz to 20 kHz. These vectors contained the correction values in the frequency domain and can be applied to any input signal. To apply this corrections a Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) was used in MATLAB to fit the values to the given input. Figure 5.8 presents an example of the normalized third octave moving average RIR acquired with a loudspeaker (blue line), the same RIR but acquired with a signal that was filtered (red line), and the filter frequency values obtained with the inversion of the original RIR (black line).

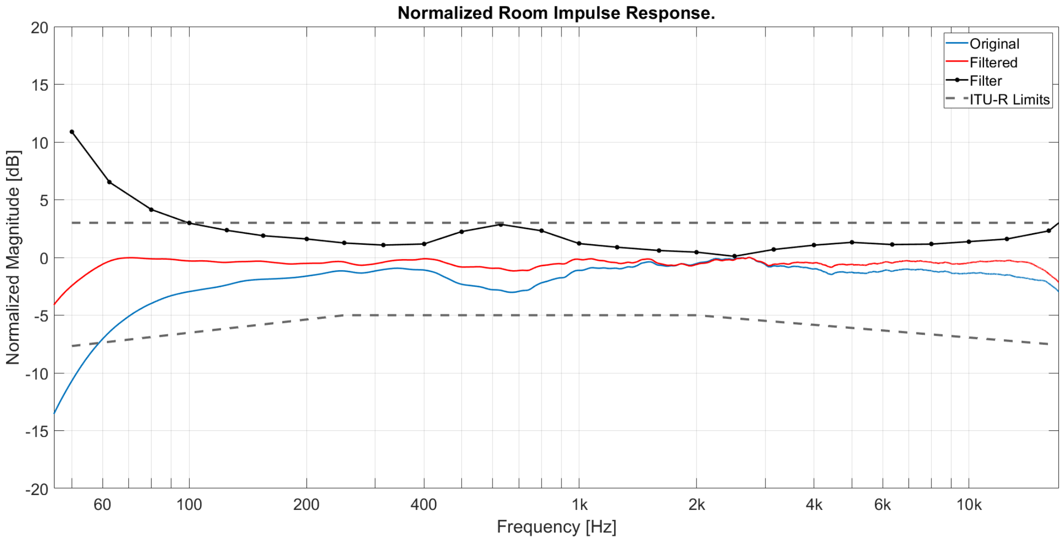


Figure 5.8: *Loudspeakers normalized frequency response and inverted filter. Doted lines represent ITU-R 1116-3 limits.*

Figures 5.9 and 5.10 shows the moving average of each loudspeaker's normalized frequency response without and with the filter, respectively.

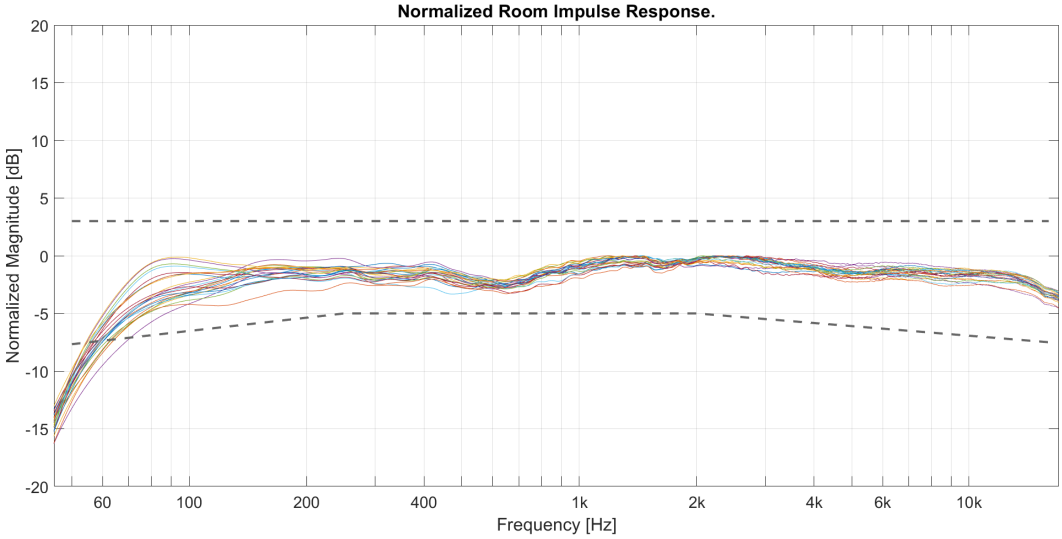


Figure 5.9: *Loudspeakers normalized frequency response (colored solid lines), doted lines represent ITU-R 1116-3 limits.*

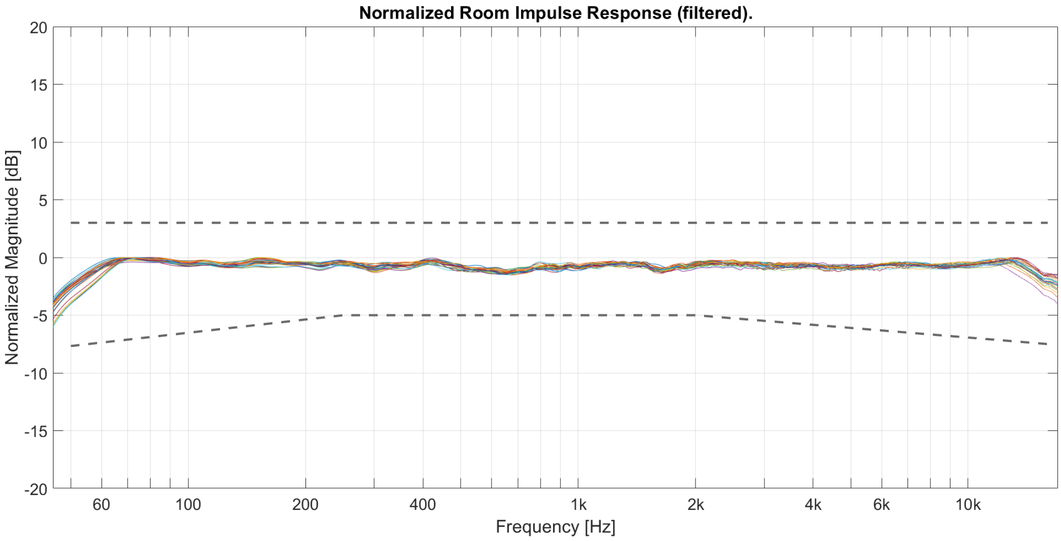


Figure 5.10: *Loudspeakers normalized frequency response with frequency filter correction (colored solid lines), doted lines represent ITU-R 1116-3 limits.*

As the amplification to each active loudspeaker is individually controlled it is possible that a same file could be reproduced at a different sound pressure level (if someone inadvertently or accidentally change the volume control button directly in the loudspeaker for example). Since the α_{rms} was already calculated and it was possible to convert a signal from FS to Pa, and consequently, to dB SPL, and *vice-versa* the individual loudspeakers' SPL were measured with a signal defined to be played at 70 dB SPL Equation 5.4.

$$\text{signal}(t) = \left(\frac{\text{signal}(t)}{\text{rms}(\text{signal}(t))} 10^{\frac{70-\text{dBperV}}{20\upmu}} \right) \Gamma_l \tag{5.4}$$

where

Γ_l is the level factor to the loudspeaker l with initial value = 1;
$\text{dBperV} = 20 \log_{10} \left(\frac{\alpha_{\text{rms}}}{20\upmu} \right)$.

The signal(t) was played through a loudspeaker l and simultaneously recorded with the microphone $\text{S}_l(t)$; the SPL of the recorded signal was calculated as follows

$$\mathrm{SPL}_l = 20\log_{10}\left(\frac{\mathrm{S}_l(t)[\mathrm{FS}]\alpha_{\mathrm{rms}}\left[\frac{\mathrm{Pa}}{\mathrm{FS}}\right]}{20[\mu\mathrm{Pa}]}\right)[\mathrm{dB}]\,, \tag{5.5}$$

Ten measurements were sequentially performed with each loudspeaker at intervals of 1 s; another iteration of measurements were performed if the measured SPL exceeded the tolerance of 0.5 dB on any of the measurements. A step of ± 0.1 [FS] is set to update Γ_l in its next iteration according to the SPL obtained.

5.3 System Characterization

The Iceberg auralization method in a four-loudspeaker system (minimum required) was evaluated for its capabilities to reproduce the intended reverberation time and the appropriate binaural cues. This section describes the system setup, and the conditions experimented with utilizing the Iceberg method. The method's accuracy at the optimal and sub-optimal positions was considered in this characterization as well the impact of the RT. Furthermore, placing a second listener inside the ring was investigated to support a more ecological situation. By the end, a complementary study for those conditions was conducted with an aided mannequin to supplement the objective data as the pandemic prevented subjective data collection.

The present study used the ITA-Toolbox [29] for signal acquisition and processing. To further enhance the accuracy of the localization estimates, a MATLAB implementation of the May and Kohlrausch [182] localization model from the Auditory Modeling Toolbox (AMT, https://www.amtoolbox.org) [287]) was also employed. The May model is specifically designed to be robust against the detrimental effects of reverberation on localization performance, making it an ideal choice for supplementing the objective data gathered in the present study. The reverberation, or the persistence of sound after its initial source has ceased, was a parameter in this test that could significantly distort the estimated location of a sound source. The May model accounts for reverberation's influence through frequency-dependent time delay parameters, enabling more accurate localization estimates in reverberant environments. By incorporating the model in our analysis, we supplemented the objective data gathered through signal processing with an additional layer of modeling that allowed a relative comparison with previous studies.

The main objective of an auralization method and its virtualization setup is to deliver appropriate spatial awareness to human listeners. The natural step for this would be to verify and validate the method. Unfortunately, special conditions were in place during the course of this study; due to COVID-19 restrictions, validation tests with participants were not feasible. The Section 5.5 extends the system verification and analysis to a targeted application of hearing aid research. Although it does not replace a subjective impression validation and analysis, it can help understand and predict the system's behavior in a typical use case for hearing research, which is the user with hearing aids.

5.3.1 Experimental Setup

The proposed method was implemented, and the tests were conducted at Eriksholm Research Centre in Denmark. The test environment was an anechoic room (IAC Acoustics) with inner dimensions of 4.3 m × 3.4 m × 2.7 m. Signals were routed through a sound card (MOTU PCIe-424) with a Firewire 440 connection to the MOTU Audio 24 I/O interface and played via loudspeakers Genelec model 8030C (Genelec Oy, Iisalmi, Finland). The well-controlled sound environment was appropriate for the assessment of small impairments in audio systems, although the acoustic properties of the room exceed the sound booths and rooms commonly encountered in audiology clinics [316].

5.3.2 Virtualized RIRs & BRIRs

A set of 72 room impulse responses and 72 binaural room impulse responses were acquired through the system separated over 5 degrees angles around the center position assuming x as lateral axis and y the front-back (mid-saggital) axis of a person inside the ring. Moreover, the same number of RIRs and BRIRs were measured at off-center positions.

The virtualized RIRs and BRIRs were acquired utilizing a logarithimc sweep signal (50-20000 Hz, 2.73 s (FFT Degree 18, Sample Frequency 96 kHz)) [194] as input. The signal was auralized to each angle in the Iceberg method for the same three spaces as in Chapter 4: a classroom (9.46 m x 6.69 m x 3.00 m) with an overall Reverberation Time RT of 0.5 s, a restaurant dining area (12.19 m x 7.71 m x 2.80 m) with an overall RT of 1.1 s, and an anechoic room (4.3 m x 3.4 m x 2.7 m) with an ideal overall RT of 0.0 s. All rooms were acoustically simulated in ODEON software V.12, that generated the ambisonics RIRs representing each mentioned source-receptor configuration. The absorption coefficients of the room surfaces are listed in Appendix E.

The initial step to acquiring the RIR and BRIR was to auralize the sweep file utilizing the Iceberg method to the desired positions (72 angles around the center) in three different room conditions. Then play it through the four loudspeakers positioned in the front (0°), left (90°), back (180°) right (270°) counter-clockwise angles. The auralized version of the sweep should correspond to the signal played in the virtual environment as the reverberation added by the anechoic room is negligible. After that, the recorded file was de-convolved with a zero-padded version of the raw sweep (See Figure 5.11).

The playback and recording utilized the maximum sampling rate supported on the AD/DA system (96,000 Hz) as the difference in time is in the µs scale. Therefore, the step size in time provided in microseconds is given by step size $= (1/96,000) * 1,000,000 = 10.42$ µs. The created sweep duration was 2.731 s (FFT Degree = 18).

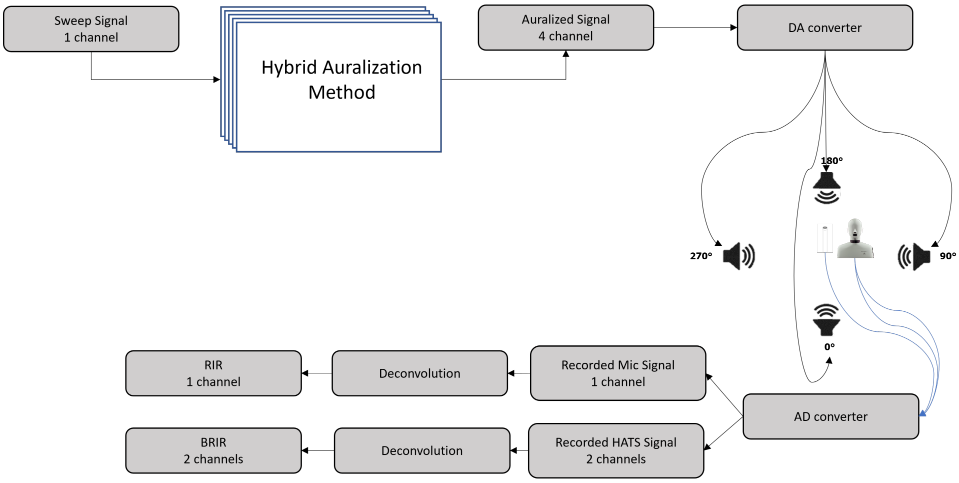


Figure 5.11: *BRIR/RIR acquisition flowchart: Iceberg auralization method.*

A manikin with artificial pinnae (HATS model 4128-C; Brüel and Kjær) was used to record the binaural files. Also, a second listener was simulated during the tests with a different manikin (KEMAR; GRAS), (See Figure 5.12). The HATS recordings were calibrated as described in Section 3.2.3 following Equations 3.1a and 3.1b.

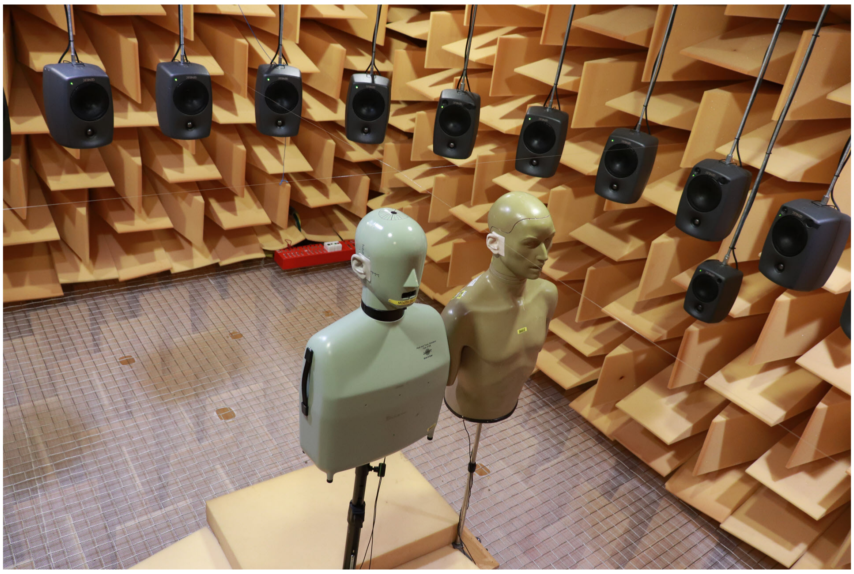


Figure 5.12: *BRIR measurement setup: B&K HATS and KEMAR positioned inside the anechoic room.*

5.3.3 Conditions

The auralized files were then recorded under the following conditions:

- Optimal position (alone and centered)
- Optimal position (centered) accompanied by a second listener
- Off center positions alone

The positions grid can be visualized in Figure 5.13.

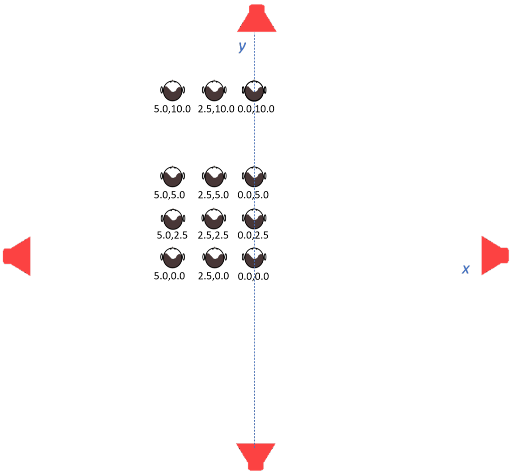


Figure 5.13: *Measurement positions: Obtained through virtualized sound sources with Iceberg method (VBAP and Ambisonics) in a four-loudspeaker setup.*

The most accurate performance is theoretically expected from optimal position. These techniques provide virtualization assuming the receptor (listener) is in the center of the loudspeaker ring [65, 241]. Adding a second listener into the reproduction area and/or moving the primary listener away from the center can challenge the system's ability to render the scene as intended. The following sections presents and discusses the system's capabilities to reproduce Iceberg auralized files by measuring the binaural cues and RT in different conditions.

5.3.4 Reverberation Time

A room's characteristic wave field pattern can affect the human perception of a reproduced sound. Room acoustics can alter attributes related to spatial perception. For example, a recorded sound has almost no chance of being correctly reproduced if the reproduction room has stronger reverberation than the recorded one. Also, reverberation overshoot can smear the perceived direction of a sound source, as early reflections would be heightened in this case [242]. The RT was calculated from impulse responses measured within the three virtualized environments (note that the simulated environments were aimed to present RT of 0, 0.5, and 1.1 seconds). Reverberation time was calculated using the ITA Toolbox. The parameters were set as follows: Frequency from 125 Hz to 16 kHz, one band per octave, and 20 dB threshold below maximum.

The reverberation time was shown to be stable in this virtualization setup. An approx. 0.08 s Overall RT can be observed for the anechoic simulation (0 s RT). That is most likely driven by the presence of the hardware inside the anechoic room: loudspeakers and wood base for the chair, although covered with foam. The overall reverberation time was measured without an omnidirectional sound source. To circumvent this limitation, the measurement was repeated utilizing all 24 loudspeakers as sound source, one at a time. The overall RT in this case was considered as the maximum value across frequencies in octave bands from 125 Hz to 16k Hz. Figure 5.14 presents the boxplot of the measured values in relation to its position inside the room. Rows represent the aimed RT (0, 0.5 and 1.1 s). The top line presents results without lateral displacement, the middle line presents the results according to a lateral displacement of 2.5 cm from the center and the bottom line presents the results according to a lateral displacement of 5 cm from the center.

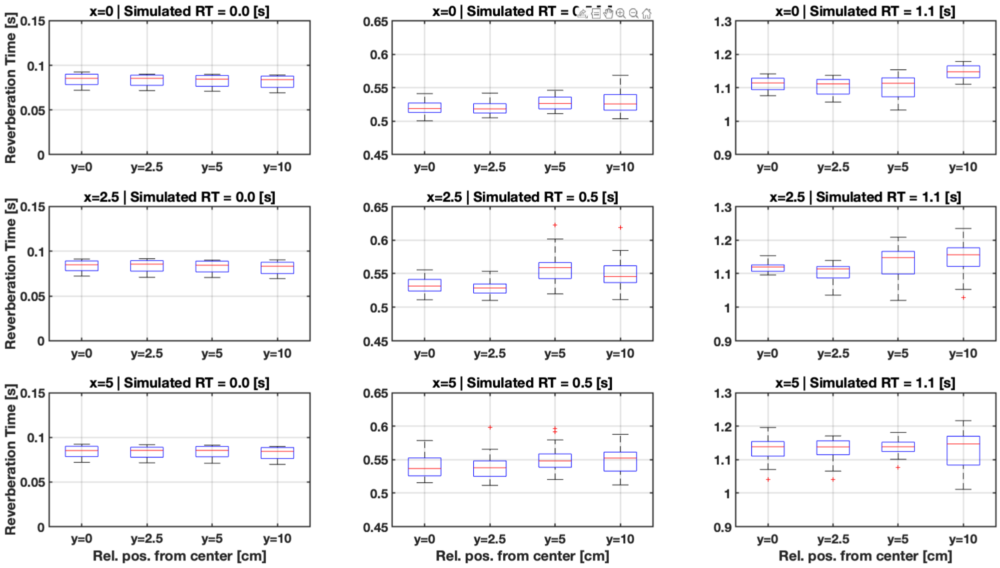


Figure 5.14: *Reverberation Time environments measured with files produced with Iceberg method and virtualized in four-loudspeakers.*

Table 5.1 present the median of the values of the overall RTs. Therefore, it is possible to notice that virtualized environment RTs', tend to be stable, and to the measured conditions, under the just noticeable difference JND of 5% [264, 265] across positions inside the room.

Table 5.1: Reverberation Time in three virtualized environments in different positions inside the loudspeaker ring.

	RT = 0	RT = 0.5	RT = 1.1
Position [cm]		Overall RT [s]	
x=0.0; y=0.0	0.085	0.519	1.114
x=0.0; y=2.5	0.085	0.519	1.111
x=0.0; y=5.0	0.085	0.526	1.113
x=0.0; y=10.0	0.084	0.526	1.147
x=2.5; y=0.0	0.085	0.531	1.120
x=2.5; y=2.5	0.086	0.529	1.114
x=2.5; y=5.0	0.084	0.559	1.148
x=2.5; y=10.0	0.083	0.546	1.157
x=5.0; y=0.0	0.085	0.537	1.139
x=5.0; y=2.5	0.085	0.538	1.138
x=5.0; y=5.0	0.085	0.548	1.138
x=5.0; y=10.0	0.084	0.552	1.147

5.4 Main Results

This section presents the results based on the mannequin positions (center and off-center) and conditions (HATS and HATS with KEMAR) to angles referenced clockwise.

5.4.1 Centered Position

5.4.1.1 Interaural Time Difference

The blue line in Figure 5.15 represents the result of the Interaural Time Difference ITD filtered with a 1 kHz low-pass filter virtualized through the proposed system.

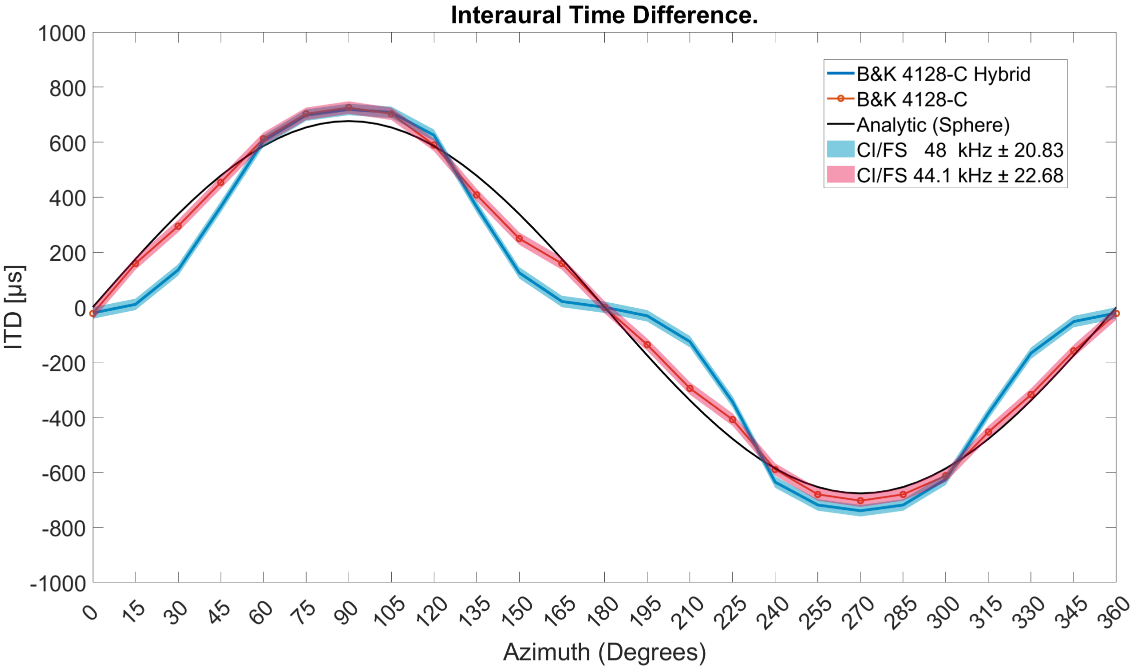


Figure 5.15: *Interaural Time Difference under 1 kHz as a function of azimuth angle for a HATS Brüel and Kjær TYPE 4128-C in the horizontal plane through a proposed Iceberg auralization method on a 4 loudspeakers setup. The red line is the ITD results with real loudspeakers (without virtualization). According to the sample rate, the blue and red shaded areas in are the confidence intervals. The black line represents the analytical ITD values.*

Wang and Brown [297] defined the analytical ITD (black line in figure) (See Equation 5.6) considering a centered, perfect sphere of 10.5 cm of radius (a) and sound propagation velocity (c) of 340 m/s, θ is the angle in radians.

$$\text{ITD} = \left(\frac{a}{c}\right) 2\sin(\theta) \tag{5.6}$$

The maximum absolute difference found is 170 µs, representing a mismatch around 15º on the given angle. The calculated average difference is 67 µs, representing a difference of around 7º in localization.

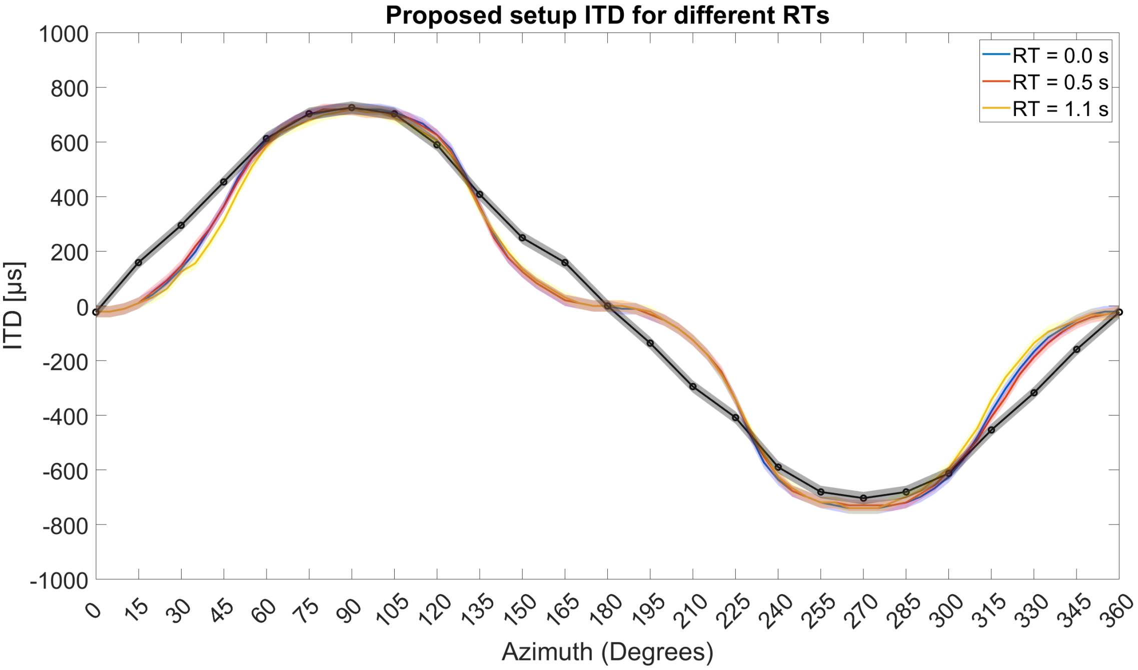


Figure 5.16: *Interaural Time Difference at 1 kHz as a function of azimuth angle for a HATS Brüel and Kjær TYPE 4128-C in the horizontal plane through a proposed Iceberg method on a 4 loudspeakers setup on three different reverberation time scenarios.*

Three different simulated rooms were measured utilizing files generated via Iceberg method to a four-loudspeaker setup, keeping the listener in the center position. Figure 5.16 presents the acquired ITDs with Iceberg method for RT = 0 s (blue), RT = 0.5 s (red), RT = 1.1 s (yellow) and the ITD for RT = 0 without using virtualization (*i.e.*, reproducing through real loudspeakers) in black. There were no substantial differences across the different reverberation times virtualized. This was expected, though, as the direct sound drives the ITD.

5.4.1.2 Interaural Level Difference

Figure 5.17 shows the ILDs (calculated following Equation 3.7) across octave bands for the angles around the center horizontally spaced 30º for better visualization. The ILDs were most affected than the ITDs, with a substantial reduction in ILD relative to the actual loudspeakers observed in the 2 kHz band. The ILD values have a similar pattern and magnitude for a significant part of the spectrum at most angles.

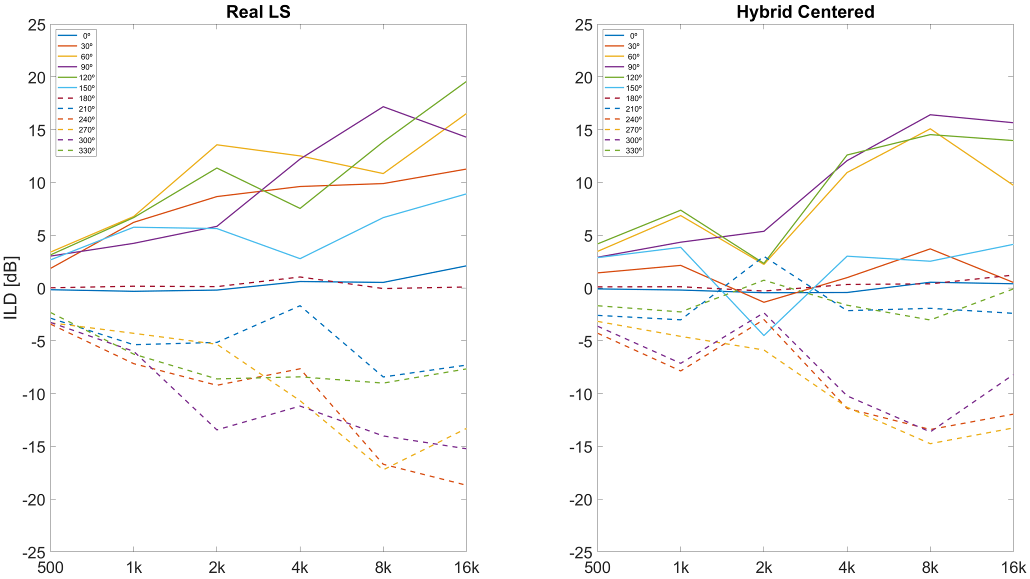


Figure 5.17: *Iceberg Interaural Level Differences as a function of octave-band center frequencies; separate lines for angles of incidence.*

Figure 5.18 presents the ILD for both setups: real loudspeakers and Iceberg method in six octave bands as a function of azimuth in spaces of 15º. The top-right corner graph shows the 2 kHz band. It shows that apart from the positions where there is an actual loudspeaker (*i.e.,* 0º, 90º, 180º, and 270º), the differences are large, greater than 10 dB at some azimuth angles.

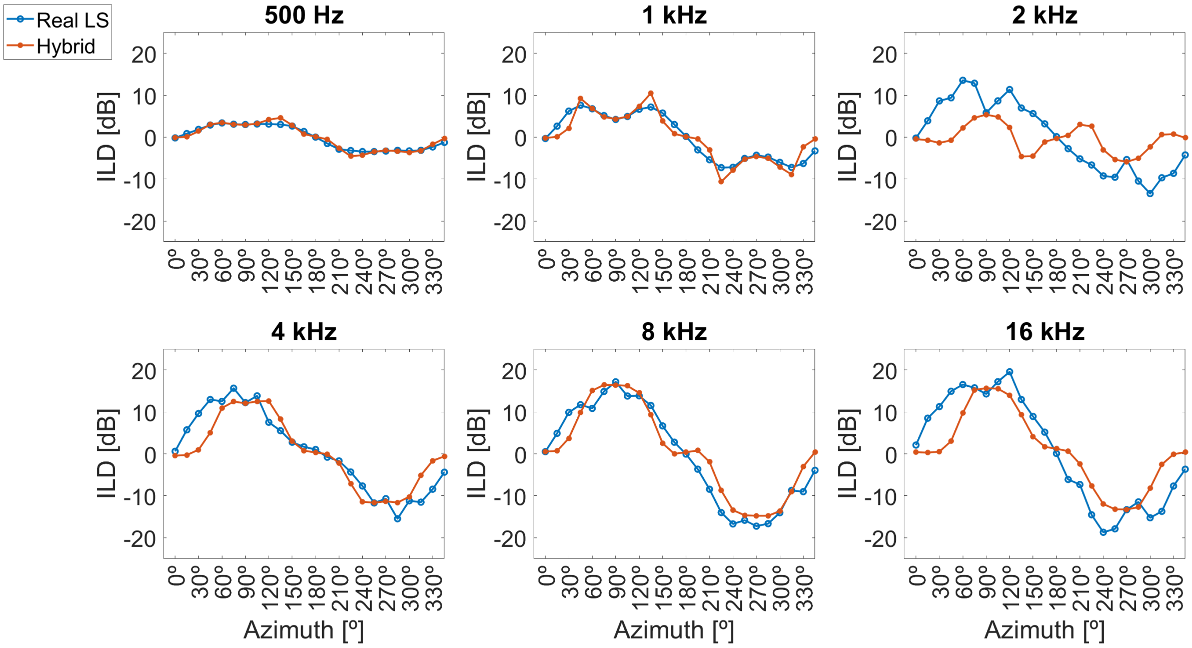


Figure 5.18: *Iceberg ILD as a function of azimuth angle. Listener alone in the center.*

Figure 5.19 shows the absolute difference in ILD between physical loudspeakers and virtual loudspeakers created by the Iceberg method.

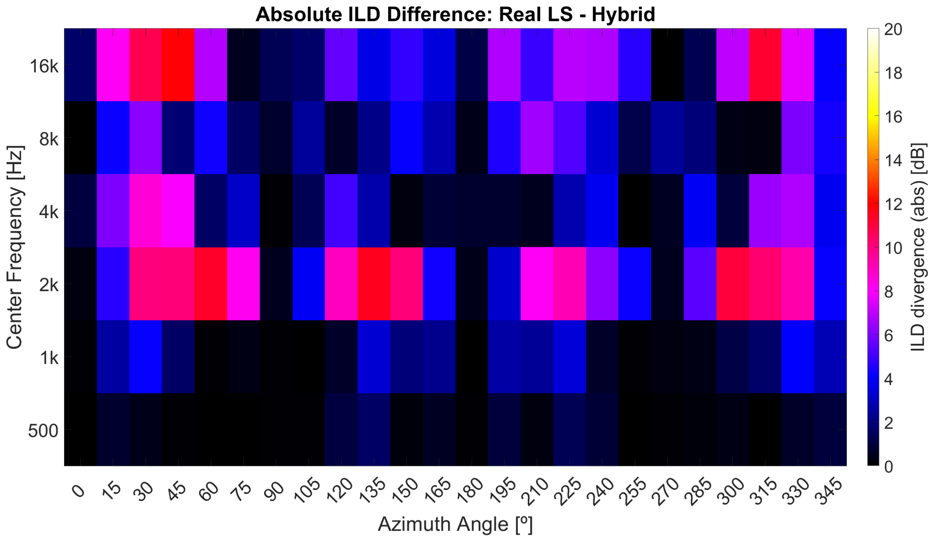


Figure 5.19: *Iceberg method: Absolute ILD Differences as a function of azimuth angle.*

Tu [291] measured just-noticeable differences (JNDs) in ILDs for normal-hearing participants using pure tones at different presentation levels. These JNDs can be used to estimate the perception of differences between ILDs obtained with physical loudspeakers and the Iceberg auralization setup to analyze if the ILD difference between setups will be perceived in a given specific frequency band. Figure 5.20 presents the values from Figure 5.19 minus the appropriate pure-tone ILD JND values. That means that positive values that exceeded the JND could be perceived not as intended; that is, a perceptible ILD deviation can cause spatial distortion [38]. The 2-kHz ILDs show up to 8º divergence across most angles. ILDs in other frequency bands (1, 2, 8, and 16 kHz) also presented values that could relate to noticeable differences (up to 4 dB), but those are mostly limited to frontal ±30° angles.

The 2 kHz mismatch can be considered a flaw in the reproduction system. The effect on sound localization and subjective impression of complex sounds involving these frequencies needs further investigation as to the scale of spatial distortion. As the ITDs and ILDs at other frequencies were relatively well preserved in the auralized system with only real loudspeakers, it is possible this flaw at 2 kHz may have a minimal effect, especially for lower frequency stimuli. System reliability should be verified first for stimuli with peak energy in the 2 kHz band or tasks requiring greater localization accuracy (e.g., with sound sources within ±30°).

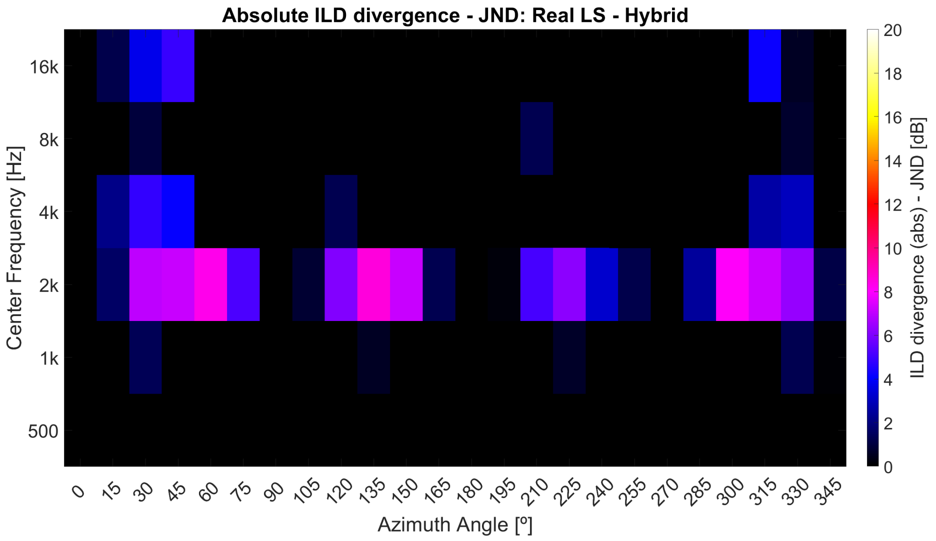


Figure 5.20: *Iceberg: Absolute ILD Differences over JND as a function of azimuth angles around the central point).*

5.4.1.3 Azimuth Estimation

The frontal azimuth angle was estimated using the binaural model by May and Kohlrausch [182]. Each BRIR was convolved with a pink noise with a duration of 2.9 s as input into the model. The mean of the azimuth of each file is stated as azimuth predicted by the model. Figure 5.21 presents the angles estimated with the May and Kohlrausch model for files auralized with the Iceberg method for an anechoic room and virtualized over the four-loudspeaker setup (blue curve), angles estimated for binaural files acquired without virtualization with real loudspeakers (red curve), and the reference (dotted black).

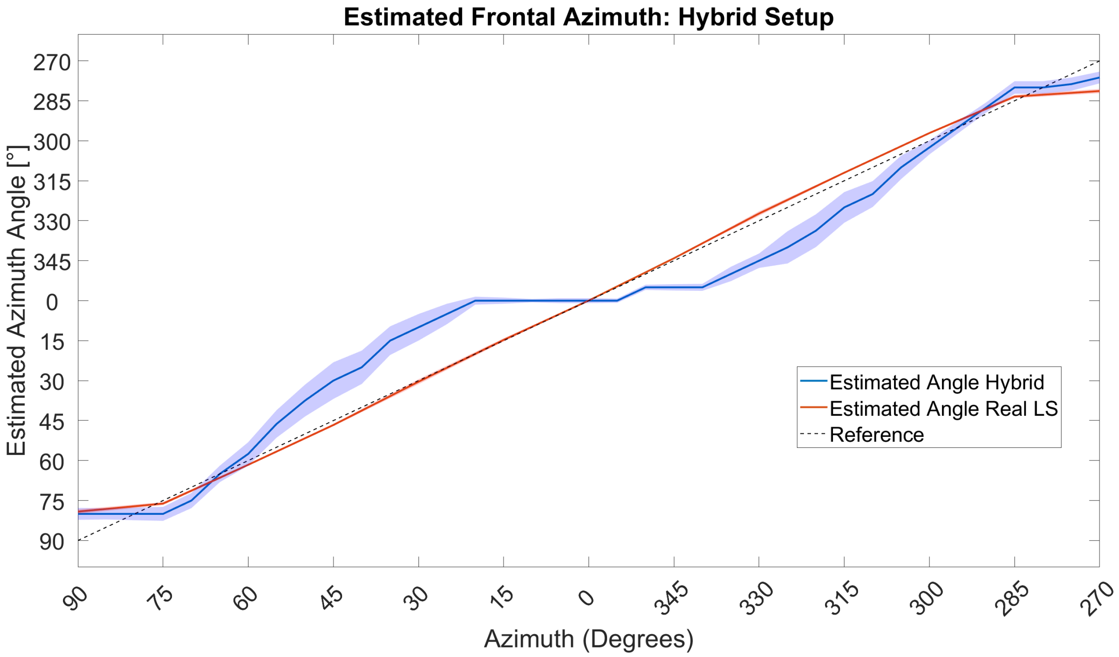


Figure 5.21: *Iceberg method: Estimated azimuth angle (model by May and Kohlrausch [182]), HATS centered, and RT = 0s.*

The model's results are in line with the analysis of the binaural cues supporting the assumption of the worst localization accuracy around ±30º difference (30º and 330º in Figure 5.21). Also, the virtualized sound tends to have more difficulty separating from the frontal angle (0º).

5.4.2 Off-Center Positions

Moving the primary listener off-center (displaced both on the x and y axes) is proposed to measure the impact of a person's head (and body) not being centered – such as when not fixated – on the system's ability to render the appropriate binaural cues.

5.4.2.1 Interaural Time Difference

Figure 5.22 presents 72 measured ITDs around the listener (5º spacing) in four different placements: at center and displaced forwards (y-axis) 2.5, 5 and 10 cm.

When displaced from the center position, the Iceberg method can cope with delivering a reasonably interaural time difference in frontal displacements up to 5 cm or considering a simultaneous misplacement lateral and frontal up to 2.5 cm. However, compared to the center position, the error increased dramatically with 10 cm displacement for frontal angles (around ± 45) up to 400 µs compared to the listener in the center.

Lateral displacement positions (2.5 and 5.0 cm) were also investigated. The ITD results for these displacements presented the same trend as seen without lateral displacement. Similar results were found when virtualizing the scenes with a reverberation time of 0.5 and 1.1 seconds. All combination results are presented in Figure 5.23 to improve readability.

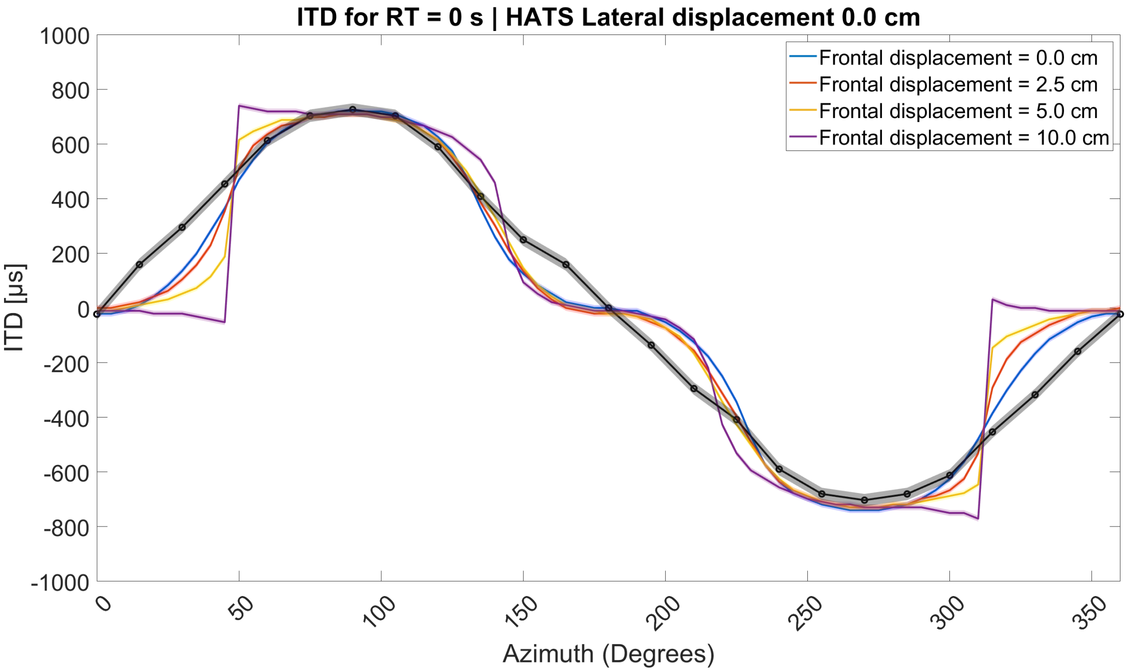


Figure 5.22: *Iceberg ITD as a function of frontal displacement: Centered listener in proposed Iceberg method in a four-loudspeaker setup.*

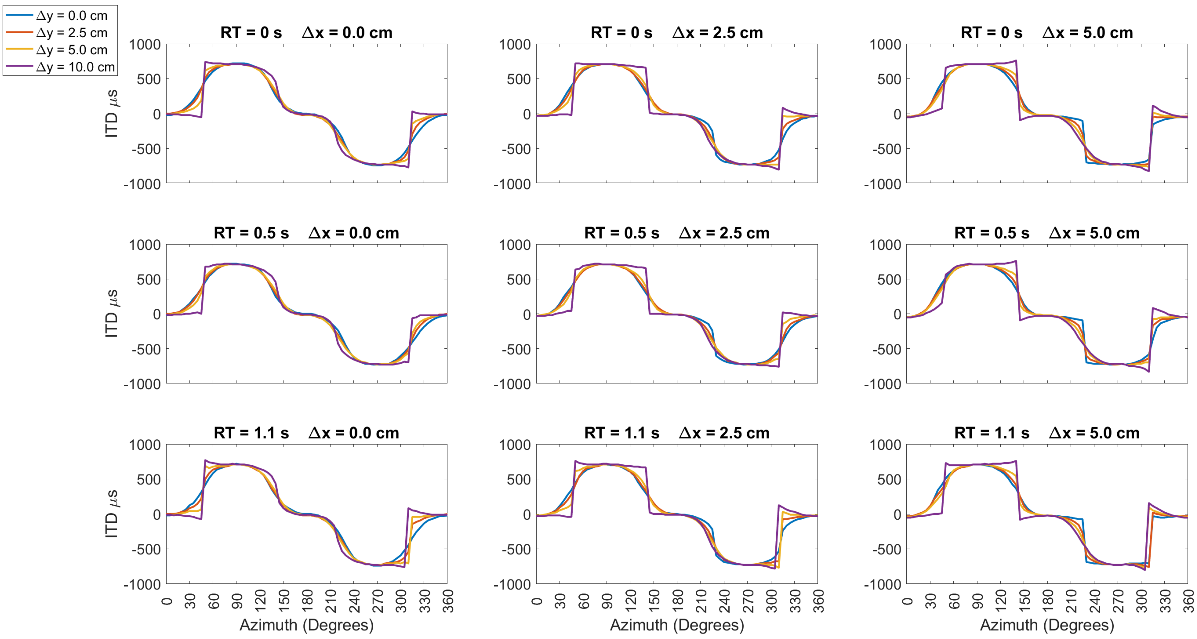


Figure 5.23: *ITD Iceberg virtualized setup: Listener displacement: listener position 2.5 cm off-center in proposed Iceberg method in a four-loudspeaker setup.*

ITDs were affected with frontal displacements depending on the amount of reverberation simulated. In the simulated dry condition the squared behavior is present with 5 cm off center, with mild reverberation the effect only appears with a displacement of 10 cm and the largest reverberation tested demonstrated the problem to virtualize sources in all off center positions. The deviation is centered to ±45º in all conditions. Lateral movements were even more affected, as expected, delivering ITDs based on loudspeaker position (the squared shape) and not by circular placement of virtualized sound sources on displacements further than 3.5 cm from the center (combining the lateral and frontal movements.)

5.4.2.2 Interaural Level Difference

Figure 5.24 presents the difference between the ILDs measured in the center and the ILDs measured in different positions for a dry room simulation (RT = 0 s). The lateral displacement (x-axis) is ordered as rows (top row = center, middle row = 2.5 cm and bottom row = 5 cm to the right). The four columns are related to the frontal displacement (y-axis) of 0 (center), 2.5, 5, and 10 cm. Note that these are additional ILD errors to the previously discussed errors introduced by the simulation itself (with the listener at center).

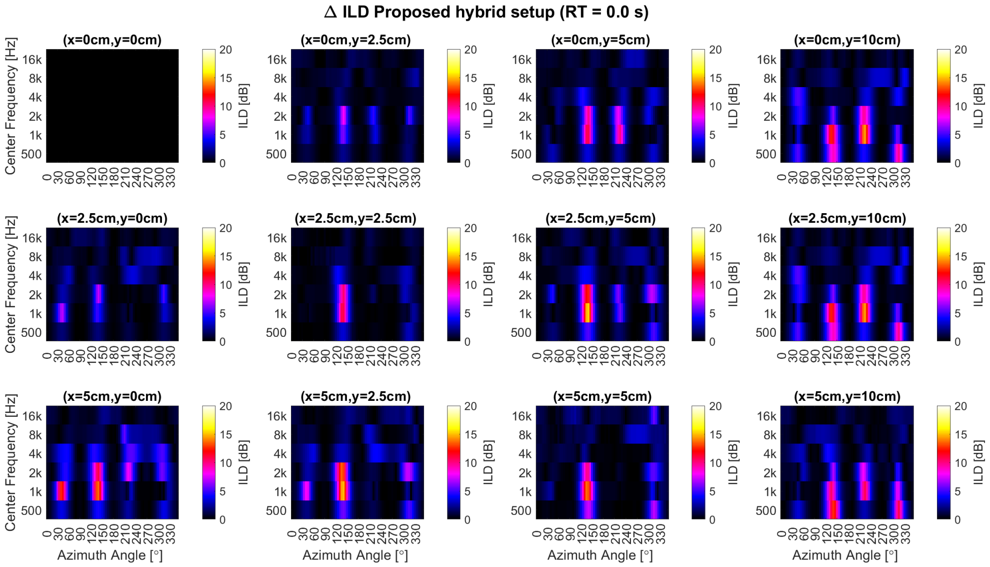


Figure 5.24: *Difference in ILD as a function of azimuth angle for a HATS Brüel and Kjær TYPE 4128-C in the horizontal plane through a proposed Iceberg auralization method on a four-loudspeakers setup (RT = 0.0 s).*

The ILDs are affected by listener displacement mostly in the mid frequencies and only at certain angles. Lateral displacement of 2.5 cm produces larger interference (up to 8 dB) for the left angles 40º and 130º. In contrast, at other angles, ILD differences are lower than 3 dB. The 5 cm displacement also presents differences up to 15 dB at these same angles and up to 8 dB differences contralaterally (220º and 320º). Frontal displacement follows a similar pattern with more differences at some of the rear angles (130º and 220º). These particular differences indicate relatively low impact on ILD cues using the Iceberg method in the simulated anechoic room (RT = 0 s). Similar results in terms of affected angles were found analyzing ILDs for the same listener positions for simulated rooms with RT = 0.5 s Figure 5.25 and RT = 1.1 s Figure 5.26. These conditions are closer to everyday situations. The increased energy of the late reflections results in lesser magnitude differences in ILD, indicating slightly better performance for more realistic simulations.

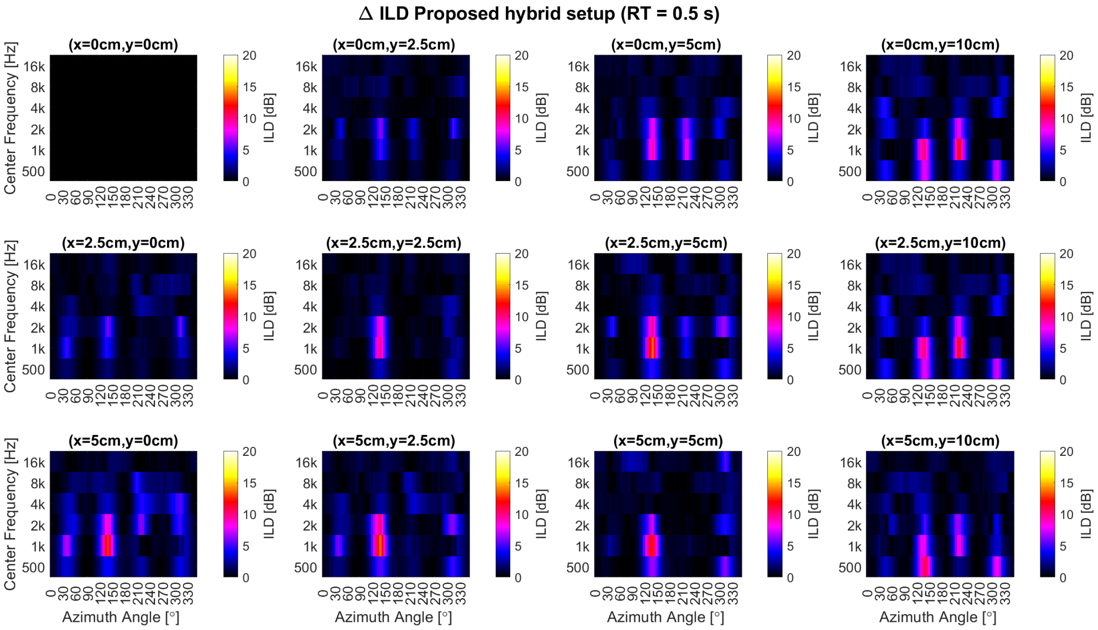


Figure 5.25: *Difference in ILD as a function of azimuth angle for a HATS Brüel and Kjær TYPE 4128-C in the horizontal plane through a proposed Iceberg method on a four-loudspeakers setup RT = 0.5 s.*

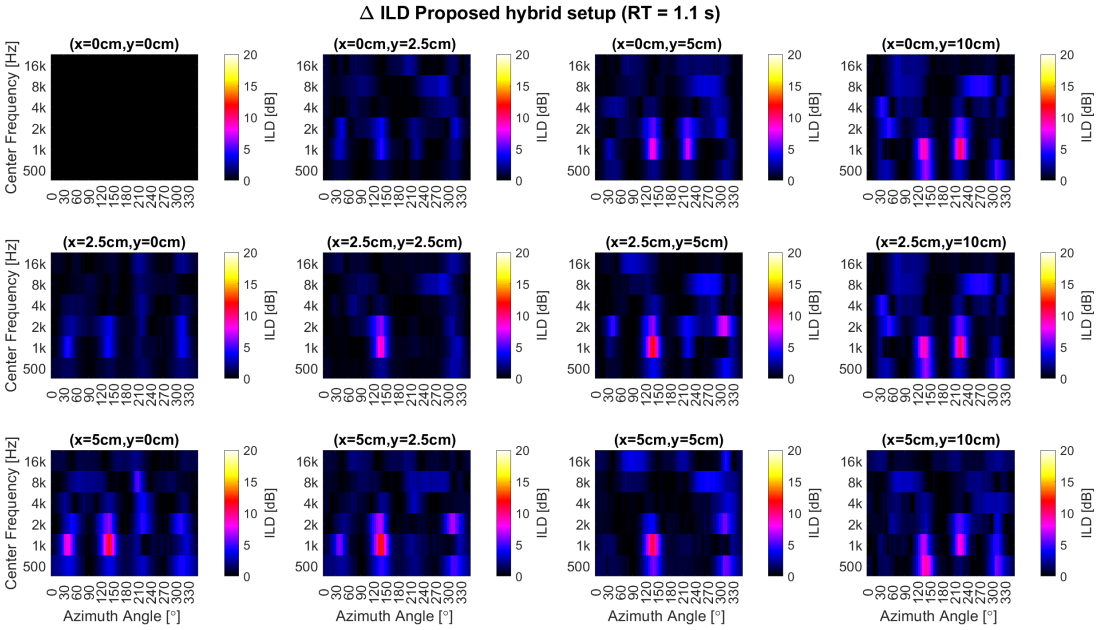


Figure 5.26: *Difference in ILD as a function of azimuth angle for a HATS Brüel and Kjær TYPE 4128-C in the horizontal plane through a proposed Iceberg method on a four-loudspeakers setup RT = 1.1 s.*

5.4.2.3 Azimuth Estimation

Using again the May *et al.* model, in the same setup as in Section 5.4.1.3 to predict the localization of a sound source, Figure 5.27 presents the predicted source locations when moving the listener along the grid positions mentioned (x = 0, 2.5 and 5 cm; y = 0, 2.5, 5 and 10 cm).The different RTs are represented by the line colors in the graphs (blue = 0.0s, red = 0.5s, yellow = 1.1s). The results indicate the system's spatial sound accuracy is dependent of the listener position. On the other hand the error is not dependent on the reverberation time. Lateral movements increase the error to the side that is getting closer to the ear, while lessened on the contra-lateral side. Frontal movements increased the number of angles that are not delivering the correct source angle (longer straight horizontal line around zero). The model estimates an maximum error up to ≈30º to a listener within 3.5 cm from the center (combining Lateral anf frontal displacement).

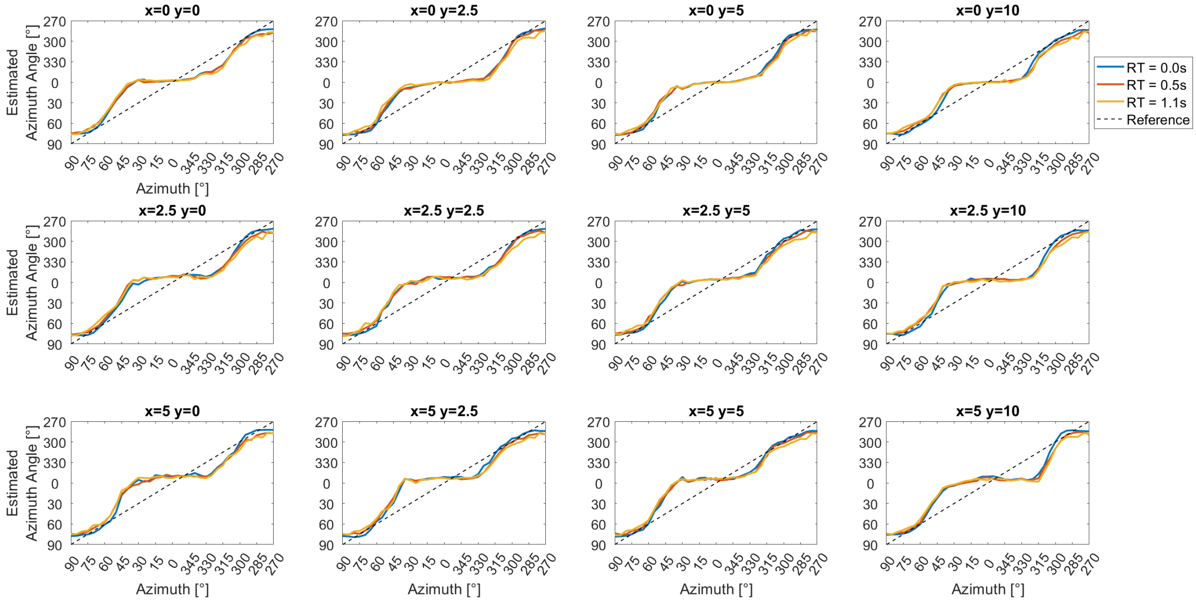


Figure 5.27: *Estimated (model by May and Kohlrausch [182]) frontal azimuth angle at different positions inside the loudspeaker ring as function of the target angle.*

The errors when comparing the angles estimated on displaced positions to the estimated to the center position are lessened with the increment of reverberation to the majority of the angles.

5.4.3 Centered Accompanied by a Second Listener

The Binaural cues were investigated, adding a second listener to the scene and maintaining the first in the center (sweet spot). The second listener was positioned in three different lateral (x-axis) distances on the left from the center:

- 50 cm (simulating shoulder to shoulder).
- 75 cm.
- 100 cm.

5.4.3.1 Interaural Time Difference

The upper row of Figure 5.28 shows the ITDs considering the setup with the HATS alone at the center (blue line) and also with the presence of a second listener positioned at the right side at three different distances from the center and the reference. The ITDs in black were computed with no virtualization as a reference.

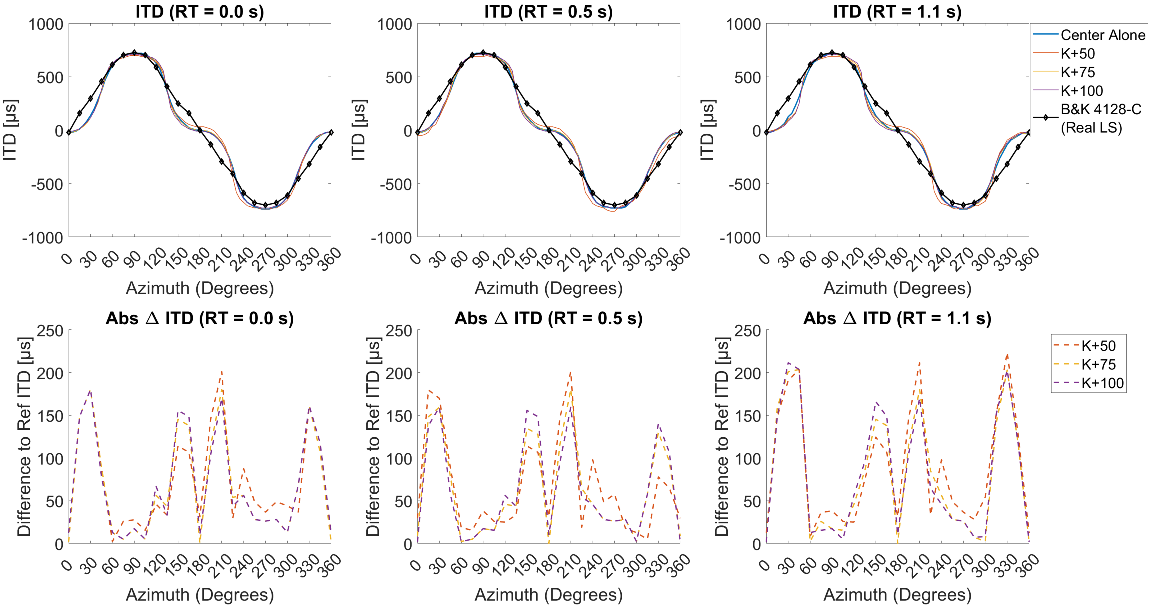


Figure 5.28: *ITDs and absolute ITD differences as a function of angle for multiple configurations with (colored lines) and without a second listener (black line).*

There was a small difference (≈ 15 µs) as the second listener is placed at the closest position (50 cm) considering rear and right angles. The absolute difference has a maximum of 201 µs, equivalent to approximated 15° in the source position (see bottom row of Figure 5.28).

5.4.3.2 Interaural Level Difference

Figure 5.29 presents the difference (Δ ILD) between the ILD computed from the BRIRs collected with and without a second listener inside the ring. The panel rows top to bottom show Δ ILDs for simulated rooms with RTs of 0, 0.5 and 1.1 s. The columns represent the different distances between the centred and second listener, from 50 to 100 cm, left to right.

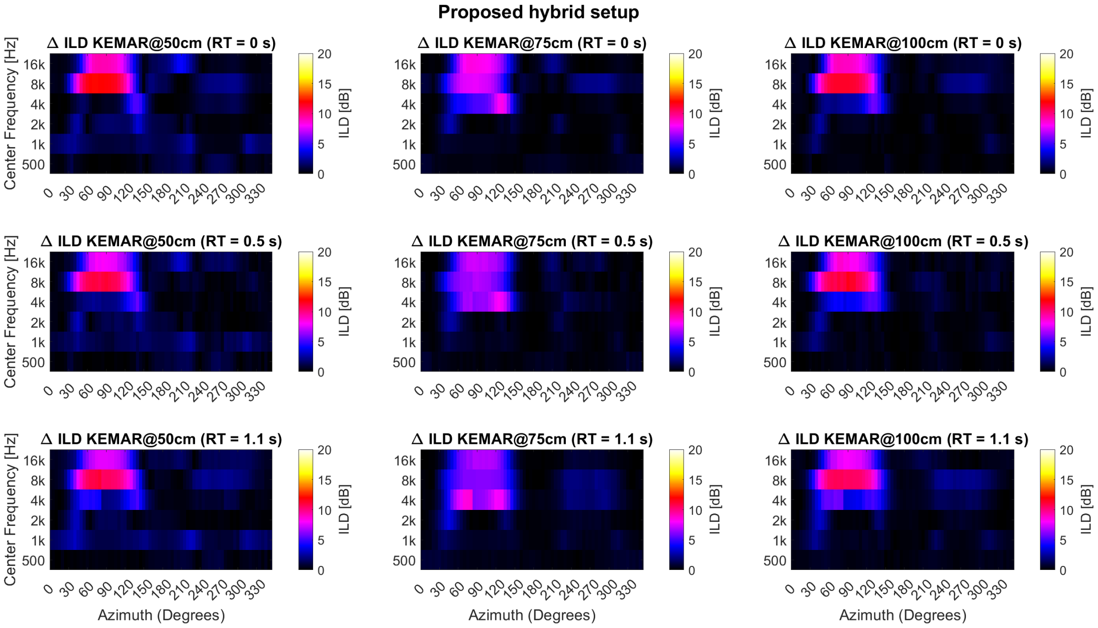


Figure 5.29: *Interaural level differences averaged octave bands as a function of azimuth angle for a HATS Brüel and Kjær TYPE 4128-C in the horizontal plane through a proposed Iceberg method on a four-loudspeakers setup.*

The results show that adding a second listener impacts the ILD for the angles shadowed by the second listener. The effect is more pronounced in the higher octave bands (8 kHz and 16 kHz), reaching approx. 14 dB, especially at the closest and farthest distances (50 cm and 100 cm). Although there is less impact of having a second listener at an intermediary distance (approx. 9 dB), it still produces noticeable ILD changes in the 4 kHz band. The Δ ILD produced by the presence of a second listener is expected as a result of natural acoustic shadowing. The analysis of ILD around the listener is also important in the hemifield opposite to the second listener (*i.e.*, 180-360°). The auralization methods that rely on the full set of loudspeakers to form the sound pressure (*e.g.*, Ambisonics) can present noise to the side with a free path as a physical object, preventing the sound wave from forming accordingly at the center (sweet spot).

Although the Iceberg method is partially composed by first-order Ambisonics, which is a method that requires all loudspeakers combined to form the appropriate auralization, the part that VBAP performs presents the sound only through the indicated quadrant, not requiring the other loudspeakers to be active. That extends the system's robustness with a limited number of loudspeakers and frequency limit (not being dependent on the Ambisonics order).

5.4.3.3 Azimuth Estimation

Figure 5.30 depicts the frontal azimuth angles estimated by the May *et al.* for pink noise inputs of 2.9 seconds. The pink noise is convolved to the recorded BRIRs. The BRIRs, by its time, were recorded utilizing files generated by the Iceberg auralization method in a four loudspeaker setup. The columns in the top row present graphs with the average estimated angle to a centered listener accompanied by a second lister at 50 cm (light blue curve), 75 cm (red curve), 100 cm (yellow curve) according to room simulation (denoted by reverberation time). The shaded area corresponds to the standard deviation. The top-left graph also presents the estimated angles to the real loudspeaker condition without virtualization (blue dotted line). Finally, the bottom row graphs show the differences between the estimated azimuth angle and the target angle (the estimated error).

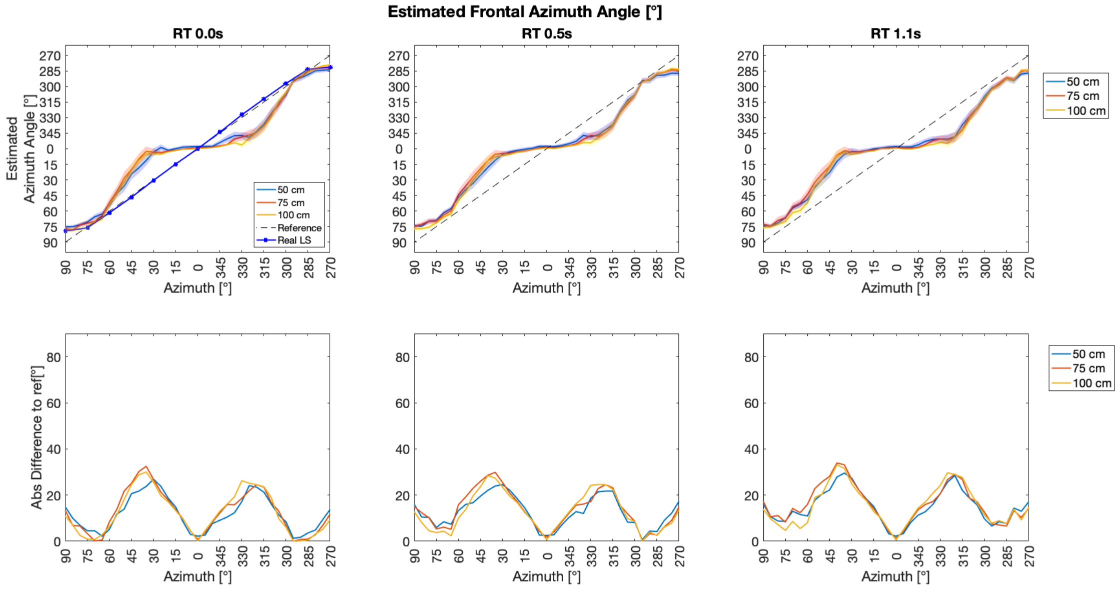


Figure 5.30: *Top line = Estimated localization error with presence of a second listener; bottom line = difference to reference. Columns depicts different RTs, line colors different second listener positions.*

According to the model's results, this difference reveals that the sound created via the Iceberg method and virtualized via a four-loudspeaker setup gives consistent localization cues even with a side listener inside the ring in the described positions. The median error is 9.9º, and the standard deviation is 8.8º. On the other hand, the distribution of these differences made clear that the setup of four loudspeakers has more difficulty accurately presenting the localization cues between the frontal loudspeakers, reaching up to 27º of mismatching to these positions (45º&315º). This result is in line with simulations from Grimm *et al.* [97].

The values obtained with the low number of speakers utilized (four) are in line with the expected in the literature [97] with a equivalent pattern [84]. Although it can raise a flag for experiments needing a more precise localization representation, the Iceberg method can improve simple setups' realism. Accordingly, it needs to be thoroughly investigated considering subjective listening tests, especially in the lateral angles.

Figure 5.31 presents the absolute differences between the estimated angles of arrival for the centered position alone and the centered accompanied by a second listener at 50 cm, 75 cm, and 100 cm in the three simulated rooms tested. These differences reflect the estimated influence of having the second listener inside the ring.

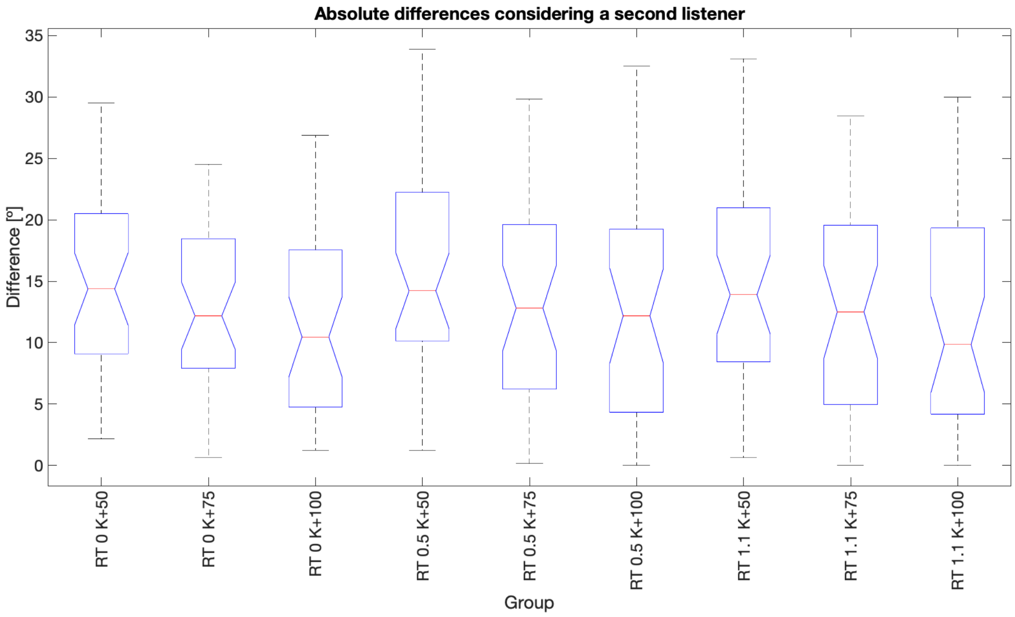


Figure 5.31: *Absolute difference to target in estimated localization considering the presence of a second listener and the reverberation time.*

The average error presents a slight increase with the proximity of the second listener. However, the effect is less perceivable in moderate RT. That fact suggests that the acoustic shadow is present.

An ANOVA analysis of the variance of the estimated absolute errors between the RT and KEMAR position groups was proposed. For the distribution with 8 degrees of freedom and a number of observations equal to 30, the tabulated value of the F Distribution of Snedecor on (p=0.05) is equal to 2.26. Thus, values greater than the tabulated F accept the null hypothesis that there is no significant difference between the means of the absolute error of the groups of angles $H_0 : \mu_i = \mu_j$. From the analysis, the results of the F statistic (presented in Table 5.2) H_0 is accepted in all groups.

Table 5.2: One way anova, columns are absolute difference between estimated and reference angles for different KEMAR positions and RTs.

Source	SS	df	MS	F	Prob F
Columns	746.5	8	93.3142	1.3755	0.2062
Error	21980.9	324	67.8423		
Total	22727.4	332			

Therefore, there is no statistical difference between the KEMAR positions for any of the evaluated RTs. That suggests the method's stability in this setup, even with a second listener considering the reverberations and positions tested.

The model has difficulty estimating the extreme lateral locations (90$^{\underline{o}}$ and 270), as even the actual loudspeakers could not reach this estimation. A comparison between the estimated angles with a listener alone in the center position acquired only with actual loudspeakers (without virtualization) and the listener in the center accompanied by a second listener acquired from virtualized files with the Iceberg model is presented in Figure 5.32.

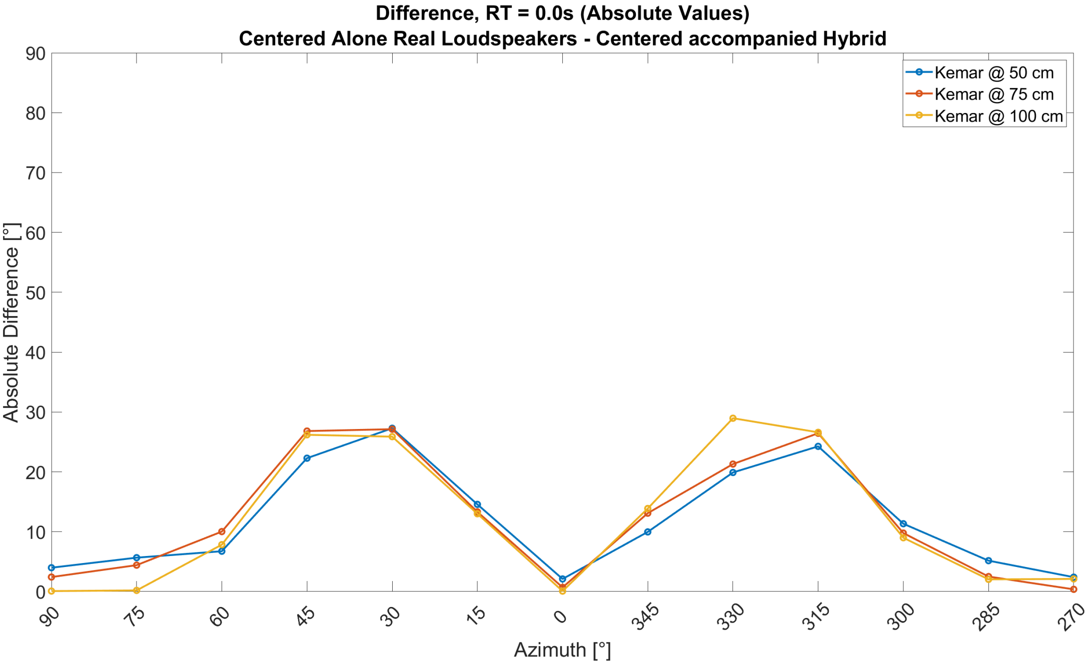


Figure 5.32: *Estimated error to RT=0 considering the estimation of real loudspeakers as basis.*

The data analyzed in this section suggests that by observing the indicated positions where the estimated difference can be significant, although comparable to similar methods listed in Table 2.2 experiments with equivalent requirements (e.g., Chapter 4) can benefit from applying the Iceberg method. The method will fairly reproduce sounds with the presence of a second listener and increase the sense of immersiveness while reproducing spatialized sound with only four loudspeakers. Subjective tests are needed to investigate further the system's spatial rendering performance.

5.5 Supplementary Test Results

A concern about virtualization processes is how reliable they are when an extra layer of signal processing is added to the experimental setup [97, 213]. That is, how the sound acquisition through a hearing device microphone and its signal processing would be affected by virtualization relative to a simple loudspeaker reproduction or the real life situation [98]. This section describes a comparison of the binaural cues with and without hearing aids. The RIRs were collected in the same positions as presented in Section 5.4.1 and Section 5.4.2. Further, inspired in the study of Simon *et al.* [276] the robustness of the virtualization setup outside the sweet spot was evaluated.

Oticon Opn S1 Mini RITE hearing aids with open domes were coupled to each ear of the HATS manikin (See Figure 5.33). Modern hearing devices like these present a series of signal processing features that can affect the analysis depending on the brand or model. In order to ensure compatibility of the results to other devices, specific features were not enabled. The devices were programmed in the fitting software to compensate for the hearing loss of the N3 moderate standard audiogram [35]; its beamformer sensitivity was set to omnidirectional, and the noise reduction set to off. The hearing level of the audiogram is presented in Table 5.3. The opened domes were chosen to set the virtualization system's most difficult signal mix condition. The signal played through the system is not attenuated as the ear is not occluded. The amplified signal from the hearing device is received at the eardrum (microphone) 8.1 ms after the original signal.

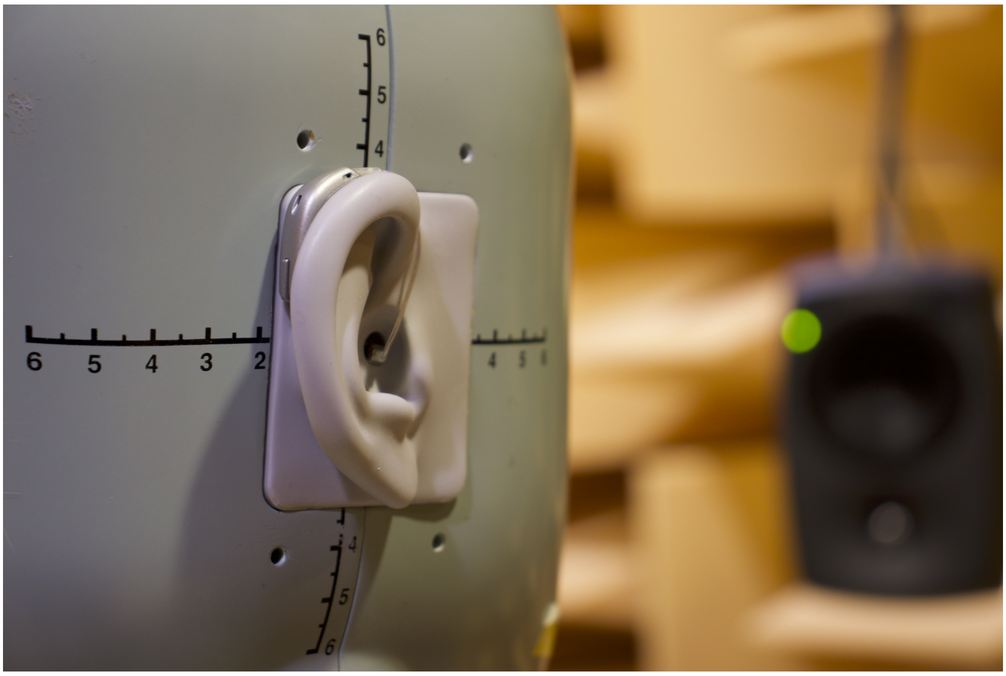


Figure 5.33: *HATS wearing the Oticon Open-S 1 Mini RITE.*

Table 5.3: Hearing Level in dB according to the proposed Standard Audiograms for the Flat and Moderately sloping group [35].

Nº	Category	Frequency									
		250	375	500	750	1000	1500	2000	3000	4000	6000
N3	Moderate	35	35	35	35	40	45	50	55	60	65

5.5.1 Centered Position (Aided)

The system was tested by measuring the BRIR with the listener (manikin) in the center. The calculated binaural cues are presented at incidence angles separated by 15° at 1.35 m from the listener.

5.5.1.1 Interaural Time Difference

The ITD results (See Figure 5.34) in the aided condition were very similar to the unaided condition 5.4.1.1.The maximum absolute difference found is 170 µs, representing a mismatch around 15º on the given angle (same as previously measured unaided ITD difference). The calculated average difference is 67 µs, representing a difference of around 7º in localization.

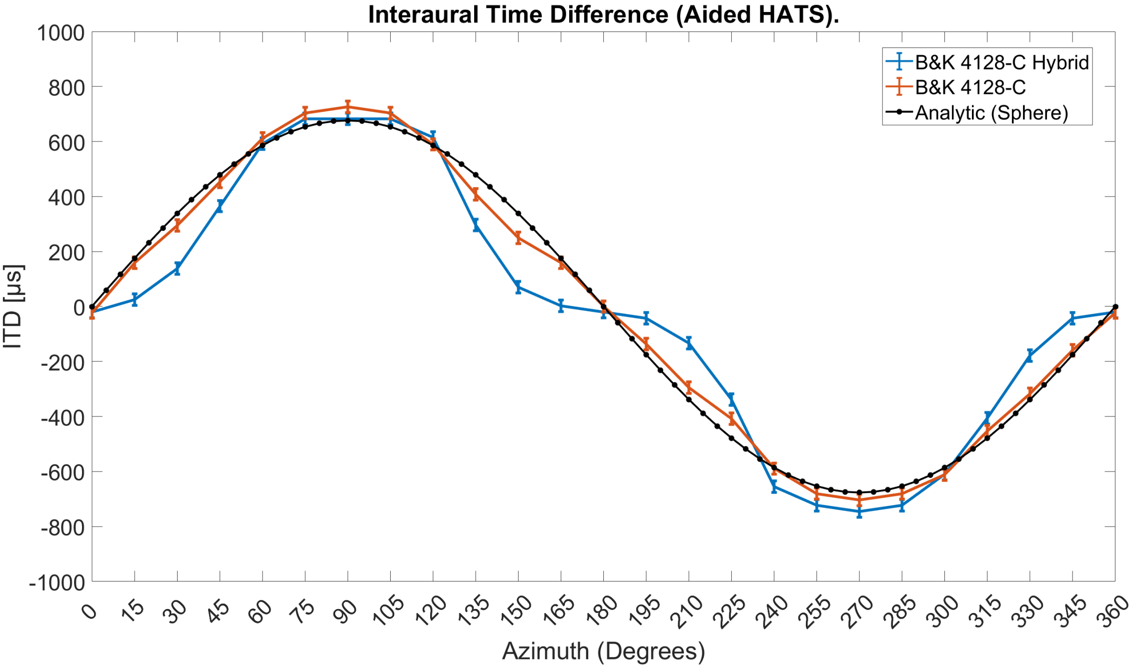


Figure 5.34: *Interaural Time Difference under 1 kHz at the proposed Iceberg method as a function of azimuth angle for a HATS Brüel and Kjær TYPE 4128-C in the horizontal plane wearing a pair of hearing aids in omnidirectional mode (blue line). The red line is the ITD results with real loudspeakers (without virtualization). The black line represents the analytical ITD values.*

Figure 5.34 depicts higher differences concentrated on specific regions: angles around ±30º to front and back. The similarity to the unaided condition is expected as the devices are not blocking the sound wave or increasing the path more in one ear than the other (*i.e.*, there is only a static group delay added to the system). Therefore the sound reaches the HATS microphones with hearing aids proportionally as in the previous unaided condition.

5.5.1.2 Interaural Level Difference

Figure 5.35 shows the effect on the ILD to the centered position. Although in higher octave bands (8 and 16 kHz) the difference between ILD on an aided HATS with real loudspeakers to the aided HATS utilizing the Iceberg method is a bit larger than unaided (See Figure 5.18), the effect on the 2 kHz band is considerably smaller. That can be due to the added delay in the signal, which can diminish the possible comb filtering by the Iceberg method at this specific frequency region, especially for the angles between two loudspeakers (where there is a larger distance between real loudspeakers and the virtualized sound source).

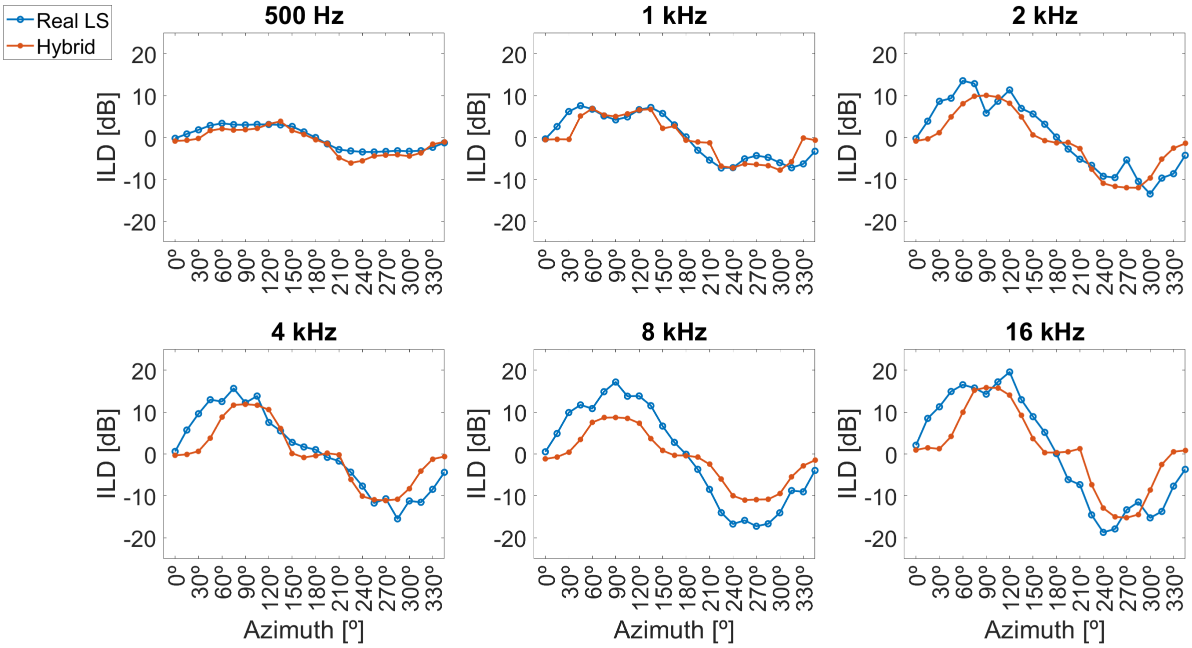


Figure 5.35: *Interaural Level Differences as a function of octave-band center frequencies. Angles around the central point.*

5.5.1.3 Azimuth Estimation

Figure 5.36 presents the angles estimated using the May and Kohlrausch's [182] model for files auralized with the Iceberg method and virtualized over the four-loudspeakers setup (blue curve), angles estimated for binaural files acquired without virtualization with real loudspeakers (red curve), and the reference (dotted black). The presented model's result is in line with the analysis of the binaural cues supporting the assumption of the worst localization accuracy around ±30º (30º and 330º in Figure 5.21). Some differences bigger than the standard deviation are noted between different RT, especially close to the lateral angles (90º and 270º). The results suggest that the added reverberation can negatively impact on localization accuracy.

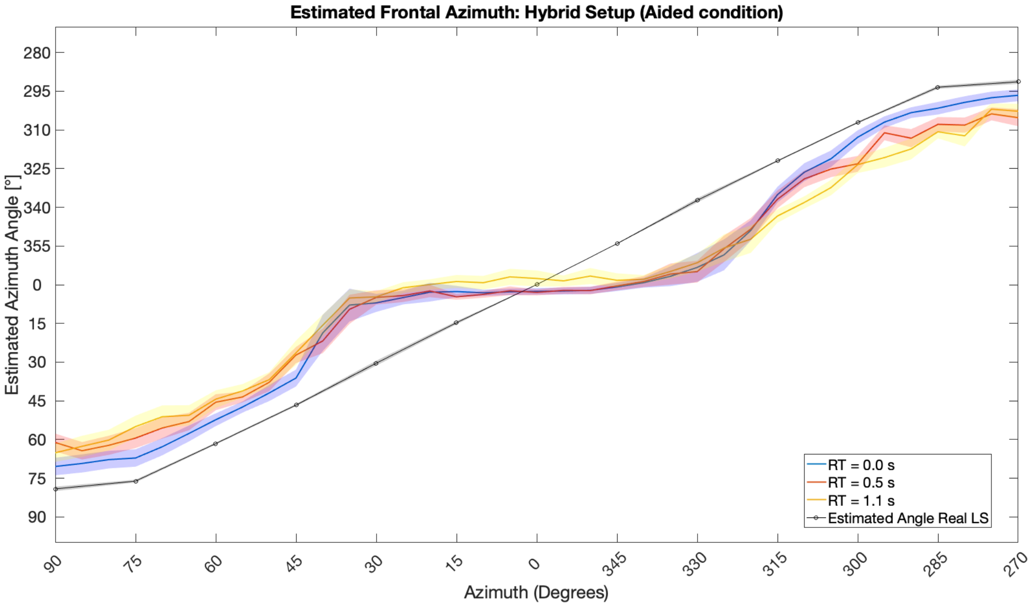


Figure 5.36: *Iceberg method: Estimated azimuth angle (model by May and Kohlrausch [182]), HATS centered and aided.*

Also, according to the figure, the virtualized sound tends to have more difficulty separating from the frontal angle (0º) denoted by the flat lines from 30º up to 340º. Figure 5.37 depicts the boxplot diagram of the absolute differences grouped by RT.

An ANOVA analysis of the variance of the estimated absolute errors between the RT and position groups was proposed. For the distribution with 2 degrees of freedom and a number of observations equal to 30, the tabulated value of the F Distribution of Snedecor on (p=0.05) is equal to 3.32.

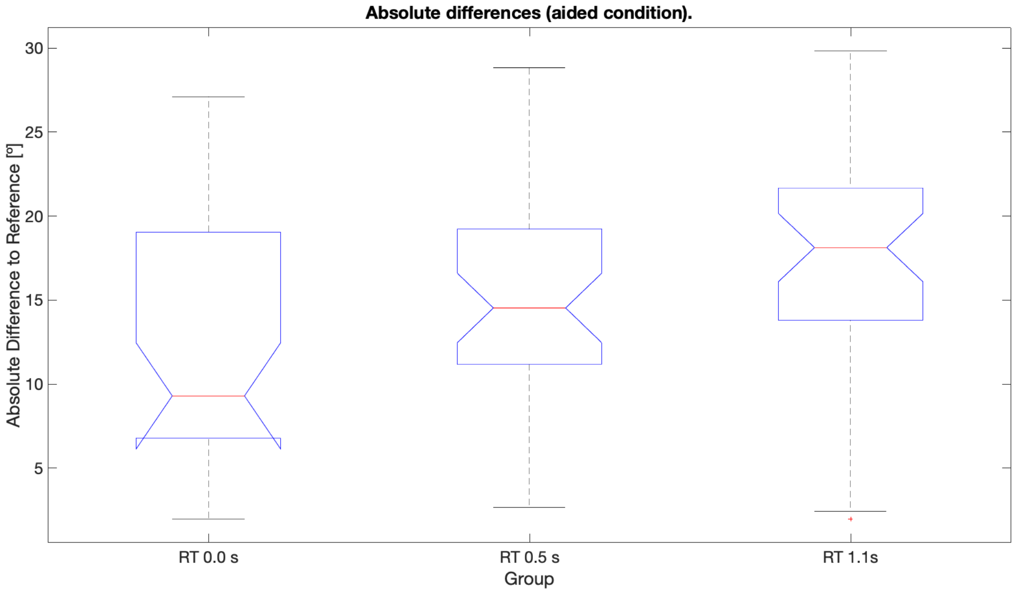


Figure 5.37: *Absolute difference to target in estimated localization in aided condition in aided condition considering different RTs.*

Thus, values greater than the tabulated F deny the null hypothesis that there is no significant difference between the means of the absolute error of the groups of angles $H_0 : \mu_i = \mu_j$. From the analysis, the results of the F statistic (presented in Table 5.4 H_0 is rejected and the hypothesis alternative $H1 : \mu_i \neq \mu_j$ is accepted (F=5.68).

Table 5.4: One way anova, columns are absolute difference between estimated and reference angles for different positions and RTs.

Source	SS	df	MS	F	Prob F
Columns	520.77	2	260.386	5.68	0.0045
Error	4947.29	108	45.808		
Total	5468.06	110			

To identify in which sets of means the discrepancy is statistically significant, Tukey's multiple comparison test was performed and the result is shown in Figure 5.38.

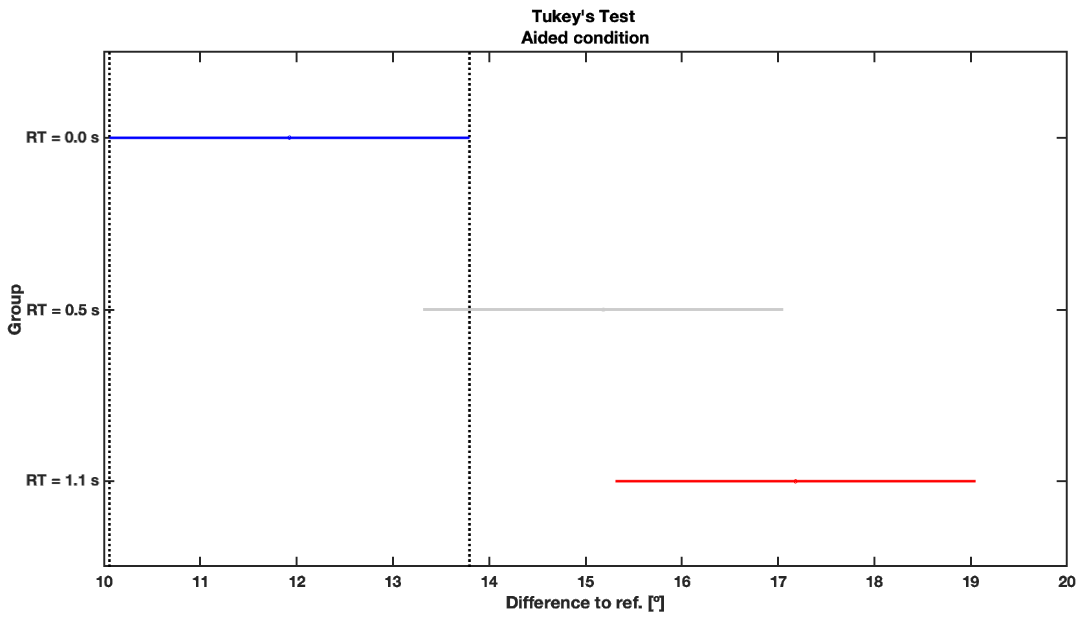


Figure 5.38: *Tukey test to compare means in aided condition. Group mean in RT 1.1s presented significant difference from mean in group RT 0.0 s*

This reflected a trend towards an increase in the estimated location error when there is signal amplification through the hearing aid, which did not occur in the similar condition without the aid seen in Section 5.4.1.3.

5.5.2 Off-center Positions (Aided)

The listener was moved from the center position to simulate a displaced test participant wearing hearing aids. The BRIRs were measured in the positions described in Section 5.3.3, and the results were analyzed in this section.

5.5.2.1 Interaural Time Difference

Figure 5.39 presents the ITD for the different angles around the listener as the listener is displaced to different positions according to the specified grid.

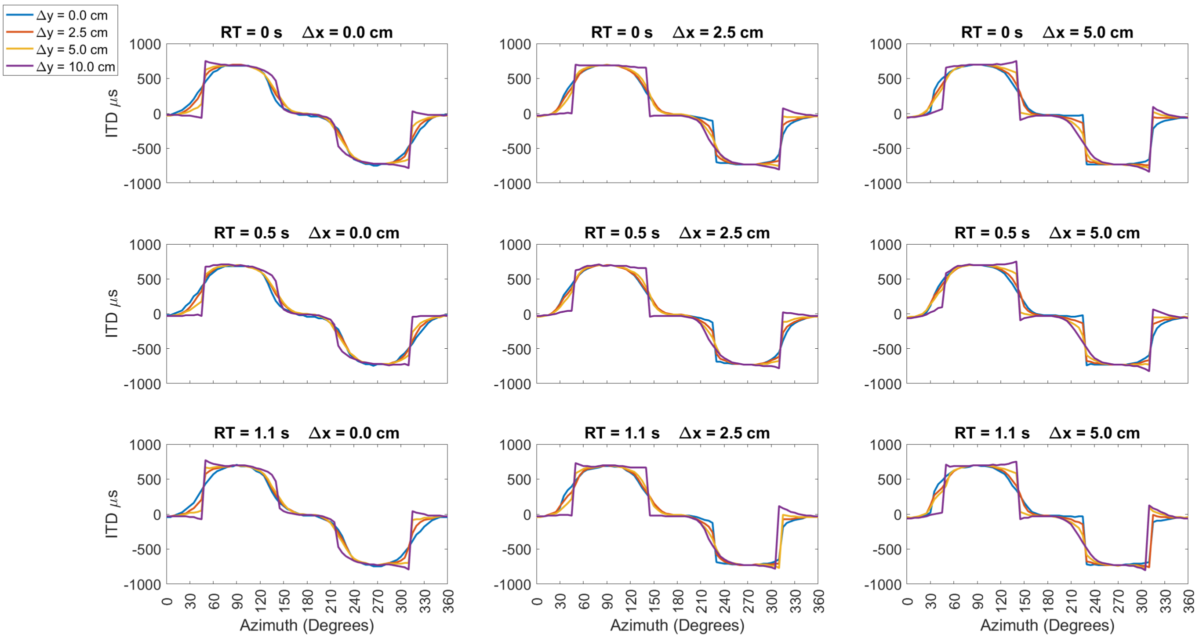


Figure 5.39: *Interaural Time Differences as a function of octave-band center frequencies. Angles around the central point.*

when it moves 5 cm to front it starts to blur more the correct ITD for the frontal angles. Especially around ±45 degrees in the frontal hemisphere where the ITD indicates that the sound is coming from 90º or 0º angles. Further than this distance, also the rear ±45 are affected, pointing to the break of the panning illusion. Compared to the unaided condition (Section 5.4.2.1), this condition is slightly more sensitive to displacements

Although the ITD analysis is angle-dependent, the results in the Table 5.5 indicates that the displacement limitations can be overall mapped to indicate the maximum distance. Table 5.5 shows the maximum ITD difference according the displacement. Although the ITD analysis is angle-dependent, the results The maximum value difference can indicate the tendency of the ITD shape to be squared, representing no virtualization. That may help to identify displacement limitations can be overall mapped to indicate the maximum distance. The squared behavior occurs when the sound of one individual speaker is the main pressure contribution, arriving too early to one of HATS ears because of the HATS's position.

Table 5.5: Maximum ΔITD relative to the center position according to displacement, lines refer to lateral displacement and columns refer to frontal displacement.

RT = 0.0 s

Displacement [cm]	0.0	2.5	5.0	10.0
0.0	0 [µs]	88 [µs]	182 [µs]	374 [µs]
2.5	233 [µs]	239 [µs]	364 [µs]	472 [µs]
5	317 [µs]	353 [µs]	399[µs]	566 [µs]

RT = 0.5 s

Displacement [cm]	0.0	2.5	5.0	10.0
0.0	0 [µs]	97 [µs]	229 [µs]	386 [µs]
2.5	213 [µs]	157 [µs]	313 [µs]	472 [µs]
5	317 [µs]	282 [µs]	389 [µs]	566 [µs]

RT = 1.1 s

Displacement [cm]	0.0	2.5	5.0	10.0
0.0	0 [µs]	140 [µs]	299 [µs]	341 [µs]
2.5	236 [µs]	310 [µs]	372 [µs]	437 [µs]
5	283 [µs]	372 [µs]	380 [µs]	520 [µs]

In this case frontal displacements up to 2.5 centimeters are not presenting the square behavior and a maximum ΔITD, of 140 µs (RT= 1.1 s), considering the centered position as a reference. Lateral movements are more affected, starting to present the squared behavior in the transition angles between the rear loudspeaker and the right angle (230º) and the right loudspeaker and the front (310º). This pattern seems not be RT dependent, which is expected due to ITD's nature.

5.5.2.2 Interaural Level Difference

Figure 5.40 presents the ILD, considering the simulation of an anechoic environment (RT = 0s), on 24 angles around the listener as the listener is displaced to different positions according to the specified grid. Compared to the normal condition, although it presents the same pattern, the aided condition has lesser differences between ILDs across more angles and frequencies.

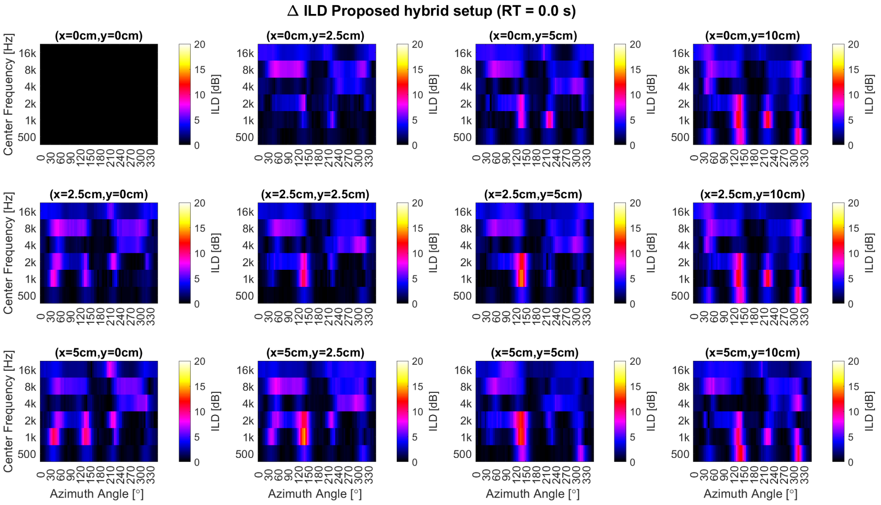


Figure 5.40: *Difference in ILD as a function of azimuth angle for a B & K 4128-C. Iceberg method, horizontal plane in a 4 loudspeakers setup (RT = 0.0 s).*

The differences were also lessened as the RT increased, as can be seen in Figures 5.41 and 5.42. This result shows that increasing the reverberation can positively affect the ILD error in off center positions (reducing the differences to the ILD in the center).

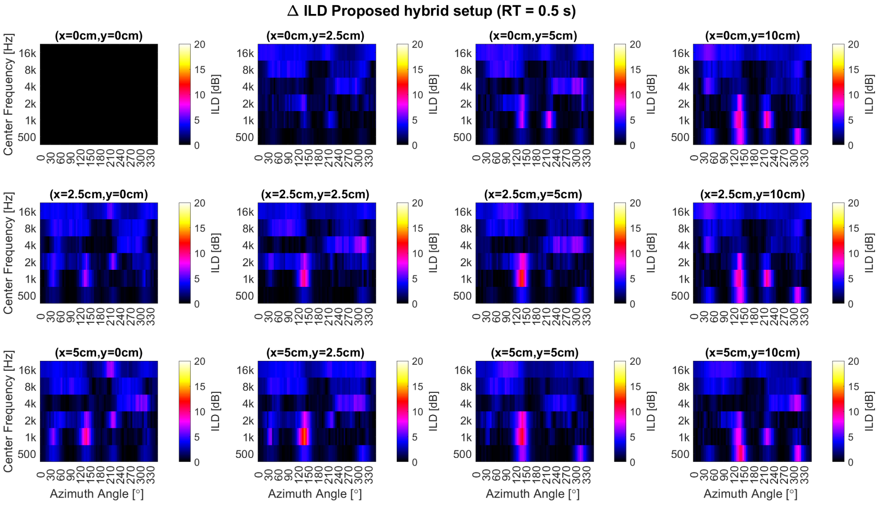


Figure 5.41: *Difference in ILD as a function of azimuth angle for a HATS Brüel and Kjær TYPE 4128-C in the horizontal plane through a proposed Iceberg method on a four-loudspeaker setup RT = 0.5 s.*

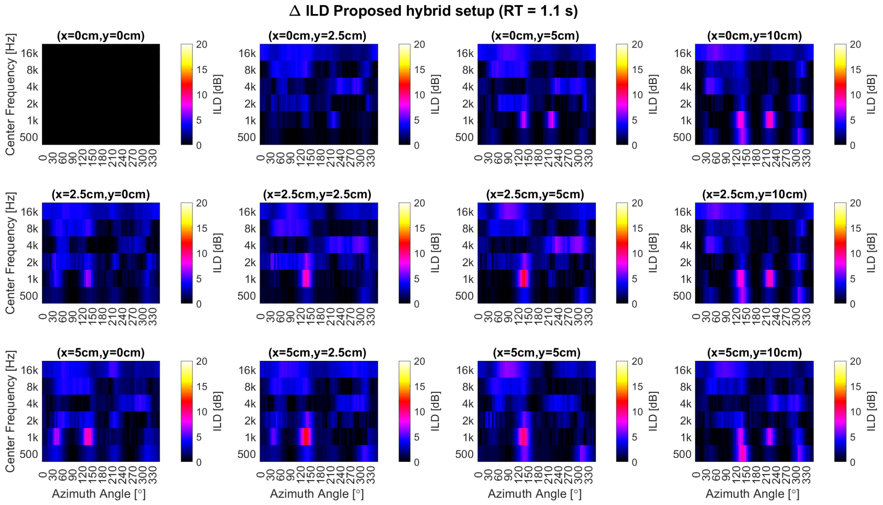


Figure 5.42: *Difference in ILD as a function of azimuth angle for a HATS Brüel and Kjær TYPE 4128-C in the horizontal plane through a proposed Iceberg method on a four-loudspeakers setup RT = 1.1 s.*

5.5.2.3 Azimuth Estimation

Figure 5.43 presents the estimated azimuth angle [182] according to the position of the listener. The different RTs are represented by the line colors in the graphs (blue = 0.0s, red = 0.5s, yellow = 1.1s). The results demonstrate that the Iceberg method present less accuracy in the reproducing sound in frontal angles, especially ±30° and ±330°. The lateral discrepancy is smaller and also noted with real loudspeakers, what can imply that the model used has some difficulty to assess that region.

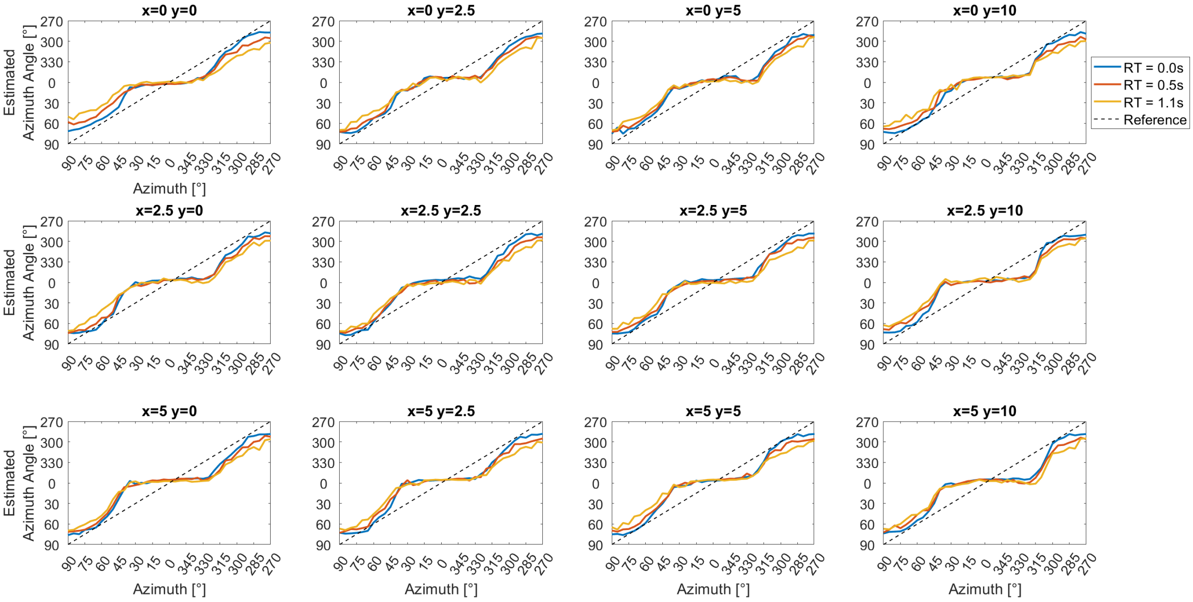


Figure 5.43: *Estimated frontal azimuth angle at different positions inside the loudspeaker ring as function of the target angle (aided condition). Model by May and Kohlrausch [182].*

According to the model's results, to an aided listener, the localization error is up to 30 degrees within a frontal or lateral displacement of 5 centimeters. In the case of 10 cm of displacement, the virtualization will fail, presenting the squared behavior on the contralateral side. The increase of reverberation tends to maintain the maximum error magnitude, although increasing the spread to more angles. That means the lateral side close to the loudspeaker will present the sound source position less in the desired position but more in the loudspeaker's physical position. Medium reverberations are less affected by displacement, meaning extreme cases should drive extra care with listener positioning.

5.6 Discussion

This chapter proposed a new hybrid auralization method (Iceberg) for a virtualization setup composed of 4 loudspeakers at a 1.35 m distance from the center. This setup is a relatively limited one intended as a feasible alternative to the much more expensive and complicated arrangements proposed and used in the reviewed literature (See Section 2.3).

The innovation factor of the Iceberg method is the usage of a room acoustic parameter, called center time, used to compute the transition point between early and late reflections. The Iceberg's channel mixing and distribution automation are generalized to any RIR collected or converted in Ambisonics' first order.

Implemented in MATLAB, the Iceberg auralization algorithm can generate .wav files that were virtualized in a setup with four loudspeakers (90-degree spacing around the listener). Three simulated sound scenarios were predefined and simulated using acoustic modeling software generating RIRs in Ambisonics format. The setup provided appropriate reverberation times even when the listener was away from the center position. Regarding binaural cues, in the optimal position, the maximum deviation in ITD was 170 µs, corresponding to a shift of approximately 15º for sources around ± 30º in front and back of the centered listener. The considerable distance between loudspeakers is the most likely cause of this deviation.

In contrast with Chapter 3, the Iceberg method could not reproduce the ITDs with the same accuracy as VBAP in the sweet spot position. However, it presented a better performance than Ambisonics. The high accuracy of VBAP can be attributed to the number of loudspeakers, 24, which lessen their physical distance, therefore, its maximum error. However, even with the bigger number of loudspeakers, Ambisonics was truncated at first-order, thus not having the benefit of more sound sources.

There were also deviations in ILD, mainly at the same angles. However, the ILD deviations were most significant in the 2 kHz octave band. The actual effects of this difference in signals that encompass these frequencies should be further investigated in validation tests. The ILDs also denoted patterns with better representation through VBAP than first-order Ambisonics in Section 3.3.2.2. The Iceberg method with four loudspeakers presents a pattern with ILDs closer to actual loudspeakers than pure Ambisonics, but again not as accurate as VBAP in 24 channels. This is characterized mainly to a difference in the 2 kHz octave band. That needs further investigation and consideration in experiments requiring ILDs accuracy at that frequency band.

Overall the results for the binaural cues reproduction via the Iceberg method in four loudspeakers are better than a pure Ambisonics first-order but worse than VBAP (considering 24 loudspeakers). Therefore the Iceberg method can be considered an option when the number of loudspeakers is limited or the need for a sense of realism is higher. The Iceberg method combines relative accuracy with a sense of immersion. The maximum estimated localization uncertainty was around 30 degrees to the Iceberg method in the minimal configuration of four loudspeakers. The different amounts of reverberation tested did not impact the results. Although the estimated localization was imperfect, the method's performance was in line with similar VBAP implementations [97]. The results were similar to the aided condition, with ILDs indicating better cue reproduction in the 2 kHz octave band. This improvement was not translated into a better-estimated angle, getting about the same results. A slight variation was identified, and a statistically significant difference was found between different RTs, especially in the lateral angles. This deviation needs to be further evaluated with other models and also with subjective validation, especially as the model results presented unexpected results in these angles for non virtualized sound sources.

A second listener was introduced at the side of the primary listener while maintaining the listener in the optimal position to simulate a condition where there is a need for social interaction or presence in a test. In this case, the binaural cues provided by the Iceberg auralization method virtualized in a four loudspeakers setup were compared to a baseline without virtualization (actual loudspeakers). Also, the model of May and Kohlrausch [182] was applied to predict frontal localization accuracy. Three distances were tested with the three simulated rooms (different RTs). There was the expected acoustical shadow at angles blocked by the second listener but not to the remaining sound source locations around the listener. That can be considered a measure of rendering robustness; the second listener did not break the virtualization of binaural cues by scrambling the sound pressure summation.

Regarding sub-optimal positions to unaided HATS, the Section 5.4.2 presented surprising results. The the virtualized effect was affected differently to the displaced HATS according to the amount of RT. In the dry RT (0 s) displacements up to 5 cm when moving forward did not present the undesired effect. The mild reverberation (0.5 s) got the undesired effect only with 10 cm from the center and the large RT up to 5. The large reverberation (1.1 s) was heavily affected presenting the squared behavior on all off center displacements. Lateral movements were affected in similar way for all the RTs tested, presenting the effect on displacements further than 3.5 cm from center. The ILDs presented the shadowing effect as expected, increasing the distortions with the distance and reducing it with the increase of reverberation. The combination came to an estimated azimuth angle in practice not dependent of RT and an error off $\approx 30^{\circ}$ with displacements up to 2.5 cm. As the displacement from center increases the maximum estimated error increases but also moved, meaning that the virtualization is affected but still would produce a virtualized effect. In the aided condition (Section 5.5.2) the off-center ITDs indicated a maximum frontal displacement on the Iceberg method under 10 cm from the center in unaided condition and under 5 cm in aided condition. The ILDs were also impacted by distance, but to a lower extent, while the different RT affected less the ITDs and more the ILDs. The ILDs present a smearing behavior, lowering the error, with higher RT. That behavior suggests an equivalent compensation on the error predicted by the model. Within that distance limit, the maximum error predicted was around 30 degrees for all RTs, agreeing at the end with the non-aided condition.

When the listener is away from the center, the Iceberg method virtualized using the four-loudspeaker setup increases the deviations in binaural cues compared to the cues at those sound-source angles, with a near complete loss of gradient in cues (*i.e.*, either zero or extreme values) occurring when the listener was 10 cm in front of the center position. ITDs for this condition revealed minor differences across the tested reverberation conditions. The values indicated that the files created by the method and reproduced on the four loudspeakers configuration produce similar ITDs as the baseline condition, having the week point in the 30 degrees. The absolute ITD values align with similar experiments found in the literature for VBAP configurations without a second listener [241]. The acoustic shadow is indicated by an increase of the Delta ITD difference around 270 degrees (left side), especially to the closest position, similar to the finding with pure VBAP Chapter 3. Also, the difference in ILDs (ΔILDs) showed that the presence is well captured in higher frequencies. All RTs conditions and positions demonstrated to capture differences in ILD to the left side of the mannequin. A ΔILD is expected as a result of natural acoustic shadowing produced by the presence of a second listener. The benefit of the Iceberg method is that the VBAP is not limited in frequency by aliasing in the higher frequencies and does not require all to be active loudspeakers simultaneously as pure Ambisonics. The way the division is done in the Iceberg method brings the Ambisonics' responsibility to the time domain, defining the method as more natural to physical presence between loudspeakers and the listener. That extends the system's robustness with a limited number of loudspeakers and frequency limit (not being dependent on the Ambisonics order).

The predicted error for a second listener compared to the Iceberg baseline condition (listener centered alone). The method presents a deviation of around 10 degrees in all RTs when the second listener is in the shoulder-to-shoulder situation, the closest position (50 cm). As the difference is at the second listener position, it is possible to argue that the Iceberg energy balance is advantageous, not entirely depending on the four loudspeakers' summation.

Therefore, compared to VBAP or Ambisonics, the Iceberg method is a suitable option in terms of localization that adds the benefit of immersiveness in a modest hardware.

5.6.1 Subjective impressions

The auralization was compared by the author and his supervisor to VBAP and Ambisonics in subjective listening sessions. The experiment was not performed systematically, as the Covid-19 emergency rules imposed a series of restrictions and these impressions are the initial opinion. The speech signals were auralized via Iceberg, VBAP, and Ambisonics and reproduced in an anechoic room. The rooms were simulated in Odeon software v.12 with reverberation time equivalent to 0.5 and 1.3 seconds. Both agreed that the sound direction from VBAP is easily identifiable but with poor immersiveness, as all the reverberation came from a specific side (2 loudspeakers). Ambisonics offered a more immersive experience with all loudspeakers active simultaneously, but the localization was very difficult; a "blurred" position seems to be a trending description. The Iceberg system provided a sound localization close to VBAP while maintaining the immersiveness.

The Iceberg method, upon a trade-off on spatial localization, allows for the reproduction of sounds that can be easily manipulated regarding sound-source direction, sound pressure level, reverberation time, and simultaneous sound sources. That makes it possible to create or reproduce specific virtual sound scenarios with high reproducibility. Thus, researchers can conduct auditory tests with increased ecological validity in spaces that usually do not count with numerous loudspeakers, as is common in clinics, universities, and small companies. Notwithstanding these benefits, some limitations challenged the method with a small number of loudspeakers. These limitations impose some constraints on its use in terms of the spatial localization of sound sources.

5.6.2 Advantages and Limitations

A fundamental advantage of the proposed Iceberg method is the minimum number of loudspeakers required (four). Furthermore, its compatibility with any RIR in low- or high-order Ambisonics already collected. That is possible as an RIR in HOA can be easily scaled down to first-order Ambisonics and its sound spatial properties composed with any given sound via the algorithm [237]. Furthermore, an essential part of the method's definition and an additional advantage is the automation of the definition of the amount of energy from the RIR that corresponds to the specific auralization technique. That automation performed by the Central Time room acoustics parameter allows a smooth transition between the direct sound and early reflections portion and the late reflections of the RIR, resulting in a potentially more natural sound while maintaining control over the incidence direction.

The auralization method is designed for a virtualization setup of 4 loudspeakers. However, it is possible to use it with more loudspeakers, reducing the eventual limitations on spatial accuracy. Furthermore, although not within the scope of this thesis, the method, using the VBAP technique, would allow the possibility for dynamically moving sound sources around the listener.

5.6.3 Study limitations and Future Work

The initial aim of this study was to investigate the correlation between objective parameters related to spatial sound, particularly those psychoacoustically motivated by auralization methods, and subjective responses to these methods. However, due to the Covid-19 pandemic, tests with participants were not possible due to the risk of infection as mandated by government rules. As a result, the study is limited to verifying objective parameters. Therefore, section 5.5 was included to explore the system capabilities within a relevant context for hearing research, although without subjective tests involving participants. In future work, structured validation with participants would be of value to the field, allowing for adjustments and the measurement of the effectiveness of this method in real-world auditory tests. Additionally, future implementations of this method could include improvements such as guided sound source movements around the listener, with simultaneous updates of VBAP and Ambisonics weights defined by time constants, and the ability to pan with intensity using techniques such as Vector-Based Intensity Panning (VBIP), which could be tailored to specific cases with different loudspeaker arrangements or stimuli frequency content and potentially merged with VBAP depending on the type of stimuli and specific frequencies.

5.7 Concluding Remarks

Tests that require hearing aids can be performed, considering some constraints, utilizing the proposed Iceberg method. These tests aimed to verify the impact of the auralization method through a simple setup (four loudspeakers) to the virtualized spatial impression by analyzing the binaural cues and their deviations from actual sound-sources loudspeakers. This is an important step, although not discounting the importance of validation with test participants.

To a centered listener, the verified deviation in binaural cues presented limitations of around 30º degrees in localization (through ITD) with reasonably matching ILDs. The system's reliability is compromised as the listener is moved out from the sweet spot, but less so than when unaided, possibly due to comb filtering or the addition of compression into the signal path. Small movements up to 2.5 cm generated errors within a JND, meaning they likely would not be perceived as distortions or artifacts. Thus, tests with people that require sound sources positioned in spaces larger than 30º can benefit from this Iceberg method that incorporates spatial awareness and immersiveness.

Chapter 6

Conclusion

Throughout the course of this study, a new auralization method called Iceberg was conceptualized and compared to well-known methods, including VBAP and first-order Ambisonics, using objective parameters. The Iceberg method is innovative in that it uses TS to find the transition point between early and late reflections in order to split the Ambisonics impulse responses and adequately distribute them. VBAP is responsible for localization cues in this proposed method, while Ambisonics contributes to the sense of immersion. In the center position, the Iceberg method was found to be in line with the localization accuracy of other methods while also adding to the sense of immersion. Also, a second listener added to the side did not present undesired effects to the auralization. Additionally, it was found that virtualization of sound sources with Ambisonics can implicate limitations on a participant's behavior due to its sweet spot in a listening-in-noise test. However, these limitations can be circumvented and extended to Iceberg, resulting in subjective responses that align with behavioral performance in speech intelligibility tests and increasing the localization accuracy.

6.1 Iceberg

In the previous chapter, we conducted a thorough analysis comparing the performance of the Iceberg method to the results presented in Chapter 3 and the relevant literature in Chapter 2. This comparison included evaluating the Iceberg method's performance at the center position, at various off-center positions, and in the presence of a second listener. The results showed that the Iceberg method was able to provide the designed overall reverberation times of 0 seconds, 0.5 seconds, and 1.1 seconds across all measured positions. Additionally, the differences between the reverberation times were below the JND 5% threshold.

When comparing values to the ones obtained with a HATS in the center without virtualization, it is noteworthy that the Iceberg method uses 20 fewer loudspeakers than this VBAP and Ambisonics configuration. The Iceberg method exhibited lower accuracy in reproducing ITDs at the sweet spot position than VBAP, but it performed better than first-order Ambisonics. We also observed detrimental deviations in ILDs, with values exceeding 4 dB, particularly at the same angles as the ITDs. The most significant ILD deviations occurred in the 2 kHz octave band, which could influence the perceived localization accuracy. Further investigation through validation tests is necessary to fully understand the extent of these differences between the methods. Regarding overall binaural cue reproduction, the Iceberg method using four loudspeakers was superior to pure first-order Ambisonics but less accurate than VBAP with 24 loudspeakers. The Iceberg presented a maximum estimated localization error of around 30 degrees for angles plus minus 40 degrees from the center while the listener is centered. Although this magnitude matches the similar methods in Table 2.2, the binaural cues were pointed to a lower estimate (around 15 degrees). Therefore further studies with perceptual evaluation are highly encouraged. In the Aided condition, we observed that the ITD was not affected at the center position, and the ILD was closer to the VBAP condition with 24 loudspeakers. However, this improvement was not reflected in the model estimate, which still showed maximum deviations of around ±30°.

At off-center positions, the Iceberg method showed slight variations in localization estimates, particularly in lateral angles, which were found to be statistically significant when comparing different reverberation times. This variation is likely due to the method's spatial limitation, known as the sweet spot, as discussed in the Chapter 2. When the reverberation time was 0 s or 1.1 s, the sweet spot was more limited in terms of displacement from the center (up to 3.5 cm). This means that these conditions were more prone to breaking virtualization when sound sources were virtualized on the contralateral side of the displacement. In contrast, the mild condition (0.5 s) maintained this up to 5 cm. A sweet spot is generally smaller in first-order Ambisonics compared to VBAP with a 24 loudspeaker setup, as identified in Chapter 3. However, it is important to note that objective parameters may not always correspond directly to subjective impressions. Despite this, the Iceberg method with four loudspeakers was found to perform similarly to VBAP (with 24 loudspeakers) in terms of binaural cue reproduction. The model estimates also showed that, within a combined displacement of up to 3.5 cm in both lateral and frontal directions, the maximum error would be less than 30 degrees, indicating the presence of virtualization (*i.e.*, the sound being physically composed of more than just the nearest speaker). It is therefore recommended to evaluate this deviation further using other models and subjective validation tests.

The results in Section 5.4.3.3, the condition with the listener in the center, showed that the presence of a second listener did not negatively affect the performance of the Iceberg method in all conditions of reverberation tested. No statistical difference in the means of estimated error was identified when considering the three RT conditions and the three KEMAR positions. The binaural cues errors followed the same trend as the Alone version, meaning that ITDs pointed to an error around 15 degrees, but with ILDs having absolute values with differences exceeding 4 dB (JND), which can probably explain the 30º error estimated by the model in the worst position (*i.e.*, the angle of the virtualized sound source at ±45º). Based on these results, the Iceberg method can be viable for virtualization setups with limited loudspeakers or when a higher sense of realism is desired.

6.2 General Discussion

In this work, we explored the use of auralization methods in hearing research as a means of improving the ecological validity of acoustic environments. The use of virtualized sound fields has become increasingly popular in laboratory tests. However, it is essential to understand the limitations of these methods in order to ensure unbiased results [97]. Our literature review (Chapter 2) identified the need for auralization methods that can be implemented in smaller-scale setups, and our initial evaluations focused on the spatial accuracy of several fundamental auralization methods, as well as their potential use in tasks involving multiple listeners. A collaborative study allowed us to test one of these techniques with real participants, and our findings highlighted both the limitations and potential improvements of using Ambisonics for conducting listening effort tests. Based on this experience and our knowledge of room acoustics and auralization, we proposed a new hybrid method called Iceberg, which combines the strengths of Ambisonics and VBAP and can be implemented using just four loudspeakers. This proposed method offers a low-cost option for auralization that could increase its adoption among researchers worldwide.

In Chapter 3, the VBAP and Ambisonics auralization methods were objectively characterized and compared in terms of binaural cues for the center and off-center positions. This investigation provided a foundation for combining the methods and further highlighted the strengths of each technique: localization in VBAP and immersiveness in Ambisonics. Objective parameters extracted from BRIRs and RIRs were examined for a single listener and in the presence of a second listener in the room. The results showed that the presence of a second listener did not significantly impact the performance of VBAP. At the same time, Ambisonics was less effective in reproducing the examined cues, especially with a second listener present. This information was crucial in developing the proposed Iceberg auralization method, which combines the strengths of both VBAP and Ambisonics to create a hybrid method suitable for use with simple setups such as four loudspeakers.

The results of the collaborative study described in Chapter 4 demonstrate the feasibility of using a virtualization method to deliver a hearing test with a certain level of spatial resolution and immersion across different room simulations and signal-to-noise ratios. This study suggests that virtualization methods have the potential to provide realistic acoustic environments for hearing tests, allowing researchers to vary the acoustic demands of a task and potentially improve ecological validity. Additionally, the significant correlation between participants' subjective perception of effort and their speech recognition performance highlights the importance of considering listening effort in hearing research. However, the limitations and potential solutions identified in this study also highlight the need for further investigation into virtualization methods in hearing research, including developing new auralization methods that address these limitations.

In Chapter 5, we presented the development of a new auralization method called Iceberg, which was designed to be compatible with small-scale virtualization setups using only four loudspeakers. Previous hybrid methods that combine Ambisonics and VBAP have been developed, but the innovative aspect of the Iceberg method is its approach to handling and combining the different methods to virtualize sounds while delivering appropriate spatial cues. This feature is achieved by identifying a transition point in the RIR using the Central Time parameter from the omnidirectional channel of an Ambisonics RIR. This automated process allows the user to input any Ambisonics RIR, along with the desired presentation angle(s) and sound file(s), to be auralized using the VBAP and Ambisonics methods merged into a final multi-channel .wav file for presentation over a four-loudspeaker system. One of the benefits of this approach is that it does not require any additional parameters, such as those generated by a simulation program, and can be used with any Ambisonics RIR, including those in higher-order format that must be converted to an appropriate order for the number of loudspeakers. Overall, the development of the Iceberg method illustrates the potential for adapting existing technology to meet the needs of smaller-scale virtualization setups while still delivering realistic spatial cues. This approach could support the broader adoption of auralization in hearing research and encourage researchers to utilize virtualized sound fields in their protocols.

6.2.1 Iceberg capabilities

The auralization method proposed in this work combines the use of Ambisonics RIRs and VBAP to balance the acoustic energy in two spatial domains: the perception of sound localization and the perception of immersion. This results in a file that captures the characteristics of a given sound as if it were played in the desired environment. The method can be reproduced with at least four loudspeakers but is scalable to a more extensive array of loudspeakers of any size greater than four, theoretically increasing its efficiency. In addition, multiple sound sources can be virtualized and merged at presentation to create more complex environments. The input to Iceberg includes Ambisonics RIRs corresponding to specific source-and-receptor positions and the sounds to be virtualized, preferably recorded in (near) anechoic conditions. The method can pan the source around the listener, as the VBAP component is independent of Ambisonics. However, it is recommended that RIRs be generated for specific angles when using room acoustic software to generate the Ambisonics RIRs. One benefit of this method is that it can reproduce sounds above the cut-off frequency associated with lower-order Ambisonics due to its use of VBAP, which is initially not frequency limited [241]. VBAP is responsible for the delivery of both direct sound and early reflections. Additionally, the default properties are defined to work with normalized RIRs, enabling the researcher to specify the sound pressure level of the auralized files.

6.2.2 Iceberg & Second Joint Listener

Testing with a second listener inside the loudspeaker ring helps illuminate the potential for this virtualization system in different tasks and human-interaction situations [143, 202, 230, 234]. A system that allows these tasks and situations needs to deliver the appropriate sound properties for the sound to be perceived as coming from the intended position [97]. Ambisonics was shown to be not effective in this test, as the shadow caused by a second listener prevented higher frequency spatial information from being correctly presented, distorting the sound field (especially in low-order Ambisonics). Vector-based solutions can have less impact as the sound is physically formed from two (or three in 3D setups) loudspeakers in the same quadrant. That means that the interference will happen only at angles where the acoustic shadow of a physical object would naturally interfere in a non-virtualized reproduction. In Chapter 5 BRIRs were acquired with files generated by the Iceberg method and reproduced via a modest setup composed of four loudspeakers in the presence of a second listener. It could be observed that it did not disturb the sound field, as the (primary) listener in the center position received the appropriate binaural cues. The system designed to reproduce files virtualized with Iceberg method managed to perform competitively with systems with more loudspeakers rendering pure methods (See Table 2.2).

6.2.3 Iceberg: Listener Wearing Hearing Aids

Adding the possibility of allowing participants to use hearing devices is another crucial step in making auditory tests with auralized files accessible to more researchers [134, 144]. It has been observed that hearing aid signals can influence the intelligibility and clarity of speech in virtualized sound fields [7, 97, 99, 103, 161, 188, 213, 276]. When the hearing aid signals are not appropriately aligned with the characteristics of the virtualized sound field, listeners may struggle to comprehend spoken words or sentences [98]. This issue can be exacerbated when the virtualized sound field includes noise or other distractions that can interfere with speech perception or when the hearing aid signals fail to amplify or enhance the speech signal to an adequate degree [98, 137]. Suppose the hearing aid signals are not correctly capturing the sound field and, therefore, not correcting it to the individual needs and preferences of the listener. In that case, the listener may experience difficulty using the virtualized sound field comfortably and effectively. Swept signals were auralized by the Iceberg method, played through the system, recorded with a manikin wearing hearing aids, and deconvolved. The resulting BRIRs were analyzed in terms of binaural cues and compared to the same signals from actual loudspeakers. The localization error was estimated by May and Kohlrausch's probabilistic model for robust sound source localization based on a binaural auditory front end. This model estimates the location of a sound source using binaural cues such as interaural level differences and interaural time differences extracted from the signals received by the two ears. By combining these cues in a probabilistic framework, the model can robustly estimate the location of the sound source, even in noisy or distracting environments. Evaluation of the model suggests its potential for use in practical contexts such as in hearing aids or virtual reality systems. Results obtained using the Iceberg method with an aided HATS showed similar performance to the unaided results with the listener positioned in the sweet spot, indicating suitable performance (see Section 5.5).

6.2.4 Iceberg Limitations

The virtualization system playing files auralized with the Iceberg method has been found to be less effective outside of the sweet spot, as the binaural cues are not correctly rendered. This mismatch, which occurs for more than 2.5 cm displacements, can be mitigated by keeping the listener centered in the virtualized sound field. While this is a significant limitation, the method can still be applied with simple measures such as a modest head restraint, reducing the setup requirements compared to other classical methods. One major limitation of the Iceberg method is its spatial resolution capabilities. It is recommended for scenarios with a minimum of 30º of separation between sound sources (it can be lower if closer to loudspeakers, although it should be checked for the error distribution). Furthermore, the distance to the sound source should be equal to the radius of the loudspeaker array, as Ambisonics and VBAP cannot define sources inside the array. VBAP can only pan between physical sound sources. These limitations should be considered when using the Iceberg method to create virtualized sound fields.

6.3 General Conclusion

As computational capacity increases, using more complex and natural sound scenarios in auditory research becomes feasible and desirable. This technology allows for testing new features, sensors, and algorithms in controlled conditions with increasing realism and ecological validity. Even clinical tests can benefit from auralization, allowing for investigations in different scenarios with varying acoustics (*e.g.*, in a speech-in-noise test). The spatial-cue performance of the Iceberg auralization method, reproducing files through a system of four loudspeakers, is mainly sufficient for these types of tests. It is essential to understand the constraints of auralization methods, Iceberg included, which are tied to the virtualization setup and should be chosen by researchers based on their needs and the available hardware. However, utilizing the Iceberg, virtualization can be conducted by auditory research groups that cannot afford or house expensive anechoic chambers with tens or hundreds of loudspeakers and sophisticated hardware and need more freedom than using headphones. The method presented in this work serves as an additional tool for researchers to consider.

6.4 Main Contributions

In this work, we have presented a novel auralization method called Iceberg, designed to create virtualized sound scenarios for use in auditory research. The main contributions of this work are:

1. The development of a hybrid auralization method that combines two psychoacoustic virtualization methods to balance the energy of an RIR and output a multi-channel file for presentation.

2. The implementation of an effective, simple, and partially automated auralization method that allows for the creation of reasonably realistic virtualized sound scenarios with a modest setup.

3. The exploration of the use and limitations of auralization methods in auditory research, including the suggestion that the Iceberg method has the potential to be a helpful tool for testing new features, sensors, and algorithms in controlled conditions with increasing realism and ecological validity.

4. We researched the limitations and feasibility of using Ambisonics in the context of speech intelligibility with normal-hearing listeners.

5. Identifying the potential for the Iceberg method to be applied in a range of practical contexts, including in hearing aids and virtual reality systems.

Bibliography

[1] Aguirre, S. L. (2017). Implementação e avaliação de um sistema de virtualização de fontes sonoras (in portuguese). Master, Programa de Pós-Graduação em Engenharia Mecânica, Universidade Federal de Santa Catarina.
(Cited on pages 41 and 49)

[2] Aguirre, S. L., Bramsløw, L., Lunner, T., and Whitmer, W. M. (2019). Spatial cue distortions within a virtualized sound field caused by an additional listener. In *Proceedings of the 23rd International Congress on Acoustics : integrating 4th EAA Euroregio 2019*, pages 6537–6544, Berlin, Germany. ICA International Congress on Acoustics, Deutsche Gesellschaft für Akustik.
(Cited on page 94)

[3] Aguirre, S. L., Seifi-Ala, T., Bramsløw, L., Graversen, C., Hadley, L. V., Naylor, G., and Whitmer, W. M. (2021). Combination study 3. http://hear-eco.eu/combination-study-3/. (accessed: 24.11.2021).
(Cited on page 97)

[4] Agus, T. R., Akeroyd, M. A., Gatehouse, S., and Warden, D. (2009). Informational masking in young and elderly listeners for speech masked by simultaneous speech and noise. *The Journal of the Acoustical Society of America*, 126(4):1926–1940.
(Cited on page 40)

[5] Ahnert Feistel Media Group (2011). Ease enhanced acoustic simulator for engineers. https://www.afmg.eu/en/ease-enhanced-acoustic-simulator-engineers. Last checked on: Nov 28, 2021.
(Cited on page 27)

[6] Ahrens, A., Marschall, M., and Dau, T. (2017). Measuring speech intelligibility with speech and noise interferers in a loudspeaker-based virtual sound environment. *The Journal of the Acoustical Society of America*, 141(5):3510–3510.
(Cited on pages 16, 42 and 52)

[7] Ahrens, A., Marschall, M., and Dau, T. (2019). Measuring and modeling speech intelligibility in real and loudspeaker-based virtual sound environments. *Hearing Research*, 377:307–317.
(Cited on pages 42, 53, 99 and 191)

[8] Ahrens, A., Marschall, M., and Dau, T. (2020). The effect of spatial energy spread on sound image size and speech intelligibility. *The Journal of the Acoustical Society of America*, 147(3):1368–1378.
(Cited on page 42)

[9] Akeroyd, M. A. (2006). The psychoacoustics of binaural hearing. *International Journal of Audiology*, 45(sup1):25–33.
(Cited on pages 9 and 17)

[10] Alfandari Menase, D. (2022). *Motivation and fatigue effects in pupillometric measures of listening effort*. PhD thesis, University of Nottingham.
(Cited on page 33)

[11] Algazi, V. R., Duda, R. O., and Thompson, D. M. (2004). Motion-tracked binaural sound. *Journal of the Audio Engineering Society*, 52(11):1142–1156.
(Cited on page 23)

[12] Alhanbali, S., Dawes, P., Millman, R. E., and Munro, K. J. (2019). Measures of Listening Effort Are Multidimensional. *Ear and hearing*.
(Cited on pages 51 and 118)

[13] Alpert, M. I., Alpert, J. I., and Maltz, E. N. (2005). Purchase occasion influence on the role of music in advertising. *Journal of business research*, 58(3):369–376.
(Cited on page 8)

[14] Arau-Puchades, H. (1988). An improved reverberation formula. *Acta Acustica united with Acustica*, 65(4):163–180.
(Cited on page 34)

[15] Archontis Politis (2020). Higher Order Ambisonics (HOA) library.
(Cited on page 107)

[16] Arlinger, S. (2003). Negative consequences of uncorrected hearing loss—a review. *International Journal of Audiology*, 42(sup2):17–20.
(Cited on pages 1 and 55)

[17] Aspöck, L., Pausch, F., Stienen, J., Berzborn, M., Kohnen, M., Fels, J., and Vorländer, M. (2018). Application of virtual acoustic environments in the scope of auditory research. In *XXVIII Encontro da Sociedade Brasileira de Acústica, SOBRAC, Porto Alegre, Brazil*. SOBRAC.
(Cited on pages 16 and 42)

[18] Attenborough, K. (2007). *Sound Propagation in the Atmosphere*, pages 113–147. Springer New York, New York, NY.
(Cited on page 10)

[19] Baldan, S., Lachambre, H., Delle Monache, S., and Boussard, P. (2015). Physically informed car engine sound synthesis for virtual and augmented environments. In *2015 IEEE 2nd VR Workshop on Sonic Interactions for Virtual Environments (SIVE)*, pages 1–6. IEEE.
(Cited on page 20)

[20] Barron, M. (1971). The subjective effects of first reflections in concert halls—the need for lateral reflections. *Journal of Sound and Vibration*, 15(4):475–494.
(Cited on page 37)

[21] Barron, M. and Marshall, A. (1981). Spatial impression due to early lateral reflections in concert halls: The derivation of a physical measure. *Journal of Sound and Vibration*, 77(2):211–232.
(Cited on pages 9, 37 and 38)

[22] Bates, E., Kearney, G., Furlong, D., and Boland, F. (2007). Localization accuracy of advanced spatialisation techniques in small concert halls. *The Journal of the Acoustical Society of America*, 121.
(Cited on pages 47, 49 and 53)

[23] Benesty, J., Sondhi, M., and Huang, Y. (2008). *Springer Handbook of Speech Processing*. Springer Handbook of Speech Processing. Springer-Verlag Berlin Heidelberg. bibtex: Benesty2008.
(Cited on page 24)

[24] Berkhout, A. J. (1988). a holographic approach to acoustic control. *Journal of the Audio Engineering Society*, 36(12):977–995.
(Cited on page 31)

[25] Berkhout, A. J., de Vries, D., and Vogel, P. (1993). Acoustic control by wave field synthesis. *The Journal of the Acoustical Society of America*, 93(5):2764–2778.
(Cited on page 31)

[26] Bertet, S., Daniel, J., Parizet, E., and Warusfel, O. (2009). Influence of microphone and loudspeaker setup on perceived higher order ambisonics reproduced sound field. *Proceedings of Ambisonics Symposium*. cited By 3.
(Cited on page 31)

[27] Bertet, S., Daniel, J., Parizet, E., and Warusfel, O. (2013). Investigation on localisation accuracy for first and higher order ambisonics reproduced sound sources. *Acta Acustica united with Acustica*, 99:642 – 657.
(Cited on pages 31 and 77)

[28] Bertoli, S. and Bodmer, D. (2014). Novel sounds as a psychophysiological measure of listening effort in older listeners with and without hearing loss. *Clinical Neurophysiology*.
(Cited on page 50)

[29] Berzborn, M., Bomhardt, R., Klein, J., Richter, J.-G., and Vorländer, M. (2017). The ITA-Toolbox: An Open Source MATLAB Toolbox for Acoustic Measurements and Signal Processing. In *43th Annual German Congress on Acoustics, Kiel (Germany), 6 Mar 2017 - 9 Mar 2017*, volume 43, pages 222–225.
(Cited on pages 60, 107 and 138)

[30] Best, V., Kalluri, S., McLachlan, S., Valentine, S., Edwards, B., and Carlile, S. (2010). A comparison of cic and bte hearing aids for three-dimensional localization of speech. *International Journal of Audiology*, 49(10):723–732.
(Cited on page 42)

[31] Best, V., Keidser, G., Buchholz, J. M., and Freeston, K. (2015). An examination of speech reception thresholds measured in a simulated reverberant cafeteria environment. *International Journal of Audiology*.
(Cited on pages 42 and 52)

[32] Best, V., Marrone, N., Mason, C. R., and Kidd, G. (2012). The influence of non-spatial factors on measures of spatial release from masking. *The Journal of the Acoustical Society of America*, 131(4):3103–3110. bibtex: best2012.
(Cited on page 33)

[33] Bidelman, G. M., Davis, M. K., and Pridgen, M. H. (2018). Brainstem-cortical functional connectivity for speech is differentially challenged by noise and reverberation. *Hearing Research*.
(Cited on page 50)

[34] Bigg, G. R. (2015). *The science of icebergs*, page 21–124. Cambridge University Press.
(Cited on page 128)

[35] Bisgaard, N., Vlaming, M. S. M. G., and Dahlquist, M. (2010). Standard audiograms for the iec 60118-15 measurement procedure. *Trends in Amplification*, 14(2):113–120.
(Cited on pages 163 and 164)

[36] Blackstock, D. (2000). *Fundamentals of Physical Acoustics*. A Wiley-Interscience publication. Wiley.
(Cited on page 228)

[37] Blauert, J. (1969). Sound localization in the median plane. *Acta Acustica united with Acustica*, 22(4):205–213.
(Cited on page 10)

[38] Blauert, J. (1997). *Spatial hearing: the psychophysics of human sound localization*. MIT press.
(Cited on pages 9, 13, 14, 16, 18, 40, 73, 93, 126 and 149)

[39] Blauert, J. (2005). *Communication acoustics*. Springer-Verlag Berlin Heidelberg, 1 edition.
(Cited on pages 2, 3, 10, 13, 20, 76 and 98)

[40] Blauert, J. (2013). *The technology of binaural listening*. Springer.
(Cited on pages 9, 20, 22, 33 and 76)

[41] Blauert, J., Lehnert, H., Sahrhage, J., and Strauss, H. (2000). An interactive virtual-environment generator for psychoacoustic research. i: Architecture and implementation. *Acta Acustica united with Acustica*, 86:94–102.
(Cited on pages 1 and 57)

[42] Bock, T. M. and Keele, Jr., D. B. D. (1986). The effects of interaural crosstalk on stereo reproduction and minimizing interaural crosstalk in nearfield monitoring by the use of a physical barrier: part 1. *Journal of the Audio Engineering Society*.
(Cited on page 43)

[43] Bradley, J. S. (1986). Speech intelligibility studies in classrooms. *The Journal of the Acoustical Society of America*, 80(3):846–854.
(Cited on page 127)

[44] Bradley, J. S. and Soulodre, G. A. (1995). Objective measures of listener envelopment. *The Journal of the Acoustical Society of America*, 98(5):2590–2597.
(Cited on pages 36, 38 and 58)

[45] Brandão, E. (2018). *Acústica de salas: Projeto e modelagem.* Editora Blucher, São Paulo.
(Cited on pages 16, 19, 33, 36, 38 and 128)

[46] Brandao, E., Morgado, G., and Fonseca, W. (2020). A ray tracing engine integrated with blender and with uncertainty estimation: Description and initial results. *Building Acoustics*, 28:1–20.
(Cited on page 27)

[47] Breebaart, J., van de Par, S., Kohlrausch, A., and Schuijers, E. (2004). High-quality parametric spatial audio coding at low bitrates. *Journal of the Audio Engineering Society.*
(Cited on page 57)

[48] Breebaart, J., Van de Par, S., Kohlrausch, A., and Schuijers, E. (2005). Parametric coding of stereo audio. *EURASIP Journal on Advances in Signal Processing*, pages 1–18.
(Cited on page 57)

[49] Brinkmann, F., Aspöck, L., Ackermann, D., Lepa, S., Vorländer, M., and Weinzierl, S. (2019). A round robin on room acoustical simulation and auralization. *The Journal of the Acoustical Society of America*, 145(4):2746–2760.
(Cited on page 21)

[50] Brinkmann, F., Aspöck, L., Ackermann, D., Opdam, R., Vorländer, M., and Weinzierl, S. (2021). A benchmark for room acoustical simulation. concept and database. *Applied Acoustics*, 176:107867.
(Cited on page 21)

[51] Brinkmann, F., Lindau, A., and Weinzierl, S. (2017). On the authenticity of individual dynamic binaural synthesis. *The Journal of the Acoustical Society of America*, 142(4):1784–1795.
(Cited on page 14)

[52] Brown, C. and Duda, R. (1998). A structural model for binaural sound synthesis. *IEEE Transactions on Speech and Audio Processing*, 6(5):476–488.
(Cited on page 12)

[53] Brown, V. A. and Strand, J. F. (2019). Noise increases listening effort in normal-hearing young adults, regardless of working memory capacity. *Language, Cognition and Neuroscience.*
(Cited on pages 51 and 118)

[54] Brungart, D. S., Cohen, J., Cord, M., Zion, D., and Kalluri, S. (2014). Assessment of auditory spatial awareness in complex listening environments. *The Journal of the Acoustical Society of America*, 136(4):1808–1820.
(Cited on page 18)

[55] Buchholz, J. M. and Best, V. (2020). Speech detection and localization in a reverberant multitalker environment by normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 147(3):1469–1477.
(Cited on page 42)

[56] Byrnes, H. (1984). The role of listening comprehension: A theoretical base. *Foreign language annals*, 17(4):317.
(Cited on page 8)

[57] Campanini, S. and Farina, A. (2008). A new audacity feature: room objective acustical parameters calculation module.
(Cited on page 36)

[58] Choi, I., Shinn-Cunningham, B. G., Chon, S. B., and Sung, K.-m. (2008). Objective measurement of perceived auditory quality in multichannel audio compression coding systems. *Journal of the Audio Engineering Society*, 56(1/2):3–17.
(Cited on page 73)

[59] Claus Lynge Christensen, Gry Bælum Nielsen, J. H. R. (2008). Danish acoustical society round robin on room acoustic computer modelling. https://odeon.dk/learn/articles/auralisation/. Last checked on: Nov 28, 2021.
(Cited on pages 18, 27, 68, 107 and 129)

[60] Cooper, D. H. and Bauck, J. L. (1989). prospects for transaural recording. *Journal of the Audio Engineering Society*, 37(1/2):3–19.
(Cited on page 22)

[61] Cubick, J. and Dau, T. (2016). Validation of a virtual sound environment system for testing hearing aids. *Acta Acustica united with Acustica*.
(Cited on pages 42, 53, 56 and 120)

[62] Cuevas-Rodríguez, M., Picinali, L., González-Toledo, D., Garre, C., de la Rubia-Cuestas, E., Molina-Tanco, L., and Reyes-Lecuona, A. (2019). 3d tune-in toolkit: An open-source library for real-time binaural spatialisation. *PloS one*, 14(3):e0211899.
(Cited on pages 16, 20 and 121)

[63] Cunningham, L. L. and Tucci, D. L. (2017). Hearing loss in adults. *New England Journal of Medicine*, 377(25):2465–2473.
(Cited on pages 1 and 55)

[64] Daniel, J. (2000). *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia (In French)*. PhD thesis, University of Paris VI.
(Cited on pages 31 and 99)

[65] Daniel, J. and Moreau, S. (2004). Further study of sound field coding with higher order ambisonics. In *Audio Engineering Society Convention 116*.
(Cited on pages 30, 32, 99, 121 and 142)

[66] Davies, W. J., Bruce, N. S., and Murphy, J. E. (2014). Soundscape reproduction and synthesis. *Acta Acustica united with Acustica*, 100(2):285–292.
(Cited on page 42)

[67] Dietrich, P., Masiero, B., Müller-Trapet, M., Pollow, M., and Scharrer, R. (2010). Matlab toolbox for the comprehension of acoustic measurement and signal processing. In *Fortschritte der Akustik – DAGA*.
(Cited on page 107)

[68] Dreier, C. and Vorländer, M. (2020). Psychoacoustic optimisation of aircraft noise-challenges and limits. In *Inter-Noise and Noise-Con Congress and Conference Proceedings*, volume 261, pages 2379–2386. Institute of Noise Control Engineering.
(Cited on page 20)

[69] Dreier, C. and Vorländer, M. (2021). Aircraft noise—auralization-based assessment of weather-dependent effects on loudness and sharpness. *The Journal of the Acoustical Society of America*, 149(5):3565–3575.
(Cited on page 20)

[70] Duda, R., Avendano, C., and Algazi, V. (1999). An adaptable ellipsoidal head model for the interaural time difference. In *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258)*, volume 2, pages 965–968 vol.2.
(Cited on page 12)

[71] Dunne, R., Desai, D., and Heyns, P. S. (2021). Development of an acoustic material property database and universal airflow resistivity model. *Applied Acoustics*, 173:107730.
(Cited on page 21)

[72] Eddins, D. A. and Hall, J. W. (2010). Binaural processing and auditory asymmetries. In Gordon-Salant, S., Frisina, R. D., Popper, A. N., and Fay, R. R., editors, *The Aging Auditory System*, pages 135–165. Springer New York, New York, NY.
(Cited on page 11)

[73] Epain, N., Guillon, P., Kan, A., Kosobrodov, R., Sun, D., Jin, C., and Van Schaik, A. (2010). Objective evaluation of a three-dimensional sound field reproduction system. In Burgess, M., Davey, J., Don, C., and McMinn, T., editors, *Proceedings of 20th International Congress on Acoustics, ICA 2010*, volume 2, pages 949–955. International Congress on Acoustics (ICA).
(Cited on page 31)

[74] Eyring, C. F. (1930). Reverberation time in “dead” rooms. *The Journal of the Acoustical Society of America*, 1(2A):168–168.
(Cited on page 34)

[75] Farina, A. (2000). Simultaneous measurement of impulse response and distortion with a swept-sine technique. *Journal of The Audio Engineering Society*.
(Cited on page 63)

[76] Farina, A., Glasgal, R., Armelloni, E., and Torger, A. (2001). ambiophonic principles for the recording and reproduction of surround sound for music. *journal of the audio engineering society.*
(Cited on pages 43 and 45)

[77] Favrot, S. and Buchholz, J. (2009). Validation of a loudspeaker-based room auralization system using speech intelligibility measures. In *Audio Engineering Society Convention Papers*, volume Preprint 7763, page 7763. Praesens Verlag. 126th Audio Engineering Society Convention, AES126 ; Conference date: 07-05-2009 Through 10-05-2009.
(Cited on pages 21 and 99)

[78] Favrot, S., Marschall, M., Käsbach, J., Buchholz, J., and Weller, T. (2011). Mixed-order ambisonics recording and playback for improving horizontal directionality. In *Proceeding of the audio engineering society 131st convention*. 131st AES Convention ; Conference date: 20-10-2011 Through 23-10-2011.
(Cited on page 99)

[79] Favrot, S. E., Buchholz, J., and Dau, T. (2010). *A loudspeaker-based room auralization system for auditory research.* phdthesis, Technical University of Denmark.
(Cited on pages 1, 42, 43, 45, 56, 57, 120 and 127)

[80] Fichna, S., Biberger, T., Seeber, B. U., and Ewert, S. D. (2021). Effect of acoustic scene complexity and visual scene representation on auditory perception in virtual audio-visual environments. *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA).*
(Cited on page 42)

[81] Fintor, E., Aspöck, L., Fels, J., and Schlittmeier, S. (2021). The role of spatial separation of two talkers auditory stimuli in the listener's memory of running speech: listening effort in a non-noisy conversational setting. *International Journal of Audiology.*
(Cited on page 57)

[82] Fitzroy, D. (1959). Reverberation formula which seems to be more accurate with nonuniform distribution of absorption. *The Journal of the Acoustical Society of America*, 31(7):893–897.
(Cited on page 34)

[83] Francis, A. L. and Love, J. (2020). Listening effort: Are we measuring cognition or affect, or both? *WIREs Cognitive Science*, 11(1):e1514.
(Cited on page 98)

[84] Frank, M. (2014). Localization using different amplitude-panning methods in the frontal hhorizontal plane. In *Proceedings of the EAA Joint Symposium on Auralization and Ambisonics 2014.*
(Cited on pages 45, 49 and 160)

[85] Frank, M. and Zotter, F. (2008). Localization experiments using different 2d ambisonics decoders. In *25th Tonmeistertagung-VDT International Convention, Leipzig.*
(Cited on pages 45 and 49)

[86] Franklin, W. S. (1903). Derivation of equation of decaying sound in a room and definition of open window equivalent of absorbing power. *Phys. Rev. (Series I)*, 16:372–374.
(Cited on page 34)

[87] Fraser, S., Gagné, J. P., Alepins, M., and Dubois, P. (2010). Evaluating the effort expended to understand speech in noise using a dual-task paradigm: The effects of providing visual speech cues. *Journal of Speech, Language, and Hearing Research*.
(Cited on page 50)

[88] Furuya, H., Fujimoto, K., Young Ji, C., and Higa, N. (2001). Arrival direction of late sound and listener envelopment. *Applied Acoustics*, 62(2):125–136.
(Cited on page 37)

[89] Gandemer, L., Parseihian, G., Bourdin, C., and Kronland-Martinet, R. (2018). Perception of Surrounding Sound Source Trajectories in the Horizontal Plane: A Comparison of VBAP and Basic-Decoded HOA. *Acta Acustica united with Acustica*, pages 338–350.
(Cited on pages 40, 57 and 122)

[90] Gelfand, S. and Gelfand, S. (2004). *Hearing: An Introduction to Psychological and Physiological Acoustics, Fourth Edition*. Taylor & Francis.
(Cited on page 76)

[91] Gerzon, M. A. (1985). Ambisonics in multichannel broadcasting and video. *AES: Journal of the Audio Engineering Society*.
(Cited on pages 23 and 58)

[92] Giguere, C. and Woodland, P. C. (1994). A computational model of the auditory periphery for speech and hearing research. i. ascending path. *The Journal of the Acoustical Society of America*, 95(1):331–342.
(Cited on page 8)

[93] Gil Carvajal, J., Cubick, J., Santurette, S., and Dau, T. (2016). Spatial hearing with incongruent visual or auditory room cues. *Scientific Reports*, 6.
(Cited on pages 16 and 42)

[94] Glasgal, R. (2001). the ambiophone derivation of a recording methodology optimized for ambiophonic reproduction. *journal of the audio engineering society*.
(Cited on page 43)

[95] Glasgal, R. and Yates, K. (1995). *Ambiophonics: Beyond Surround Sound to Virtual Sonic Reality*. Ambiophonics Institute.
(Cited on page 43)

[96] Gomes, L., Fonseca, W. D., de Carvalho, D. M. L., and Mareze, P. H. (2020). Rendering binaural signals for moving sources. In *Reproduced Sound 2020*.
(Cited on page 20)

[97] Grimm, G., Ewert, S., and Hohmann, V. (2015a). Evaluation of spatial audio reproduction schemes for application in hearing aid research. *Acta Acustica united with Acustica*, 101(4):842–854.
(Cited on pages 18, 21, 31, 41, 49, 53, 94, 117, 121, 160, 163, 177, 187, 190 and 191)

[98] Grimm, G., Kollmeier, B., and Hohmann, V. (2016a). Spatial Acoustic Scenarios in Multichannel Loudspeaker Systems for Hearing Aid Evaluation. *Journal of the American Academy of Audiology*, 27(7):557–566.
(Cited on pages 56, 163 and 191)

[99] Grimm, G., Kollmeier, B., and Hohmann, V. (2016b). Spatial Acoustic Scenarios in Multichannel Loudspeaker Systems for Hearing Aid Evaluation. *Journal of the American Academy of Audiology*.
(Cited on page 191)

[100] Grimm, G., Luberadzka, J., Herzke, T., and Hohmann, V. (2015b). Toolbox for acoustic scene creation and rendering (TASCAR): Render methods and research applications. *Proceedings of the Linux Audio Conference*.
(Cited on page 42)

[101] Grimm, G., Luberadzka, J., and Hohmann, V. (2018). Virtual acoustic environments for comprehensive evaluation of model-based hearing devices *. *International Journal of Audiology*.
(Cited on pages 42 and 120)

[102] Grimm, G., Luberadzka, J., and Hohmann, V. (2019). A toolbox for rendering virtual acoustic environments in the context of audiology. *Acta Acustica united with Acustica*, 105:566–578.
(Cited on pages 1, 42 and 57)

[103] Guastavino, C., Katz, B., Polack, J.-D., Levitin, D., and Dubois, D. (2004). Ecological validity of soundscape reproduction. *Acta Acustica united with Acustica*, 50.
(Cited on pages 42 and 191)

[104] Guastavino, C. and Katz, B. F. G. (2004). Perceptual evaluation of multi-dimensional spatial audio reproduction. *The Journal of the Acoustical Society of America*, 116:1105–1115.
(Cited on pages 57, 86 and 122)

[105] Guastavino, C., Larcher, V., Catusseau, G., and Boussard, P. (2007). Spatial audio quality evaluation: comparing transaural, ambisonics and stereo. In *Proceedings of the 13th International Conference on Auditory Display. Montréal Canada*. Georgia Institute of Technology.
(Cited on pages 86 and 122)

[106] Hacihabiboglu, H., De Sena, E., Cvetkovic, Z., Johnston, J., and Smith III, J. O. (2017). Perceptual spatial audio recording, simulation, and rendering: An overview of spatial-audio techniques based on psychoacoustics. *IEEE Signal Processing Magazine*, 34(3):36–54.
(Cited on pages 20, 22 and 23)

[107] Hamdan, E. C. and Fletcher, M. D. (2022). A compact two-loudspeaker virtual sound reproduction system for clinical testing of spatial hearing with hearing-assistive devices. *Frontiers in Neuroscience*, 15.
(Cited on pages 42, 47 and 49)

[108] Hammershøi, D. and Møller, H. (1992). Fundamentals of binaural technology. In *Fundamentals of Binaural Technology*.
(Cited on pages 12 and 16)

[109] Hammershøi, D. and Møller, H. (2005). Binaural technique — basic methods for recording, synthesis, and reproduction. In Blauert, J., editor, *Communication Acoustics*, pages 223–254. Springer Berlin Heidelberg, Berlin, Heidelberg.
(Cited on page 16)

[110] Harris, P., Nagy, S., and Vardaxis, N. (2018). *Mosby's Dictionary of Medicine, Nursing and Health Professions - Revised 3rd Anz Edition*. Elsevier Health Sciences Apac.
(Cited on page 7)

[111] Havelock, D. I., Kuwano, S., and Vorlander, M. (2008). *Handbook of signal processing in acoustics*. Springer, New York.
(Cited on page 98)

[112] Hazrati, O. and Loizou, P. C. (2012). The combined effects of reverberation and noise on speech intelligibility by cochlear implant listeners. *International Journal of Audiology*.
(Cited on page 98)

[113] He, J. (2016). *Spatial Audio Reproduction with Primary Ambient Extraction*. SpringerBriefs in Electrical and Computer Engineering. Springer Singapore.
(Cited on page 18)

[114] Hecker, S. (1984). Music for advertising effect. *Psychology & Marketing*, 1(3-4):3–8.
(Cited on page 8)

[115] Hendrickx, E., Stitt, P., Messonnier, J.-C., Lyzwa, J.-M., Katz, B. F., and de Boishéraud, C. (2017). Improvement of externalization by listener and source movement using a "binauralized" microphone array. *Journal of the audio engineering society*, 65(7/8):589–599.
(Cited on page 23)

[116] Hendrikse, M. M. E., Llorach, G., Hohmann, V., and Grimm, G. (2019). Movement and gaze behavior in virtual audiovisual listening environments resembling everyday life. *Trends in Hearing*, 23.
(Cited on pages 1 and 57)

[117] Hiyama, K., Komiyama, S., and Hamasaki, K. (2002). The minimum number of loudspeakers and its arrangement for reproducing the spatial impression of diffuse sound field. *Journal of the Audio Engineering Society*.
(Cited on page 41)

[118] Hohmann, V., Paluch, R., Krueger, M., Meis, M., and Grimm, G. (2020). The virtual reality lab: Realization and application of virtual sound environments. *Ear & Hearing*, 41:31S–38S.
(Cited on pages 1 and 57)

[119] Holman, J. A., Drummond, A., and Naylor, G. (2021). Hearing aids reduce daily-life fatigue and increase social activity: a longitudinal study. *medRxiv*.
(Cited on pages 1 and 56)

[120] Holube, I., Fredelake, S., Vlaming, M., and Kollmeier, B. (2010). Development and analysis of an international speech test signal (ists). *International Journal of Audiology*, 49(12):891–903.
(Cited on page 131)

[121] Holube, I., Haeder, K., Imbery, C., and Weber, R. (2016). Subjective Listening Effort and Electrodermal Activity in Listening Situations with Reverberation and Noise. *Trends in hearing*.
(Cited on pages 99 and 118)

[122] Hong, J. Y., He, J., Lam, B., Gupta, R., and Gan, W.-S. (2017). Spatial audio for soundscape design: Recording and reproduction. *Applied Sciences*, 7(6).
(Cited on page 18)

[123] Hornsby, B. W. (2013). The effects of hearing aid use on listening effort and mental fatigue associated with sustained speech processing demands. *Ear and hearing*, 34(5):523–534.
(Cited on page 33)

[124] Howard, D. and Angus, J. (2009). *Acoustics and Psychoacoustics 4th Edition*. Oxford: Focal Press, 4th edition.
(Cited on page 73)

[125] Huisman, T., Ahrens, A., and MacDonald, E. (2021). Ambisonics sound source localization with varying amount of visual information in virtual reality. *Frontiers in Virtual Reality*, 2.
(Cited on pages 16 and 49)

[126] International Telecommunications Union - Radiocommunications Sector (ITU-R) (2015). Methods for the subjective assessment of small impairments in audio systems. Technical report, International Telecommunications Union, Geneva.
(Cited on pages 42, 61, 105 and 132)

[127] ISO (2009). 3382-1: Acoustics - measurement of room acoustic parameters. part 1 : Performance spaces. ISO 1:2009, ISO.
(Cited on pages 33, 34, 35 and 70)

[128] Jäncke, L. (2008). Music, memory and emotion. *Journal of biology*, 7(6):1–5.
(Cited on page 8)

[129] Jin, C., Corderoy, A., Carlile, S., and van Schaik, A. (2004). Contrasting monaural and interaural spectral cues for human sound localization. *The Journal of the Acoustical Society of America*, 115(6):3124–3141.
(Cited on page 11)

[130] Jot, J.-M., Wardle, S., and Larcher, V. (1998). approaches to binaural synthesis. *Journal of the Audio Engineering Society*.
(Cited on page 27)

[131] Kang, S. and Kim, S.-H. K. (1996). Realistic audio teleconferencing using binaural and auralization techniques. *Etri Journal*, 18:41–51.
(Cited on page 23)

[132] Katz, B. F. G. and Noisternig, M. (2014). A comparative study of interaural time delay estimation methods. *The Journal of the Acoustical Society of America*, 135(6):3530–3540.
(Cited on page 70)

[133] Keet, V. (1968). The influence of early lateral reflections on the spatial impression. *Proc. 6th Int. Cong. Acoust., Tokyo*, 2.
(Cited on page 39)

[134] Keidser, G., Naylor, G., Brungart, D. S., Caduff, A., Campos, J., Carlile, S., Carpenter, M. G., Grimm, G., Hohmann, V., Holube, I., Launer, S., Lunner, T., Mehra, R., Rapport, F., Slaney, M., and Smeds, K. (2020). The quest for ecological validity in hearing science: what it is, why it matters, and how to advance it. *Ear and Hearing*, 41(S1):5S–19S.
(Cited on pages 53, 56, 98, 122 and 191)

[135] Kestens, K., Degeest, S., and Keppler, H. (2021). The effect of cognition on the aided benefit in terms of speech understanding and listening effort obtained with digital hearing aids: A systematic review. *American Journal of Audiology*, 30(1):190–210.
(Cited on page 98)

[136] Kirsch, C., Poppitz, J., Wendt, T., van de Par, S., and Ewert, S. D. (2021). Computationally efficient spatial rendering of late reverberation in virtual acoustic environments. *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*.
(Cited on page 42)

[137] Klatte, M., Lachmann, T., Meis, M., et al. (2010). Effects of noise and reverberation on speech perception and listening comprehension of children and adults in a classroom-like setting. *Noise and Health*, 12(49):270.
(Cited on pages 1 and 191)

[138] Kleiner, M., Dalenbäck, B.-I., and Svensson, P. (1993). Auralization-an overview. *Journal of the Audio Engineering Society*, 41(11):861–875.
(Cited on page 19)

[139] Klemenz, M. (2005). Sound synthesis of starting electric railbound vehicles and the influence of consonance on sound quality. *Acta acustica united with acustica*, 91(4):779–788.
(Cited on page 20)

[140] Klockgether, S. and van de Par, S. (2016). Just noticeable differences of spatial cues in echoic and anechoic acoustical environments. *The Journal of the Acoustical Society of America*, 140(4):EL352–EL357.
(Cited on pages 93 and 94)

[141] Kobayashi, M., Ueno, K., and Ise, S. (2015). The Effects of Spatialized Sounds on the Sense of Presence in Auditory Virtual Environments: A Psychological and Physiological Study. *Presence: Teleoperators and Virtual Environments*, 24(2):163–174.
(Cited on page 16)

[142] Koehnke, J. and Besing, J. (1996). A procedure for testing speech intelligibility in a virtual listening environment. *Ear and Hearing*, 17(3):211–217. cited By 59.
(Cited on page 40)

[143] Koelewijn, T., Zekveld, A. A., Festen, J. M., and Kramer, S. E. (2012). Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear and Hearing*, 33(2):291–300.
(Cited on page 190)

[144] Kramer, S. E., Bhuiyan, T., Bramsløw, L., Fiedler, L., Graversen, C., Hadley, L. V., Innes-Brown, H., Naylor, G., Richter, M., Saunders, G. H., Versfeld, N. J., Wendt, D., Whitmer, W. M., and Zekveld, A. A. (2020). Innovative hearing aid research on ecological conditions and outcome measures: The hear-eco project.
(Cited on page 191)

[145] Kramer, S. E., Kapteyn, T. S., Festen, J. M., and Tobi, H. (1996). The relationships between self-reported hearing disability and measures of auditory disability. *Audiology*, 35(5):277–287.
(Cited on page 56)

[146] Krokstad, A., Strom, S., and Sørsdal, S. (1968). Calculating the acoustical room response by the use of a ray tracing technique. *Journal of Sound and Vibration*, 8(1):118–125.
(Cited on page 19)

[147] Krueger, M., Schulte, M., Brand, T., and Holube, I. (2017). Development of an adaptive scaling method for subjective listening effort. *The Journal of the Acoustical Society of America*.
(Cited on page 50)

[148] Kuttruff, H. (2009). *Room Acoustics, Fifth Edition*. Taylor & Francis.
(Cited on pages 19, 34, 68 and 127)

[149] Kwak, C., Han, W., Lee, J., Kim, J., and Kim, S. (2018). Effect of noise and reverberation on speech recognition and listening effort for older adults. *Geriatrics and Gerontology International*.
(Cited on pages 50, 99 and 118)

[150] Laitinen, M.-V. and Pulkki, V. (2009). Binaural reproduction for directional audio coding. In *2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 337–340.
(Cited on page 41)

[151] Lau, M. K., Hicks, C., Kroll, T., and Zupancic, S. (2019). Effect of auditory task type on physiological and subjective measures of listening effort in individuals with normal hearing. *Journal of Speech, Language, and Hearing Research*.
(Cited on pages 50 and 52)

[152] Lau, S.-T., Pichora-Fuller, M., Li, K., Singh, G., and Campos, J. (2016). Effects of hearing loss on dual-task performance in an audiovisual virtual reality simulation of listening while walking. *Journal of the American Academy of Audiology*, 27.
(Cited on page 57)

[153] Letowski, T. and Letowski, S. (2011). Localization error accuracy and precision of auditory localization. In Strumillo, P., editor, *Advances in Sound Localization*, chapter 4, pages 55–78. Intech, Oxford.
(Cited on pages 9, 10 and 17)

[154] Levy, S. M. (2012). Section 9 - calculations to determine the effectiveness and control of thermal and sound transmission. In Levy, S. M., editor, *Construction Calculations Manual*, pages 503–544. Butterworth-Heinemann, Boston.
(Cited on page 60)

[155] Lindau, A. and Brinkmann, F. (2012). perceptual evaluation of headphone compensation in binaural synthesis based on non-individual recordings. *journal of the audio engineering society*, 60(1/2):54–62.
(Cited on page 14)

[156] Lindau, A., Kosanke, L., and Weinzierl, S. (2010). perceptual evaluation of physical predictors of the mixing time in binaural room impulse responses. *Journal of the Audio Engineering Society*.
(Cited on page 126)

[157] Lindemann, W. (1986). Extension of a binaural cross-correlation model by contralateral inhibition. i. simulation of lateralization for stationary signals. *The Journal of the Acoustical Society of America*, 80 6:1608–22.
(Cited on page 45)

[158] Liu, Z., Fard, M., and Jazar, R. (2015). Development of an acoustic material database for vehicle interior trims. Technical report, SAE Technical Paper.
(Cited on page 21)

[159] Llopis, H. S., Pind, F., and Jeong, C.-H. (2020). Development of an auditory virtual reality system based on pre-computed b-format impulse responses for building design evaluation. *Building and Environment*, 169:106553.
(Cited on page 133)

[160] Llorach, G., Evans, A., Blat, J., Grimm, G., and Hohmann, V. (2016). Web-based live speech-driven lip-sync. In *2016 8th International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES)*, pages 1–4.
(Cited on pages 1 and 57)

[161] Llorach, G., Grimm, G., Hendrikse, M. M., and Hohmann, V. (2018). Towards Realistic Immersive Audiovisual Simulations for Hearing Research. In *Proceedings of the 2018 Workshop on Audio-Visual Scene Understanding for Immersive Multimedia*, pages 33–40.
(Cited on pages 1, 18, 27, 40, 53, 56, 57, 58 and 191)

[162] Llorca-Bofí, J., Dreier, C., Heck, J., and Vorländer, M. (2022). Urban sound auralization and visualization framework;case study at ihtapark. *Sustainability*, 14(4).
(Cited on page 20)

[163] Loizou, P. C. (2007). *Speech enhancement: theory and practice*. CRC press.
(Cited on page 52)

[164] Lokki, T. and Savioja, L. (2008). Virtual acoustics. In Havelock, D., Kuwano, S., and Vorländer, M., editors, *Handbook of Signal Processing in Acoustics*, pages 761–771. Springer New York, New York, NY.
(Cited on page 20)

[165] Long, M. (2014). *Architectural Acoustics*. Elsevier Science.
(Cited on pages 16, 19 and 42)

[166] Lopez, J. J., Gutierrez, P., Cobos, M., and Aguilera, E. (2014). Sound distance perception comparison between Wave Field Synthesis and Vector Base Amplitude Panning. In *ISCCSP 2014 - 2014 6th International Symposium on Communications, Control and Signal Processing, Proceedings*.
(Cited on pages 21 and 121)

[167] Lovedee-Turner, M. and Murphy, D. (2018). Application of machine learning for the spatial analysis of binaural room impulse responses. *Applied Sciences*, 8(1).
(Cited on page 15)

[168] Lund, K. D., Ahrens, A., and Dau, T. (2020). A method for evaluating audio-visual scene analysis in multi-talker environments. In *Proceedings of the International Symposium on Auditory and Audiological Research*, volume 7, pages 357–364. The Danavox Jubilee Foundation. International Symposium on Auditory and Audiological Research ISAAR2019.
(Cited on page 42)

[169] Lundbeck, M., Grimm, G., Hohmann, V., Laugesen, S., and Neher, T. (2017). Sensitivity to Angular and Radial Source Movements as a Function of Acoustic Complexity in Normal and Impaired Hearing. *Trends in Hearing*, 21:2331–2165.
(Cited on pages 33 and 57)

[170] Lyon, R. F. (2017). *Human and Machine Hearing Extracting Meaning from Sound*. a. Cambridge University Press.
(Cited on page 10)

[171] Magezi, D. A. (2015). Linear mixed-effects models for within-participant psychology experiments: an introductory tutorial and free, graphical user interface (lmmgui). *Frontiers in Psychology*, 6:2.
(Cited on page 113)

[172] Malham, D. G. and Myatt, A. (1995). 3-d sound spatialization using ambisonic techniques. *Computer Music Journal*, 19(4):58–70.
(Cited on page 29)

[173] Mansour, N., Marschall, M., May, T., Westermann, A., and Dau, T. (2021a). Speech intelligibility in a realistic virtual sound environment. *The Journal of the Acoustical Society of America*, 149(4):2791–2801.
(Cited on page 99)

[174] Mansour, N., Westermann, A., Marschall, M., May, T., Dau, T., and Buchholz, J. (2021b). Guided ecological momentary assessment in real and virtual sound environments. *Acoustical Society of America. Journal*, 150(4):2695–2704.
(Cited on pages 16 and 42)

[175] Marentakis, G., Zotter, F., and Frank, M. (2014). Vector-base and ambisonic amplitude panning: A comparison using pop, classical, and contemporary spatial music. *Acta Acustica united with Acustica.*
(Cited on pages 57 and 86)

[176] Marrone, N., Mason, C. R., and Kidd, G. (2008). The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms. *The Journal of the Acoustical Society of America*, 124(5):3064–3075.
(Cited on pages 16 and 40)

[177] Marschall, M. (2014). Capturing and reproducing realistic acoustic scenes for hearing research. *PhD Thesis - Technical University of Denmark.*
(Cited on pages 40, 53, 99 and 120)

[178] Masiero, B. (2012). *Individualized Binaural Technology. Measurement, Equalization and Perceptual Evaluation.* PhD thesis, RWTH Aachen University.
(Cited on page 14)

[179] Masiero, B. and Fels, J. (2011). Perceptually robust headphone equalization for binaural reproduction. In *Audio Engineering Society Convention 130*. Audio Engineering Society.
(Cited on page 22)

[180] Masiero, B. and Vorlaender, M. (2011). Spatial Audio Reproduction Methods for Virtual Reality. In *42º Congreso Español de Acústica Encuentro Ibérico de Acústica - European Symposium on Environmental Acoustics and on Buildings Acoustically Sustainable*, pages 1–12, Cáceres.
(Cited on pages 23, 24, 58 and 86)

[181] Matthen, M. (2016). Effort and displeasure in people who are hard of hearing. *Ear and Hearing*, 37 Suppl 1.
(Cited on page 57)

[182] May, T., van de Par, S., and Kohlrausch, A. (2011). A probabilistic model for robust localization based on a binaural auditory front-end. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(1):1–13.
(Cited on pages xix, 6, 138, 150, 156, 166, 167, 174 and 178)

[183] Meesawat, K. and Hammershoi, D. (2003). The time when the reverberation tail in a binaural room impulse response begins. In *Audio Engineering Society Convention 115*. Audio Engineering Society.
(Cited on page 15)

[184] Menase, D. A., Richter, M., Wendt, D., Fiedler, L., and Naylor, G. (2022). Task-induced mental fatigue and motivation influence listening effort as measured by the pupil dilation in a speech-in-noise task. *medRxiv.*
(Cited on page 33)

[185] Michael, V. and Vorländer, M. (2008). *Auralization. Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality.* Springer.
(Cited on page 98)

[186] Miles, K., McMahon, C., Boisvert, I., Ibrahim, R., de Lissa, P., Graham, P., and Lyxell, B. (2017). Objective Assessment of Listening Effort: Coregistration of Pupillometry and EEG. *Trends in Hearing.*
(Cited on pages 50, 51 and 118)

[187] Millington, G. (1932). A modified formula for reverberation. *The Journal of the Acoustical Society of America*, 4(1A):69–82.
(Cited on page 34)

[188] Minnaar, P., Favrot, S., and Buchholz, J. (2010). Improving hearing aids through listening tests in a virtual sound environment. *Hearing Journal*, 63(10):40–44.
(Cited on pages 1, 16, 42, 121 and 191)

[189] Møller, H., Sørensen, M. F., Hammershøi, D., and Jensen, C. B. (1995). Head-related transfer functions of human subjects. *Journal of the Audio Engineering Society*, 43(5):300–321.
(Cited on page 12)

[190] Monaghan, J. J., Krumbholz, K., and Seeber, B. U. (2013). Factors affecting the use of envelope interaural time differences in reverberationa). *The Journal of the Acoustical Society of America*, 133(4):2288–2300. bibtex: Monaghan2013.
(Cited on page 33)

[191] Moore, B. C. J. and Tan, C.-T. (2004). development and validation of a method for predicting the perceived naturalness of sounds subjected to spectral distortion. *Journal of the Audio Engineering Society*, 52(9):900–914.
(Cited on page 41)

[192] Moore, T. M. and Picou, E. M. (2018). A potential bias in subjective ratings of mental effort. *Journal of Speech, Language, and Hearing Research.*
(Cited on pages 50, 51 and 118)

[193] Mueller, M. F., Kegel, A., Schimmel, S. M., Dillier, N., and Hofbauer, M. (2012). Localization of virtual sound sources with bilateral hearing aids in realistic acoustical scenes. *The Journal of the Acoustical Society of America*, 131(6):4732–4742.
(Cited on page 16)

[194] Müller, S. and Massarani, P. (2001). Transfer-function measurement with sweeps. *Journal of the Audio Engineering Society*, 49:443–471.
(Cited on pages 63 and 140)

[195] Murta, B. (2019). *Plataforma para ensaios de percepção sonora com fontes distribuídas aplicável a dispositivos auditivos: perSONA (in Portuguese).* PhD thesis, Federal University of Santa Catarina.
(Cited on pages 1 and 57)

[196] Murta, B., Chiea, R., Mourão, G., Pinheiro, M. M., Cordioli, J., Paul, S., and Costa, M. (2019). Cci-mobile: Development of software based tools for speech perception assessment and training with hearing impaired brazilian population. In *CONFERENCE on Implantable Auditory Prostheses (CIAP), Lake Tahoe, California, US.*
(Cited on page 18)

[197] Møller, H. (1992). Fundamentals of binaural technology. *Applied Acoustics*, 36(3-4):171–218.
(Cited on pages 15 and 73)

[198] Nachbar, C., Zotter, F., Deleflie, E., and Sontacchi, A. (2011). Ambix – a suggested ambisonics format.
(Cited on page 103)

[199] Narbutt, M., Allen, A., Skoglund, J., Chinen, M., and Hines, A. (2018). Ambiqual - a full reference objective quality metric for ambisonic spatial audio. In *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6.
(Cited on page 27)

[200] Naugolnykh, K. A., Ostrovsky, L. A., Sapozhnikov, O. A., and Hamilton, M. F. (2000). Nonlinear wave processes in acoustics.
(Cited on page 9)

[201] Naylor, G. M. (1993). Odeon—another hybrid room acoustical model. *Applied Acoustics*, 38(2-4):131–143.
(Cited on page 68)

[202] Neuhoff, J. (2021). *Ecological psychoacoustics*. Brill.
(Cited on page 190)

[203] Neuman, A. C., Wroblewski, M., Hajicek, J., and Rubinstein, A. (2010). Combined effects of noise and reverberation on speech recognition performance of normal-hearing children and adults. *Ear and Hearing*.
(Cited on pages 99 and 118)

[204] Nicola, P. and Chiara, V. (2019). Impact of Background Noise Fluctuation and Reverberation on Response Time in a Speech Reception Task. *Journal of Speech, Language, and Hearing Research*, 62(11):4179–4195.
(Cited on pages 50, 99 and 118)

[205] Nielsen, J. and Dau, T. (2011). The danish hearing in noise test. *International journal of audiology*, 50:202–8.
(Cited on pages 101 and 102)

[206] Nocke, C. and Mellert, V. (2002). Brief review on in situ measurement techniques of impedance or absorption. In *Forum Acusticum, Sevilla.*
(Cited on page 21)

[207] Novo, P. (2005). Auditory virtual environments. In Blauert, J., editor, *Communication Acoustics*, pages 277–297. Springer Berlin Heidelberg, Berlin, Heidelberg.
(Cited on page 57)

[208] Obleser, J., Wöstmann, M., Hellbernd, N., Wilsch, A., and Maess, B. (2012). Adverse listening conditions and memory load drive a common alpha oscillatory network. *Journal of Neuroscience*, 32(36):12376–12383.
(Cited on page 111)

[209] Ohlenforst, B., Wendt, D., Kramer, S. E., Naylor, G., Zekveld, A. A., and Lunner, T. (2018). Impact of SNR, masker type and noise reduction processing on sentence recognition performance and listening effort as indicated by the pupil dilation response. *Hearing Research*.
(Cited on pages 50 and 101)

[210] Ohlenforst, B., Zekveld, A. A., Jansma, E. P., Wang, Y., Naylor, G., Lorens, A., Lunner, T., and Kramer, S. E. (2017a). Effects of hearing impairment and hearing aid amplification on listening effort: A systematic review. *Ear and hearing*, 38(3):267—281.
(Cited on page 98)

[211] Ohlenforst, B., Zekveld, A. A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., Versfeld, N. J., and Kramer, S. E. (2017b). Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. *Hearing Research*, 351:68–79.
(Cited on page 50)

[212] Oreinos, C. and Buchholz, J. (2014). Validation of realistic acoustic environments for listening tests using directional hearing aids. In *2014 14th International Workshop on Acoustic Signal Enhancement (IWAENC)*, pages 188–192.
(Cited on pages 41 and 120)

[213] Oreinos, C. and Buchholz, J. M. (2015). Objective analysis of ambisonics for hearing aid applications: Effect of listener's head, room reverberation, and directional microphones. *The Journal of the Acoustical Society of America*.
(Cited on pages 18, 41, 53, 163 and 191)

[214] Palacino, J., Nicol, R., Emerit, M., and Gros, L. (2012). Perceptual assessment of binaural decoding of first-order ambisonics. In *Acoustics 2012*.
(Cited on page 21)

[215] Parsehian, G., Gandemer, L., Bourdin, C., and Kronland Martinet, R. (2015). Design and perceptual evaluation of a fully immersive three-dimensional sound spatialization system. In *3rd International Conference on Spatial Audio (ICSA 2015)*, Graz, Austria.
(Cited on page 42)

[216] Paul, S. (13-15 maio 2014). A fisiologia da audição como base para fenômenos auditivos. In *Proceedings of the 12th AES Brazil Conference*, São Paulo, SP.
(Cited on page 9)

[217] Pausch, F., Aspöck, L., Vorländer, M., and Fels, J. (2018). An Extended Binaural Real-Time Auralization System With an Interface to Research Hearing Aids for Experiments on Subjects With Hearing Loss. *Trends in Hearing*.
(Cited on pages 16, 44, 45, 120 and 121)

[218] Pausch, F., Behler, G., and Fels, J. (2020). Scalar - a surrounding spherical cap loudspeaker array for flexible generation and evaluation of virtual acoustic environments. *Acta Acust.*, 4(5):19.
(Cited on pages 1 and 57)

[219] Pausch, F. and Fels, J. (2019). Mobilab – a mobile laboratory for on-site listening experiments in virtual acoustic environments. *bioRxiv*.
(Cited on pages 1 and 57)

[220] Pausch, F. and Fels, J. (2020). Localization performance in a binaural real-time auralization system extended to research hearing aids. *Trends in hearing*, 24:1–18.
(Cited on pages 1, 42 and 57)

[221] Pelzer, S., Masiero, B., and Vorländer, M. (2014). 3D Reproduction of Room Auralizations by Combining Intensity Panning, Crosstalk Cancellation and Ambisonics. *Proceedings of the EAA Joint Symposium on Auralization and Ambisonics*.
(Cited on pages 44, 45, 86 and 127)

[222] Peng, Z. E. and Litovsky, R. Y. (2021). The role of interaural differences, head shadow, and binaural redundancy in binaural intelligibility benefits among school-aged children. *Trends in Hearing*, 25.
(Cited on page 77)

[223] Petersen, E. B., Wöstmann, M., Obleser, J., Stenfelt, S., and Lunner, T. (2015). Hearing loss impacts neural alpha oscillations under adverse listening conditions. *Frontiers in Psychology*.
(Cited on page 50)

[224] Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., and Wingfield, A. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). In *Ear and Hearing*.
(Cited on pages 50, 51, 52, 57 and 118)

[225] Picou, E. M., Gordon, J., and Ricketts, T. A. (2016). The effects of noise and reverberation on listening effort in adults with normal hearing. *Ear and Hearing*.
(Cited on pages 50 and 99)

[226] Picou, E. M., Moore, T. M., and Ricketts, T. A. (2017). The effects of directional processing on objective and subjective listening effort. *Journal of Speech, Language, and Hearing Research*.
(Cited on pages 1 and 51)

[227] Picou, E. M., Ricketts, T., and Hornsby, B. (2013). How hearing aids, background noise, and visual cues influence objective listening effort. *Ear and Hearing*, 34:e52–e64.
(Cited on pages 50 and 56)

[228] Picou, E. M. and Ricketts, T. A. (2014). Increasing motivation changes subjective reports of listening effort and choice of coping strategy. *International Journal of Audiology*, 53(6):418–426.
(Cited on page 50)

[229] Picou, E. M. and Ricketts, T. A. (2018). The relationship between speech recognition, behavioural listening effort, and subjective ratings. *International Journal of Audiology*.
(Cited on pages 51 and 118)

[230] Pielage, H., Zekveld, A. A., Saunders, G. H., Versfeld, N. J., Lunner, T., and Kramer, S. E. (2021). The Presence of Another Individual Influences Listening Effort, But Not Performance. *Ear & Hearing*.
(Cited on pages 40, 57, 82 and 190)

[231] Pieren, R. (2018). *Auralization of Environmental Acoustical Sceneries: Synthesis of Road Traffic, Railway and Wind Turbine Noise*. PhD thesis, Delft University of Technology.
(Cited on page 20)

[232] Pieren, R., Heutschi, K., Wunderli, J. M., Snellen, M., and Simons, D. G. (2017). Auralization of railway noise: Emission synthesis of rolling and impact noise. *Applied Acoustics*, 127:34–45.
(Cited on page 20)

[233] Pinheiro, J. C. and Bates, D. M. (2000). Linear mixed-effects models: basic concepts and examples. *Mixed-effects models in S and S-Plus*, pages 3–56.
(Cited on page 113)

[234] Plain, B., Pielage, H., Richter, M., Bhuiyan, T., Lunner, T., Kramer, S., and Zekveld, A. (2021). Social observation increases the cardiovascular response of hearing-impaired listeners during a speech reception task. *Hearing Research*, page 108334.
(Cited on pages 57 and 190)

[235] Plinge, A., Schlecht, S. J., Thiergart, O., Robotham, T., Rummukainen, O., and Habets, E. A. P. (2018). six-degrees-of-freedom binaural audio reproduction of first-order ambisonics with distance information. *Journal of the audio engineering society*.
(Cited on page 27)

[236] Poletti, M. A. (2005). Three-dimensional surround sound systems based on spherical harmonics. *journal of the audio engineering society*, 53(11):1004–1025.
(Cited on page 31)

[237] Politis, A. (2016). *Microphone array processing for parametric spatial audio techniques*. Doctoral thesis, School of Electrical Engineering.
(Cited on pages 130, 131 and 181)

[238] Politis, A., McCormack, L., and Pulkki, V. (2017). Enhancement of ambisonic binaural reproduction using directional audio coding with optimal adaptive mixing. In *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 379–383.
(Cited on page 27)

[239] Pollow, M. (2015). *Directivity Patterns for Room Acoustical Measurements and Simulations*. Aachener Beiträge zur Technischen Akustik. Logos Verlag Berlin GmbH.
(Cited on pages 28 and 29)

[240] Portela, M. S. (2008). Caracterização de fontes sonoras e aplicação na auralização de ambientes. Mestrado, Universidade Federal de Santa Catarina.
(Cited on page 13)

[241] Pulkki, V. (1997). Virtual sound source positioning using vector base amplitude panning. *Journal of the Audio Engineering Society*, 45(6).
(Cited on pages 23, 24, 26, 58, 93, 142, 179 and 190)

[242] Pulkki, V. and Karjalainen, M. (2015). *Communication Acoustics: An Introduction to Speech, Audio and Psychoacoustics*. Wiley.
(Cited on pages 10, 17, 70, 73 and 143)

[243] Pulkki, V., Politis, A., Laitinen, M.-V., Vilkamo, J., and Ahonen, J. (2017). First-order directional audio coding (dirac). In *Parametric Time-Frequency Domain Spatial Audio*, chapter 5, pages 89–140. John Wiley & Sons, Ltd.
(Cited on pages 44, 45 and 127)

[244] Purdy, M. (1991). Listening and community: The role of listening in community formation. *International Journal of Listening*, 5(1):51–67.
(Cited on page 8)

[245] Queiroz, M., Iazzetta, F., Kon, F., Gomes, M. H. A., Figueiredo, F. L., Masiero, B. S., Dias, L., Torres, M. H. C., and Thomaz, L. F. (2008). Acmus: an open, integrated platform for room acoustics research. *J. Braz. Comput. Soc.*, 14(3):87–103.
(Cited on page 36)

[246] Rayleigh, L. (1907). Xii. on our perception of sound direction.
(Cited on page 10)

[247] Reichardt, W., Alim, O. A., and Schmidt, W. (1975). Definition and basis of making an objective evaluation to distinguish between useful and useless clarity defining musical performances. *Acta Acustica united with Acustica*, 32(3):126–137.
(Cited on page 36)

[248] Rennies, J., Brand, T., and Kollmeier, B. (2011). Prediction of the influence of reverberation on binaural speech intelligibility in noise and in quiet. *The Journal of the Acoustical Society of America*, 130(5):2999–3012.
(Cited on page 40)

[249] Rennies, J., Schepker, H., Holube, I., and Kollmeier, B. (2014). Listening effort and speech intelligibility in listening situations affected by noise and reverberation. *The Journal of the Acoustical Society of America*.
(Cited on page 50)

[250] Roffler, S. K. and Butler, R. A. (1968). Factors that influence the localization of sound in the vertical plane. *The Journal of the Acoustical Society of America*, 43(6):1255–1259.
(Cited on pages 10 and 33)

[251] Roginska, A. (2017). Binaural audio through headphones. In *Immersive Sound*, pages 88–123. Routledge.
(Cited on pages 40 and 53)

[252] Romanov, M., Berghold, P., Frank, M., Rudrich, D., Zaunschirm, M., and Zotter, F. (2017). Implementation and evaluation of a low-cost headtracker for binaural synthesis. *Journal of the audio engineering society*.
(Cited on page 23)

[253] Rose, J., Nelson, P., Rafaely, B., and Takeuchi, T. (2002). Sweet spot size of virtual acoustic imaging systems at asymmetric listener locations. *The Journal of the Acoustical Society of America*, 112(5):1992–2002.
(Cited on pages 31 and 121)

[254] Rossing, T. D. (2007). *Springer Handbook of Acoustics*. Springer Handbook of Acoustics. Springer-Verlag Berlin Heidelberg, Stanford, CA, 2 edition.
(Cited on pages 16, 19, 33, 36, 38 and 98)

[255] Rudenko, O. and Soluian, S. (1975). The theoretical principles of nonlinear acoustics. *Moscow Izdatel Nauka*.
(Cited on page 9)

[256] Rumsey, F. (2013). *Spatial Audio*. Focal Press, Burlington, MA, 2 edition.
(Cited on pages 30 and 98)

[257] Ruotolo, F., Maffei, L., Di Gabriele, M., Iachini, T., Masullo, M., Ruggiero, G., and Senese, V. P. (2013). Immersive virtual reality and environmental noise assessment: An innovative audio–visual approach. *Environmental Impact Assessment Review*, 41:10–20.
(Cited on page 16)

[258] Sabine, W. (1922). *Collected Papers on Acoustics*. Harvard University Press.
(Cited on page 34)

[259] Savioja, L., Huopaniemi, J., Lokki, T., and Vaananen, R. (1999). Creating interactive virtual acoustic environments. *Journal of the Audio Engineering Society*, 47:675–705.
(Cited on pages 1 and 57)

[260] Schepker, H., Haeder, K., Rennies, J., and Holube, I. (2016). Perceived listening effort and speech intelligibility in reverberation and noise for hearing-impaired listeners. *International Journal of Audiology*.
(Cited on pages 1 and 50)

[261] Schröder, D. (2011). *Physically Based Real-Time Auralization of Interactive Virtual Environments*. Aachener Beiträge zur Technischen Akustik. Logos Verlag Berlin.
(Cited on page 28)

[262] Schroeder, M. and Atal, B. (1963). Computer simulation of sound transmission in rooms. *Proceedings of the IEEE*, 51(3):536–537.
(Cited on page 22)

[263] Schroeder, M., Atal, B., and Bird, C. (1962). Digital computers in room acoustics. *Proc. 4th ICA, Copenhagen M*, 21.
(Cited on page 19)

[264] Schroeder, M. R. (1965). New method of measuring reverberation time. *The Journal of the Acoustical Society of America*, 37(3):409–412.
(Cited on page 144)

[265] Schroeder, M. R. (1979). Integrated impulse method measuring sound decay without using impulses. *The Journal of the Acoustical Society of America*, 66(2):497–500.
(Cited on page 144)

[266] Schröder, D., Pohl, A., Drechsler, S., Svensson, U. P., Vorländer, M., and Stephenson, U. M. (2013). openmat - management of acoustic material (meta-)properties using an open source database format. In *Proceedings of the AIA-DAGA 2013*.
(Cited on page 21)

[267] Schröder, D., Wefers, F., Pelzer, S., Rausch, D., Vorlaender, M., and Kuhlen, T. (2010). Virtual reality system at rwth aachen university. In *Proceedings of the International Symposium on Room Acoustics (ISRA)*.
(Cited on page 56)

[268] Seeber, B. U., Baumann, U., and Fastl, H. (2004). Localization ability with bimodal hearing aids and bilateral cochlear implants. *The Journal of the Acoustical Society of America*, 116(3):1698–1709.
(Cited on page 40)

[269] Seeber, B. U., Kerber, S., and Hafter, E. R. (2010). A system to simulate and reproduce audio–visual environments for spatial hearing research. *Hearing research*, 260(1):1–10.
(Cited on page 56)

[270] Seikel, J., King, D., and Drumright, D. (2015). *Anatomy & Physiology for Speech, Language, and Hearing*. Cengage Learning.
(Cited on page 14)

[271] Sette, W. J. (1933). A new reverberation time formula. *The Journal of the Acoustical Society of America*, 4(3):193–210.
(Cited on page 34)

[272] Shavit-Cohen, K. and Zion Golumbic, E. (2019). The dynamics of attention shifts among concurrent speech in a naturalistic multi-speaker virtual environment. *Frontiers in Human Neuroscience*, 13:386.
(Cited on pages 1 and 57)

[273] Shojaei, E., Ashayeri, H., Jafari, Z., Dast, M., and Kamali, K. (2016). Effect of signal to noise ratio on the speech perception ability of older adults. *Medical journal of the Islamic Republic of Iran*, 30:342.
(Cited on pages 1 and 55)

[274] Silzle, A., Kosmidis, D., Felix Greco, G., Beer, D., and Betz, L. (2016). The influence of microphone directivity on the level calibration and equalization of 3d loudspeakers setups. In *29th Tonmeistertagung - VDT International Convention 2016*.
(Cited on page 21)

[275] Simon, L. S. R., Dillier, N., and Wüthrich, H. (2021). Comparison of 3D audio reproduction methods using hearing devices. *Journal of the Audio Engineering Society*, 68(12):899–909.
(Cited on pages 21, 46, 93, 94 and 121)

[276] Simon, L. S. R., Wuethrich, H., and Dillier, N. (2017). Comparison of higher-order ambisonics, vector- and distance-based amplitude panning using a hearing device beamformer. In *Proceedings of 4th International Conference on Spatial Audio, Graz, Austria*.
(Cited on pages 20, 21, 23, 117, 121, 163 and 191)

[277] Simón Gálvez, M., Menzies, D., Fazi, F., de Campos, T., and Hilton, A. (2015). Listener tracking stereo for object based audio reproduction. In *Tecniacustica 2016 (Valencia)-European Symposium in Virtual Acoustics and Ambisonics*.
(Cited on page 27)

[278] Skudrzyk, E. (1971). *The foundations of acoustics: basic mathematics and basic acoustics*. Springer-Verlag.
(Cited on page 229)

[279] Solvang, A. (2008). Spectral impairment of two-dimensional higher order ambisonics. *J. Audio Eng. Soc*, 56(4):267–279.
(Cited on page 94)

[280] Spandöck, F. (1934). Akustische modellversuche. *Annalen der Physik*, 412(4):345–360.
(Cited on page 19)

[281] Spors, S., Teutsch, H., Kuntz, A., and Rabenstein, R. (2004). Sound field synthesis. In Huang, Y. and Benesty, J., editors, *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, pages 323–344. Springer US, Boston, MA.
(Cited on page 31)

[282] Spors, S., Wierstorf, H., Raake, A., Melchior, F., Frank, M., and Zotter, F. (2013). Spatial sound with loudspeakers and its perception: A review of the current state.
(Cited on pages 21, 23, 27, 40, 42 and 53)

[283] Stitt, P., Bertet, S., and Van Walstijn, M. (2013). Perceptual investigation of image placement with ambisonics for non-centred listeners. In *Proc. of the 16th Int. Conference on Digital Audio Effects (DAFx-13), Maynooth, Ireland*.
(Cited on pages 21, 46 and 49)

[284] Strauss, H. (1998). Implementing doppler shifts for virtual auditory environments. *Journal of the Audio Engineering Society*.
(Cited on page 20)

[285] Strumiłło, P. (2011). *Advances in Sound Localization*. a. InTech.
(Cited on pages 9 and 10)

[286] Sudarsono, A. S., Lam, Y. W., and Davies, W. J. (2016). The effect of sound level on perception of reproduced soundscapes. *Applied Acoustics*, 110:53–60.
(Cited on page 42)

[287] Søndergaard, P. and Majdak, P. (2013). The auditory modeling toolbox. In Blauert, J., editor, *The Technology of Binaural Listening*, pages 33–56. Springer, Berlin, Heidelberg.
(Cited on page 138)

[288] Tenenbaum, R. A., Camilo, T. S., Torres, J. C. B., and Gerges, S. N. (2007). Hybrid method for numerical simulation of room acoustics with auralization: part 1-theoretical and numerical aspects. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 29:211–221.
(Cited on page 68)

[289] Tremblay, P., Brisson, V., and Deschamps, I. (2020). Brain aging and speech perception: Effects of background noise and talker variability. *NeuroImage*, 227:117675.
(Cited on pages 1 and 55)

[290] Treviño, J., Okamoto, T., Iwaya, Y., and Suzuki, Y. (2011). Evaluation of a new ambisonic decoder for irregular loudspeaker arrays using interaural cues. In *Ambisonics Symposium*.
(Cited on page 94)

[291] Tu, W., Hu, R., Wang, H., and Chen, W. (2010). Measurement and analysis of just noticeable difference of interaural level difference cue. *2010 International Conference on Multimedia Technology*, pages 1–3.
(Cited on page 148)

[292] Van Wanrooij, M. M. and Van Opstal, A. J. (2004). Contribution of head shadow and pinna cues to chronic monaural sound localization. *Journal of Neuroscience*, 24(17):4163–4171.
(Cited on page 11)

[293] Vorländer, M. (2007). *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. RWTHedition. Springer Berlin Heidelberg.
(Cited on pages 2, 14, 15, 18, 19, 20, 21, 22, 33, 40, 42, 53 and 121)

[294] Vorländer, M. (2008). Virtual Acoustics: Opportunities and limits of spatial sound reproduction for audiology. *Hausdeshoerens-Oldenburg*.
(Cited on page 56)

[295] Vorländer, M. (2014). Virtual acoustics. *Archives of Acoustics*, vol. 39(No 3):307–318.
(Cited on page 40)

[296] Wallach, H. (1938). On sound localization. *The Journal of the Acoustical Society of America*, 10(1):83–83.
(Cited on page 10)

[297] Wang, D. and Brown, G. J. (2006). Binaural sound localization. In *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*, pages 147–185. Wiley.
(Cited on page 146)

[298] Wanner, L., Blat, J., Dasiopoulou, S., Domínguez, M., Llorach, G., Mille, S., Sukno, F., Kamateri, E., Vrochidis, S., Kompatsiaris, I., André, E., Lingenfelser, F., Mehlmann, G., Stam, A., Stellingwerff, L., Vieru, B., Lamel, L., Minker, W., Pragst, L., and Ultes, S. (2016). Towards a multimedia knowledge-based agent with social competence and human interaction capabilities. In *Proceedings of the 1st International Workshop on Multimedia Analysis and Retrieval for Multimodal Interaction*, MARMI '16, page 21–26, New York, NY, USA. Association for Computing Machinery.
(Cited on pages 1 and 57)

[299] Ward, D. B. and Abhayapala, T. D. (2001). Reproduction of a plane-wave sound field using an array of loudspeakers. *IEEE Transactions on Speech and Audio Processing*, 9(6):697–707.
(Cited on pages 31, 77 and 86)

[300] Wendt, D., Dau, T., and Hjortkjær, J. (2016). Impact of background noise and sentence complexity on processing demands during sentence comprehension. *Frontiers in Psychology*.
(Cited on page 51)

[301] Wendt, D., Hietkamp, R. K., and Lunner, T. (2017). Impact of noise and noise reduction on processing effort: A pupillometry study. *Ear and Hearing*.
(Cited on pages 50 and 101)

[302] Wendt, D., Koelewijn, T., Książek, P., Kramer, S. E., and Lunner, T. (2018). Toward a more comprehensive understanding of the impact of masker type and signal-to-noise ratio on the pupillary response while performing a speech-in-noise test. *Hearing Research*, pages 1–12.
(Cited on pages 50, 101 and 102)

[303] Westermann, A. and Buchholz, J. M. (2017). The effect of nearby maskers on speech intelligibility in reverberant, multi-talker environments. *The Journal of the Acoustical Society of America*, 141(3):2214–2223.
(Cited on pages 42 and 99)

[304] Whitmer, W. M. and Akeroyd, M. A. (2013). The sensitivity of hearing-impaired adults to acoustic attributes in simulated rooms. *Proceedings of Meetings on Acoustics*, 19(1):015109.
(Cited on pages 1, 18 and 50)

[305] Whitmer, W. M., Seeber, B. U., and Akeroyd, M. A. (2012). Apparent auditory source width insensitivity in older hearing-impaired individuals. *The Journal of the Acoustical Society of America*, 132(1):369–379.
(Cited on pages 16, 18 and 40)

[306] Wightman, F. L. and Kistler, D. J. (1992). The dominant role of low-frequency interaural time differences in sound localization. *The Journal of the Acoustical Society of America*, 91(3):1648–1661.
(Cited on page 10)

[307] Wightman, F. L. and Kistler, D. J. (1997). Monaural sound localization revisited. *The Journal of the Acoustical Society of America*, 101(2):1050–1063.
(Cited on page 11)

[308] Wilcox, R. (2004). Inferences based on a skipped correlation coefficient. *Journal of Applied Statistics*, 31(2):131–143.
(Cited on page 116)

[309] Williams, G. (1999). *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. Academic Press.
(Cited on pages 28 and 229)

[310] Wisniewski, M. G., Thompson, E. R., and Iyer, N. (2017). Theta- and alpha-power enhancements in the electroencephalogram as an auditory delayed match-to-sample task becomes impossibly difficult. *Psychophysiology*, 54(12):1916–1928.
(Cited on page 111)

[311] Wisniewski, M. G., Thompson, E. R., Iyer, N., Estepp, J. R., Goder-Reiser, M. N., and Sullivan, S. C. (2015). Frontal midline θ power as an index of listening effort. *Neuroreport*, 26(2):94—99.
(Cited on page 111)

[312] Wong, G. S. K. (1986). Speed of sound in standard air. *The Journal of the Acoustical Society of America*, 79(5):1359–1366.
(Cited on page 15)

[313] Wöstmann, M., Lim, S.-J., and Obleser, J. (2017). The Human Neural Alpha Response to Speech is a Proxy of Attentional Control. *Cerebral Cortex*, 27(6):3307–3317.
(Cited on page 111)

[314] Xie, B. (2013). *Head-related transfer function and virtual auditory display*. J. Ross Publishing.
(Cited on pages 22 and 70)

[315] Yost, W. (2013). *Fundamentals of Hearing: An Introduction*. Brill.
(Cited on page 8)

[316] Zapata Rodriguez, V., Jeong, C.-H., Hoffmann, I., Cho, W.-H., Beldam, M.-B., and Harte, J. (2019). Acoustic conditions of clinic rooms for sound field audiometry. In *Proceedings of 23rd International Congress on Acoustics*, pages 4654–59. Deutsche Gesellschaft für Akustik. 23rd International Congress on Acoustics , ICA 2019 ; Conference date: 09-09-2019 Through 13-09-2019.
(Cited on pages 122 and 139)

[317] Zekveld, A., Kramer, S., and Festen, J. (2011). Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response. *Ear and hearing*, 32:498–510.
(Cited on pages 1 and 55)

[318] Zekveld, A. A. and Kramer, S. E. (2014). Cognitive processing load across a wide range of listening conditions: Insights from pupillometry. *Psychophysiology*.
(Cited on pages 50 and 112)

[319] Zekveld, A. A., Kramer, S. E., and Festen, J. M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear and Hearing*.
(Cited on pages 50 and 118)

[320] Zhang, W., Samarasinghe, P., Chen, H., and Abhayapala, T. (2017). Surround by Sound: A Review of Spatial Audio Recording and Reproduction. *Applied Sciences*, 7(5):532.
(Cited on pages 19, 20, 21, 27 and 40)

[321] Zobel, B. H., Wagner, A., Sanders, L. D., and Başkent, D. (2019). Spatial release from informational masking declines with age: Evidence from a detection task in a virtual separation paradigm. *The Journal of the Acoustical Society of America*, 146(1):548–566.
(Cited on pages 16 and 40)

[322] Ľuboš Hládek, Ewert, S. D., and Seeber, B. U. (2021). Communication conditions in virtual acoustic scenes in an underground station.
(Cited on page 42)

[323] Şaher, K., Rindel, J. H., Nijs, L., and Van Der Voorden, M. (2005). Impacts of reverberation time, absorption location and background noise on listening conditions in multi source environment. In *Forum Acusticum Budapest 2005: 4th European Congress on Acustics*.
(Cited on page 50)

Appendix A

ITDs Ambisonics

Figure A.1, depicts ITDs for measurements with a listener (HATS manikin) in the center with Ambisonics (black line), in nine off-center positions combinations accompanied by a second listener (KEMAR) and alone in those three off center positions.

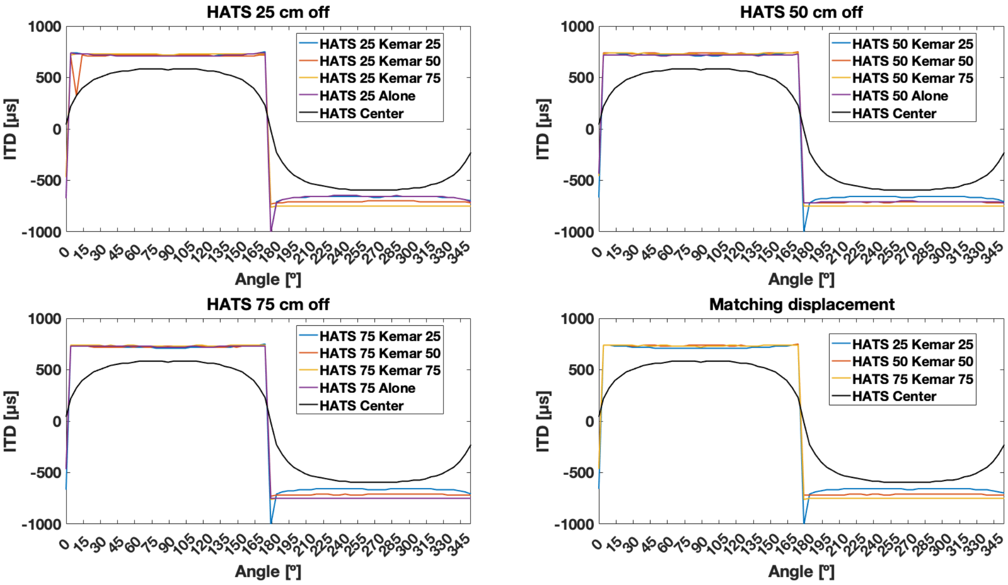


Figure A.1: *ITD as a function of source angle in Ambisonics virtualized setup. Top left HATS displacement = 25 cm, top right HATS displacement = 50 cm, bottom left HATS displacement = 75 cm, bottom right HATS displacement matching KEMAR displacement.*

Appendix B

Delta ILD Ambisonics

Figures B.1, B.2, and B.3, present the differences in ILD between center and off-center listener positions utilizing 24 loudspeakers to render an Ambisonics with a second listener present inside the ring of loudspeakers. In the figures, the number following H indicates the position of the main listener, while the numbers after K indicate the position of the second listener.

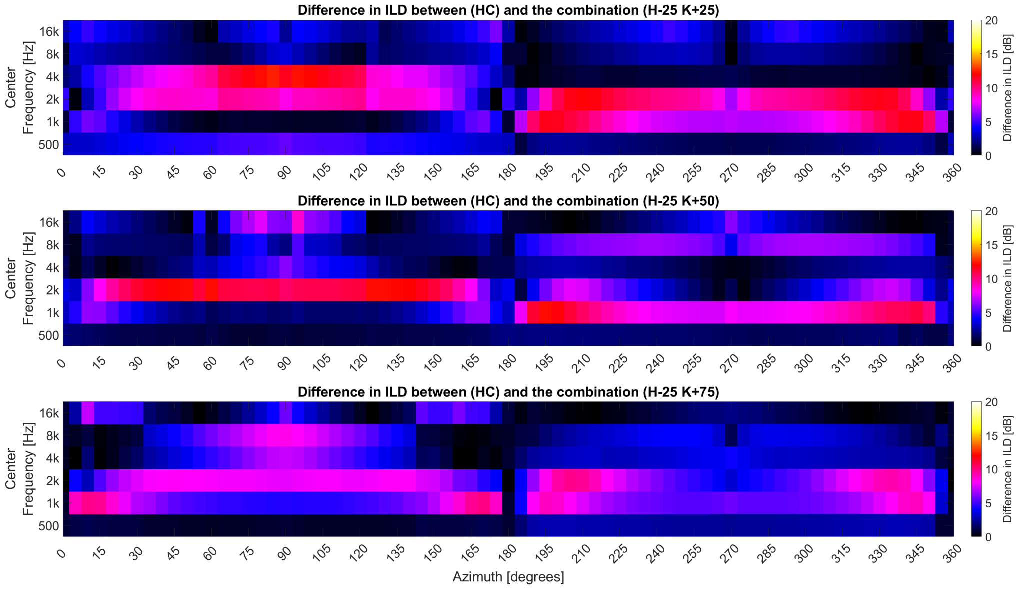


Figure B.1: *Differences in the ILD between centered setup and off-center setups: HATS at 25 cm to the right with: KEMAR at 25 cm to the left (top); KEMAR at 50 cm to the left (middle); KEMAR 75 cm to the left (bottom).*

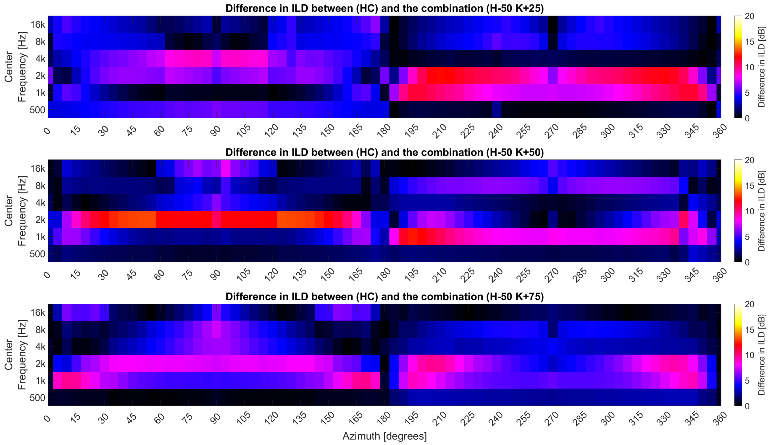


Figure B.2: *Differences in the ILD between centered setup and off-center setups: HATS at 50 cm to the right with: KEMAR at 25 cm to the left (top); KEMAR at 50 cm to the left (middle); KEMAR 75 cm to the left (bottom).*

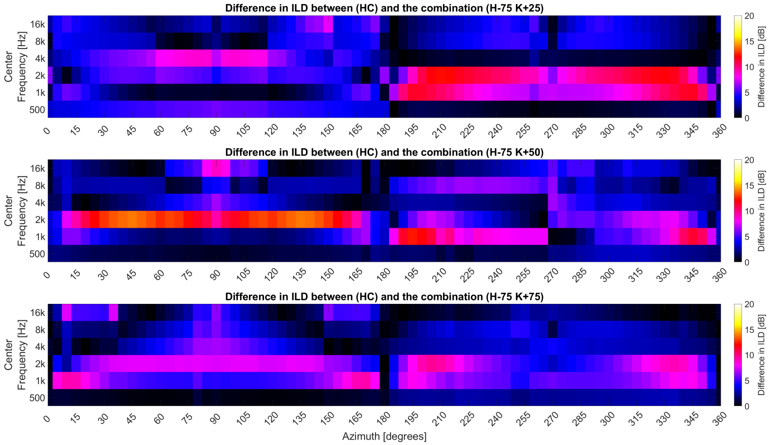


Figure B.3: *Differences in the ILD between centered setup and off-center setups: HATS at 75 cm to the right with: KEMAR at 25 cm to the left (top); KEMAR at 50 cm to the left (middle); KEMAR 75 cm to the left (bottom).*

Appendix C

Wave Equation and Spherical Harmonic Representation

Spherical harmonics (SH) represent spatial variations of an orthogonal set of solutions in the Laplace equation (orthonormal basis) when the solution is expressed in spherical coordinates, thus giving the spatial representation of weighted sums in spherical forms that represents a signal (space and frequency dependent).

C.1 Wave Equation in Spherical Coordinates

Expressing the wave equation in spherical coordinates (r, ϕ, θ) [36] we have:

$$\frac{\partial^2 p}{\partial r^2} + \frac{2}{r}\frac{\partial p}{\partial r} + \frac{1}{r^2 \sin(\theta)}\frac{\partial}{\partial \theta}\left(\sin(\theta)\frac{\partial p}{\partial \theta}\right) + \frac{1}{r^2 \sin^2(\phi)}\frac{\partial^2 p}{\partial \phi^2} - \frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2} = 0\,, \quad \text{(C.1)}$$

C.2 Separation of the Variables

The differential equation solution tool called *separation of variables* can be used for the Equation C.1, being formulated from the product of three space dependent variables and a time dependent variable:

$$p(r, \theta, \phi, t) = R(r)\Theta(\theta)\Phi(\phi)T(t)\,. \quad \text{(C.2)}$$

With the separation of the variables, according to Skudrzyk [278], there are four homogeneous differential equations:

$$\frac{\mathrm{d}^2\,\Phi}{\mathrm{d}\phi}+m^2=0\,, \tag{C.3a}$$

$$\frac{1}{\sin\theta}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\sin\theta\frac{\mathrm{d}\Theta}{\mathrm{d}\theta}\right)+\left[n(n+1)-\frac{m^2}{\sin^2\theta}\right]\Theta=0\,, \tag{C.3b}$$

$$\frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2\frac{\mathrm{d}R}{\mathrm{d}r}\right)+k^2R-\frac{n(n+1)}{r^2}R=0\,, \tag{C.3c}$$

$$\frac{1}{c^2}\frac{\mathrm{d}^2\,T}{\mathrm{d}t^2}+k^2T=0\,. \tag{C.3d}$$

where m and n integers, the general solutions to the equations are

$$\Phi(\phi)=\Phi_1\,\mathrm{e}^{\mathrm{j}m\phi}+\Phi_2\,\mathrm{e}^{-\mathrm{j}m\phi}\,, \tag{C.4a}$$

$$\Theta(\theta)=\Theta_1P_n^m(\cos(\theta))+\Theta_2Q_n^m(\cos(\theta))\,, \tag{C.4b}$$

$$R(r)=R_1h_n^{(1)}(kr)+R_2h_n^{(2)}(kr)\,, \tag{C.4c}$$

$$T(\omega)=T_1\,\mathrm{e}^{\mathrm{j}\omega t}+T_2\,\mathrm{e}^{-\mathrm{j}\omega t}\,, \tag{C.4d}$$

where $h_n^{(1)}(x)$ and $h_n^{(2)}(x)$ are the first and second-kind spherical Hankel functions that represent convergent and divergent waves depending on the signal agreed for the time and $P_n^m(x)$ and $Q_n^m(x)$ are the associated Legendre functions of the first and second type.

Due to the singularities in the poles of Legendre's associated functions at $\theta=0$ and $\theta=\pi$ the term Θ_2 is treated as null, and for simplification, you can use the positive m variable or negative, so the term Φ_2 is also null. According to Williams [309], for there to be no singularities in the poles of Legendre's associated functions, the n index must be an integer. Still, considering causal systems, the term T_2 in C.4d is equal to 0 given the convention used.

The associated Legendre functions of the first type defined for positive degrees m are

$$P_n^m(x) = (1)^m (1-x^2)^{\frac{m}{2}} \frac{\mathrm{d}^m}{\mathrm{d}x^m} P_n(x)\,. \tag{C.5}$$

Meanwhile, the functions for negative degrees $-m$ are given by

$$P_n^{-m} = (-1)^m \frac{(n-m)!}{(n+m)!} P_n^m(x)\,, \tag{C.6}$$

P_n being the Legendre Polynomial given by

$$P_n(x) = \frac{1}{2^n n!} \frac{\mathrm{d}^n}{\mathrm{d}x^n} (x^2-1)^n\,. \tag{C.7}$$

C.3 Spherical Harmonics

Equations C.4a and C.4b admit periodic solutions in angular coordinates, and combined are called spherical harmonics of order n and degree m defined by

$$Y_n^m(\theta,\phi) = \sqrt{\frac{(2n+1)}{4\pi}\frac{(n-m)!}{(n+m)!}} P_n^m(\cos(\theta))\,\mathrm{e}^{\mathrm{j}m\phi}\,. \tag{C.8}$$

The negative order SH functions are obtained through the relation

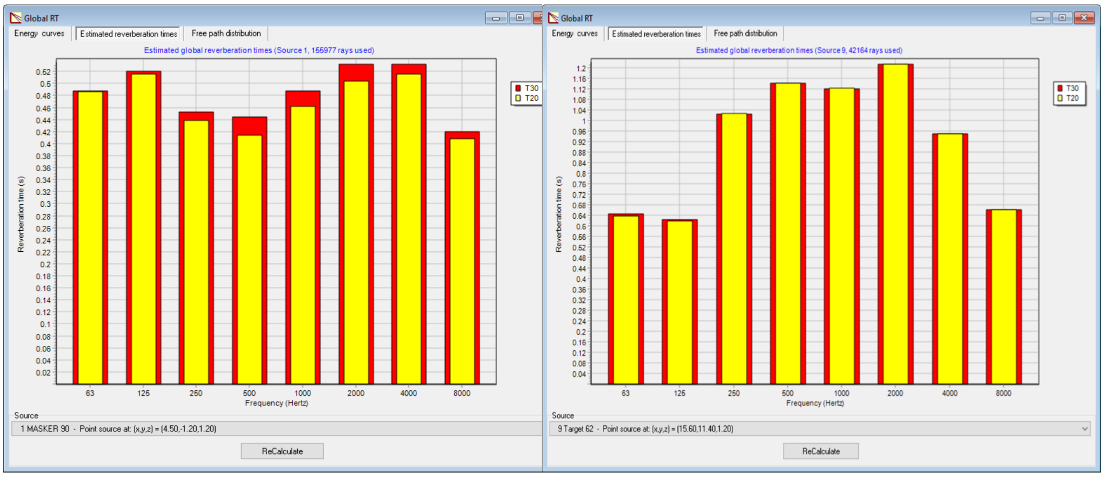
$$Y_n^m(\theta,\phi) = (-1)^m \cdot (Y_n^- m(\theta,\phi))^*\,, \tag{C.9}$$

where * denotes the conjugate complex, and demonstrates that only the phase changes between the positive and negative degrees of the function. Thus the magnitude is commonly expressed with the radius and the phase in terms of color or color scale, as in Figure 2.9.

Appendix D

Reverberation time in Acoustic Simulation

The reverberation time for the classroom and restaurant are presented in Figure D.1



Source
1 MASKER 90 - Point source at: (x,y,z) = (4.50,-1.20,1.20)

Source
9 Target 62 - Point source at: (x,y,z) = (15.60,11.40,1.20)

(a) **(b)**

Figure D.1: *Reverberation time (a) Classroom (b) Restaurant in octave bands*

Appendix E

Alpha Coefficients

Figures E.1, E.2, and E.3 presents the absorption coefficients according to the frequency introduced in the ODEON software to simulate the environments.

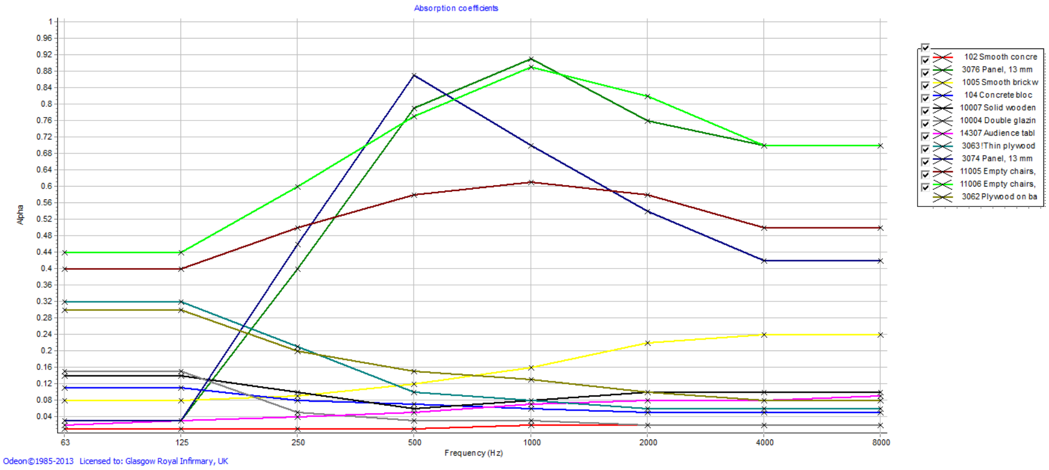


Figure E.1: *Classroom alpha coefficients (ODEON software).*

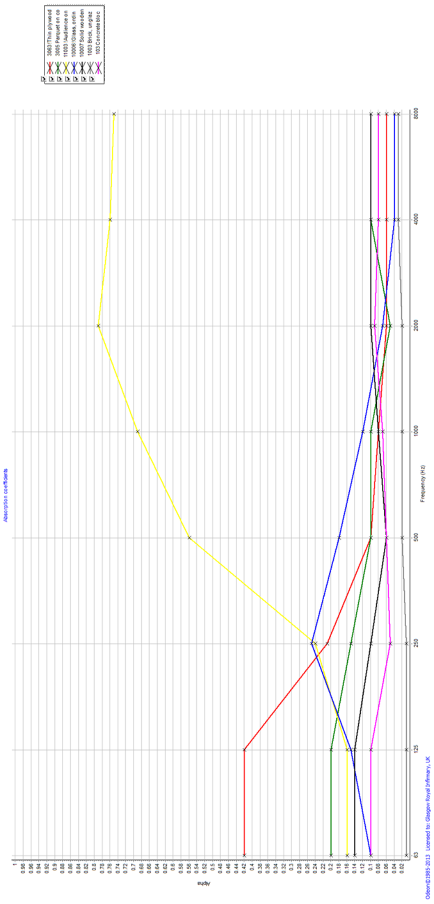


Figure E.2: *Restaurant alpha coefficients (ODEON software).*

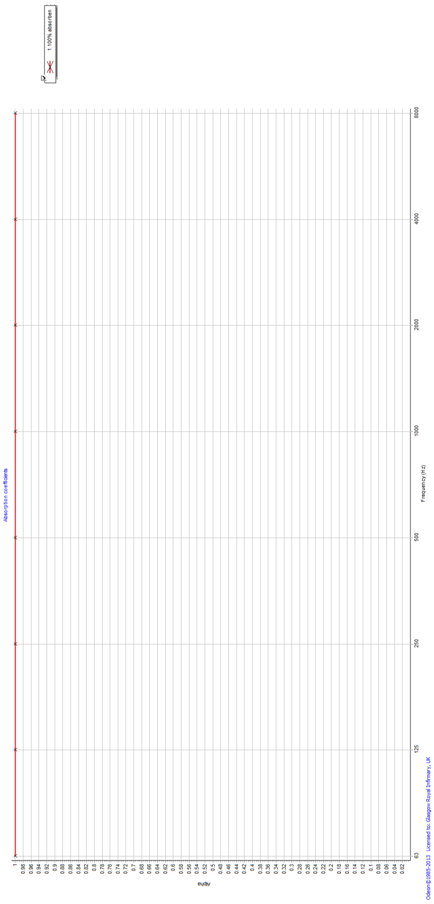
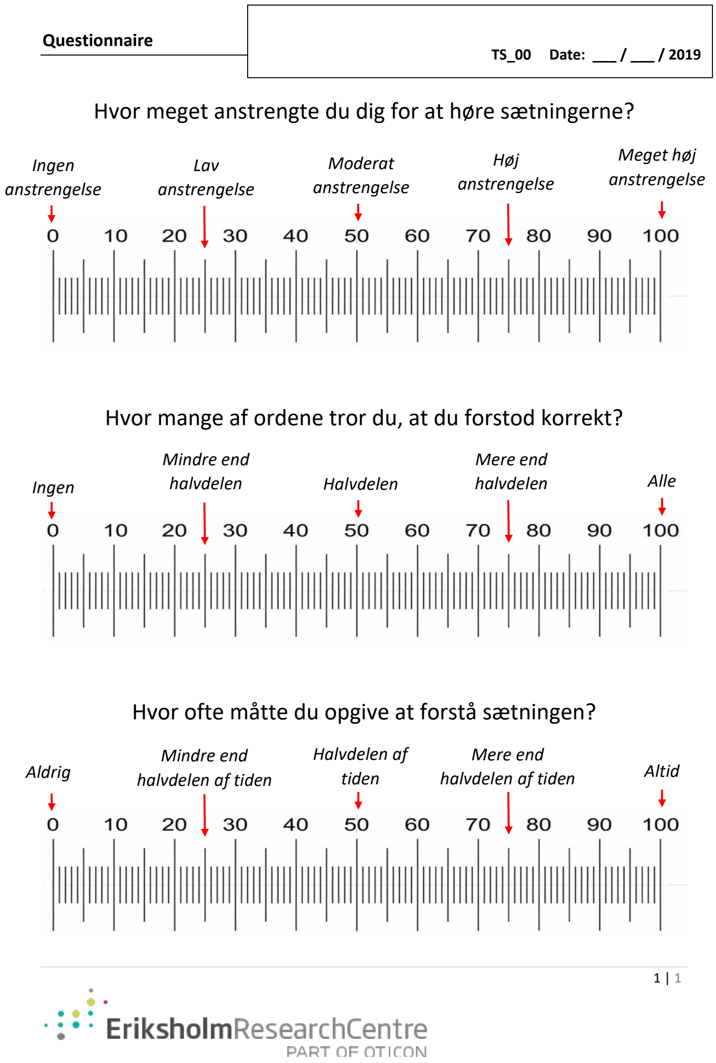


Figure E.3: *Anechoic room alpha coefficients.*

Appendix F

Questionnaire



Questionnaire

TS_00 Date: ___ / ___ / 2019

Hvor meget anstrengte du dig for at høre sætningerne?

Ingen anstrengelse · *Lav anstrengelse* · *Moderat anstrengelse* · *Høj anstrengelse* · *Meget høj anstrengelse*

0 10 20 30 40 50 60 70 80 90 100

Hvor mange af ordene tror du, at du forstod korrekt?

Ingen · *Mindre end halvdelen* · *Halvdelen* · *Mere end halvdelen* · *Alle*

0 10 20 30 40 50 60 70 80 90 100

Hvor ofte måtte du opgive at forstå sætningen?

Aldrig · *Mindre end halvdelen af tiden* · *Halvdelen af tiden* · *Mere end halvdelen af tiden* · *Altid*

0 10 20 30 40 50 60 70 80 90 100

1 | 1

ErikshɔlmResearchCentre
PART OF OTICON