POLITECNICO DI TORINO

Master's Degree in Cinema and Media Engineering

Master's Degree Thesis

Physical Acoustical Validation of the Audio Space Lab at the Polytechnic of Turin

Supervisors

Prof. Arianna ASTOLFI

Dott.ssa Angela GUASTAMACCHIA

Prof. Louena SHTREPI

Candidate

Riccardo LACQUA

July 2024

A mia mamma, la linfa vitale che mi ha permesso di arrivare fino a qua,

A mio papà, la mia corazza nel posto giusto al momento giusto,

Ai miei angeli custodi, Nonno Franco e Nonna Anna,

A mio fratello, compagno di viaggio per la vita.

Abstract

Hearing loss, affecting about 5% of the world's population, is the partial or complete inability to perceive sounds. Hearing Aids (HAs) can help mitigate this problem. However, HAs require fine-tuning, usually done through simple tests in the lab or after in-field trials by users in daily life, involving several audiological visits to achieve a satisfactory fit.

Using augmented reality and immersive audio to recreate daily auditory scenarios in the lab can expedite the fitting of HAs. However, establishing such advanced laboratories is expensive and complex.

This study aims to validate a low-cost immersive audiovisual laboratory suitable for hospitals and clinics, namely the Audio Space Lab (ASL) at the Polytechnic University of Turin, Italy. The ASL is equipped with a spherical system of 16 speakers driven by 3° order ambisonics spatial audio rendering and synchronized with a Head-Mounted Display (HMD). The ASL reproduces a sound field at the center of the array, i.e., the sweet spot, along with a 360° stereoscopic visual scene. However, this study focused only on the validation of the spatial audio playback. The study addresses solely the physical validation of audio reproduction by performing dual comparisons between different conditions. The evaluation was based on objective metrics that model human binaural listening, such as Interaural Time Difference (ITD), Interaural Level Difference (ILD), and Inter-Aural Cross-Correlation (IACC).

The validation was divided into two main phases: intra-ASL and inter-ASL validation. The intra-ASL validation focused on comparing different conditions recreated within the laboratory to select the best-performing configuration. While, during the inter-ASL validation, the acoustic characteristics of a real classroom were compared with those measured in the ASL when the same classroom was acoustically virtualized to verify the accuracy of the final audio reproduction.

In the intra-ASL validation, the influence of the chair headrest and the HMD on the sound field was examined. The ITD and ILD analyses favored a chair without a headrest and showed that the HMD does not introduce significant changes to the sound field. Additionally, room acoustic treatment was tested under three conditions: adding absorbing panels behind specific speakers, covering a window with an absorbing curtain, and reinforcing the corners and walls adjacent to the window with panels. Analyses of the ITD and ILD indicated that the first acoustic condition provided the most noticeable improvement. Measurements were carried out to assess how the accuracy of the reproduced sound field deteriorates with increasing distance from the sweet spot, particularly at distances of 0, 10, and 20 cm.

Results from ILD and ITD analyses showed poorer performance as the distance from the sweet spot increased. However, at distances within 10 cm from the sweet spot, the reconstructed sound field can still be considered as sufficiently accurate.

Inter-ASL validation showed that the ASL can adequately reproduce real environments, but not perfectly. ITD and ILD analyses showed differences between the real and virtual classroom within the just noticeable difference limits, while IACC analyses showed values above those limits.

In conclusion, the physical acoustical validation yielded overall satisfactory results. Future work will focus on the perceptual validation of the ASL to confirm the current results and determine whether the IACC inaccuracies are perceivable by humans.

Table of Contents

Chapter 1

Introduction

With this work, we aim to validate the physical aspects of immersive audio in the Audio Space Lab, using comparative measurements conducted both within the lab and externally but reproduced within it. The ultimate goal of validating this system is to implement it in a clinical setting for the calibration of cochlear implants and hearing aids. The intent is to calibrate these devices directly at the clinical stage using a laboratory with a spherical array of speakers that can accurately reproduce real-world soundscapes. This approach would eliminate the need for cochlear implant and hearing aid users to test the devices externally and then return for calibration. Instead, the calibration would be performed directly in the clinic through the auditory, and possibly visual, reproduction of real-life scenes.

- **Chapter 2 - State of the Art**
  This chapter describes the current state of reproduction systems in a spatialized setting. It includes a report on studies conducted so far in clinical settings, whether planned for clinics or already carried out in specialized hospitals. Validations of spatial reproduction systems relevant to our purpose are analyzed. Details on how these systems were validated, the test procedures used, and the parameters considered for proper validation are provided. Additionally, the most significant metrics for our validation are described.

- **Chapter 3 - Audio Space Lab**
  This chapter describes the structure of the laboratory where the validation of the spatialized audio playback system, the Audio Space Lab, located in the Department of Energy at the Polytechnic University of Turin (Italy), took place. It details the construction of the loudspeaker array, the acoustic treatment of the lab, and the configuration of the spatial reproduction system. Additionally, it describes the initial validation of the laboratory that was carried out earlier.

- **Chapter 4 - Validation**
  This chapter deals with the the physical validation of the laboratory, describing the measurements made. The validation includes Intra-ASL measurements, conducted entirely within the laboratory by generating sound in situ, and Inter-ASL measurements, conducted within the laboratory but reproducing recordings made at an earlier time in a university classroom (1T classroom of the Polytechnic University of Turin). Both types of validation are analyzed from a physical point of view, and the results are reported.

- **Chapter 5 - Discussions and Conclusions**
  This chapter describes the validation process, outlining the steps taken to reach the conclusions. The results obtained are discussed, and recommendations for modifications to improve the system's efficiency are provided. Analyses and future work are suggested to achieve complete laboratory validation, enabling the use of this system directly in specialized clinics for the calibration of hearing aids and cochlear implants.

Chapter 2

State of the art

2.1 Spatial Audio Reproduction Systems for Clinical Applications

Studies with the same intent as ours or approaching our results were considered. This includes studies conducted in clinics or hospitals that used a spatial audio system similar to the one we used and involved people with hearing disabilities.

In the study [1], researchers evaluated sound localization in the horizontal plane amidst noise to assess the sound recognition ability of normal-hearing individuals and those with unilateral and bilateral cochlear implants (CI). This study was intended for use in specialized clinics and was significant because it was the first to demonstrate the improvement of bilateral cochlear implants over unilateral ones, with comparisons made against normal-hearing participants.

The experiments were conducted using a "Simulated Open Field Environment," where 36 speakers were arranged at 10-degree intervals, covering the entire 360-degree horizontal plane. The speakers between -120 and +120 degrees azimuth were concealed behind acoustically transparent curtains. Participants sat facing the speaker at 0 degrees azimuth, with their ears aligned to the speaker height and their heads resting on a headrest. Sounds were randomly played from speakers positioned at 0, ±10, ±20, ±40, ±60, and ±80 degrees. Participants identified the sound source by pointing a trackball in the perceived direction and pressing the left button for sounds perceived in front and the right button for sounds perceived from behind.

Target sounds were presented both in quiet conditions and with diffuse background noise. Six normal-hearing individuals and 14 CI users (4 unilateral and 10 bilateral) took part in the study. The target sounds consisted of trains of six pulses, each lasting 10ms, separated by 120ms pauses, resulting in a total duration of 660ms. These sounds also featured 2ms ramps.

The background noise was a stationary uniform exciting noise (UEN), matching the passband of the target pulse trains. When both target sound and background noise were played, their playback was not simultaneous but staggered slightly to improve separation. Additionally, Speech Reception Threshold (SRT) was measured for all CI users using five different lists of phrases, with the median SRT results representing overall speech performance in noise.

For normal-hearing subjects, results showed that in silent conditions, localization responses were near the actual target sound locations. With background noise, responses were similarly accurate, although they shifted closer to the median plane as the signal-to-noise ratio (SNR) decreased.

Subjects with unilateral CI displayed poor localization responses in both conditions. The decline in performance with background noise was less pronounced than for normal-hearing subjects, as localization was already poor in silent conditions.

Conversely, bilateral CI subjects' localization ability closely resembled that of normal-hearing subjects, showing very good results in silent conditions and reduced performance in noise, improving as the SNR decreased. Although their localization accuracy was not as high as that of normal-hearing subjects, it was significantly better than that of unilateral CI users.

In contrast, an alternative approach to spatial sound reproduction was proposed in [2]. Rather than using a conventional spatial loudspeaker array system, which can be costly and large, the study explored reproducing the spatial audio system using only two loudspeakers in a compact virtual acoustic setup. This system was designed for clinic use and validated within a clinical audiology booth.

The experiment aimed to assess the accuracy of hearing aids with this system and to evaluate the impact of involuntary head movements typical in clinical settings, despite participants being instructed to remain still during tests.

Two environments were selected for comparison: an ideal anechoic chamber, representing a highly acoustically treated environment, and a clinical audiology booth, representing a typical testing environment.

The experiment utilized a KEMAR head and torso simulator fitted with hearing aids, mounted on a stand to facilitate movement and rotation. A HTC VIVE tracker was employed to record head movements, which was detached and reattached for each measurement.

The system featured a 2-speaker, 6-channel array arranged in a semicircle, with each speaker positioned at ear height, 1.5 meters from the center of KEMAR's head in the anechoic chamber and 1 meter in the booth due to space constraints.

Results indicated that the system could accurately reproduce hearing aid signals, even in the presence of minimal head rotations (approximately 2 degrees). It also performed well with rotations up to 10 degrees, although adjustments were needed to accommodate changes in head position.

These findings highlight the system's potential for clinical adaptation, enabling the reproduction of complex acoustic environments with a minimal number of speakers. For enhanced clinical application, the study suggested increasing the number of speakers to a maximum of 6, which would improve control over the sound field at each location despite increasing system costs.

In another study, [3] investigated the accuracy of sound source localization, specifically identifying left or right side, among children with normal hearing and those fitted with bilateral and unilateral cochlear implants (CI). This study is among the first to explore localization abilities in children with CI, although further research is needed to replicate these experiments in clinical settings that simulate more complex real-world environments while maintaining similar results.

The experiments were conducted in an acoustically treated booth using an array of 15 speakers placed at 10-degree intervals within the frontal hemisphere (-70° to +70°). Each speaker was positioned at ear level, 1.2 meters from the center of the child's head. Children were seated facing the center speaker (0°), and each speaker was associated with a visual image to aid localization tasks.

For stimuli, after initial attempts using 3 bursts of 25ms pink noise, the word "baseball" was employed as a verbal stimulus, found to be more effective. Twenty-one children with bilateral CI and 7 with normal hearing (NH) participated in the study. Among the children with CI, 11 were reassessed after 7 months to evaluate localization improvements following extended bilateral CI use.

During the localization task, children used a computer mouse to select the visual image corresponding to the heard speaker. Visual feedback was provided with the selected image flashing after each selection. Localization tests were initially conducted bilaterally for children with CI and subsequently unilaterally (with the first implanted device).

To compare results across different groups (NH, bilateral CI, and unilateral CI), individual scatter plots were analyzed for localization accuracy, and graphs displaying relative Root Mean Square (RMS) errors were generated.

The analysis revealed superior localization accuracy among NH subjects, with errors ranging from 9° to 26°, compared to bilateral CI subjects, who exhibited errors ranging from 19° to 56°. Nevertheless, both groups demonstrated comparable performance, albeit with slightly degraded accuracy in the CI group. The study also highlighted the advantage of bilateral over unilateral implantation in achieving better localization, as evidenced by significantly lower RMS errors.

Furthermore, improvement in localization ability was observed in children tested after 7 months of bilateral CI use, indicating adequate development of bilateral localization skills even after initial adaptation to unilateral CI.

[4] conducted a study to evaluate the sound localization abilities of individuals with cochlear implants (CI), focusing on bilateral versus unilateral localization and exploring correlations between localization and speech intelligibility in the presence of interference. The experiments were carried out across multiple clinical sites using a portable testing apparatus.

Tests took place in soundproof booths equipped with an array of 8 matched speakers arranged in a semicircular structure with a one-meter radius, spanning from -70° to +70° azimuth. Each speaker was positioned 20° apart and numbered sequentially from 1 to 8 (left to right).

Stimuli consisted of four 170ms bursts of pink noise with 10ms rise and fall times and a 50ms inter-stimulus interval. During each listening mode, stimuli were randomly presented 20 times per speaker. Participants were instructed to keep their heads straight and minimize head movements. Background noise was introduced from the 0°, 90°, or -90° azimuth positions corresponding to front, right, or left locations, respectively.

Seventeen subjects with CI participated in the study, undergoing unilateral (left and right) and bilateral testing in three different modes, randomized for each subject. Tests were repeated at 3-month intervals over 6 months to assess localization improvements with bilateral experience.

For localization assessments, participants indicated the number of the speaker perceived as the sound source after each test without immediate feedback. They then made a forced-choice selection from 8 options.

Speech intelligibility in noise was evaluated using the Bamford-Kowal-Bench-speech-in-noise (BKB-SIN) test, which comprises 36 lists balanced for difficulty. Each list includes sentences with keywords presented alongside noise from four speakers.

Statistical analysis included the Kruskal-Wallis test for assessing left and right hemisphere lateralization separately across all listening modes, and Root Mean Square (RMS) calculation to evaluate localization accuracy across both hemispheres.

Results indicated that 82% of participants showed improved discrimination of the involved hemisphere (right/left), with 47% achieving better localization accuracy within the targeted hemisphere. However, improvements were less pronounced when the target sound and interference were completely coincident. Significant enhancements in bilateral localization were observed over the 3- and 6-month test intervals.

Speech intelligibility scores correlated closely with localization abilities, indicating that those who could better discern speech in noise also exhibited improved localization of sound sources. As expected, bilateral CI users generally outperformed unilateral CI users across all measures.

[5] investigated auditory spatial localization and discrimination in individuals with unilateral deafness and unilateral cochlear implants (CI), conducted within hospital settings. The study aimed to explore the feasibility of conducting such tests directly in clinics to streamline implementation time and costs.

Testing took place in a soundproof anechoic room equipped with 47 speakers arranged in a semicircular layout with a radius of 2.35m, spanning from -98° to +98° azimuth at 4.3° intervals. Each speaker was numbered for participant reference, and the entire array was covered with acoustically transparent black gauze to obscure speaker positions.

Additionally, the setup included an array of 188 white light-emitting diodes (LEDs) mounted at eye level in 1° azimuthal steps. These LEDs were linked to 51 printed circuit boards atop the speakers, each board integrating four infrared-sensitive (IR) phototransistors. A custom IR flashlight served as a targeting system for participants to indicate their perceived sound source direction, with the corresponding LED flashing upon selection.

Stimuli consisted of high- and low-frequency Gaussian noises to facilitate Interaural Time Difference (ITD) and Interaural Level Difference (ILD) calculations, respectively. These noises spanned two-octave bands and lasted 500ms.

Eighteen participants with unilateral CI (with normal hearing in the opposite ear) were involved, their results compared against 120 normal-hearing individuals matched by age groups. Auditory localization was tested at six azimuthal positions: ±4°, ±30°, and ±60°, each tested five times in random sequence. Participants were allowed to move their heads freely and reposition themselves facing 0° before indicating the sound direction with the IR flashlight.

Analysis focused on ITD and ILD discrimination, assessed through Just Noticeable Difference (JND) calculations. Notably, only six participants demonstrated significant ITD utilization for low-frequency discrimination, while ILD sensitivity did not exhibit consistent changes with sound source location.

Statistical analysis, including ANOVA, evaluated localization performance across selected positions. Results indicated that eight participants showed robust localization skills, while the remaining ten exhibited varying degrees of difficulty.

Overall, the study suggests that CI can enhance auditory localization, although effectiveness varies among individuals with unilateral CI. Challenges persist in localizing sounds directly in front for both CI and normal-hearing ears, highlighting differences in spatial discrimination capabilities despite improved localization abilities facilitated by CI.

In [6], a theoretical framework for a real-time extended aural binaural system is proposed, aimed at creating virtual acoustic environments for calibrating hearing aids (HA) and enhancing their real-world performance without actual experiments or participant involvement.

The paper begins with a review of existing sound field reproduction systems and their techniques, highlighting inherent errors and shortcomings identified through various metrics.

The concept introduces a real-time binaural extended aural system capable of simulating complex acoustic environments using a 3D model. This involves simulating acoustic parameters within a controlled laboratory environment with stable acoustic conditions, typically achieved in an acoustically treated booth suitable for auditory testing. The setup includes a motion tracker to synchronize scene simulation with user movements, enhancing realism. Sound sources are reproduced through a set of four loudspeakers strategically positioned at azimuths of n*45° (where n=1, 3, 5, 7) and 1.2 meters from the listener's auditory center axis.

Room Impulse Responses (BRIR) and Hearing Aid Impulse Responses (HARRIR) are computed, with playback signals simulated based on Head-Related Transfer Functions (HRTF) and Hearing Aid Response Transfer Functions (HARTF). These functions are derived from measurements conducted either on an adult artificial head or individually mapped across a dense spatial grid.

Although no experimental data is presented, the study outlines promising outcomes for implementing such a system in clinical settings with spatial constraints. The proposed system aims to bridge the gap between hearing aid calibration in controlled laboratory conditions and their performance in real-world settings, often falling short of expectations.

In [7], a portable virtual reality (VR) system designed for assessing localization abilities in individuals with bilateral Cochlear Implants (CI) is introduced. This system utilizes a VR headset to overcome the limitations of traditional oversized speaker setups, making it suitable for testing in various settings such as clinics, homes, or offices.

Localization tests were conducted using HRTF (Head-Related Transfer Function) data recorded at 0° elevation across 72 azimuth positions (0° to 360°), with each position modeled using finite impulse response (FIR) filters tailored for each ear.

The VR environment, developed using Unity, featured a virtual room with 13 speakers positioned from -90° to +90° at 15° intervals, all placed at 0° elevation. An Oculus VR headset was employed for immersive visualization and included head-tracking sensors.

Participants included four individuals with bilateral CI and 12 with normal hearing (NH). A 5-second white noise served as the stimulus, randomly played from one of the 13 speaker locations. Participants indicated the perceived direction by turning their heads towards the sound source and maintained the position for 2 seconds to record their response. Visual feedback confirmed the recorded response.

NH subjects underwent 6 test sessions, while CI participants completed 9 sessions under three conditions: using the first implant, the second implant, and both implants simultaneously.

Key metrics analyzed included signal response time, left/right discrimination accuracy, percent correct identification, and root mean square error (RMS). NH subjects demonstrated faster response times and superior left/right discrimination (98% accuracy) compared to CI subjects, particularly when using bilateral CI or the first implant alone. Percent correct identification showed no significant differences between NH and CI subjects, although NH participants generally outperformed CI participants. RMS values indicated NH subjects had more accurate localization (12.4° RMS) compared to CI subjects (23.81° RMS).

Overall, the study highlighted differences in localization abilities between NH subjects and those with CI, with NH subjects exhibiting superior performance. However, CI subjects showed improved localization when using bilateral implants or both implants simultaneously, underscoring the system's utility in assessing CI outcomes.

In [8], an experiment focusing on sound source localization in the frontal plane using an augmented reality system was conducted, demonstrating significant relevance for clinical applications, particularly in subjects with Cochlear Implants (CI).

The experiment took place at the Audiology Pavilion of the Institute for Maternal and Child Health IRCCS "Burlo Garofolo." The setup included 13 speakers arranged in a semicircle with a radius of 1.4m, spanning from -90° to 90° at 15° intervals. An Oculus VR headset was utilized to monitor head movements, projecting a natural landscape created in Unity3D. Participants were encouraged to move their heads freely and use a laser pointer to indicate the perceived direction of sound sources.

Thirty-seven individuals with normal hearing participated, exposed to a 200ms pink noise stimulus with specified onset and offset ramps and pauses between stimuli. The experiment comprised two conditions: real and virtual. In the real condition, participants saw and heard the actual speaker array and pointed with a laser pointer. Conversely, in the virtual condition, participants wore the VR headset, experiencing a simulated environment without direct visibility of speakers, again using the laser pointer for localization. Each position was tested five times, totaling 65 positions for identification.

Statistical analyses focused on variance and mean error between target and perceived positions. Results indicated no significant differences in mean head divergence between conditions. However, head divergence was slightly higher in the virtual condition compared to the real condition.

The findings suggest that visibility of speakers in the real condition led to better localization accuracy, consistent with previous research demonstrating a normal 1° rightward shift.

A correlation was noted between head divergence and localization error, as some participants instinctively turned their heads towards perceived sources, impacting accuracy. Nonetheless, this head movement did not influence localization significantly. The presence of the VR headset may have introduced minor localization errors, aligning with expected outcomes. Overall, the study underscores the potential of such systems in clinical settings for physiological studies on auditory systems and validations for cochlear implants and hearing aids.

2.2 Validation of Spatial Audio Reproduction System

To validate a spatial sound reproduction system, two types of validations can be considered: physical validation and perceptual validation. In this study, we focused on the physical validation of the Audio Space Lab, which will be covered in detail in later chapters. Perceptual validation is left as future work to achieve a complete validation of the lab.

2.2.1 Physical Validation

The physical validation consists of an objective evaluation through the analysis of acoustic parameters, without the use of human subjects. Different measurement tools are used to calculate monaural parameters, using directional microphones, and binaural parameters, using head-torso simulators. This section also describes the configuration of the playback system, as well as the techniques for encoding and decoding audio information. A complete description of the physical space where the system is installed and its operation is provided.

The analyzed parameters are then divided into monaural and binaural categories, which are the main cues for sound localization:

- **Binaural**: these cues are based on localizing a sound within the frontal plane. This type of cue relies on timing differences, known as Interaural Time Differences (ITD), where a sound arrives at one ear before the other, and intensity differences, known as Interaural Level Differences (ILD), where a sound is perceived louder in one ear compared to the other.

  These cues enable the determination of whether a sound originates from the right or left. One more parameter considered is the Interaural Cross Correlation (IACC). This parameter was utilized in our validation, particularly with the use of a head-torso simulator (HATS).

- **Monoaural**: these cues are based on the localization of a sound along the vertical plane. This is achieved through the shape of the outer ear, which alters the spectral characteristics of the sound. The brain utilizes these cues to deduce information about the vertical position of a sound source and its front-back identification. The most relevant parameters are reverberation time (RT) and speech clarity (C50).

2.2.2 Perceptual Validation

The perceptual validation, on the other hand, refers to an evaluation performed with human subjects, with subjective results based on their lived experiences. The same measurements made in physical validation are then re-presented to people to assess the immersiveness of the system. This assessment can be conducted through questionnaires or by calculating speech intelligibility. This section also lists the number of participants in the experiment and their biographical characteristics. The stimuli used as targets are described, as well as how they are used. If there is a localization part, its operation is explained, along with an explanation of the feedback used.

Subjective evaluation can then be implemented through the use of questionnaires or the calculation of speech intelligibility:

- **Questionnaires**: these questionnaires are provided to be filled out by the user after the end of the experiment. They include closed or open-ended questions about the words heard during the experiment (if not already requested during the test), to facilitate the analysis of speech intelligibility. Additionally, users are asked a series of questions to rate their enjoyment of the test, assessing how immersed they felt in the system and whether it compared to a real environment. An open section may also be included for additional comments and suggestions aimed at improving the implementation of the system.

- **Speech Intelligibility**: refers to speech intelligibility, which measures how well an individual can understand words or text that are the focus of the experiment. This factor is assessed through the calculation of the Speech Reception Threshold (SRT), which determines the lowest intensity level at which an individual can correctly repeat familiar two-syllable words, known as base words, more than half the time.

  Another parameter used is the Speech Transmission Index (STI), which quantifies various effects of the transmission channel on the intelligibility of messages from a speaker to a listener. This parameter is valuable because it considers the transmission channel's impact on speech intelligibility independently of the speaker or listener.

2.2.3 State of the Art of Spatial Audio Reproduction Systems Validation

To proceed with the validation of the Audio Space Lab, relevant papers on the topic were reviewed.

Paper	**Physical Metrics**				**Perceptual Metrics**		
	Room Acoustic Parameters						
	Monoaural			**Binaural**			
	EDT	**RT30**	**C50**	**IACC**	**SRT**	**Localization**	**Distance**
[9]		X	X	X	X	X	
[10]		RT20				X	X
[11]	X	X	X		X		
[12]	X	X	C80	X			
[13]	X	X	C80				
[14]					X		
[15]					X	X	
[16]							X
[17]		RT60					X
[18]						X	
[19]							X
[20]							X
[21]						X	
[22]						X	
[23]						X	
[24]						X	
[25]						X	
[26]							X
[27]						X	

Table 2.1: Metrics used by each paper.

These papers have attempted to establish a foundation for validating spatial playback systems in confined environments by comparing real and spatialized environments. Papers conducting validations similar to what is envisioned for our case study were analyzed, focusing on benchmark metrics used in these validations.

Among the analyzed articles, in addition to the physical validation, the perceptual validation carried out is also reported. Although it is not the focus of our study, it is included to lay the foundation for the future perceptual analysis to be conducted by the Audio Space Lab.

The most relevant papers are presented in Table 2.1, highlighting the predominant metrics considered. As shown in Table 2.1, the reviewed articles generally evaluated similar metrics for their validations, with a few exceptions. For instance, [10] used RT20 instead of RT30 for Reverberation Time, while [17] employed RT60. Additionally, [12] and [13] utilized C80 instead of C50 for Clarity Factor.

Not all studies focused on IACC (Interaural Cross Correlation), with only a few papers including this metric in their analyses. Many studies concentrated on sound source localization, with some also assessing perceived distance to the source. Generally, studies emphasizing localization did not delve into physical acoustic parameters but rather focused on perceptual analysis.

In the study by Cubick [9], a comparative analysis was conducted between a real classroom environment and a Virtual Sound Environment (VSE) created using a spherical array of 29 speakers. The real classroom, seating 40 individuals, was simulated using the LoRA system within ODEON software.

The VSE utilized 29 speakers arranged in a spherical configuration: a horizontal ring of 16 speakers at ear level of seated listeners, two rings of 6 speakers at ±45° elevation, and one ceiling-mounted speaker above the array center. Non-linear (NLS) and Higher Order Ambisonic (HOA) methods were employed for sound field rendering.

Physical validations included Reverberation Time (RT30), Clarity (C50), and Interaural Cross Correlation (IACC), derived from Room Impulse Responses (RIR) captured via logarithmic sine sweeps. RIR measurements were obtained at 32 positions and averaged across 25 source positions at distances of 2m or more.

The study involved 8 normal hearing (NH) participants, evaluating speech intelligibility using the Danish Dantale II speech-noise Test in both real and virtual environments. Stimuli consisted of 160 sentences presented under noisy conditions at distances of 2m and 5m from the listener. Speech Reception Thresholds (SRT) were measured to assess intelligibility.

Results indicated comparable RT30 and C50 values between the VSE and ODEON simulations, while IACC values were lower in the VSE than in the real classroom. This discrepancy was attributed to spatial dispersion introduced by NLS and imperfect sound field reproduction with HOA.

Furthermore, SRT values varied slightly between the VSE and real classroom, with HOA encoding showing higher values at 2m distances, and greater spatial diffusion observed in the real environment at 5m. Statistical analyses (ANOVA) confirmed that NLS encoding generally provided closer approximations to real-world conditions in terms of speech intelligibility.

In addition to SRT measurements, participants were asked to indicate perceived sound source locations on a response sheet during tests, revealing that HA affected spatial perception, causing sounds to appear less distinct and sometimes reverberant in the VSE.

Overall, the study underscored the potential of VSE systems in evaluating hearing aid signal processing capabilities under realistic conditions, despite some discrepancies in spatial perception compared to real environments.

In the study conducted by Fargeot [10], an experiment comparing real and virtual environments was undertaken using three different acoustic settings. The experiment utilized three empty acoustic rooms, each containing two speakers positioned at distances of 2m and 4m from the listener's position.

Acoustic measurements were performed using Room Impulse Responses (RIRs) captured with a 10-second sine-sweep signal in a semi-anechoic chamber, recorded through a spherical array of 32 microphones. The RIRs were encoded with fourth-order Higher Order Ambisonics (HOA) and decoded using Basic optimization. Reproduction of stimuli was achieved through a spherical array comprising 42 speakers arranged on a 3.8m radius geodesic sphere.

Reverberation Time (RT20) was calculated and presented in octave bands for the three acoustic environments, alongside other variables such as perceived distance and source height. Data analysis was conducted using a linear mixed model. The absolute error in azimuth and elevation was graphically represented across different listening conditions (real vs. virtual) and acoustic environments. Another graph depicted perceived distance as a function of listening conditions and acoustic environment.

Participants, totaling 21 normal-hearing individuals, underwent three listening sessions: one for familiarization and one for each condition (real and virtual) for stimulus type. Stimuli included a vocal segment (3s), a guitar excerpt (1min), and a white noise sequence (1s). In the Real condition, stimuli were presented through room speakers, while in the Virtual condition, stimuli were convolved with the respective RIRs and played through the virtual system.

Participants used a controller as a laser pointer within a virtual space, adjusting a transparent hemisphere to indicate perceived distance and marking the position and amplitude of the perceived sound source. This "open loop" localization mode repeated stimuli until completion of the task.

In the real condition, higher localization errors were observed for sources at 4m (2.4°) compared to 2m (1.5°). Conversely, no significant distance effect was noted in the virtual condition. Elevation errors were notably higher in the virtual condition (8.5° absolute error) compared to the real condition (1.9°). Source width showed differences between 2m and 4m sources in the real condition but not in the virtual condition.

Overall, the experiment highlighted challenges in sound source localization under auralization conditions, particularly in terms of angular accuracy. The extent of localization errors appeared to correlate with the reverberant characteristics of the acoustic environments tested.

In [11], the validation of auralization techniques involved comparing speech intelligibility tests conducted in both real and virtual university classrooms. Two classrooms were selected: one with acoustic treatment and another without, each evaluated at three distinct listening positions along the centerline.

The source, simulating a professor, was positioned behind a desk, while noise was placed at the back of the classroom between two listening positions. Both classrooms were digitally modeled and auralized using CATT-Acoustic software. Speech and noise reproduction utilized speakers in both real and virtual environments, with virtual tests conducted in a soundproof booth using headphones.

Physical measurements included metrics such as RT (Reverberation Time), EDT (Early Decay Time), and C50 (Clarity Index). Directionality was analyzed through polar diagrams of the source in octave bands. Graphs depicting predicted versus actual values of reverberation time, decay time, U50 (Useful-to-Detrimental Energy Ratio), and clarity were plotted for comparison across listening positions. Impulse responses were also graphically compared between predicted and measured values for both classrooms. Additional graphs showed the percentage of predicted and actual speech intelligibility across the three listening positions in both real and virtual environments. Parameters like U50, C50, and SNR (Signal-to-Noise Ratio) were analyzed collectively to assess variations in speech intelligibility across different listening conditions.

Analysis revealed that the classroom with acoustic treatment exhibited significantly lower RT and EDT compared to the untreated classroom, showing good agreement between predicted and measured values. However, discrepancies were noted in C50 measurements: predictions tended to overestimate values in the treated classroom and underestimate them in the untreated one. Similar trends were observed with U50 metrics.

The experiment involved eight participants with normal hearing, subjected to a Modified Rhyme Test across octave frequencies from 250 to 8000 Hz in both real and virtual classrooms. Background noise from simulated conversations was used during the tests. Results showed varying speech intelligibility levels between virtual and real environments, influenced by acoustic treatment and background noise conditions. Differences were more pronounced in classrooms with higher absorption rates or significant noise levels.

In conclusion, the study demonstrated that virtual classrooms can reliably simulate speech intelligibility tests under conditions of low absorption and minimal noise. However, the study highlighted the sensitivity of auralization to noise levels, with higher variability in virtual classrooms when significant noise was present.

The study's findings underscore the importance of accurate acoustic modeling in auralization studies, particularly in environments with varying acoustic properties.

In [12], the study introduces a high-order ambisonic auralization (LoRA) system designed to integrate acoustic environment models (specifically a classroom and a concert hall) with a loudspeaker-based auralization technique.

ODEON software, known for its built-in environmental acoustic models, was utilized to simulate these environments. Room Impulse Responses (RIRs) were computed based on receiver positions within the rooms.

The LoRA system employs RIRs' direct sound, first reflections, and last reflections components, facilitating seamless data exchange with ODEON's ambient data through a dedicated toolbox.

Research has demonstrated that higher ambisonic orders are crucial for achieving precise spatial reproduction of individual sound sources. Two distinct environments were simulated using ODEON: a classroom and a concert hall. Eight RIRs were generated for each environment across various source-receiver configurations. Multichannel impulse responses derived from the LoRA toolbox were analyzed using a speaker array configuration comprising 29 speakers, with a maximum ambisonic order of 4.

Four listening positions were selected for evaluation: one at the center of the speaker array and three others positioned at varying distances. Monoaural parameters such as Reverberation Time (T30), Early Decay Time (EDT), Clarity (C80), Gain (G) across seven octave bands, and the Speech Transmission Index (STI) were considered. Additionally, Binaural parameters, including Interaural Cross-Correlation (IACC), were examined.

Validation involved graphical representation of impulse responses in octave bands from 63 Hz to 8 kHz, emphasizing relative energy normalization. Reverberation time and decay time were depicted with average and standard deviation graphs across seven octave bands from 125 Hz to 8 kHz, ranging from 0.5s to 2s. Clarity and gain were similarly represented across the same frequency bands. IACC analysis was presented through separate graphs for early and late components across octave bands.

Results indicated that the LoRA system effectively preserved the temporal and spectral characteristics of RIRs without significant distortion, thereby maintaining the integrity of room acoustic parameters. Spatial properties of the room impulse response were accurately reproduced by the system, contingent upon the deployment of a sufficient number of loudspeakers to achieve low IACC values, particularly in high-frequency bands.

The study concluded that the LoRA system represents an efficient integration of room acoustic models with loudspeaker arrays, facilitating accurate perceptual simulations of acoustic environments. Furthermore, it highlighted the impact of ambisonic order and speaker count on localization accuracy, noting that while higher orders enhance spatial fidelity, the presence of reflections and reverberations can influence auditory localization accuracy by altering source amplitude perception.

In [13], a comparative analysis of speech intelligibility tests is conducted between CATT-Acoustic and ODEON, two software platforms used for room modeling and sound auralization. The study aims to determine which software provides more effective auralization capabilities.

The experiment involved two 95-seat classrooms with differing acoustic characteristics: one with poor acoustic treatment and high reverberation, and the other with high sound absorption and low reverberation. Both classrooms were equipped with forced ventilation systems generating background noise. Virtual models of these classrooms were created based on real measurements using AUTOCAD and imported into both CATT-Acoustic and ODEON.

Acoustic characterization included the calculation of Room Impulse Responses (RIRs) at three receiver positions. Key room acoustic parameters such as Reverberation Time (RT), Early Decay Time (EDT), Clarity (C80), and Sound Pressure Level (Lp) were evaluated in both real and virtual conditions. Results were analyzed across octave bands from 125 Hz to 4 kHz.

Tables and graphs presented the free-field sound pressure levels, sound power levels, and detailed comparisons of predicted versus measured RT, EDT, clarity, and sound pressure level for both software platforms across different listening positions and classrooms.

The study found close agreement between real and virtual conditions for RT, EDT, and C80, indicating that both CATT-Acoustic and ODEON effectively simulated these parameters. Speech intelligibility tests using the Modified Rhyme Test (MRT) were conducted in both environments with and without added noise (simulated conversation noise from a dodecahedral omnidirectional loudspeaker).

Graphical representations of Speech Intelligibility (SI) coefficients showed average measured and predicted results across all listening positions in both real and virtual classrooms under noise-off and noise-on conditions. Interestingly, the study noted that virtual models generally aligned closely with real-world results, particularly under noise-off conditions, except for anomalous findings in the low-reverberation real classroom.

Overall, the study concluded that both CATT-Acoustic and ODEON performed equivalently in terms of modeling classrooms, predicting acoustic parameters, and auralizing speech intelligibility tests. The results emphasized the reliability of these software tools for simulating complex acoustic environments and their potential applications in architectural acoustics and audio engineering research.

In [14], a novel validation approach was introduced utilizing the LoRA speaker system to create virtual auditory environments through both single speaker and Ambisonic (first-order and HOA) methods. The experiment was conducted within an acoustically treated room equipped with a 3D array of 29 speakers. A virtual classroom environment was simulated using ODEON, with specific positions assigned for the source and receiver.

The receiver was situated at 0° azimuth and 0° elevation within the room, facing the source.

Sound stimuli were delivered using a single speaker and decoded using fourth-order Ambisonics (HOA) and first-order Ambisonics. Low-mid frequencies used "basic" decoding, while high frequencies employed "max-RE" decoding.

Azimuth and elevation characteristics were assessed using polar graphs, detailing directionality, latency, and attenuation of direct sound and initial reflections.

Nine participants with normal hearing took part in the study, undergoing speech intelligibility tests based on the "Danish Hagerman Dantale II" protocol. Each sentence consisted of five words structured as 'Noun' + 'Verb' + 'Number' + 'Adjective' + 'Noun'. A diffuse noise, shaped to match speech characteristics, accompanied the sentences at a fixed 60 dB SPL level. Speech Reception Thresholds (SRT) were determined for each subject using two sets of 10 sentences. Participants selected the correct set of five words from 10 options presented on a tablet, with a "don't know" option available for uncertain responses.

Graphical representations included mean and standard deviation plots of SRT for direct sound only, as well as combined direct sound and initial reflections. Additionally, a comparison graph juxtaposed the mean intelligibility scores and psychometric curves derived from the data.

Analysis revealed significant dependence of intelligibility scores on the chosen auralization method. The highest scores were achieved with the single speaker technique, while first-order Ambisonics yielded the lowest scores. Overall, the study found that the intelligibility threshold (SNR at 55% word correctness) was lower with the single speaker method compared to Ambisonics, and decreased as Ambisonic order decreased.

Conclusively, the study supports conducting speech intelligibility experiments using the LoRA system with either the single speaker or HOA techniques, as both effectively reproduce early reflections essential for intelligibility. However, HOA methods are preferable for more complex auditory scenes due to their ability to simulate a broader range of acoustic environments.

In [15], a validation of virtual acoustic environments began with an anechoic room and a reverberation room. Two distinct experiments were conducted: one focused on speech intelligibility, while the other examined front-to-back source localization.

ODEON was employed to create these virtual environments. Seven measured impulse responses of both Head-Related Transfer Functions (HRTFs) were utilized in the anechoic chamber to generate anechoic stimuli. An artificial head was positioned at the center of a thin metal ring with a 1m radius, housing 24 speakers arranged at 15° intervals in the horizontal plane. Both HRTFs were adapted in ODEON and subsequently used as Binaural Room Impulse Responses (BRIRs) in the reverberant room.

Within the reverberation room, seven BRIRs of both HRTFs were convolved with anechoic samples to prepare auditory stimuli for perception tests related to sound source localization. Additionally, three impulse responses (at 0°, -90°, -180°) were convolved with speech test material for speech comprehension tests. Three additional BRIRs were used to prepare stimuli for speech intelligibility tests.

All tests were conducted exclusively in the reverberation chamber, utilizing the ODEON system to predict acoustic parameters. The experimental configurations involved placing 7 speakers in a semicircle on the horizontal plane to the right of each subject, positioned at a height of 1.2m from the ground.

Analysis and findings were based on HRTFs, presenting graphs that depicted the average performance of 8 subjects along with relative standard deviations across 5 conditions: Own Ears (OE), Artificial Head Measurement with Headphones (AHM), Artificial Head Measurement without Headphones (AHS), Hearing Aid Measurement with Headphones (HAM), and Hearing Aid Measurement without Headphones (HAS). Additionally, a graph illustrated the relationship between direct and reverberant sound (from speakers positioned 1m from the head) in the right ear across 1/3-octave bands, correlating with the angle of direct sound.

The study involved 22 participants with normal hearing, aged 19 to 43 years, who took part in the experiments within the anechoic chamber. Eight of these participants also participated in listening tests within the reverberation chamber. Throughout the experiments, subjects maintained their head orientation at 0°.

The test material consisted of sentences from the VU test set, spoken by a male speaker with each sentence lasting 200ms and featuring rise and fall times of 50ms. Stimuli were presented randomly, with 6 repetitions of each sound source per test, resulting in 42 responses per test. Each subject completed two tests.

The graph highlighting the Speech Reception Threshold (SRT) coefficient showcased the average SRT values in both environments (anechoic and reverberation room) across different HRTF conditions (OE, AHM, AHS, HAM, HAS).

Significant differences were observed in speech intelligibility tests across various spatial sound scenarios within the anechoic environment. Spatial separation of signal and noise sources significantly enhanced speech comprehension. Speech intelligibility was notably superior in the anechoic environment compared to the reverberant room across all spatial sound scenarios.

Direct sound levels varied based on sound source location, whereas late reverberation levels exhibited nearly location-independent characteristics.

Regarding localization tasks, participants were tasked with identifying which of seven labeled sources (real or virtual) emitted a sound signal and reporting the source number to the operator. Communication between participants and operators occurred via an auxiliary microphone-speaker system.

Analysis of tests conducted in the anechoic room revealed substantial differences in sound localization abilities between using one's own ears (OE) and using headphones (AHM and HAM), both within the anechoic and reverberation rooms. Localization errors with the artificial head (AHM) ranged up to 20% in terms of forward-backward errors, representing the largest standard deviation among all tests conducted. Interestingly, reverberation had no discernible impact on forward-backward localization performance.

Overall, better localization outcomes were generally observed in the reverberation room compared to the anechoic room, likely due to high sound diffusion, which renders reverberant reflections location-independent.

In summary, the study found overall strong agreement between simulated (virtual) and measured (real) data, validating the effectiveness of the ODEON system in creating and assessing virtual acoustic environments.

Research presented in [16] investigated perceived distance within a LoRA-based environmental auralization system for generating virtual auditory environments using Ambisonic HOA auralization via loudspeakers.

Experiments were conducted in an acoustically damped room equipped with a three-dimensional spherical array of 29 speakers having a radius of 1.8m. Three distinct environments were studied: a classroom, a large auditorium, and an anechoic room. Each environment featured 15 receiver positions situated at various distances from an omnidirectional sound source. Eight of these positions were designated for exploring distance perception, while seven were used to assess HOA accuracy.

In some instances, Binaural Room Impulse Responses (BRIRs), generated using a HATS (head and torso simulator), were employed at receiver positions within real rooms. These BRIRs were also captured in the anechoic chamber for comparison. Virtual reproductions of the classroom and auditorium were created using ODEON software, incorporating corresponding BRIR calculations.

The sound source was consistently positioned azimuthally at 90° relative to the receiver to ensure significant binaural differences. Testing at 0° azimuth was omitted based on preliminary findings indicating indiscernible stimuli variations at different distances. Frequency responses of each speaker in the array were equalized to maintain uniformity.

Seven participants with normal hearing were involved in the study. Danish Speech Test sentences were used as auditory stimuli. Each experimental condition comprised three blocks, with each block containing three repetitions of randomized distance configurations. Participants initially heard phrases played at the closest and farthest distances, providing feedback on perceived distance after each repetition.

To illustrate trends in perceived distance, mean values and standard deviations of logarithmic apparent distance were graphed against logarithmic physical distance for each participant across all conditions.

Results indicated no significant disparity in the perceived quality or realism of distance perception between the LoRA system employing room simulation and the binaural system utilizing recordings with the HATS simulator. However, the study highlighted a tendency for the binaural auralization technique to underestimate distances compared to the speaker-based method.

The findings suggest that speaker-based auralization within the LoRA system can effectively replicate auditory environments where distance perception aligns closely with real-world counterparts. Nonetheless, the study's statistical limitations may have precluded definitive conclusions regarding the comparative efficacy of the two auralization techniques.

In [17], an experiment was conducted to compare responses between a real environment and its virtual counterpart. The study employed a movie theater and a listening booth as the real environments, both of which were then recreated virtually using OGRE gaming software. The virtual models were operated via a PC running a VR application, and participants used Oculus Rift HMD visors to experience the virtual environments while electrostatic headphones provided audio playback. Within the listening booth, a 5.1 surround sound system comprising 5 speakers and a subwoofer was utilized.

Reverberation time (RT60) served as a metric, graphically represented across octave bands from 10 to 8 kHz.

The experiment involved listening sessions with short music excerpts across 5 scenarios: one in a "neutral room" using headphones, and four in either the movie theater or listening booth, experienced in both real and virtual modes. A total of thirty participants with normal hearing participated, including 21 audio experts and 27 individuals previously involved in listening tests. Two sets of stimuli were used: one featuring pink noise, castanets, and drum sounds, and another comprising fifteen musical excerpts from DVDs like "Mercedes-Benz Signature Sound" and "BR Klangdimensionen," each lasting 10 seconds.

Participants rated their overall auditory experience (OLE) during the music excerpts across different environments. Additionally, the quality of experience (QoE), which assesses user pleasure and satisfaction, was evaluated. Perceived sound distance in virtual versus real environments was also investigated.

Results indicated significantly lower OLE ratings for VR sessions compared to real-world sessions. VR sessions also required more time to complete than their real-world counterparts. Participants perceived sound distances to be greater in virtual environments than in corresponding real-world settings. Moreover, the listening booth generally provided better results than the cinema, likely due to improved sound perception in a smaller space.

In summary, the study highlighted discernible differences between real and virtual environments, with VR generally yielding slightly lower ratings compared to real-world experiences.

This comparison underscores ongoing challenges and opportunities for enhancing immersive virtual auditory environments.

Consider also the article by [18], where a validation of a compact immersive sound space is presented using a system comprising 42 speakers employing various techniques such as HOA and VBAP to assess their accuracy in localizing sound sources. This study aligns with our validation objectives.

The system consists of a geodesic sphere arrangement of 42 speakers spanning a 3m diameter, positioned 1.5m from the central user within a room measuring 5x4x4 meters. The room is acoustically treated with a 5cm layer of wood, providing a reverberation time of 300ms at medium frequencies. Experimental conditions included real playback (Directed Altspeaker HP), VBAP, HOA3, and HOA5.

Results were analyzed using graphs depicting lateral angular error, combining boxplots, histograms, and mean magnitude errors for each rendering condition. Significant differences were observed between the HP condition and others, as well as between VBAP and other conditions, with no significant difference between HOA3 and HOA5.

The experiment involved thirty participants with hearing impairments tasked with indicating the perceived sound direction by hand and confirming their choice using a button. Head and hand movements were tracked using a 6-degree-of-freedom magnetic position and orientation sensor. Stimuli consisted of three 40ms pulses of white noise with specific intervals to prevent user head movements.

Each experiment was divided into four blocks of 60 trials, with each block lasting approximately 5 minutes and corresponding to a different rendering condition. Forward/backward and up/down confusion errors, lateral and polar localization errors, and average percentages of each type of localization error were reported.

Overall, virtual conditions exhibited higher forward/backward and up/down confusion percentages compared to real sources, but similar combined error percentages. Azimuth localization was more accurate with real sources than virtual ones, while elevation perception showed better performance with VBAP compared to HOA techniques.

In conclusion, the study highlighted challenges in 3D spatial localization within virtual environments across different rendering techniques, with VBAP demonstrating advantages over HOA techniques, particularly in elevation perception.

In [19], while no experiments involving the localization or validation of real/virtual environments were conducted, the study discusses current findings from research on auditory perception of distance. The paper explores factors influencing accurate distance perception, including the use of compressive power functions which approximate psychophysical distance functions.

The analysis includes a histogram plot illustrating perceived distance estimates compared to physical distances of sound sources, employing power functions with adjusted parameters or variances.

The importance of intensity in accurate distance perception is emphasized, following the inverse square law which entails a 6 dB decrease in sound pressure for each doubling of distance from the source.

Recent studies have highlighted the role of reverberant energy in distance perception, revealing its impact alongside intensity clues.

Unlike experiments in anechoic chambers where only intensity is present as a parameter, studies now demonstrate the dependence of accurate distance perception on the ratio of direct to reverberant energy. This ratio significantly affects listeners' ability to gauge source distance under conditions of low or moderate reverberation.

Additional parameters such as Interaural Time Difference (ITD) and Interaural Level Difference (ILD) are discussed, noting their effectiveness primarily at short distances. ITD, invariant to distance, aids in determining the lateral position of the source, while ILD magnitude assists in estimating distance.

Acoustic parallax is mentioned as another factor influencing distance perception, particularly when the source is close enough to introduce significant differences between the angles of the source to the left and right ears. However, intensity and the direct-reverberant ratio are found to be more influential than acoustic parallax.

Visual input also contributes significantly to perceptual distance accuracy; the presence of visual targets enhances auditory distance estimation and reduces judgment variability, especially when multiple visual targets are available.

Conversely, background noise in environments has shown varying effects: it tends to increase the perceived distance of speech but decrease the perceived distance of non-speech sounds.

In conclusion, findings from the studies cited in [19] suggest that perceived distance is often a biased estimate of physical distance. Sound sources are typically underestimated in distance perception, although sources within approximately 1 meter are generally overestimated. This underestimation may serve as a safety margin in real-world scenarios to avoid collisions with objects.

In [20], four main aspects of auditory distance perception are comprehensively discussed: signal processing, developmental aspects, consequences of visual and auditory impairments, and the underlying neurological basis. A synthesis of previous research findings is presented, focusing on these facets of distance perception.

The paper also delves into the advancements in binaural technology, which enable the simulation of realistic auditory environments through headphones, a method employed in prior studies for presenting auditory stimuli to listeners.

One critical metric discussed for assessing distance perception is the direct-reverberant energy ratio (DRR). While reverberation can degrade azimuth localization, it plays a crucial role in distance judgment. DRR decreases as the distance from the sound source increases, enhancing the perceived distance.

This factor holds significance for sound sources both near (peripersonal space) and far (extrapersonal space), and for sounds originating from frontal as well as lateral directions. The combination of DRR and sound level typically provides more accurate distance information, although the presence of reverberation can influence distance judgments. Additionally, spectral cues contribute to the perception of sound source distance.

In close proximity, auditory distance judgments are more precise when the sound is presented laterally to the listener. While Interaural Time Difference (ITD) changes minimally with distance, Interaural Level Difference (ILD) varies significantly in the near acoustic field, providing distance cues up to approximately 1 meter, beyond which it becomes less distance-sensitive. It is noted that headphone use can sometimes create a sensation of sound being perceived inside the head rather than externally.

Factors enhancing auditory distance perception include familiarity with the stimuli used as targets and prolonged exposure to them, which improves distance judgment accuracy. However, the presence of concurrent visual stimuli can bias perceived distance toward the visual stimulus, particularly when there is a temporal disparity between auditory and visual inputs. Hence, maintaining temporal synchrony between auditory and visual cues is crucial in such situations. The type of room where experiments are conducted also influences perceived auditory distance.

Studies cited in [20] indicate that children as young as 6 months can distinguish near and distant objects based on auditory cues, and by 9 months, they integrate visual and auditory information to assess depth and calibrate auditory spatial perception.

Regarding auditory distance perception in blind individuals, research suggests they excel in relative distance discrimination rather than absolute distance. Some blind individuals utilize echolocation, a technique involving self-generated sounds to perceive distances to silent objects based on environmental feedback. Echolocation skills can be trained, and blind individuals often exhibit superior echolocation abilities compared to sighted individuals.

Modern hearing aids, despite including amplitude compression, do not appear to significantly impair distance discrimination when auditory cues such as sound level and DRR are preserved.

The paper also discusses neural mechanisms involved in distance discrimination, highlighting ongoing research gaps concerning the effects of partial visual impairment, occlusive objects, background noise, and multiple sound sources on auditory distance perception.

In conclusion, [20] underscores the progress made in understanding auditory distance perception while identifying persistent research gaps that warrant further exploration.

In [21], a study was conducted to validate a hybrid computational method designed to recreate accurate sound localization signals for listeners in virtual environments, comparing them to real-world counterparts. This method combines predictive room impulse response (RIR) calculation with measured or simulated Head-Related Transfer Functions (HRTFs). The virtual room simulations were performed using ODEON software.

For HRTF calculation, measurements were taken using an artificial head positioned at the center of an anechoic chamber. Thirteen speakers were placed in a ring configuration around the head, spaced at 15° intervals from -90° to +90° azimuth at a distance of 1m from the center. These measurements formed the basis for generating sound samples used in the listening tests under anechoic conditions, convolved with dry binaural impulse responses for three selected stimuli. Subsequently, this data was adapted in ODEON to compute corresponding Binaural Room Impulse Responses (BRIRs) in a reverberant environment, employing a hybrid method combining image source and ray tracing techniques.

Listening tests were conducted in both anechoic and reverberant chambers. Participants sat at the center of the speaker ring, with their ears aligned at the speaker height (1.2m above the floor). Two testing sessions were conducted: monaural and binaural, focusing on users' final localization of sound sources.

Seven individuals with normal hearing participated in the study. Three types of stimuli were used: a one-third-octave noise band centered at 500 Hz lasting 200ms, another band centered at 3150 Hz with the same duration, and a broadband telephone ringing sound lasting 1 second, encompassing both low and high-frequency components. Each stimulus was randomly presented three times per speaker, resulting in 39 localization tasks per test condition.

Localization investigations were limited to the horizontal frontal plane due to the study's emphasis on binaural cue utilization. Participants reported their localization responses via microphone communication immediately after hearing each target stimulus.

The study observed minimal localization performance degradation when using headphones, particularly with high-frequency narrowband stimuli. Localization of low-frequency narrowband signals or wideband telephone signals showed no or minimal decrease in performance. Overall, in virtual scenarios, localization performance approached that observed in natural environments.

Results indicated that headphone-based experiments slightly reduced localization accuracy compared to measurements using only impulse responses, especially with high-frequency narrowband signals. Reverberation had negligible effects on localization accuracy when the speaker-receiver distance was 1m. However, at a greater distance (2.4m), performance deteriorated significantly, suggesting that decreased direct sound-to-reverberation ratios impact localization more than reverberation itself.

The study concluded that the hybrid computational approach effectively predicts BRIRs necessary for directional sound localization tests within the horizontal frontal plane, demonstrating good agreement between simulated and measured results.

In [22], a study compared localization abilities using high-frequency and low-frequency sounds reproduced from virtual sources.

Two experiments were conducted to investigate how listeners perceive and combine auditory cues from different frequency ranges.

Binaural Impulse Responses (BRIRs) were measured using an electronic dummy head (KEMAR) positioned in a room. The KEMAR was seated on an office chair with its ears at a height of approximately 1.5m from the floor and positioned 0.5m from the nearest wall. BRIRs were recorded for various side angles (-90° to +90° in 15° increments) and distances (0.15m, 0.40m, 1.00m, and 1.70m).

Seven participants with normal hearing took part in the experiments. The stimuli consisted of pink noise pulses lasting 250ms, with 2ms cosine-square ramps at the onset and offset. Experiment 1 involved participants localizing either low-frequency ("Lo") or high-frequency ("Hi") noise presented separately, filtered from the broadband pink noise. In Experiment 2, participants localized combined "Lo" and "Hi" noises presented simultaneously.

During each trial, a new noise token was played, and participants were instructed to localize the sound source by indicating its lateral angle using a graphical user interface (GUI) with a mouse.

The results were analyzed to show the average lateral response angle plotted against the actual lateral angle of the stimulus for both low-frequency (Lo) and high-frequency (Hi) sounds. Overall, participants showed less accuracy in localizing sources further from the median plane. Responses for sources at lateral angles of 60° and beyond exhibited greater variability compared to those closer to the median plane.

The study also found that in a reverberant environment, perceived source directions systematically shifted with simulated distances. Moreover, the accuracy of localization depended significantly on the frequency content of the stimuli. Higher-frequency cues provided more reliable localization information than lower-frequency cues.

Additionally, reverberant energy tended to distort perceived source positions towards the median plane, especially as the distance between the listener and the source increased and the Direct-to-Reverberant (D/R) energy ratio decreased.

In conclusion, the study highlighted that listeners may misinterpret large Interaural Level Differences (ILDs) from nearby sources as indicators of lateral positions, contributing to biases in response. Despite clear and reliable directional information at the onset of stimuli, reverberation introduced perceptual distortions across all noise types tested.

In [23], the study investigates the performance of four ambisonic microphones: the SoundField microphone and three HOA microphone prototypes developed at Orange laboratories, focusing on their localization abilities in the horizontal plane, both objectively and perceptually. The study begins with a localization test followed by an objective analysis of the microphones.

Measurements for all four microphones were conducted in the IRCAM anechoic chamber, spanning from -40° to 90° in elevation and 0° to 360° in azimuth with a 5° step. A setup with 48 speakers arranged in a dodecagonal structure (7.5 degrees apart, 1.5m radius) was used for testing, employing mixed decoding (basic + maxRe) to optimize sound field resolution.

Graphs depicting ITD calculations using the GaussianMaxIACC method were presented for the four microphones and synthetic systems of orders one to four, demonstrating the influence of ambisonic order and microphone characteristics on sound source lateralization.

For the localization task, participants (14 with normal hearing) were tasked with matching virtual sound sources generated by spatial audio systems to real sound sources emitted from a loudspeaker. The virtual sources could be adjusted digitally with high precision (one degree accuracy) using a knob.

Results indicated that higher-order ambisonic microphones generally improved localization accuracy, with performance varying based on source incidence and ambisonic order. Errors in localization were noted more frequently for lateral directions, leading to under-lateralization of sound sources, especially noticeable with lower-order systems.

Both objective measurements and perceptual evaluations underscored the impact of ambisonic order and microphone type on sound reproduction accuracy. While frontal and some posterior sources were well-reproduced even with lower-order systems, accuracy decreased notably for lateral sources, highlighting challenges in accurately localizing sources away from the frontal plane.

In [24], the study aimed to investigate the impact of VR goggles (HMD) on the localization accuracy of virtual sound sources in the horizontal plane, simulated using 1st, 3rd, 5th, or 11th order ambisonics. The experiment took place in the Audio-Visual Immersion Laboratory (AVIL), utilizing a 4.8m diameter anechoic chamber with 64 speakers, arranged in the horizontal plane with 15° spacing.

A 1:1 scale virtual replica of the anechoic chamber was created in UNITY3D software for the virtual environment. Spatial alignment between the real and virtual worlds was ensured using three trackers, recalibrating if discrepancies exceeded 1cm.

Twenty participants with normal hearing were involved, and auditory stimuli consisted of 240ms bursts of pink noise presented at 7.5° intervals from -90° to 90° azimuth.

Each position was repeated five times per condition across six experimental blocks: blindfolded, real environment (with and without HMD), and virtual environment (with and without HMD). The experiment spanned two sessions with a maximum duration of 2.5 hours. Participants used an HTC VIVE controller to indicate their judgments by pressing a button on the controller, maintaining consistent pointing methods throughout.

Results were analyzed and presented through various graphs, including localization errors in both virtual and real environments, the difference between response and origin angles for different ambisonic orders, and localization error comparisons with and without HMD (specifically for 3rd, 5th, and 11th-order ambisonics). Probability density plots by azimuth angle and mean absolute localization errors across ambisonic orders and conditions were also included.

Findings highlighted that 1st-order ambisonics exhibited greater localization errors compared to higher orders, with minimal differences observed between higher orders. Virtual environments showed a tendency for leftward (negative angle) biases compared to real environments. Notably, higher-order ambisonics reduced localization errors, and visual information provided by HMDs improved localization accuracy, particularly with higher-order ambisonics.

Additionally, the study noted increased perceived lateralization of stimuli when using VR goggles, particularly in the right hemisphere. While HMD effects were consistent across ambisonic orders, 1st-order ambisonics demonstrated the highest error rates, with improvements seen up to 5th-order ambisonics, but marginal gains beyond this order.

In [25], an experiment was conducted to investigate the impact of head-mounted displays (HMDs) on sound localization accuracy compared to visor-less conditions, exploring variations in visual information provided to participants.

The study utilized an acoustic system comprising 64 speakers arranged in an anechoic chamber, with 27 speakers positioned in the front hemisphere at three different heights: ear level (0° elevation) and ±28° elevation. The azimuthal distribution included thirteen ear-level speakers spaced at 15° intervals from -90° to +90° azimuth, and seven speakers each at ±28° elevation with a 30° azimuthal spacing.

Virtual environments were recreated using HTC Vive software and Blender with the SteamVR plugin. Spatial alignment between real and virtual worlds was ensured using three Vive trackers.

Binaural impulse responses (BRIRs) were recorded using a B&K head and torso simulator (HATS) from all 64 speakers, both with and without the HMD. Interaural level differences (ILD) and interaural time differences (ITD) were calculated to assess localization accuracy, specifically in the horizontal plane across varying azimuth angles.

Ten participants with normal hearing took part in the experiment. Pink noise pulses with a duration of 240ms, including 20ms cosine ramped onsets and offsets, served as auditory stimuli. Participants were instructed to orient consistently towards 0° azimuth. Twenty-seven source positions were randomly presented, each repeated five times per condition.

The experiment encompassed three main conditions: blindfolded, visual cues about the room and speaker arrangement, and vision of the environment with a laser pointer for precise pointing.

Results indicated that the addition of visual information significantly enhanced sound localization accuracy compared to blindfolded conditions. Visual cues about the environment and speaker positions further improved localization performance. The introduction of a virtual laser pointer improved elevation perception accuracy but had mixed effects on azimuth perception.

Overall, participants demonstrated higher accuracy in azimuthal localization compared to elevation. The study highlighted the influence of visual feedback in enhancing sound localization capabilities, particularly in virtual environments simulated through HMDs.

Carvajal in [26] conducted an experiment comparing the perceived distance to a sound source in three different acoustic environments: a standard IEC room, a small room with minimal reverberation, and a larger anechoic room. They evaluated sound reproduction using both loudspeakers and headphones, recording BRIR impulse responses for seven source positions in the reference room. During the experiments, stimuli were simulated from seven chosen source positions using four speakers placed at 0°, 30°, 90°, and 330° azimuth angles.

Participants, consisting of 18 normal-hearing individuals, evaluated three parameters—perceived distance, azimuth direction, and compactness of sound—using subjective scales ranging from 0 to 5. The experiment was divided into three parts: one where participants had only visual input, another with only auditory input, and a final part with combined visual and auditory clues. The stimuli comprised male speech sentences from the Danish version of the HINT test, each evaluated twice under each condition in all reference rooms.

The results indicated that perceived distance to the sound source varied significantly depending on the room environment, whereas azimuth direction and compactness were less affected. Evaluations with headphones resulted in a wider range of compactness perceptions compared to loudspeaker presentations.

Notably, externalized perception of sound, particularly in terms of distance, was influenced by room characteristics, with discrepancies more pronounced in smaller reverberant rooms.

Listeners demonstrated a tendency to localize peripheral sounds more accurately, while front and rear locations often led to confusion.

Additionally, stimuli presented in all settings showed greater perceived source amplitude for front and rear positions compared to side positions.

In the study by Buchholz [27], the researchers investigated sound localization abilities in individuals with bilateral sensorineural hearing loss (HI) compared to normal-hearing individuals (NH).

The experiment utilized a virtual simulation of a cafeteria environment within an anechoic chamber, implemented using ODEON software and the LoRA toolbox to replicate acoustic conditions realistically.

The cafeteria simulation featured background noise generated from conversations among 14 speakers placed at various locations within the virtual room, contributing to a complex auditory scene. Participants, including eight NH and 15 HI individuals, were tasked with localizing a target word ("two") spoken by a female speaker positioned at different distances and azimuth angles relative to the listener.

The experiment was structured into two parts: first, measuring the masked detection threshold for the target word to set subsequent target levels, and second, assessing localization accuracy. Participants underwent one or two sessions lasting 2-2.5 hours each.

Localization accuracy was evaluated across 16 azimuth positions in the horizontal plane at distances of 1 to 2 meters, and nine azimuth positions at 4 meters (left hemisphere only), totaling 41 target positions tested five times in randomized blocks. The study analyzed confusion rates in the front-back plane and horizontal root mean square (RMS) errors in degrees for both NH and HI participants under quiet and noisy conditions.

Results indicated that HI participants generally exhibited significantly higher RMS errors compared to NH participants across all conditions. NH participants showed higher mean RMS errors in the back compared to HI participants, particularly under quiet conditions. The study also noted a correlation between RMS errors in silent and noisy conditions, albeit decreasing with distance.

Furthermore, NH listeners demonstrated less sensitivity to distance under quiet conditions, contrasting with their performance in noisy environments where distance affected accuracy. Front-to-back confusion rates averaged 35% under quiet conditions, highlighting challenges in spatial perception for both groups.

The study suggested a potential link between localization accuracy in quiet conditions and low-frequency hearing thresholds, affecting neural encoding of interaural temporal differences critical for horizontal localization. Overall, the findings underscored significant differences in spatial auditory processing between NH and HI individuals, influenced by both auditory cues and environmental factors.

2.3 Metrics for Physical Validation

Of all the metrics used to validate the playback systems listed in this chapter, we selected those that were most significant and of utmost importance for our validation. A detailed description follows for each metric used.

2.3.1 Interaural Time Difference (ITD)

ITD is defined as the difference in time (delay/anticipation) with which a sound reaches one ear relative to the other (figure 2.1).

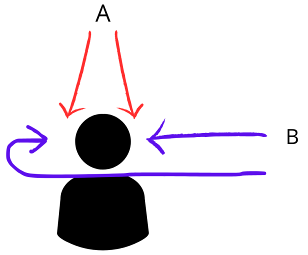


Figure 2.1: The principle behind the interaural time difference (ITD) is as follows: when a sound originates directly in front of the listener (point A), it reaches both ears simultaneously. However, when the sound source is located to the side (point B), it reaches the closer ear (typically the right ear in this diagram) before reaching the farther ear (left ear in this case).

When a sound source is directly in front of the listener (point A), both ears receive the sound simultaneously, resulting in an ITD of 0. However, if the sound source is located to the side (point B), the sound reaches the closer ear first, followed by the farther ear.

ITD increases as the sound source moves to the side, providing critical information for precise localization of sound direction, especially effective for low-frequency sounds.

According to Benichoux [28], maximum ITD values can reach about 800 microseconds at low frequencies and 600 microseconds at high frequencies. ITD varies with frequency (up to 200 microseconds for certain positions) within the frequency range relevant for judging the lateral position of a sound source.

The spatial position of a sound source cannot be accurately estimated from ITD alone; it depends on the frequency of the sound. Previous models treated the head as a rigid sphere (Kuhn [29]) or ellipsoid (Cai [30]).

In the "Head Transparent Model", where the head is approximated as a rigid sphere, ITD is calculated as the difference in path lengths to the two ears divided by the speed of sound:

$$ITD = 2a/c(\sin(\theta)) \tag{2.1}$$

where θ is the azimuth of the sound source, c is the speed of sound in the air, and a is the radius of the head. In this model, ITD does not depend on frequency.

A more realistic acoustic model, such as Rayleigh's model of a rigid sphere, shows that ITD generally decreases with increasing frequency. For high frequencies, when the wavelength is small relative to the head radius, ITD is approximated by:

$$ITD_{HF} = a/c(\theta + sin(\theta)) \tag{2.2}$$

The low-frequency limit of ITD is given by the spherical harmonic expansion of the sound field solution:

$$ITD_{LF} = 3a/c(\sin(\theta)) \tag{2.3}$$

The ratio of low-frequency to high-frequency ITD is always greater than one:

$$3(sin(\theta))/\theta + sin(\theta) \tag{2.4}$$

Thus, the size and shape of the head influence ITD differently across frequencies and sound source positions, as predicted by acoustic principles. This variability highlights how human head morphology affects the relationship between ITD and frequency, providing cues not only for azimuth but also for elevation, including discrimination of front and back locations.

In this validation, this metric was introduced for head-torso simulator (HATS).

2.3.2 Interaural Level Difference (ILD)

ILD (Interaural Level Difference) is the perceived intensity difference between one ear and the other in response to the same acoustic stimuli.

It is based on the difference in sound pressure level, influenced by the head acting as a barrier that creates an acoustic shadow. This shadow primarily affects high-frequency sounds, reducing their intensity as they reach the ear farther from the sound source.

High-frequency sound waves are smaller relative to the size of the head, making them more susceptible to the acoustic shadowing effect (see Figure 2.2). In contrast, low-frequency sounds are less affected by the head's acoustic shadow.

In [31], it is noted that the minimum detectable ILD is approximately 0.5 dB, irrespective of frequency. ILD in the far field typically does not exceed 5-6 dB, whereas in the near field, such as at 500 Hz, it can exceed 15 dB.

When considering both interaural time difference (ITD) and interaural level difference (ILD) together, they complement each other in sound localization. ITD provides crucial information for locating low-frequency sounds, while ILD is more informative for high-frequency sound localization.

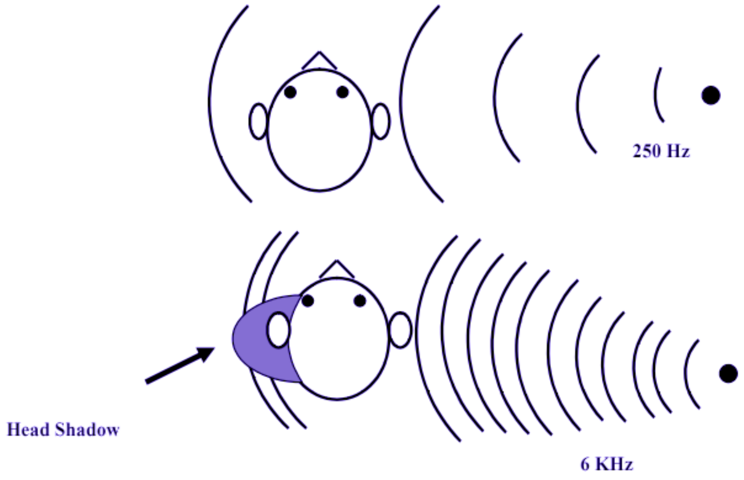


Figure 2.2: The head-shadow effect at high frequencies and ILD dependency on frequency and position. Image taken from [31].

However, as discussed in [32], ITD and ILD may provide ambiguous information about the height of a sound source. Differences in time and level can correspond to different perceived heights, making height localization less reliable. Similar ambiguities can occur in lateral sound localization.

In this validation, this metric was introduced for head-torso simulator (HATS).

2.3.3 Reverberation Time (RT)

Reverberation time (RT) is the duration it takes for sound to decay in a space after the sound source stops. It is influenced by the room's dimensions and the acoustic properties of its surfaces, which determine how sound waves are reflected or absorbed.

Spaces with longer reverberation times tend to have poorer speech intelligibility due to prolonged sound decay.

Reverberation time (RT) represents the duration for sound pressure levels to decrease by a specified amount (60dB, 30dB, 20dB) after the cessation of a sound source. RT is commonly referred to as RT60, RT30, or RT20, depending on the chosen decay criteria.

RT60 specifically denotes the time for sound pressure levels to decay by 60dB, as explained in [33]. In environments with high background noise or where the sound source emits minimal noise, detecting this decay may be challenging. In such cases, shorter decay times like RT30 (decay by 30dB) or RT20 (decay by 20dB) are used to mitigate this issue.

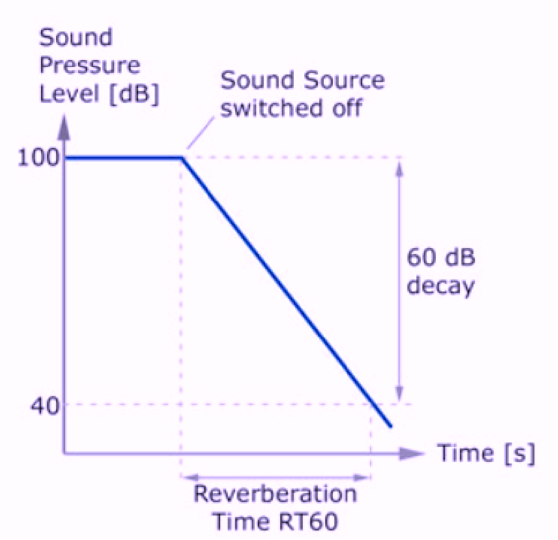


Figure 2.3: Visual representation of reverberation time, showing a decay of 60 dB. Image sourced from [33].

RT30 is calculated by doubling the decay time obtained at 30dB, while RT20 involves tripling the decay time at 20dB.

To measure RT, a sound event is reproduced, and its decay time is recorded from various positions within the room. Standard procedures recommend averaging these measurements to obtain a representative reverberation time.

2.3.4 Clarity (C50)

Speech clarity refers to how effectively speech is transmitted to listeners. In environments with high reverberation, speech intelligibility can be compromised. It is necessary to distinguish between direct sound and reflections.

As explained in [34], direct sound is the first to reach the listener and is followed by initial reflections. Initial reflections that reach the listener within 50 ms are incorporated into the direct sound and enhance listening. In contrast, reflections that arrive after 50 ms are perceived as annoying and reduce speech clarity.

Measuring speech clarity involves comparing the energy of early reflections to that of later reflections arriving after 50ms. This comparison is typically expressed in decibels. Optimal clarity values fall within the range of -1 dB to 1 dB for a metric known as C50: $-1dB \leq C50 \leq 1dB$.

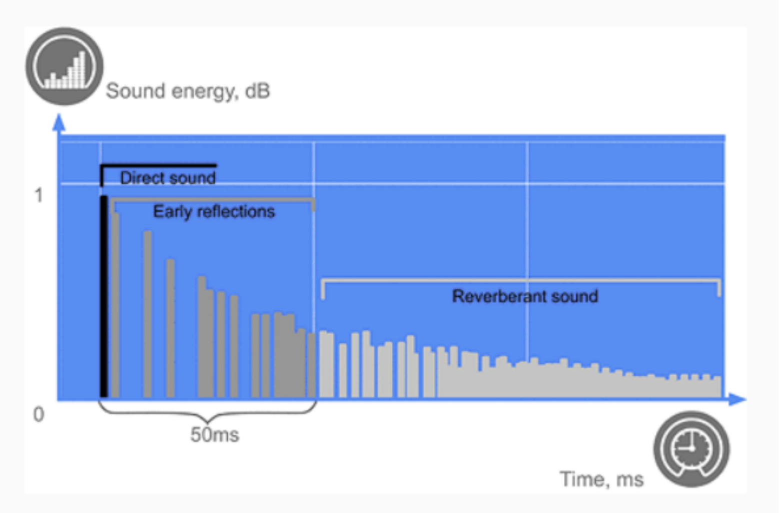


Figure 2.4: Visual representation of direct sound and initial reflections, useful for calculating C50. Image taken from [34].

Another metric, C80, extends the evaluation period to 80ms, considering early reflections as part of the direct sound. The choice between C50 and C80 depends on the nature of the sound source.

For speech, C50 is preferred due to its shorter integration time, whereas C80 is more suitable for music sources because music requires longer analysis time for auditory integration.

Another parameter that measures speech clarity is the Speech Transmission Index (STI), which reaches 1 when transmission is perfect and unimpaired. For optimal speech intelligibility, this value should exceed 0.75. The STI ranges between 0 and 1; the closer the value is to 0, the poorer the speech transmission, thus reducing speech intelligibility. Factors such as background noise, echoes, and excessively high reverberation times contribute to poorer transmission.

2.3.5 Inter Aural Cross Correlation (IACC)

The Interaural Cross-Correlation (IACC) parameter measures the difference between signals received by each ear, assessing the similarity of their impulse responses. It plays a crucial role in determining the spatial perception within an auditory scene, with values ranging from -1 to 1. An IACC of -1 indicates signals that are identical but completely out of phase between the ears, while 0 signifies no correlation between ear signals.

Mono sources positioned directly in front or behind the listener typically yield an IACC value of 1, which decreases as the source moves to the sides. High IACC values suggest a lack of spatial perception, whereas ideal values around 0.4 or 0.5 indicate enhanced spatiality and immersion in the auditory environment. Values near 0 indicate sound predominantly arriving from the lateral areas.

IACC serves as a binaural, head-related cue (whether human or simulated) that quantifies the similarity between two signals reaching the auditory system simultaneously. To calculate it, it is first necessary to define the normalized interaural cross-correlation function (IACF), expressed as follows:

$$IACF_{t_1,t_2(\tau} = \frac{\int_{t_1}^{t_2} p_l(t) * p_r(t+\tau))dt}{\sqrt{\int_{t_1}^{t_2} p_l(t)^2 dt * \int_{t_1}^{t_2} p_r(t)^2 dt}}$$

Where $p_l(t)$ represents the impulse response for the left ear, and $p_r(t)$ represents the impulse response for the right ear. Having defined this function, the IACC parameter is defined as:

$$IACC_{t_1,t_2} = \left| IACF_{t_1,t_2(\tau)} \right|$$

Here $IACF_{t_1,t_2(\tau)}$ denotes the normalized interaural cross-correlation function between $p_l(t)$ and $p_r(t)$ at time instants t_1 and t_2, considering the time delay τ, with $-1ms < \tau < +1ms$.

As mentioned earlier, IACC values will be between -1 and +1. To calculate the IACC, either a binaural microphone consisting of two symmetrical receivers (left and right) with an appropriate stand is used, or an artificial head made of materials that faithfully reproduce the acoustic parameters to which a real human head is subjected. The artificial head is the most effective method of calculation due to the shape of the earcups and ear canals, where the two microphones are placed inside them.

In this validation, this metric was introduced for head-torso simulator (HATS).

Chapter 3

Audio Space Lab

3.1 Configuration of the Spatial Audio Reproduction System

The Audio Space Lab (ASL) is located inside the Polytechnic University of Turin, at the Department of Energy, in a small room overlooking the university's courtyard. The room is 5.45m long, 2.67m wide, and 2.43m high, and it has been acoustically treated according to the criteria set out in ITU-R recommendation BS.1116-3.

The ASL has a reverberation time of about 0.17s, which is found to be an optimal value for octave bands from 0.25 Hz up to 4 kHz. The noise floor level values measured at the listening position are between NR 10 and NR 15 for frequencies up to 1 kHz, with values below 16 dB for higher octave bands.

The choice of this room as the location for the laboratory was deliberate. Choosing an anechoic chamber would have been challenging in terms of design and would have entailed very high costs. Using an "ordinary" room allows for the advantage of its reflections to mask inaccuracies in audio reproduction and make them sound natural. Additionally, the choice of a "normal," non-anechoic room is the most practical for replicating this system within specialized clinics or hospitals.

The spatial reproduction system was created using an array of 16 Genelec 8030B 2-way speakers (used for frequencies from 90 Hz to 20 kHz) and 2 Genelec 8351A 3-way speakers (used for lower frequencies from 30 to 90 Hz). These enable the reproduction of a target sound field at the center of the array, i.e., the sweet spot, using third-order ambisonic audio rendering. The speakers are arranged in a circle with a radius of 120cm from the center, representing the listening point (sweet spot), and 121.5cm from the floor. The speakers are arranged in three rings, representing three different elevation points: -45°, 0°, and +45°. There are 8 speakers on the 0° elevation ring and 4 on each of the ±45° rings. The 8 speakers on the horizontal plane are arranged with a 30° separation between each speaker, forming a complete circle around the listening point. In contrast, the other 8 speakers placed at ±45° elevation are arranged at ±45° and ±135° from the horizontal plane, tilted with their axes pointing toward the center of the speaker sphere.

The greater number of speakers on the horizontal plane is due to the higher accuracy of human hearing when interfacing with sounds in this plane and the limited spatial perception achieved when working with the elevation of a sound.

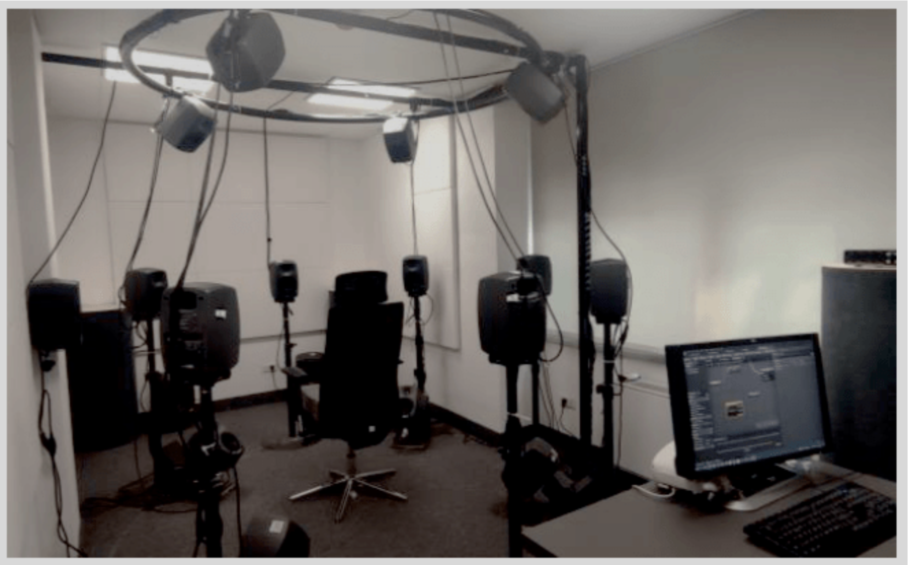


Figure 3.1: Sound Reproduction System of Audio Space Lab, reported by [35].

As shown in Figure 3.1, the loudspeakers arranged on the horizontal plane are mounted on Genelec 8000-409B solid steel adjustable stands. The upper speakers are arranged using brackets around an aluminum circle attached to the ceiling, while the lower ones are fixed on 45-degree inclined iron planes connected to Genelec 8000-409B steel adjustable stands.

All speakers are connected via XLR cables to the 32-channel Antelope Orion32 sound card, which is driven directly from a high-end desktop PC. The speaker signal is processed on the PC using the commercial DAW software Bidule, organized in blocks and wires.

As seen in the figure 3.1, a chair is placed at the center of the listening point for the user. The chair allows rotation around its axis and features a headrest to keep the user's head still. The height of the chair is preset so that users of different heights can achieve proper spatialization of sound.

Excluding the acoustic treatment of the room, this reproduction system was designed with a budget of 20,000 euros.

3.2 Software Setup

Bidule software is used to route the signal independently according to the appropriate speakers needed for the required spatialization. It is based on a block diagram, described below:

- **3OA Player**: the player where the tracks are loaded for later playback.
- **AllRA Decoder IEM spatial plugin**: a decoder used to properly decode the 3OA track into signals suitable for the current speaker array arrangement.
- **Gain blocks**: adjust the gain of the various channels
- **MultiEQ IEM plugin blocks**: a filter bank used to equalize all speakers individually
- **Delay blocks**: used to delay the signal on certain speakers if needed
- **Orion32 ASIO Driver**: routes processed signals to the sound card

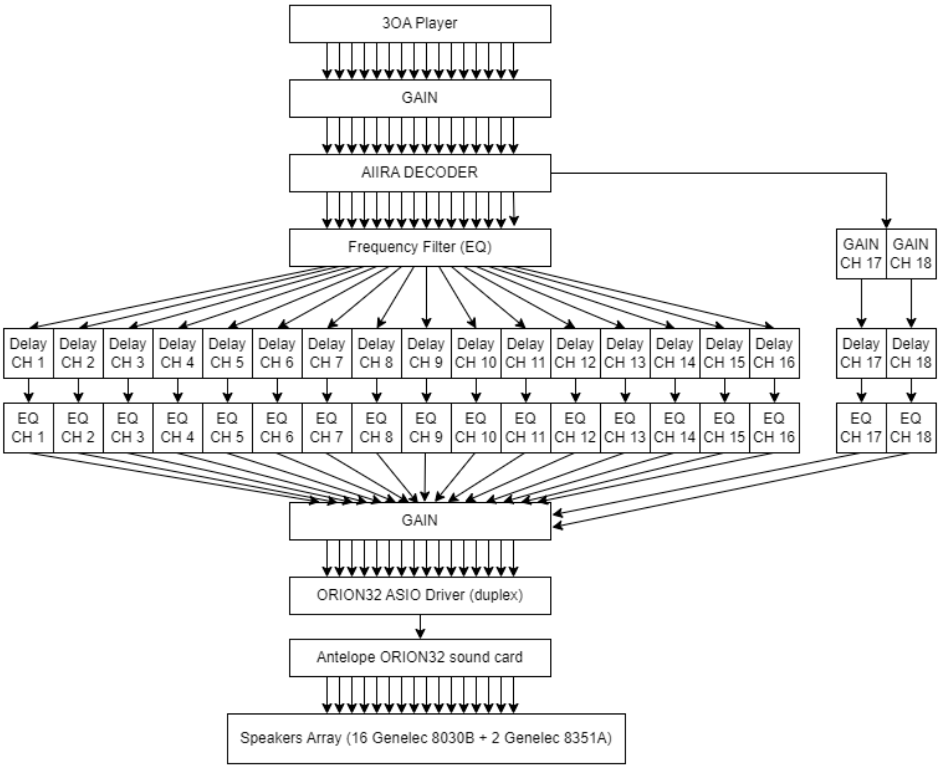


Figure 3.2: Block diagram of the Bidule playback system.

3.3 Preliminary Validation

Preliminary validation of this reproduction system, as shown in [35], has already been carried out by analyzing key acoustic metrics such as reverberation time (RT20), early decay time (EDT), and speech clarity (C50). This was done by reproducing a virtual acoustic scenario in the ASL and comparing it with the corresponding real environment.

A university lecture hall at the Polytechnic University of Turin was chosen as the environment. Five random positions within the lecture hall were selected, and measurements of the room impulse response (RIR) were made by playing a sinusoidal signal emitted by an NTi Audio Talkbox generator as the source. Each location had to be at least 2 meters away from the sound source and the other locations, and each measurement was repeated twice.

Measurements were recorded using the NTi Audio XL2 calibrated class 1 omnidirectional sound level meter (SLM) and the Zylia ZM-1 Spherical Microphone Array (SMA). The noise floor at one of the five positions was also recorded using the SLM. For virtual playback within the ASL, the 3OA tracks were played back and re-recorded using an XL2 microphone placed in the center of the speaker array. The 3OA ambisonic tracks were obtained by convolving the SMA recordings with the A2B-Zylia-3E-Jul2020 19x16 filter array.

Finally, all recordings were analyzed and compared using Matlab scripts to calculate the respective RIRs and related acoustic parameters of the analyzed room.

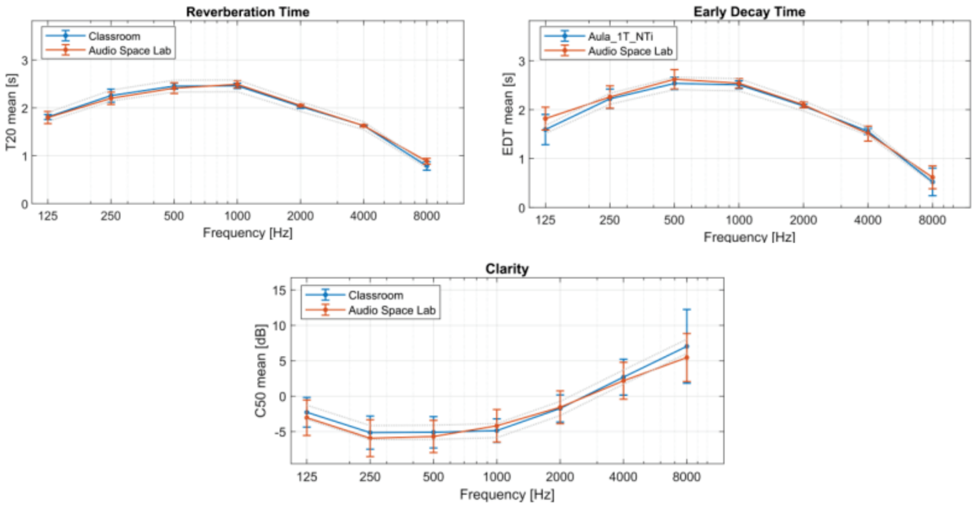


Figure 3.3: Results of the metrics analyzed in the preliminary validation: RT20, EDT, and C50, respectively, in the real environment (classroom) and in the virtual environment (ASL). The analyses are taken from [35].

After calculating all parameters for both the real and virtual conditions, frequency averages, standard deviations, and just-noticeable-difference (JND) values were determined. Overall, the analyses indicate that the virtual condition closely approximates the real condition (Figure 3.3). Specifically, the average RT20 values in the virtual room fall within the JND values of the real room for frequencies up to 8kHz, demonstrating that the reverberation time is adequately preserved even in the virtual reproduction condition. It should be noted, however, that the room does not provide significant additional reverberation at the sweet spot due to the system calibration procedure.

The average EDT values measured in the virtual room are also within the JND values of the real room, except for the two extreme frequencies analyzed (125Hz and 8kHz). The C50 clarity analysis shows that the average values in the virtual environment fall within the JND of the real environment up to frequencies of 8kHz. However, applying a 3dB estimate as JND for speech clarity under normal listening conditions, all average values would fall within the new JND values.

Preliminary validation thus yielded optimal results, indicating that the virtual condition effectively replicates the real condition. Therefore, the reproduction system can be considered a cost-effective alternative for clinics or hospitals for testing hearing aids or cochlear implants. However, analyses involving binaural cues, such as ITD, ILD, and IACC, were not conducted, nor were perceptual tests based on speech intelligibility. The impact of a visual display device (HMD), which provides visual information to the user in addition to the sound stimulus on the reproduced scene, has also not been investigated. This could offer greater spatiality but would involve additional costs and modifications to the current software for the spatial environment.

The following chapter details the progress of the new validation, which builds on the preliminary validation described here. We transitioned from a validation that only investigated monaural parameters to one that also includes binaural parameters, which are crucial for speech intelligibility and accurate reproduction of the real environment.

Chapter 4

Physical Validation of the Audio Space Lab

Further validation of the system was conducted to confirm the preliminary findings and enhance its implementation potential in clinics or hospitals. In the validation described in this study, we specifically focused on the physical validation of the laboratory, without addressing perceptual validation.

The validation consisted of a comparative analysis between two different types of reproductions. It was divided into two parts: Intra-ASL and Inter-ASL. The first part was conducted entirely within the Audio Space Lab and involved comparing two different reproductions made with varying parameters. Various measurement tools were used (Eigenmike microphone and artificial head-torso simulator), and a sine sweep was used as a stimulus, emitted either directly from individual speakers or virtually simulating the source by setting azimuth and elevation (using Bidule's multi-Encoder plugin). This allowed for a comparison between the real and virtual cases.

The second part took place in two environments: a university lecture hall at the Polytechnic University of Turin and the Audio Space Lab. In the classroom, measurements were taken to create ecological acoustic scenes, aiming to make the scenes as realistic as possible using different measurement tools and a sine sweep as input. A comparison was then made between the best configuration chosen in the Intra-ASL phase and the actual measurements made in the classroom.

Further recordings were then made in the ASL using the same measurement instruments used in the classroom, but using as the input signal the recording made in the actual classroom with the Eigenmike microphone for all source and receiver positions. This process allowed the real classroom environment to be reproduced inside the ASL to determine whether the installed virtual playback system could faithfully replicate the real environment or if there were any discrepancies.

4.1 Validation Intra-ASL

4.1.1 Comparisons with Headrest and HMD

As a first step in the intra-ASL measurements, we needed to decide whether to keep the headrest on the swivel chair placed at the center of the speaker array or to remove it for future analyses involving human participants. Once this decision was made, we then investigated the influence of the Head-Mounted Display (HMD) to determine how it affects the sound field. This exploration is crucial because in the future perceptual validation, providing users with both visual and auditory information could be beneficial.

Methods for Headrest Comparisons:

To evaluate the impact of the headrest, measurements were conducted using a head-torso simulator seated on the chair with the headrest, outfitted with B&K 4101 headphones connected to Scadas XS Hardware, which interfaced with a tablet for recording via an app. The chair was positioned at the center of the speaker array, aligning the dummy's ears at a height of 122.5cm from the ground. Initial measurements were taken with the headrest in place. Subsequently, the headrest was removed without disturbing the chair or dummy, and measurements were repeated without it.

A 1-second sine sweep, generated using Adobe Audition with a frequency range from 20 Hz to 20 kHz, in 16-bit mono at a sampling rate of 48,000 Hz, served as the stimulus. Each sweep was followed by 1 second of silence. Virtual stimulus playback was managed via Bidule software using the Multi-Encoder plugin, which enabled selection of specific azimuth and elevation angles to direct the sound perception from the surrounding speakers.

Measurements were initially recorded in the horizontal plane (0° elevation) at 15° intervals across the entire azimuth range. Following this, various azimuth positions (0°, -45°, -90°, -120°, -135°, 180°, 135°, and 45°) were selected, and measurements were taken at different elevation angles ranging from -60° to 60° in 15° increments for each azimuth. To avoid redundancy, each measurement was recorded once, resulting in a total of 88 unique recordings across both test conditions.

Results for Headrest Comparisons:

Graphs for ITD and ILD, both for the condition with a headrest and without a headrest, were calculated from Binaural Impulse Response (BIR) signals recorded via the head and torso simulator equipped with B&K 4101 headsets, using Matlab scripts. The outcomes of these analyses are presented in Figures 4.1 and 4.2.

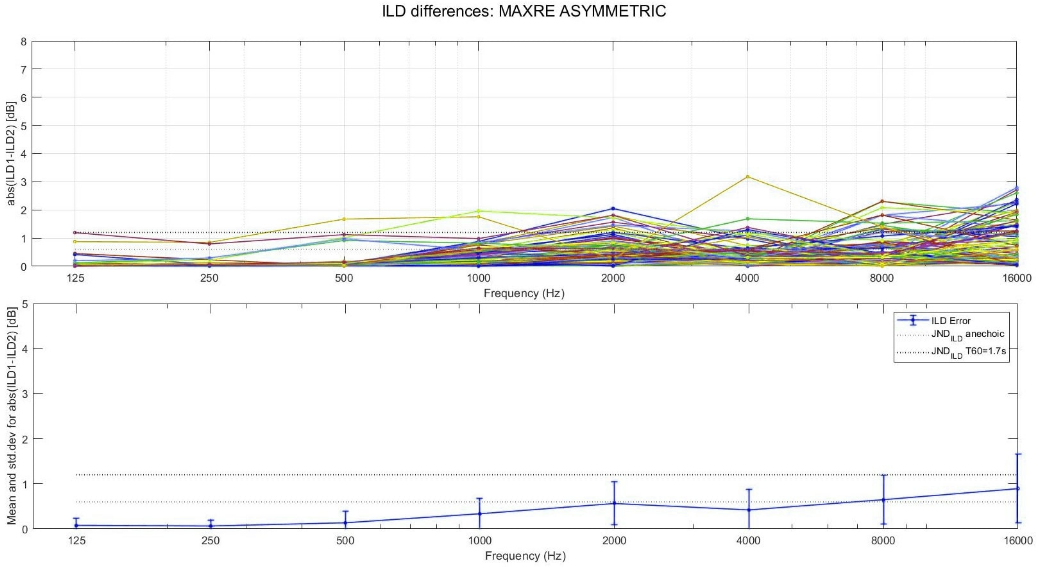


Figure 4.1: Differences in ILD between the condition with a headrest and the condition without a headrest.

The ILD graphs in Figure 4.1 do not clearly indicate a superior condition between with and without the headrest; both conditions appear equivalent.

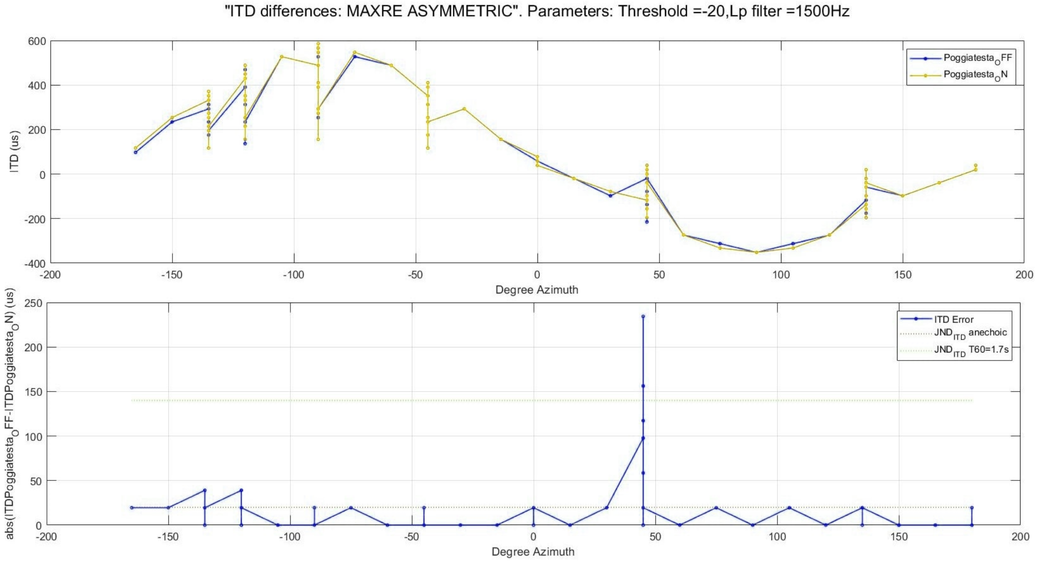


Figure 4.2: Differences in ITD between the condition with a headrest and the condition without a headrest.

However, the ITD graphs in Figure 4.2 suggest a slight improvement in timing without the headrest compared to with it, here the 45° azimuth angle exceeds the limits imposed by the JND. Ultimately, while neither condition is decisively superior, the results slightly favor the condition without the headrest.

Based on these findings, it was decided to proceed with the condition without the headrest for future perceptual analyses on speech intelligibility. However, it remains possible to reassess this decision in the future by conducting additional analyses with the headrest to validate the chosen approach.

Methods for HMD Comparisons:

To assess whether the head-mounted display (HMD) introduced reflections and affected ITD and ILD, new measurements were conducted using a dummy seated on the swivel chair, equipped with B&K 4101 headphones connected to Scadas XS Hardware, which interfaced with a tablet for recording via an app. The dummy wore the HMD on its head, with its ears positioned at a height of 122.5cm from the ground. The chair was consistently positioned at the center of the speaker array without a headrest, following the decision from previous analyses.

A sine sweep was generated using Adobe Audition, with a duration of 1 second, covering frequencies from 20 Hz to 20 kHz, in 16-bit mono format with a bitrate of 48000 Hz. Each stimulus was followed by 1 second of silence. The stimulus playback was conducted similarly to the headrest comparison, using Bidule software and the Multi-Encoder plugin, to virtually set azimuth and elevation.

Recordings were made in the horizontal plane (0° elevation) at every 15° azimuth variation across the entire plane. Subsequently, different azimuth positions (0°, -90°, -120°, 180°, 135°, and 45°) were selected, and measurements were taken while varying the elevation angle from -60° to 60° in 15° increments for each azimuth position. Eliminating duplicate recordings, a total of 72 measurements were taken with the HMD, which were compared with those previously taken without a headrest and HMD, using the same 72 positions recorded here.

Results for HMD Comparisons:

Graphs for ILD and ITD, both for the condition with HMD and without HMD (both without the headrest), were created using Matlab scripts from the Binaural Impulse Response (BIR) signals recorded via the head and torso simulator equipped with B&K 4101 headphones. The results are shown in Figures 4.3 and 4.4.

Again, as in the headrest influence analysis, we do not observe a clear difference between the conditions with HMD and without HMD. Looking at the ILD graph in Figure 4.3, it is not possible to determine which condition is superior, while from the ITD graph in Figure 4.4, the situation varies depending on the azimuth angle analyzed.

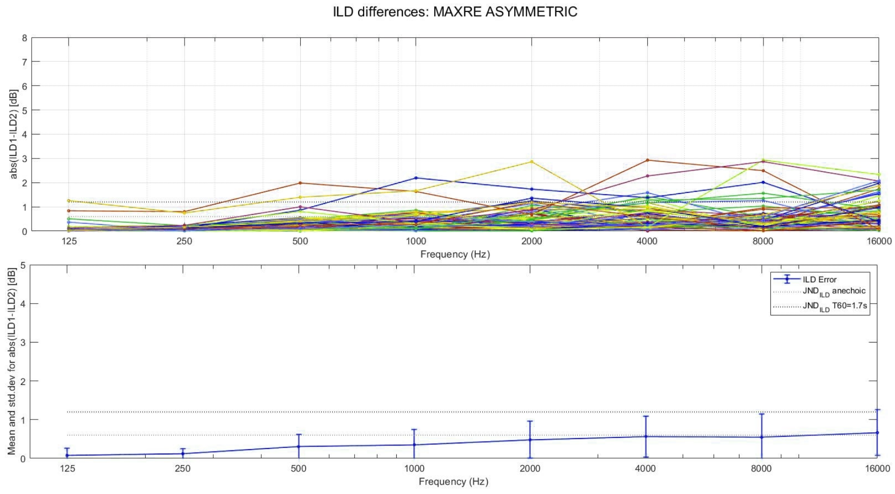


Figure 4.3: Differences in ILD between the condition with HMD and the condition without HMD (both without the headrest).

For angles with positive azimuth, the ITD is lower in the condition without the HMD, whereas for angles with negative azimuth, the ITD is better in the condition with the HMD.

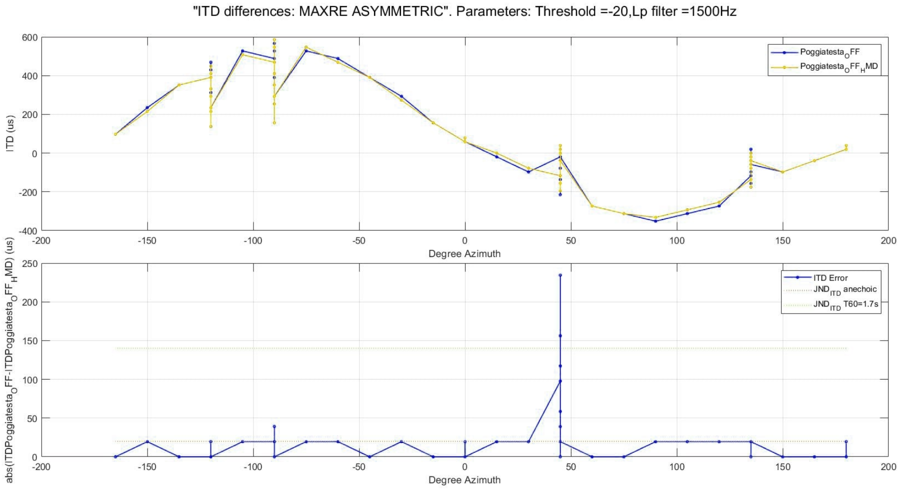


Figure 4.4: Differences in ITD between the condition with HMD and the condition without HMD (both without the headrest).

Therefore, it was concluded that the use of an HMD does not significantly affect the sound field and its parameters (ITD and ILD). Therefore, in future perceptual validation, the use of an HMD will not introduce excessive reflections that would invalidate speech intelligibility analyses.

4.1.2 Acoustic Treatment

Next, questions needed to be raised about the adequacy of the laboratory's soundproofing in its current state before proceeding with the system validation. Using the EM64 Eigenmike microphone array and a head simulator as measuring instruments, comparative measurements were performed to assess different options for enhancing the laboratory's soundproofing. The current concern is that existing reflections might degrade the perception of sound emanating from certain speakers.

Methods:

To enhance the soundproofing of the room, three conditions were tested: firstly, absorber panels were installed behind the speaker positioned at -45° azimuth, in the corner of the room adjacent to the window, and behind the speaker at 90° azimuth, on the opposite side of the room, where potentially disruptive reflections could occur. Secondly, an absorptive acoustic curtain was added to fully cover the laboratory window facing the courtyard of the Polytechnic. Lastly, additional panels were placed to reinforce the corners of the room and the sections of the laboratory wall adjoining the window. The tested acoustic conditions are depicted in Figure 4.5.

The analyses of the three conditions were conducted using an em64 Eigenmike microphone array, with lateral, frontal, and transverse polar plots analyzed, along with a head-torso simulator (HATS). Both the Eigenmike and HATS were positioned at the center of the loudspeaker array, at a height of 122.5cm from the ground. A total of 16 loudspeaker positions were tested in both real and virtual modes.

For the real-world conditions, the output speaker of the sound source was manually selected using the Bidule software's Audio-Matrix function. By choosing a channel from 1 to 16, the sound corresponding to the selected speaker was reproduced.

For the virtual condition, the Bidule software was employed again, this time utilizing the Multi-Encoder plugin (see Figure 4.6). This plugin allows setting azimuth and elevation to choose the "zone" from which the sound should be perceived, utilizing all the speakers adjacent to the selected zone.

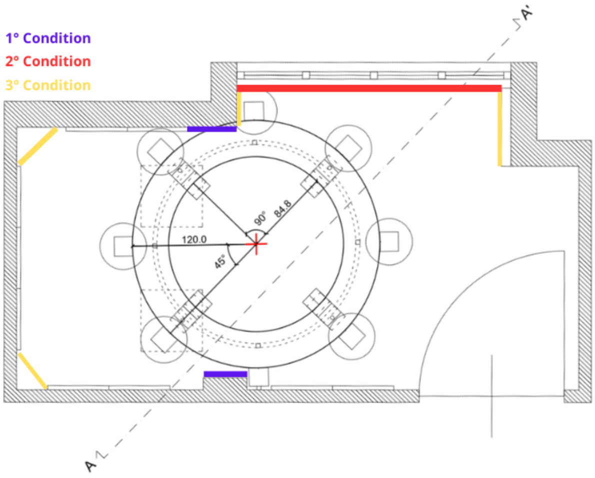
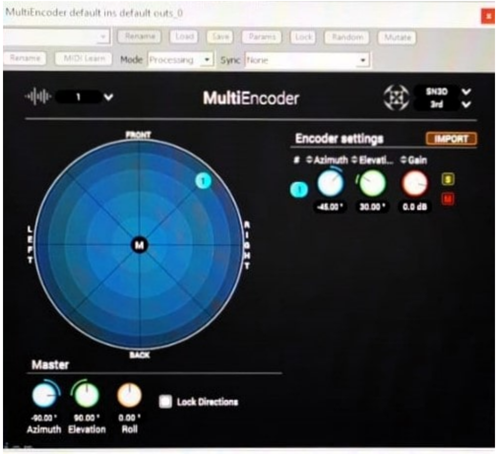


Figure 4.5: 1:60 scale floor plan of the Audio Space Lab, depicting the three tested acoustic conditions (1-blue, 2-red, 3-yellow).

A sine sweep, generated through Adobe Audition, with a duration of 1 second and frequencies ranging from 20 Hz to 20 kHz, was used as the stimulus. Each stimulus was repeated 3 times, with 1 second of silence after each repetition.


MultiEncoder default ins default outs_0

Mode Processing

Sync None

SN3D 3rd

Encoder settings

# Azimuth Elevati... Gain

1 -45.00° 30.00° 0.0 dB

Master

-90.00° Azimuth 90.00° Elevation 0.00° Roll

□ Lock Directions

Figure 4.6: Multi-Encoder plugin of the Bidule software, allowing you to set the azimuth and elevation of the desired sound source.

Results:

Polar diagrams for all three acoustic conditions tested were created using Matlab scripts from the impulse responses (IR) calculated from recordings made with the

Eigenmike.

Analyzing the virtual polar plots across the three acoustic conditions, an improvement is evident when comparing the pre-acoustic treatment plots with the post-acoustic treatment plots that include the addition of acoustic panels. Transitioning from the condition with just the panels to the condition with the added curtain shows small improvements in acoustic conditions, though they are not significant. However, in the last tested condition, which includes additional panels in the critical corners and around the curtain in addition to those already added in the first condition, no significant improvements in terms of polar diagrams are evident.

Therefore, it was decided to proceed with the validation using the first acoustic condition, which involved the use of only the initial set of acoustic panels. The use of an acoustic curtain and additional panels in the corners of the room did not significantly improve the room's acoustic conditions. The improvement achieved with this acoustic condition, compared to the initial treatment used for preliminary validation, is evident.

The improvement between the pre-acoustic treatment and post-acoustic treatment conditions is shown in Figures 4.7 and 4.8. The first condition tested, which was chosen for validation, involved the presence of acoustic absorbing panels at a frequency of 1000Hz. These improvements are also evident when raising the frequency to 2000 Hz and 4000 Hz.

Applying acoustic panels inside the laboratory eliminates a significant portion of the reflections generated by sound. Except when the source is oriented at -45° and -135° azimuth, where there is no notable improvement, in all other cases tested, the localization is more accurate in the post-treatment condition, with the direct sound directed more accurately toward the angle under consideration.

Comparison analyses between all tested acoustic conditions are also shown. Comparisons of side polar plots at 1000 Hz between the first and second acoustic conditions can be seen in Figures 4.9 and 4.10. As illustrated, there is not much difference between the first and second acoustic conditions. In fact, in some cases, there is a worsening of localization (at -45°, -90°, and 135° azimuth) because reflections are present that were not present in the first condition. Therefore, the acoustic curtain does not seem to provide a significant advantage over the first condition tested with the acoustic panels. In some cases, there is a slight improvement, while in other cases, it worsens sound localization.

A comparison between the second and third acoustic conditions (Figures 4.11 and 4.12) shows that applying additional panels in the positions described above does not provide any advantage for sound localization, as the polar diagrams are almost identical. Thus, the second and third acoustic conditions do not differ much in comparison with the first.

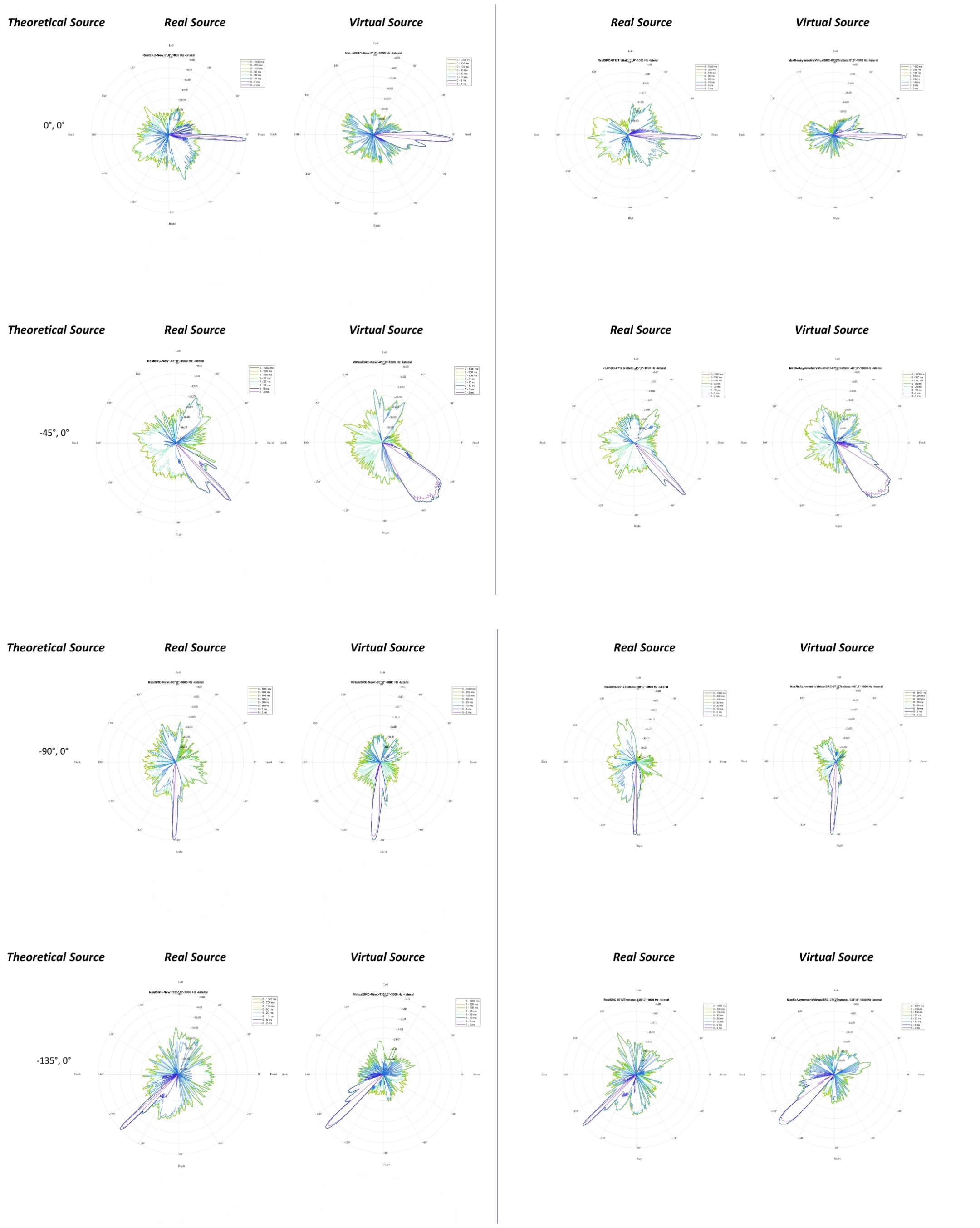


Figure 4.7: Lateral polar diagrams at 0° elevation and azimuths of 0°, -45°, -90°, and -135°, for both real and virtual environments at 1000 Hz frequency, illustrating pre-treatment conditions (left) and the first treatment condition (right).

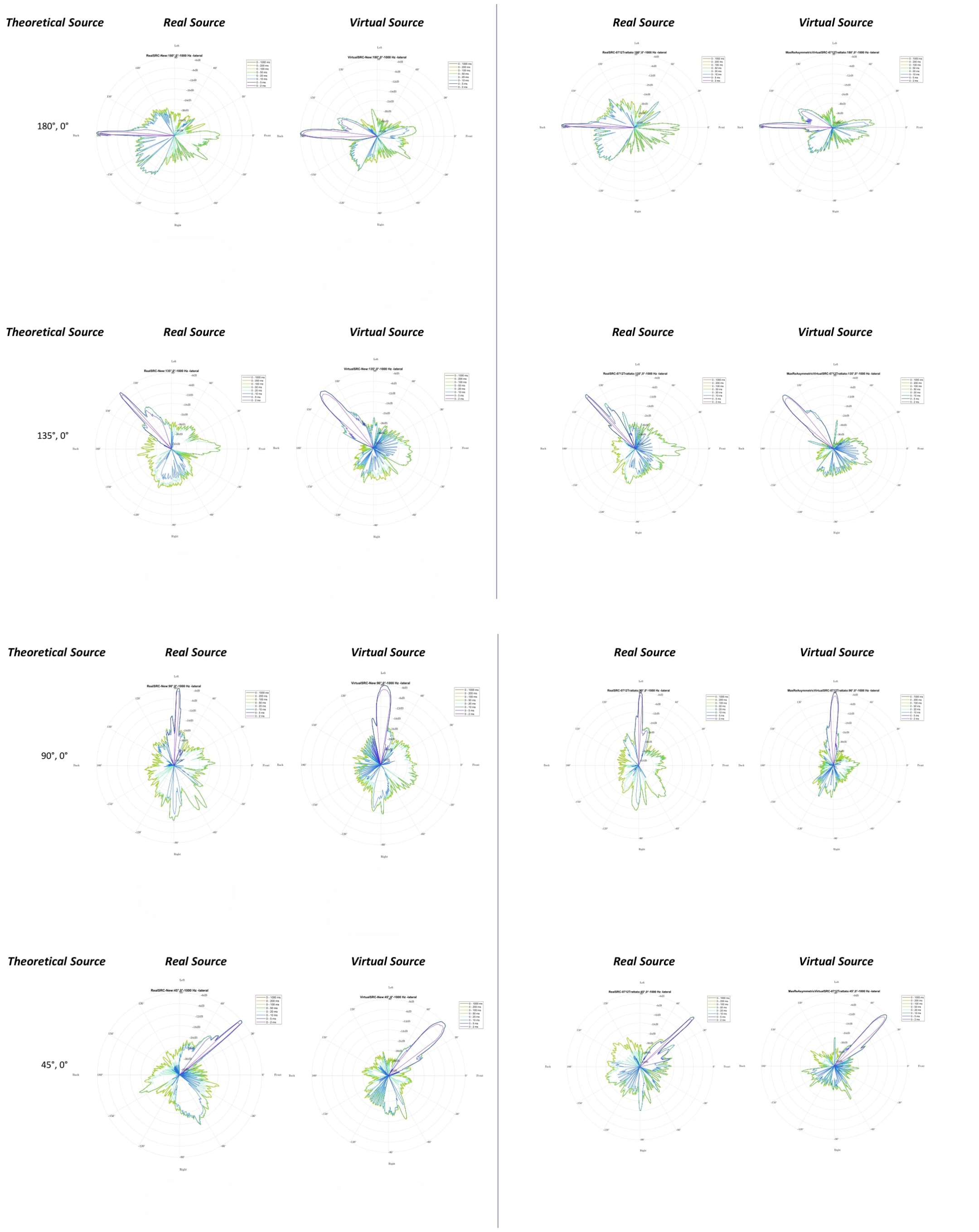


Figure 4.8: Lateral polar diagrams at 0° elevation and azimuths of 180°, 135°, 90° and 45°, and -135°, for both real and virtual environments at 1000 Hz frequency, illustrating pre-treatment conditions (left) and the first treatment condition (right).

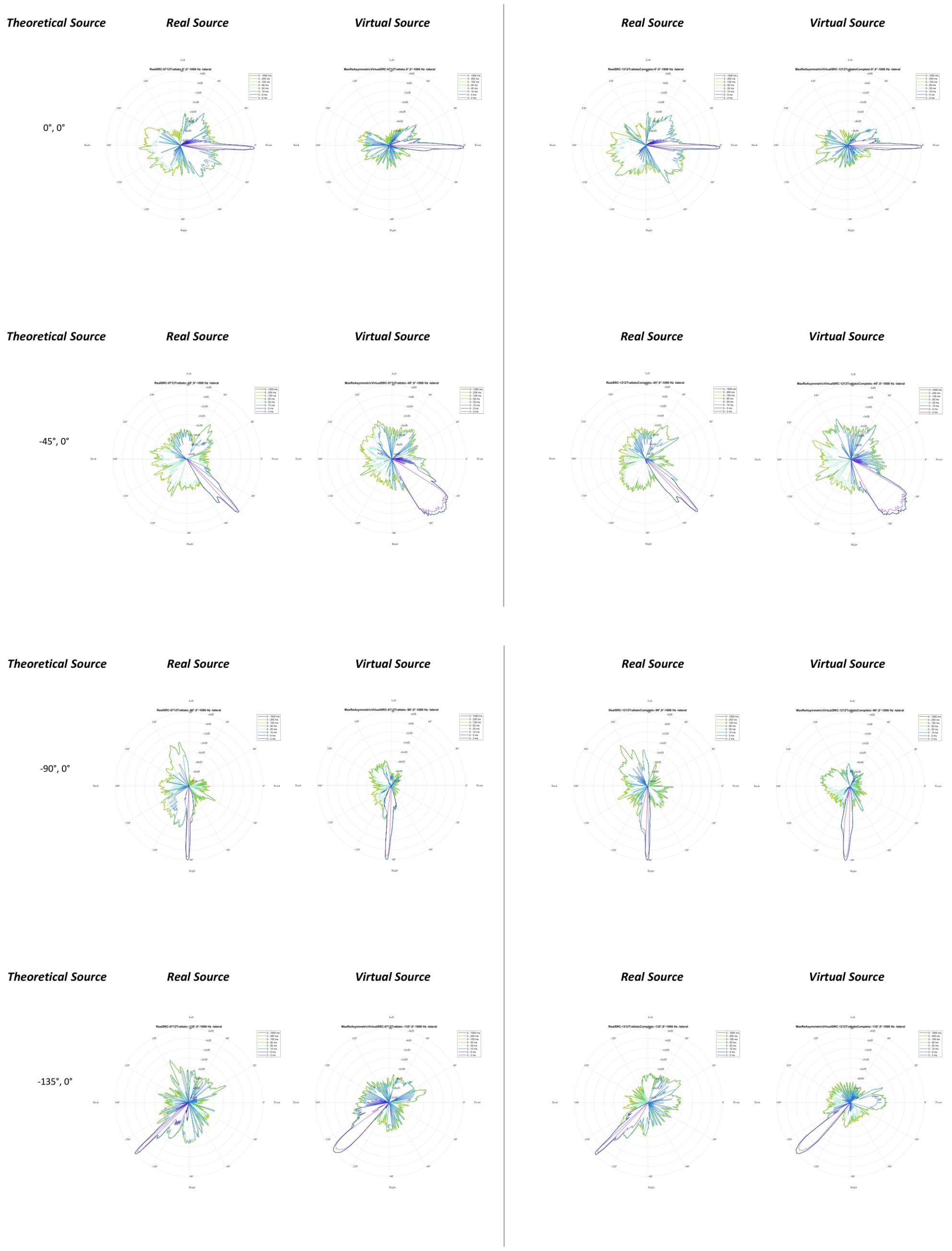


Figure 4.9: Lateral polar diagrams at 0° elevation and azimuths of 0°, -45°, -90°, and -135°, for both real and virtual environments at 1000 Hz frequency, illustrating laboratory acoustic conditions with the first treatment condition (left) and the second treatment condition (right).

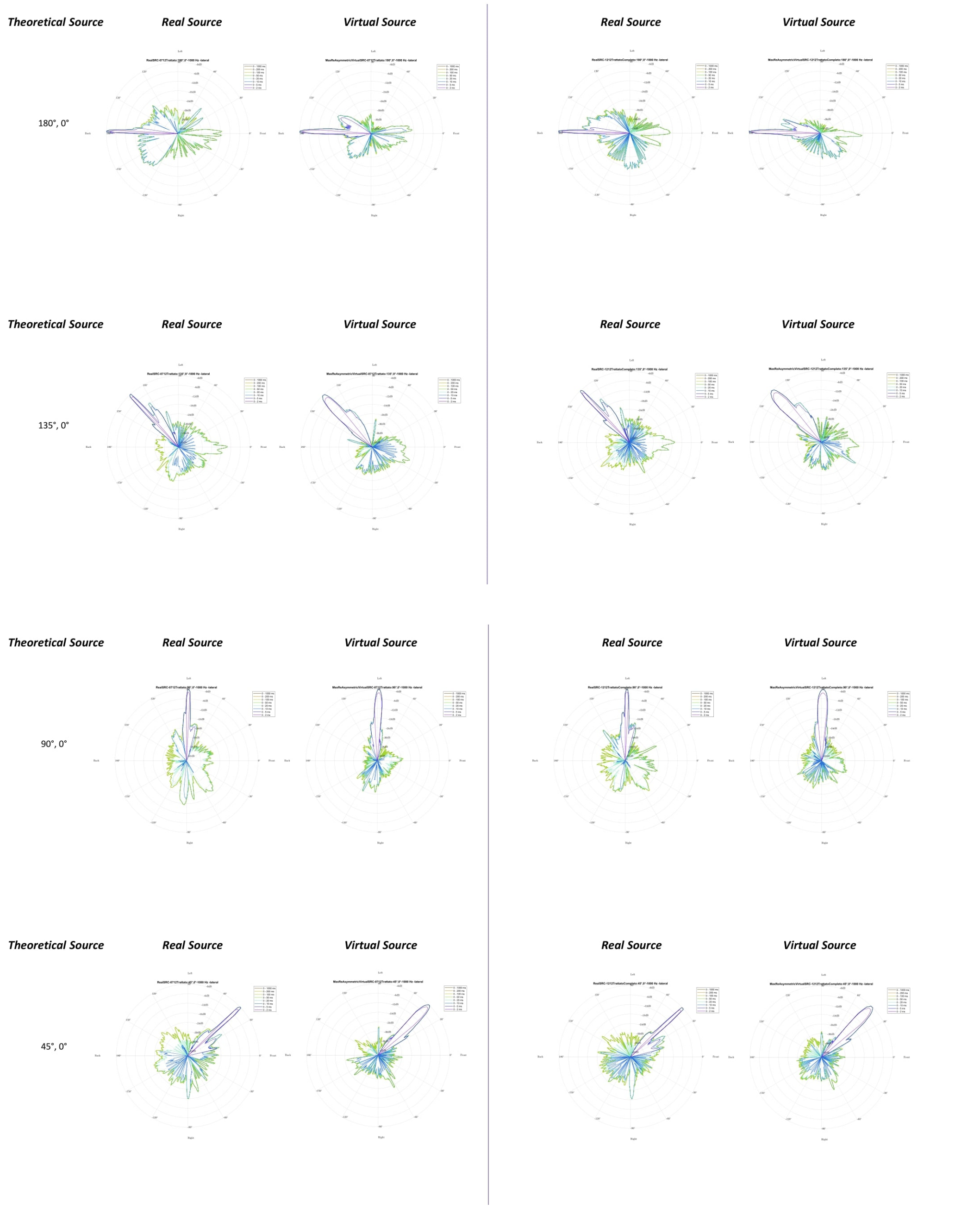


Figure 4.10: Lateral polar diagrams at 0° elevation and azimuths of 180°, 135°, 90°, and 45°, for both real and virtual environments at 1000 Hz frequency, illustrating laboratory acoustic conditions with the first treatment condition (left) and the second treatment condition (right).

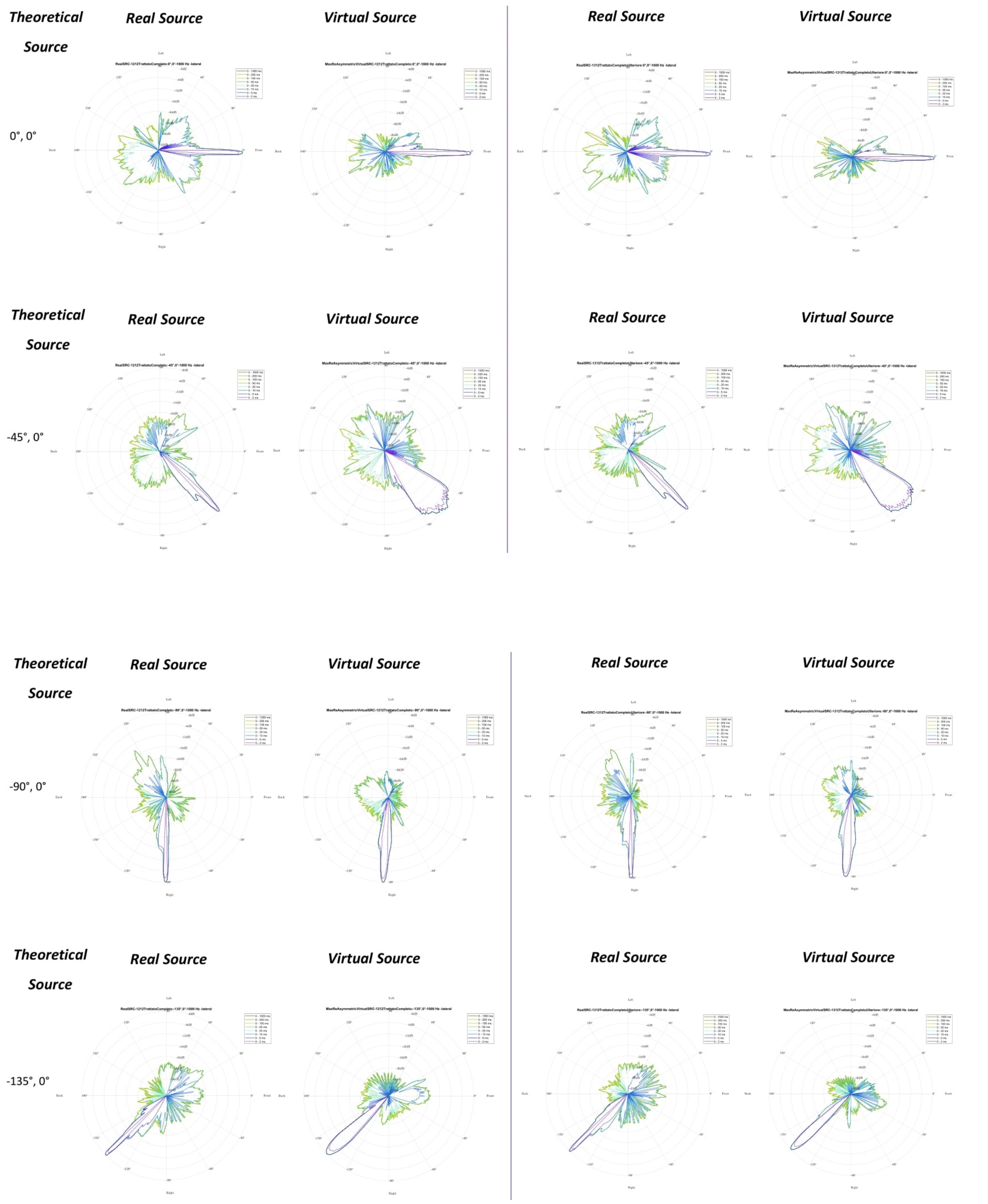


Figure 4.11: Lateral polar diagrams at 0° elevation and azimuths of 0°, -45°, -90°, and -135°, for both real and virtual environments at 1000 Hz frequency, illustrating laboratory acoustic conditions with the second treatment condition (left) and the third treatment condition (right).

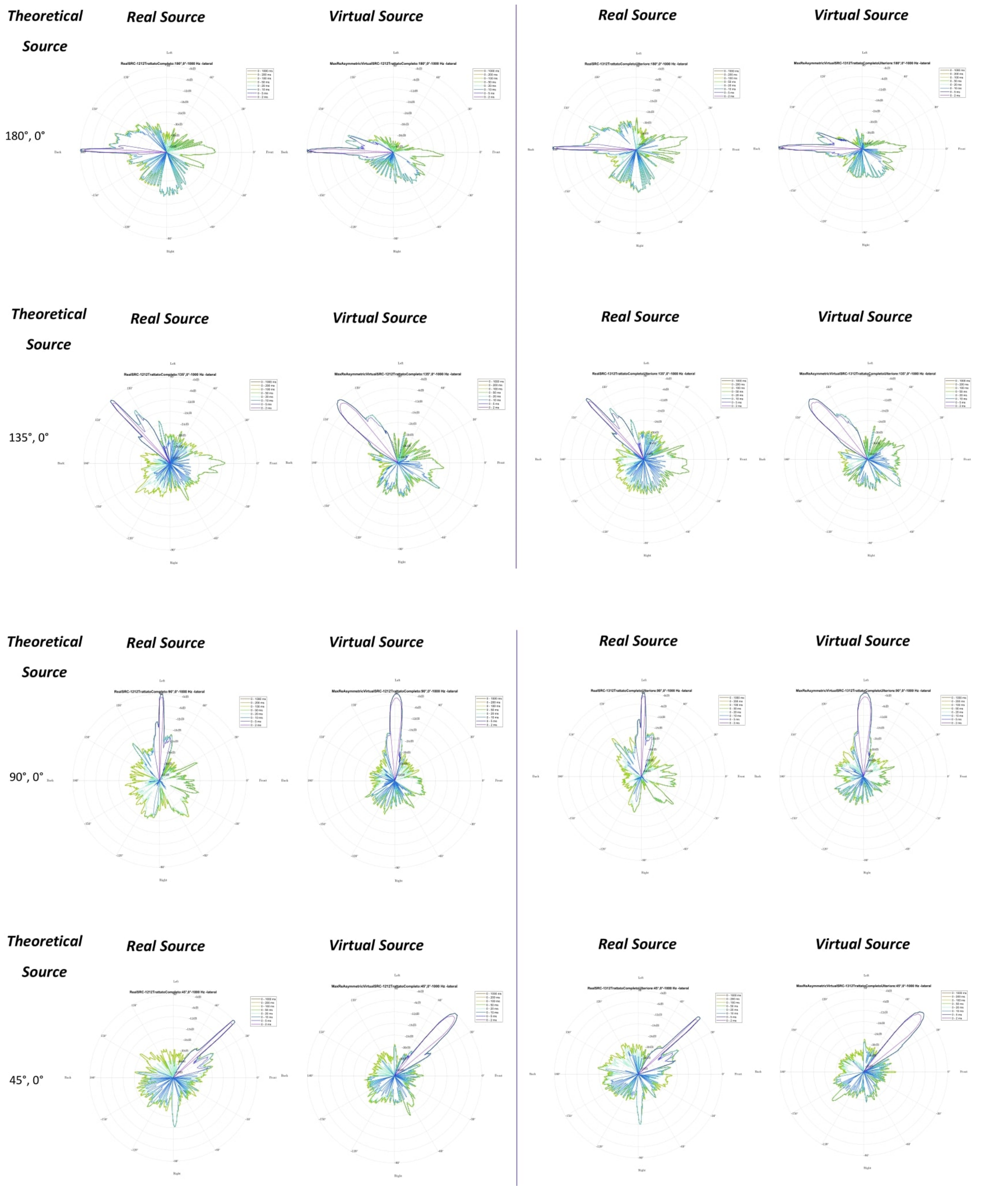


Figure 4.12: Lateral polar diagrams at 0° elevation and azimuths of 180°, 135°, 90°, and 45°, for both real and virtual environments at 1000 Hz frequency, illustrating laboratory acoustic conditions with the second treatment condition (left) and the third treatment condition (right).

Again using Matlab scripts, from the Binaural Impulse Response (BIR) signals recorded with the HATS, ITD and ILD were calculated.

ITD and ILD in the pre-treatment and post-treatment conditions with acoustic panels are reported to verify the correctness of the choice made. For calculating ITD and ILD, maxRe was chosen as the decoder; the next section will analyze the decoder choice in detail. For this purpose, the ITD and ILD pre-acoustic treatment and post-acoustic treatment, calculated with maxRe decoding, are shown in figures 4.13 and 4.14.

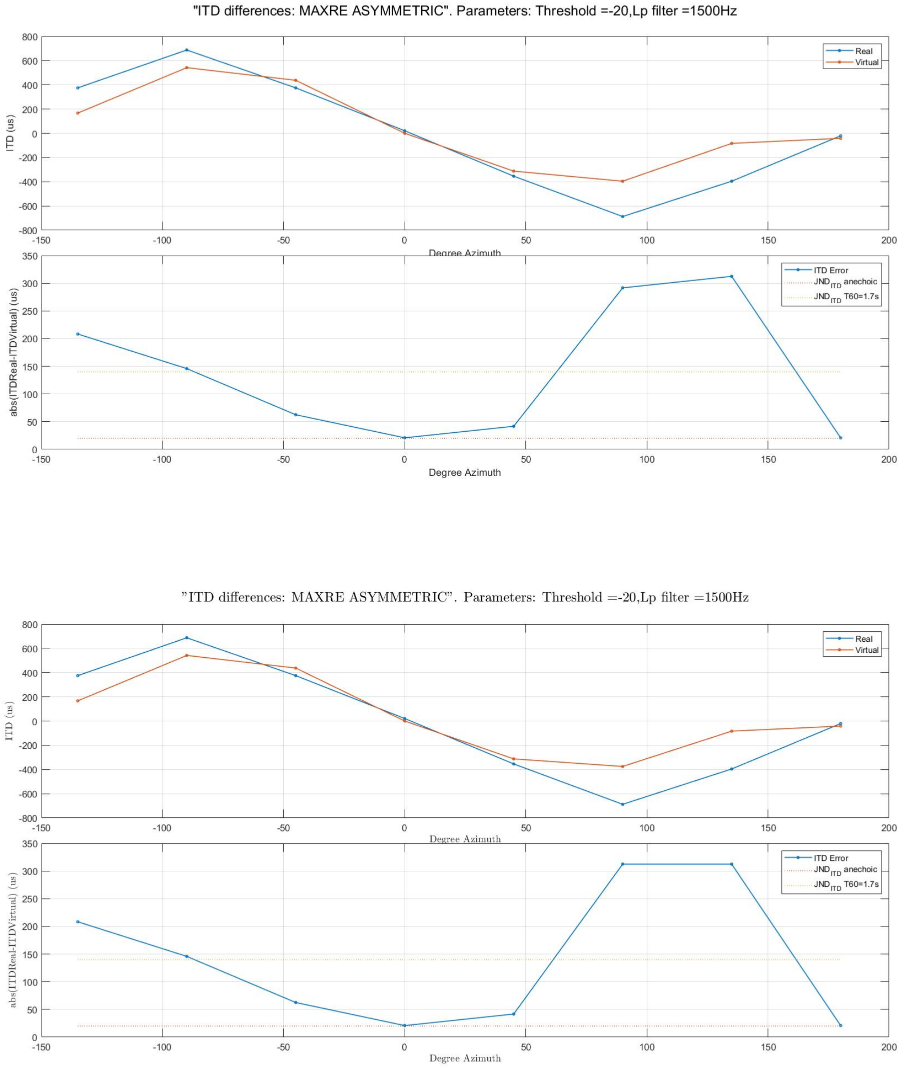


Figure 4.13: ITD differences for MaxRe decoders pre-treatment (above) and post-treatment (below).

It can be observed from figure 4.13 that the acoustic treatment affects the ITD error at the 90° azimuth angle, maintaining stability up to the 135° angle.

Overall, the ITD remains stable between the acoustic pre-treatment and post-treatment conditions, with only a variation observed in one corner.

The differences between pre-acoustic treatment ILD and post-acoustic treatment ILD are more pronounced than those for ITD. In figure 4.14, it can be observed that the ILD error decreases in the post-treatment condition, with noticeable benefits particularly at frequencies of 500 Hz, 2000 Hz, and 4000 Hz, along with a slight improvement at 1000 Hz.

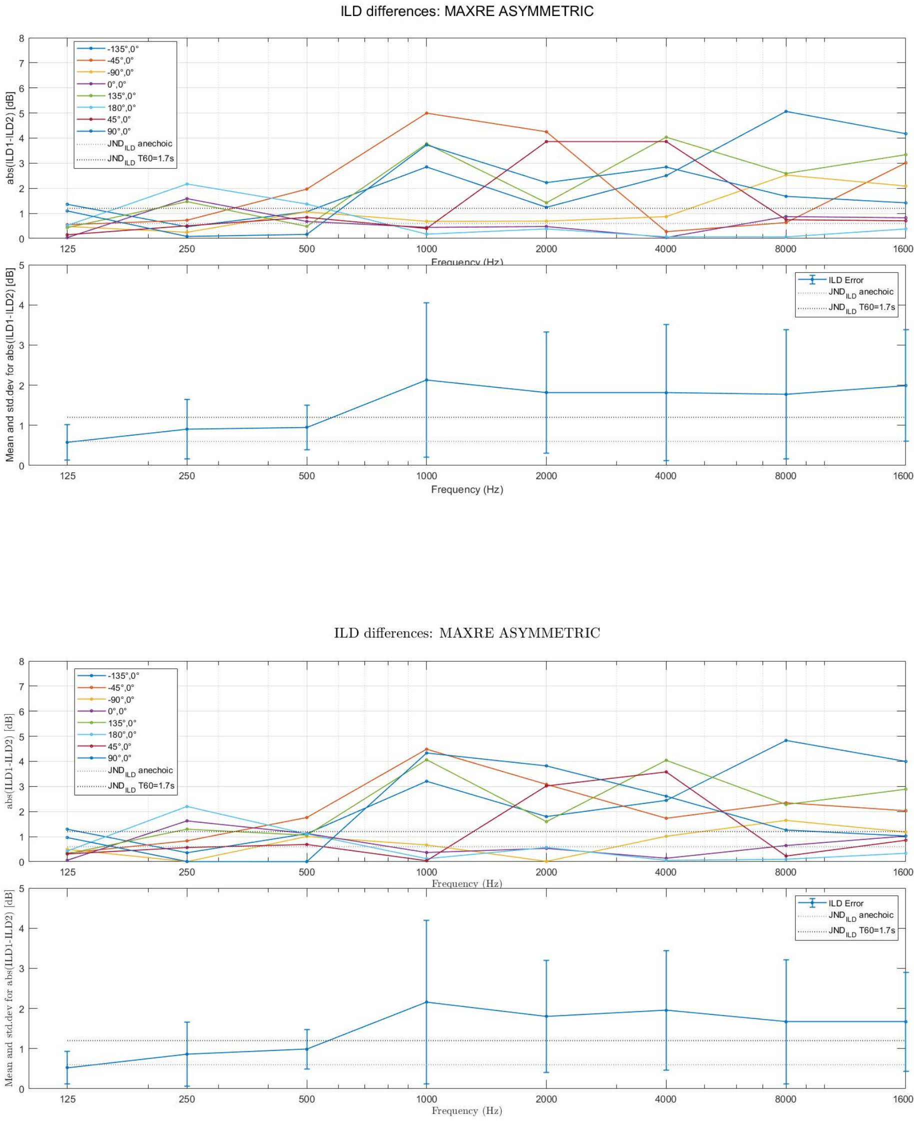


Figure 4.14: ILD differences for MaxRe decoders pre-treatment (above) and post-treatment (below).

In general, the results obtained from calculating the ITD and ILD confirm the findings shown by the polar diagrams, with greater improvements observed in the ILD compared to the ITD. This supports the decision to apply acoustic panels inside the laboratory.

4.1.3 Comparison Decoder

After deciding on the acoustic treatment to be implemented in the laboratory, the next step was to select the decoder technique (Basic or MaxRe) for use in Bidule before proceeding with validation, ensuring consistent use of the chosen decoding technique throughout all measurements.

Methods:

Two decoding techniques, MaxRe and Basic, with two different weightings for the same specific decoder (AllRA), were compared to determine which one to use for the laboratory validation. MaxRe, the traditional method used previously, set the azimuth for speakers 9 and 10 at 39° and -40° respectively, without altering the elevation parameter. In contrast, Basic decoding adjusted the weighting and changed the azimuth to 45° and -45° for speakers 9 and 10, respectively, while also modifying the elevation to -45°. These adjustments were configured within the Bidule software's AllRA Decoder block before selecting the decoding method for recordings.

To decide on the suitable decoding technique, measurements were conducted using a head-torso simulator (HATS) and an EM64 Eigenmike microphone array. Recordings were made at all 8 speaker positions in the horizontal plane (elevation 0°), both in real mode using the Bidule Audio-Matrix plugin and in virtual mode using the Multi-Encoder plugin, for each measurement instrument. This resulted in 16 recordings for each decoder type analyzed per measurement instrument.

A sine sweep generated with Adobe Audition, lasting 1 second with frequencies ranging from 20 Hz to 20 kHz, served as the stimulus. Each stimulus was repeated 3 times, with 1 second of silence following each repetition.

Results:

Polar diagrams for both MaxRe and Basic encodings were made using Matlab scripts, from the impulse responses (IR) calculated from the recordings made with the Eigenmike, to check which decoding was better for ASL validation. From the analysis of the polar diagrams shown in Figure 4.15, no clear superiority of one decoding over the other is evident; certain angles exhibit better localization with Basic encoding, while others show better results with MaxRe decoding. Overall, however, Basic encoding appears to introduce more reflections.

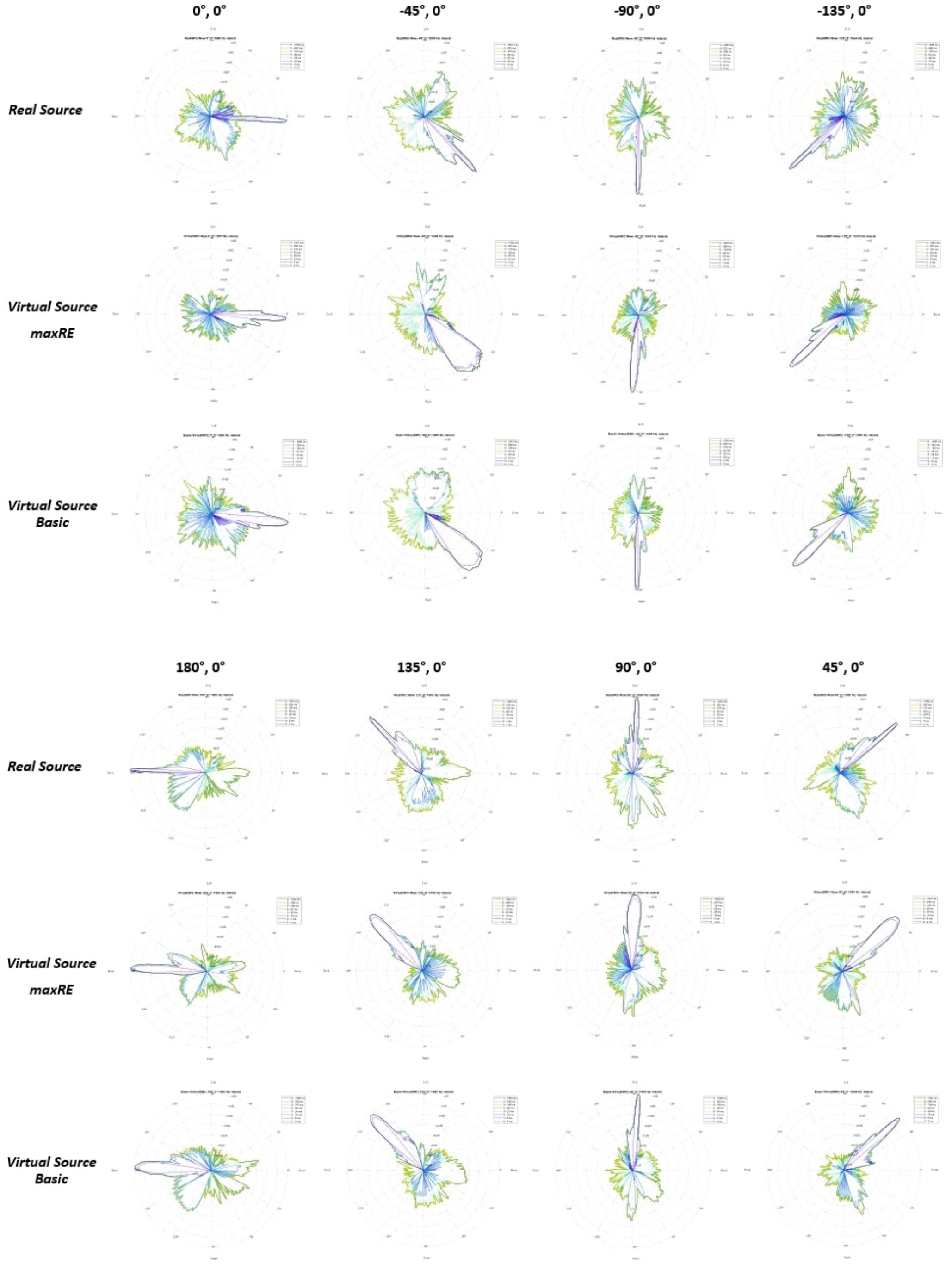


Figure 4.15: Comparison of lateral polar diagrams with MaxRe and Basic decoding for all azimuth angles at 0° elevation, at a frequency of 1000 Hz.

ITD and ILD for each type of decoding were also calculated using Matlab scripts from the Binaural Impulse Response (BIR) signals recorded with the HATS. Figure 4.16 shows the ITD for the MaxRe and Basic decoders, comparing the real and virtual cases, and then calculating the absolute difference between real ITD and virtual ITD for each azimuth angle analyzed.

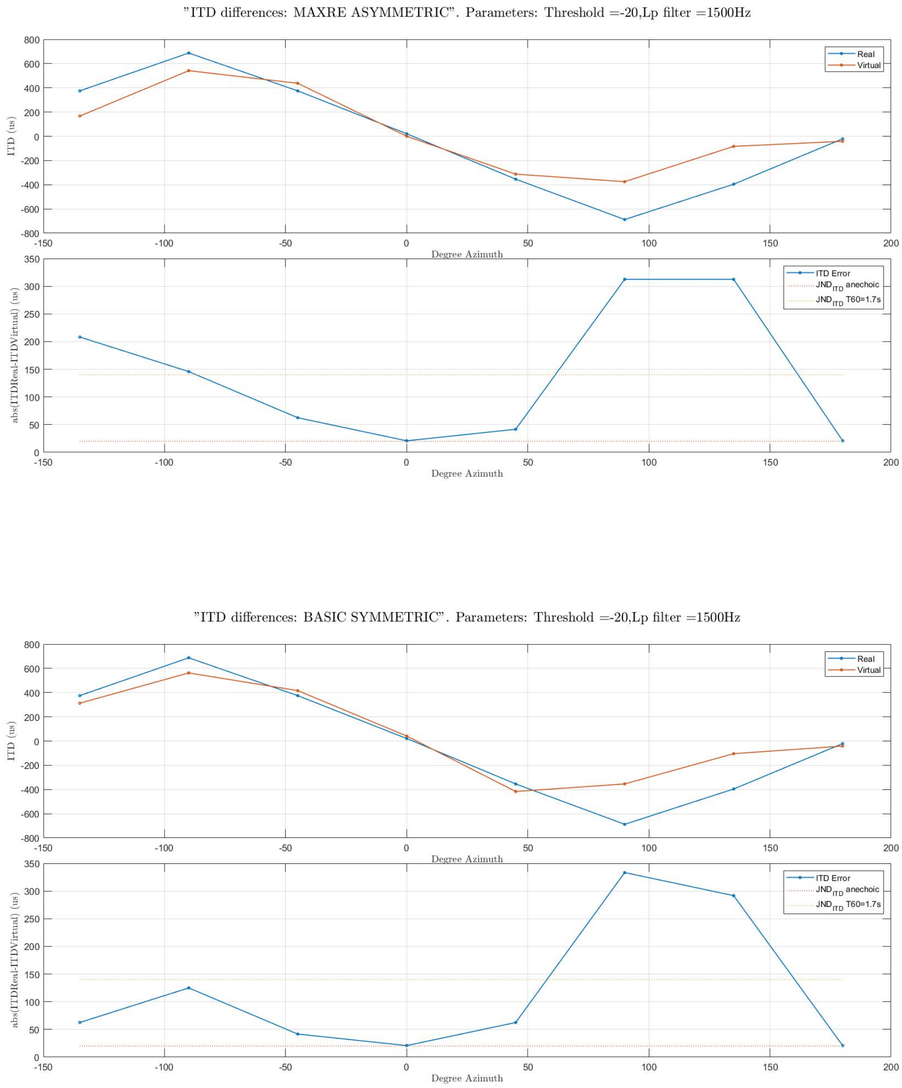


Figure 4.16: ITD differences for MaxRe (above) and Basic (below) decoders.

Once graphs were plotted for real cases and virtual cases for both decoders, the two JNDs, in the anechoic case and the case with RT60=1.7s, were considered to determine which decoding was more effective.

In contrast to the analyses performed in the previous section with the pre-treatment and post-treatment comparison of ITD and ILD, in this case, the graphs shown in Figures 4.16 and 4.17 do not clearly demonstrate a difference between the two types of decoding analyzed.

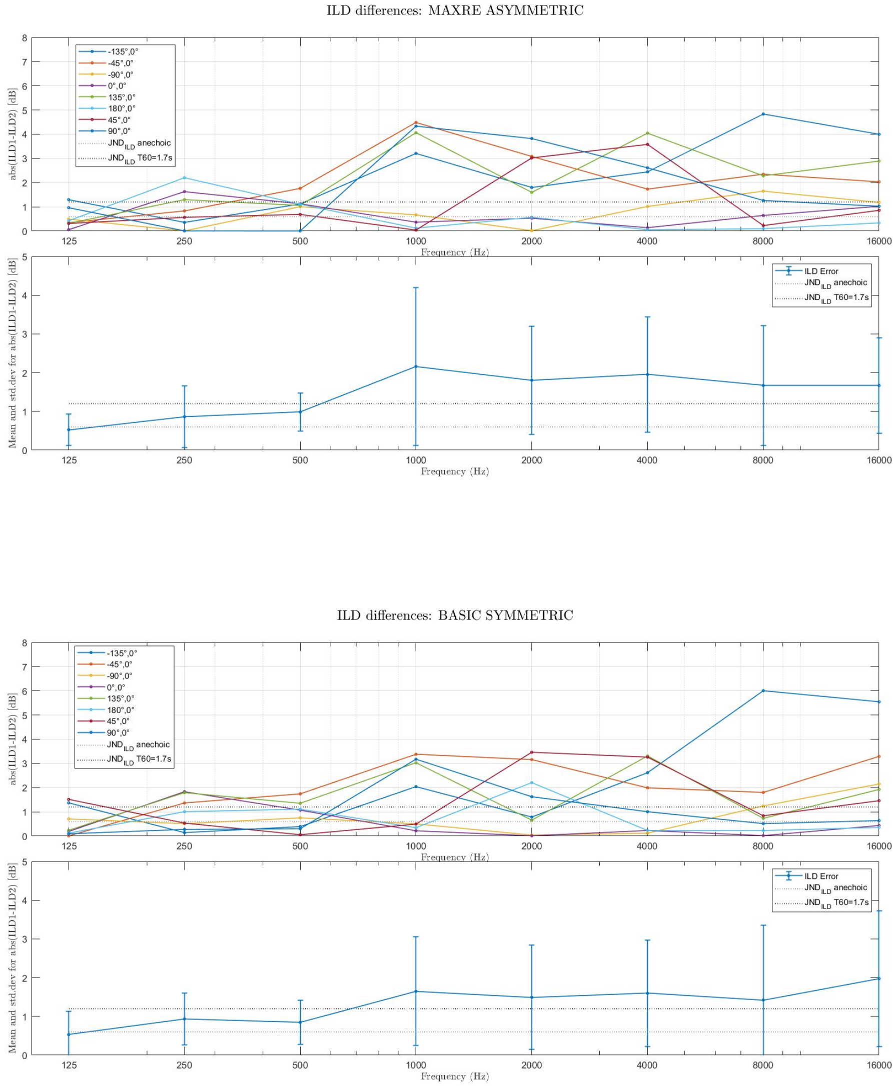


Figure 4.17: ILD differences for MaxRe (above) and Basic (below) decoders.

In the case of the ILD (Figure 4.17), there appears to be an improvement in the error with the maxRe decoder, whereas from the ITD results (Figure 4.16), it is not clear which decoder is more effective. Based on the polar diagrams analyzed above, maxRe appears to be slightly better as a decoding technique.

From previous studies, particularly highlighted in [36], it has been demonstrated that the maxRe decoding method is the most efficient, especially at higher frequencies where Basic decoding tends to perform less effectively. MaxRe decoding is generally preferred due to lower errors in pressure and intensity. Therefore, based on literature findings where maxRe decoding consistently showed superior results compared to Basic decoding, it was decided to proceed with laboratory validation using maxRe decoding

4.1.4 Sweet Spot Size Analysis

Once the best-performing acoustic solution for the playback system (first condition) and the most efficient audio decoding (MaxRe) were chosen, analyses were carried out with the decentering of the sweet spot to understand how the sound field degrades as the position of the sweet spot changes. This was conducted entirely within the Audio Space Lab using various measurement tools, listed and described below. The analyses focused on binaural cues.

Methods:

Measurements were conducted using different instruments: an EM64 Eigenmike array microphone and a Head and Torso Simulator (HATS). All instruments were positioned at the center of the speaker array, at a height of 122.5cm from the ground. Both real and virtual listening conditions were tested. Real listening conditions involved manually selecting the speaker from which the sound source was emitted, using the Audio-Matrix plugin of the Bidule software. The virtual listening conditions involved specifying azimuth and elevation angles, utilizing multiple speakers simultaneously through the Bidule software's Multi-Encoder plugin. For all measurement instruments, the sound source positions were selected from among the loudspeakers within the array for both real and virtual conditions, totaling 16 positions for each. Additionally, for the Eigenmike microphone in the virtual condition on the 0° elevation plane, additional azimuth measurements were taken, specifically with a 15° variation across the entire plane.

After taking the measurements with the chosen microphone at the center of the speaker array, the measuring instrument was moved 10cm and then 20cm to the left (in front of the speaker located at 0° azimuth and 0° elevation). Next, measurements were repeated at 16 real and 16 virtual listening positions for both shifts from the center to assess how the sound field degrades. Off-center measurements were repeated for all instruments used. A sinusoidal sweep, generated with Adobe Audition, lasting 1 second and with frequencies between 20 Hz and 20 kHz, was used as the stimulus. Each stimulus was repeated 3 times, with 1 second of silence after each repetition.

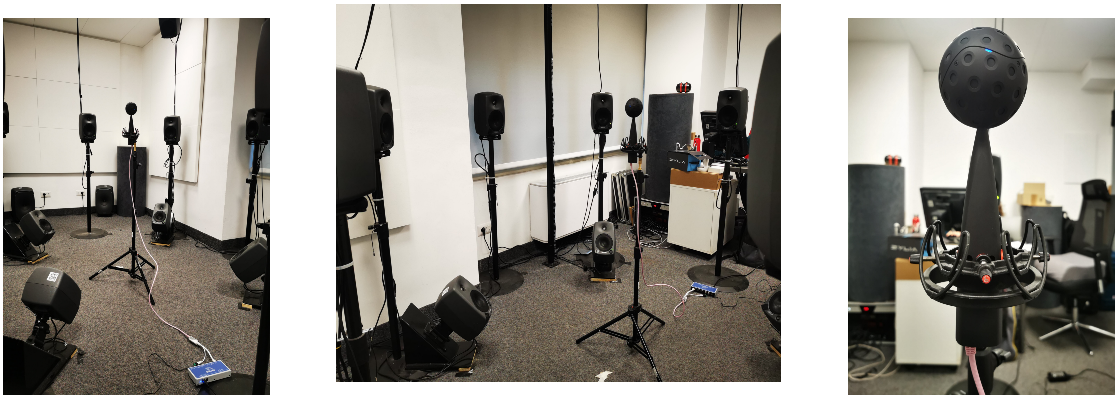


Figure 4.18: Em64 Intra-ASL.

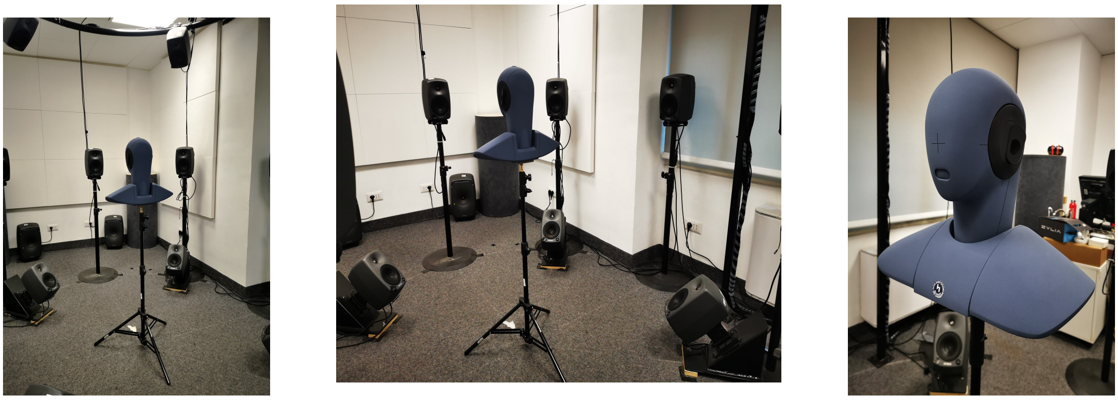


Figure 4.19: HATS Intra-ASL.

Results:

Polar diagrams for the sweet spot analyses, from the impulse responses (IR) generated from the recordings made with the Eigenmike, at all three conditions (0cm, 10cm, and 20cm) were made using Matlab scripts. Lateral polar plots were analyzed at frequencies of 1000 Hz and 2000 Hz, which are the most relevant bands for our validation. The sweet spot size analysis is crucial for our validation to understand how much a person's positioning and head displacement could influence sound perception. This is because individuals sitting at the same fixed point in the room cannot keep their heads oriented and positioned identically.

The results of lateral polar diagrams at 1000 Hz are presented for the horizontal plane with 0° elevation (figure 4.20), as well as for planes with elevations of 45° and -45° (figure 4.21). The diagrams show a noticeable degradation of the sound field as the Eigenmike is displaced, with consistent results as the frequency increases.

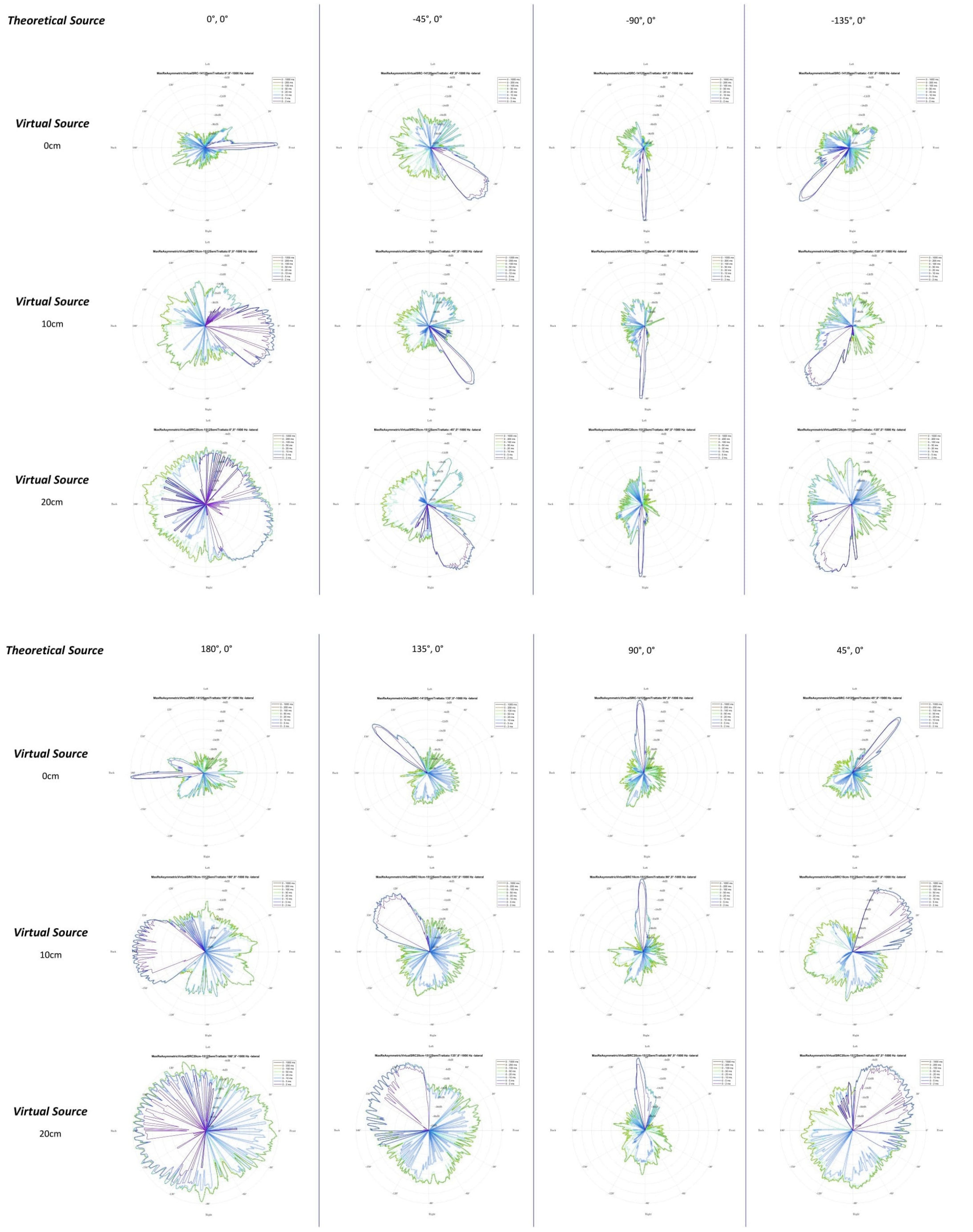


Figure 4.20: Lateral polar diagrams for measurements made with the Eigenmike at 0cm, 10cm and 20cm, at a frequency of 1000Hz, and a elevation of 0°.

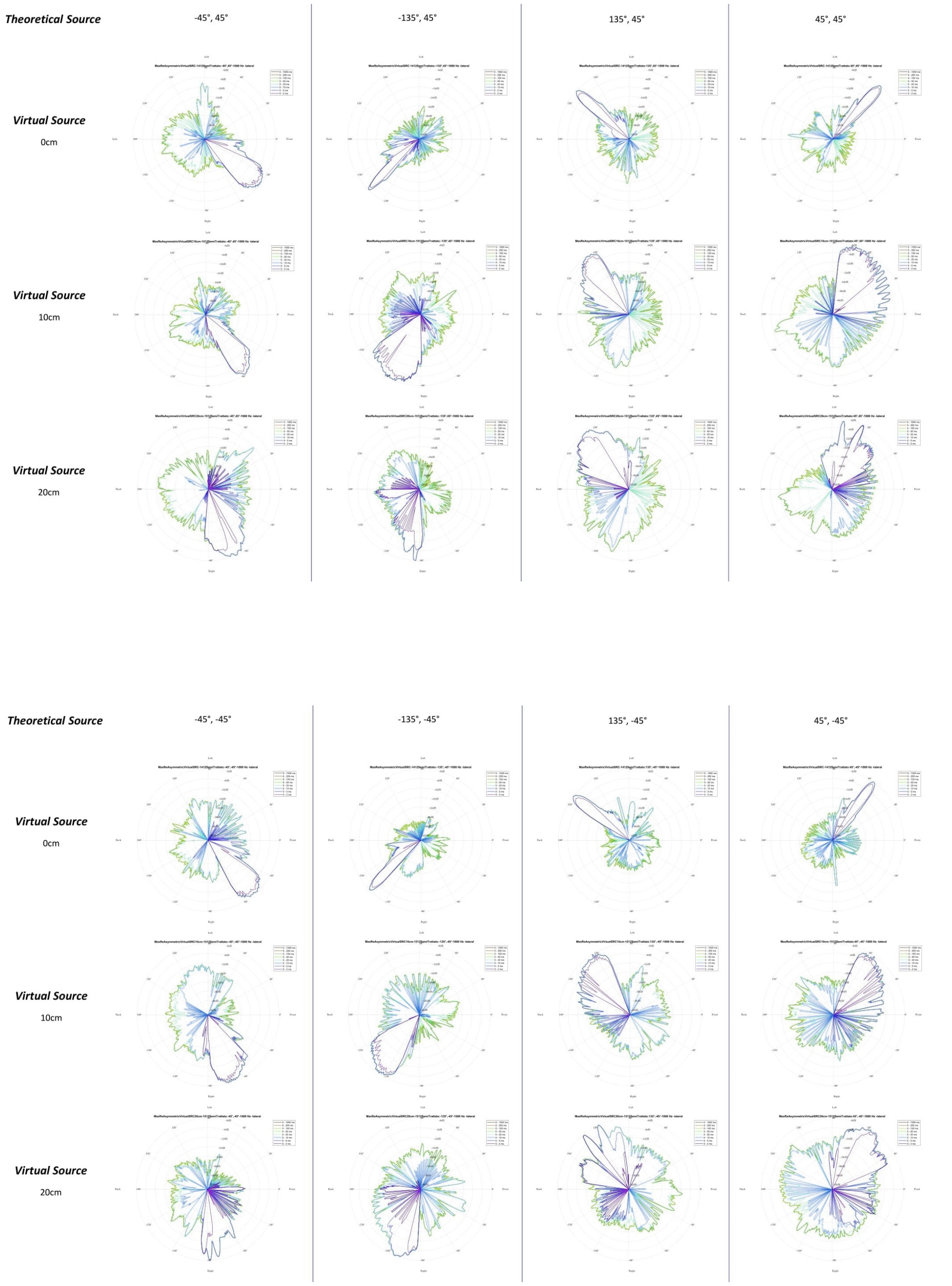


Figure 4.21: Lateral polar diagrams for measurements made with the Eigenmike at 0cm, 10cm and 20cm, at a frequency of 1000Hz, and a elevation of 45° and -45°.

In the horizontal plane with elevation 0°, good localization is maintained only at azimuth angles of -90° and 90° when displaced from the center of the loudspeaker array. In other cases, even at 10cm, the localization zone expands compared to 0cm, and at 20cm, localization becomes almost impossible with a highly expanded diagram that does not ensure accurate localization. At elevations of -45° and 45°, however, localization deteriorates regardless of the azimuth angle of the source. At both 10cm and 20cm displacements, source localization is challenging, with reflections spreading across the entire listening plane.

The same trends are observed in the lateral polar diagrams at 2000Hz, where acceptable sound source localization is maintained only at azimuth angles of -90° and 90° with elevation 0°. There is a slight improvement at elevation 0° for azimuth angles of -45° and -135°, but localization remains poor overall. At elevations of -45° and 45°, the diagrams show less jaggedness for azimuth angles of -45° and -135°, but there is still a noticeable shift in the localization zone compared to the 0cm position.

The results of the polar diagrams indicate that with a shift from the sweet spot, sound localization worsens, resulting in a degradation of the sound field. The sound field remains ideal only when there is no offset from the center of the loudspeaker array. This poses challenges for conducting tests with individuals instead of microphones or simulators placed at the center of the array. If a chair equipped with a headrest is not used, the likelihood of users maintaining a constant position without moving is very low.

Graphs of ITD, ILD, and IACC were also made using Matlab scripts from the Binaural Impulse Response (BIR) signals recorded with the HATS. The results are shown in Figures 4.22 and 4.23.

From the ITD graph (figure 4.22), the results vary depending on the azimuth angle considered. At 0° and 90° azimuth, the results are similar for all three sweet spot off-center positions. At -135° azimuth, only the 0cm (center of the array) position is closer to the real source, while at -90° it deviates more from the real case. At -45° azimuth, the 10cm and 20cm positions deviate from the real source, while the 0cm position shows good approximation. For 45° and 135° angles, the ITD at 10cm is better, while at 0cm and 20cm it gets worse, with similar results at 180°.

Overall, for ITD, good results are obtained with an off-center sweet spot up to 10cm. However, with a shift of 20cm, there is a significant deterioration in timing accuracy and a suboptimal approximation of the real case.

The ILD plot (figure 4.23) shows similar results to the ITD plot, with the results varying depending on the azimuth angle considered. However, unlike ITD, none of the sweet spot offsets correspond to the actual condition. From -135° to 0°, the 0cm position has the worst ILD, while the 10cm position is closest to the ideal real source condition.

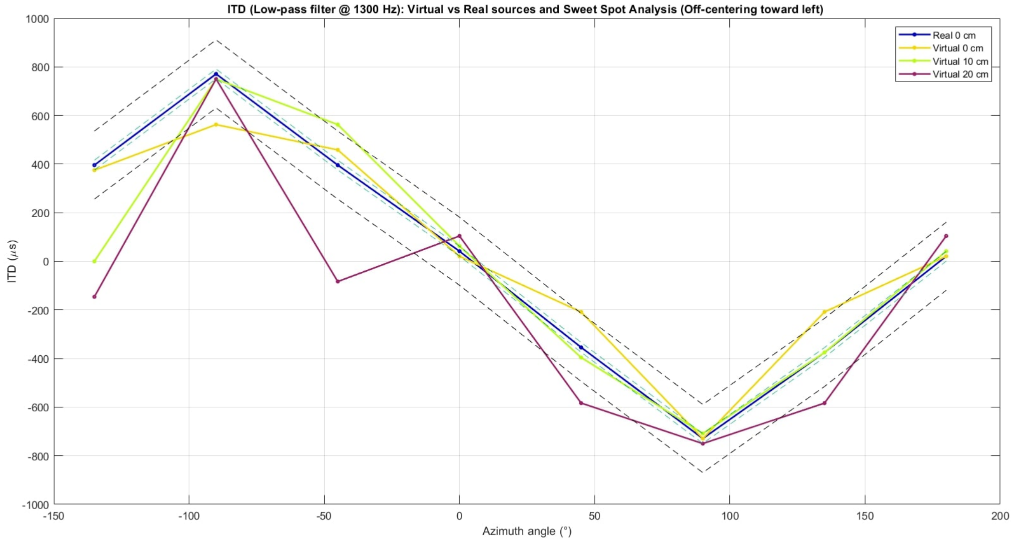


Figure 4.22: ITD comparison graphs for distance from the sweet spot of 0cm, 10cm, and 20cm.

From 90° to 180°, the sweet spot at 0cm in the center of the array performs the best, while at 20cm it only performs well at -45° and 0°.

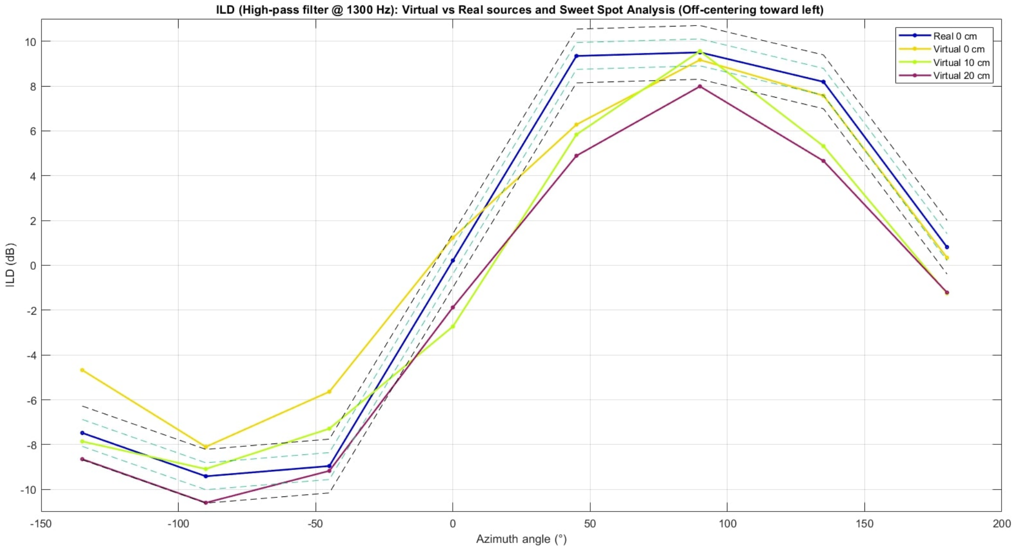


Figure 4.23: ILD comparison graphs for distance from the sweet spot of 0cm, 10cm, and 20cm.

Thus, in general, there is a good approximation up to 10cm, with the ILD being acceptably preserved and, in some cases, better than that observed at 0cm.

From all the analyses performed, the results consistently show a deterioration in all metrics as the sweet spot moves away from the center, with a degradation of the sound field. In the case of the sweet spot at 10 cm, the results do not deviate significantly from those obtained at 0 cm and remain acceptable. The ITD and ILD errors are close to the limits imposed by the JND, indicating that a shift of the sweet spot up to a maximum of 10 cm can be allowed.

4.2 Validation Inter-ASL

The Inter-ASL measurements were divided into two parts: the first part involved measurements conducted in a lecture hall at a polytechnic university, while the second part consisted of analyses conducted again within the Audio Space Lab (ASL). The measurements taken in the lecture hall were for recording ecological acoustic scenes and for ASL validation.

For validation purposes, recordings made in an outdoor environment needed to be played back later inside the ASL to assess if the laboratory could reproduce the recordings under the same acoustic conditions as those in the lecture hall where they were originally recorded.

This was done to verify that the playback system installed in the laboratory could accurately reproduce a real environment virtually, maintaining the same acoustic characteristics as those of the reference classroom. Finally, the same measures used previously for intra-ASL validation were taken and compared with each other.

4.2.1 Physical Validation

As mentioned earlier, the physical validation was divided into two parts: real and virtual, with the virtual reproduction using recordings from the real condition. The methodology applied for measuring both conditions is described below.

Methods:

Measurements for the real condition were conducted in Classroom 1T at the Polytechnic University of Turin (Italy), all on the same day. The measurements included validation of the Audio Space Lab and recording ecological acoustic scenes, utilizing impulse response measurements according to ISO 3382.

For the first set of measurements, three source positions (S1, S2, and S3) and one receiver position (R1) were selected. In the second set of measurements, additional positions were chosen as detailed below: R1 was situated in the first row of desks perpendicular to S1, positioned behind the desk.

S2 and S3 were located in the second and third rows, respectively, with S2 to the right of R1 and S3 perpendicular to R1 (see figure 4.24).

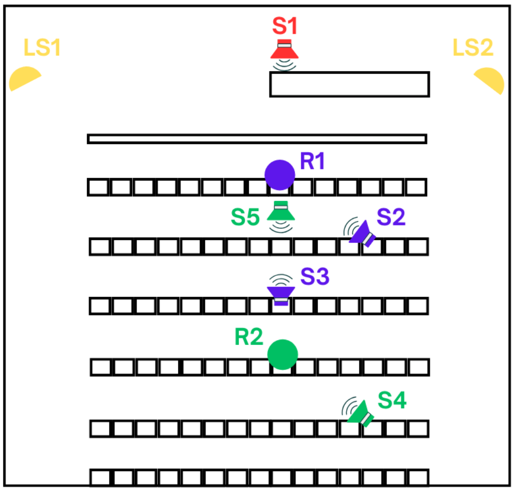


Figure 4.24: Representation of all positions for receivers (R1 and R2) and sources (S1, S2, S3, S4, and S5).

In the first set of measurements, an NTi Talkbox was used as the source, placed first in S1, then in S2, and finally in S3. The Head Acoustic (HATS) artificial simulator was used as the receiver, and was connected to a data logger for recording purposes, positioned at R1 (see figure 4.25).

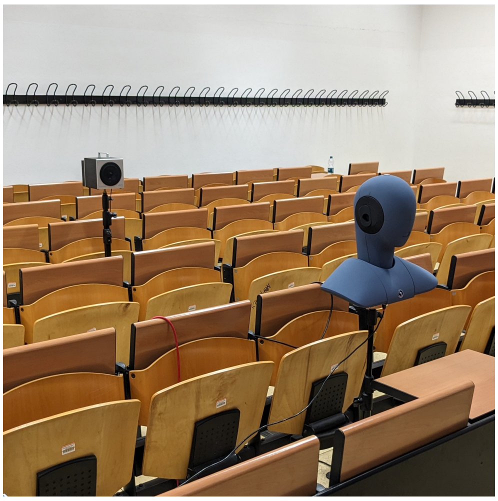


Figure 4.25: Representation of the first set of Inter-ASL measurements for collecting acoustic scenes. The HATS is positioned at R1, and the talkbox at S3.

A sine sweep was generated using Adobe Audition as the talkbox stimulus, with a sample rate of 44100 Hz, 16-bit mono, and a frequency range between 50 Hz and 16000 Hz. Each sweep lasted 5 seconds with a 5-second pause between repetitions, for a total of 20 iterations, of which 3 were always recorded. One recording was made for each talkbox position (S1, S2, and S3), for a total of 3 recordings with the HATS always placed at position R1.

In the second set of measurements for the collection of acoustic scenes, the same talkbox was used again in positions S1, S2, and S3, but with a different receiver. An em64 Eigenmike microphone array was used as the receiver, placed in position R1 (Figure 4.26), connected to an RME card, and finally to a PC. Background noise was recorded under this condition, with the projector and system on for 5 minutes. Four recordings were then made using the Eigenmike as the receiver.

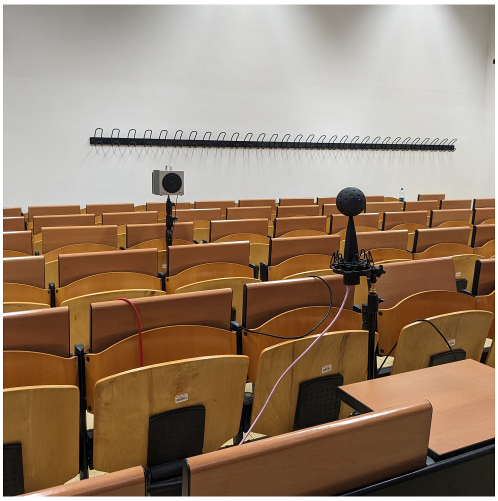


Figure 4.26: Representation of the first set of Inter-ASL measurements for collecting acoustic scenes. The Eigenmike is positioned at R1, and the talkbox at S3.

Next, a new position for the receiver (R2) was chosen, located toward the back of the classroom, in the third row from the back. The S1 position for the sources was retained, but two new positions, S4 and S5, were chosen instead of S2 and S3. S4 was placed on the same row as R2, to its right, while S5 was placed in the second row, perpendicular to R2 and S1. The same measurements were then made as with R1, S1, S2, and S3 but using R2, S1, S4, and S5. The same sine sweep used previously and loaded on the talkbox was used as the stimulus.

The first set of measurements was conducted using a HATS connected to a data logger, which acquired the recordings placed in R2 (figure 4.27), with the talkbox placed first in S1, then in S4, and finally in S5. A total of 3 recordings were made with the HATS always positioned in R2.

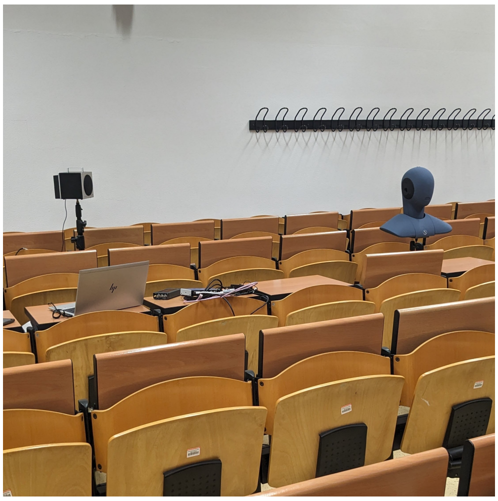


Figure 4.27: Representation of the second set of Inter-ASL measurements for collecting acoustic scenes. The HATS is positioned at R2, and the talkbox at S4.

In the second set of measurements for the collection of acoustic scenes, the same source was used again in the exact same positions (S1, S4, and S5) and conditions, using an em64 Eigenmike microphone array as the receiver, placed in position R2 (figure 4.28), connected to an RME card, and finally to a PC. Background noise was recorded in this condition, with the projector and system on for 5 minutes. Four recordings were then made using the Eigenmike as the receiver in R2.

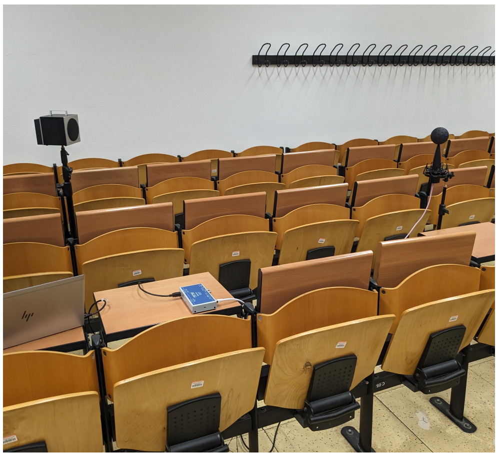


Figure 4.28: Representation of the second set of Inter-ASL measurements for collecting acoustic scenes. The Eigenmike is positioned at R2, and the talkbox at S4.

Once the recordings in the real environment were completed, further measurements were made inside the ASL using the recordings made in classroom 1T with the Eigenmike as stimulus. The em64 Eigenmike microphone array and Head Acoustic's Head Artificial Simulator (HATS) were used as measurement instruments.

For the measurements where recordings of ecological acoustic scenes were used as stimuli, all recordings made with the Eigenmike in classroom 1T were used, considering all source positions (S1, S2, S3, S4, and S5) and the two receiver positions (R1 and R2), resulting in a total of 6 recordings.

The recordings were made following the same procedure as the measurements made Intra-ASL, placing the measuring instrument in the center of the loudspeaker array, at a height of 122.5 cm from the ground. The stimuli were played back using Bidule ambisonic playback software, modifying the appropriate blocks and nodes to reproduce the recordings made with the Eigenmike in classroom 1T.

The 6 recordings of the acoustic ecological scenes were then made by both recording with the Eigenmike and recording with the HATS, for a total of 12 recordings.

Results:

The graphs of ITD, ILD and IACC, from the Binaural Impulse Response (BIR) signals recorded with the HATS, of both the 1T classroom and ASL, were made using Matlab scripts.

A comparison was conducted between the real condition, which refers to analyses performed with HATS recordings made directly in the 1T classroom, and the virtual condition, which pertains to analyses conducted with Eigenmike recordings in the 1T classroom and reproduced within the ASL by recording with the HATS.

This comparison aimed to determine whether the ASL could faithfully reproduce the acoustic environment of the 1T classroom, thereby achieving similar results between the two conditions.

The figure 4.29 shows the comparison of ITD between the real condition (classroom 1T) and the virtual condition (ASL simulating classroom 1T). It is evident from the graph that the virtual ITD closely approximates the real condition, as it is almost always within the JND.

There are only two source-receiver positions (S2-R1 and S4-R2) where the ITD slightly exceeds the JND limit. For all other positions, the reproduction is similar, if not identical, to that of the actual classroom condition.

The figure 4.30 shows the ILD comparison between the real condition (classroom 1T) and the virtual condition (ASL simulating classroom 1T). The ILD closely mirrors the behavior observed with ITD, where the virtual condition generally approximates the real condition well, remaining within the JND threshold.

However, similar to ITD, there are deviations exceeding the JND limit at two source-receiver positions (S2-R1 and S4-R2).

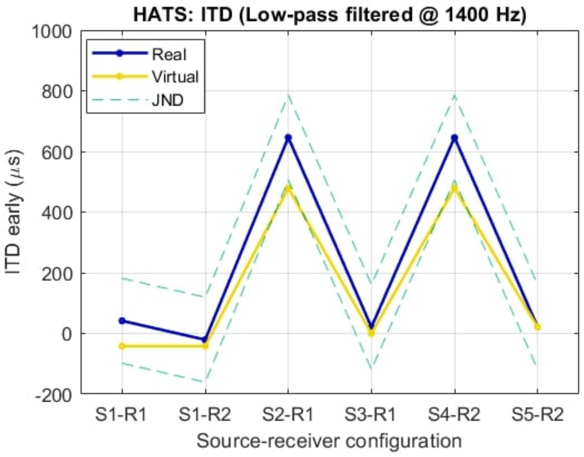


Figure 4.29: Inter-ASL ITD analysis.

A low difference in ITD and ILD between real and virtual cases indicates a good approximation of the sound source system, with accurate user perception.

Thus, both binaural ITD and ILD parameters generally show good reproduction of the real environment in the virtual setup, although with some exceptions. In two source-receiver positions, the limits imposed by the JND are exceeded for both parameters.

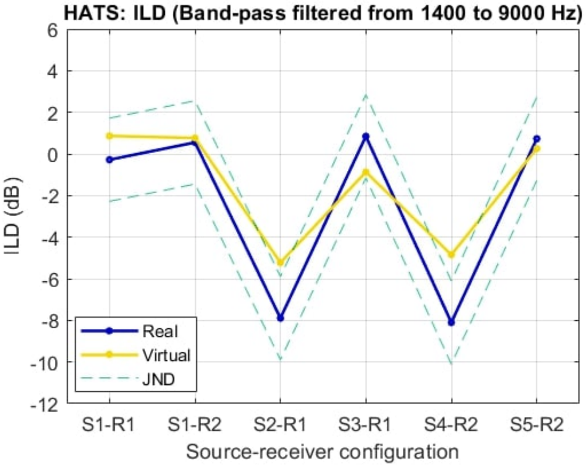


Figure 4.30: Inter-ASL ILD analysis.

Finally, the graph depicting IACC between the real condition (classroom 1T) and the virtual condition (ASL simulating classroom 1T) is shown in figure 4.31.

In contrast to ITD and ILD, the IACC parameter is not well approximated by the virtual condition.

For all source-receiver positions, the IACC value consistently exceeds the JND limit, except for position S4-R2 where it coincides with the JND limit.

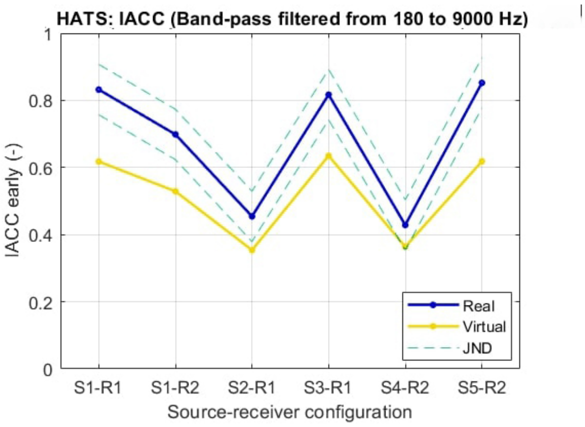


Figure 4.31: Inter-ASL IACC analysis.

Therefore, in terms of the IACC parameter, the virtual environment does not faithfully reproduce the real environment. However, there are values that are close to the optimal IACC range (0.5-0.4) and deviate from the value of 1, which indicates poor spatiality. This means that the system can accurately reproduce the position of sound sources, allowing the user to perceive sound coming from the correct location. However, it cannot recreate the same spatial sensation of sound that the user would experience in the real environment, such as perceiving a wider sound source.

This discrepancy alters the interaural cross-correlation present in the 1T classroom recordings, resulting in degraded parameter values. It remains to be determined through perceptual validation whether this discrepancy is perceptible to the human ear.

Chapter 5

Discussions and Conclusions

With this study, the physical validation of the Audio Space Lab at the Polytechnic University of Turin was conducted. Building upon the initial validation outlined in [35], additional analyses were performed to they're deeper of ecological validity of this immersive audio laboratory.

The analyses were divided into two parts: an intra-ASL section using stimuli generated within the laboratory, and an inter-ASL section involving measurements taken in a real-world environment, subsequently used as stimuli to recreate the same environment in virtual form within the laboratory.

For the intra-ASL phase, analyses were conducted on whether to use a chair with or without a headrest, and the impact of a head-mounted display (HMD) on the sound field was also tested. Polar plots, ITD, and ILD were calculated for all conditions. The analysis indicated that the condition without a headrest yielded slightly better results compared to the headrest condition, leading to the decision to use a chair without a headrest in future experiments. Regarding the HMD, no significant difference was observed between conditions, suggesting that the HMD does not introduce noticeable reflections that would affect the sound field.

The current soundproofing of the laboratory was evaluated by testing three acoustic conditions: the first involved applying acoustic absorbent panels, the second used an acoustic absorbent curtain, and the third added additional panels to critical corners. Polar diagrams, ITD, and ILD were analyzed, revealing significant improvement with the first condition of applying absorbent panels alone. The addition of an acoustic curtain or extra panels did not yield further benefits in terms of ITD, ILD, or sound localization.

Consequently, it was decided to proceed with validation using the first acoustic condition, which involved placing absorbent acoustic panels behind the speakers positioned at -45° azimuth (adjacent to the window corner) and at 90° (where unwanted reflections were noticeable).

After selecting the appropriate acoustic treatment, measurements were conducted to compare two decoding methods within the Bidule software: MaxRe and Basic decoding. Polar diagrams, ITD, and ILD were analyzed to assess the performance of each decoding method. The comparison did not reveal significant differences that would clearly justify choosing one over the other.

However, based on its widespread use and superior performance reported in the literature ([36]), MaxRe decoding was selected. Studies have demonstrated that MaxRe decoding exhibits better results at high frequencies and lower errors compared to Basic decoding.

Once the acoustic treatment and decoder to be used were chosen, comparative measurements were made to understand how the sound field degrades as the position changes from the sweet spot. Polar plots, ITD, and ILD were analyzed for microphone positions relative to the center of the loudspeaker array (0cm, 10cm, and 20cm). The analysis indicated that the polar plots worsened with displacement from the sweet spot, with sound localization deteriorating the farther the position was from the center of the array. However, positions within 10cm of the center of the array still provided acceptable results. Therefore, to maintain consistency in future tests and analyses, it is recommended that displacement from the sweet spot be limited to no more than 10cm from the center of the loudspeaker array.

The Inter-ASL validation consisted of two parts: recordings were initially made in a university classroom (classroom 1T at the Polytechnic University of Turin), followed by measurements inside the Audio Space Lab (ASL). The recordings from classroom 1T were utilized as stimuli for ASL validation and ecological acoustic scene collection. This approach differed from intra-ASL validation, where stimuli were generated within the lab itself.

The comparison between measurements taken in classroom 1T and those conducted in ASL (using classroom recordings as stimuli) aimed to verify whether the lab could faithfully reproduce the real environment virtually. ITD, ILD, and IACC were calculated, and the results showed that the virtual environment simulates the real environment closely, approaching the parameters calculated directly in the classroom for ITD and ILD, with only two source-receiver positions being at the limit imposed by the JND. In fact, having a low difference between ILD and ITD means that the system can more accurately reproduce the position of the sound sources, allowing the user to perceive the sound coming from the correct, objective position of the source.

For IACC, the virtual simulation does not reflect the real condition, with values consistently exceeding the limit imposed by the JND. However, these values were closer to 0 than in the real case, indicating that a good spatiality of the environment is still maintained. High values close to 1 refer to mono sources placed in front or behind the listener, indicating a lack of spatiality. In this case, the values ranged between 0.6 and 0.4, which is closer to the ideal IACC range of 0.4-0.5.

This means that the system can reproduce the position of the sound sources, but it does not ensure the same spatial feel of the sound that the user would experience in the real environment.

Therefore, the IACC values need further verification. Perceptual validation will help determine whether users can perceive this difference between the real and virtual cases. If they do, additional improvements will be made to the Audio Space Lab to ensure proper spatialization.

Overall, the physical acoustical validation produced satisfactory results. However, the discrepancies found during the Inter-ASL validation require further investigation through perceptual validation to determine if the differences between real and virtual cases are perceptible and impact speech intelligibility. Therefore, future work will involve conducting perceptual validation of the ASL by performing speech intelligibility tests with human subjects. This will ensure a comprehensive and verified validation of the laboratory, providing a cost-effective immersive audio lab model that can be replicated in clinics and hospitals for more efficient tuning of hearing devices.

Declaration of AI Tools Usage

During the preparation of this thesis, the artificial intelligence model developed by OpenAI ChatGPT was used for assistance in text form correction. This tool helped improve only the clarity and the grammar of the presented content. However, all ideas, interpretations, and conclusions expressed in this work are completely original.

Bibliography

[1] Stefan Kerber and Bernhard U. Seeber. «Sound localization in noise by normal-hearing listeners and cochlear implant users». In: 33.4 (July 2012), pp. 445–457. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3446659/pdf/ukmss-49881.pdf (cit. on p. 3).

[2] Eric C. Hamdan and Mark D. Fletcher. «A Compact Two-Loudspeaker Virtual Sound Reproduction System for Clinical Testing of Spatial Hearing With Hearing-Assistive Devices». In: *Frontiers in Neuroscience* 15 (2022). ISSN: 1662-453X. DOI: 10.3389/fnins.2021.725127. URL: https://www.frontiersin.org/articles/10.3389/fnins.2021.725127 (cit. on p. 4).

[3] Tina M Grieco-Calub and Ruth Y Litovsky. «Sound localization skills in children who use bilateral cochlear implants and in children with normal acoustic hearing». en. In: *Ear Hear* 31.5 (Oct. 2010), pp. 645–656. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2932831/pdf/nihms221472.pdf (cit. on p. 5).

[4] Ruth Y Litovsky, Aaron Parkinson, and Jennifer Arcaroli. «Spatial hearing and speech intelligibility in bilateral cochlear implant users». en. In: *Ear Hear* 30.4 (Aug. 2009), pp. 419–431. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2873678/pdf/nihms136730.pdf (cit. on p. 6).

[5] Alexandra Annemarie Ludwig, Sylvia Meuret, Rolf-Dieter Battmer, Michael Fuchs, Arneborg Ernst, and Marc Schönwiesner. «Auditory Spatial Discrimination and Sound Localization in Single-Sided Deaf Participants Provided with a Cochlear Implant». en. In: (Dec. 2023), pp. 1–14 (cit. on p. 7).

[6] Florian Pausch, Lukas Aspöck, Michael Vorländer, and Janina Fels. «An Extended Binaural Real-Time Auralization System With an Interface to Research Hearing Aids for Experiments on Subjects With Hearing Loss». In: *Trends in Hearing* 22 (Jan. 2018), p. 2331216518800871 (cit. on p. 7).

[7] Stephen Sechler, Alejandro Lopez Valdes, Saskia Waechter, Cristina Simoes-Franklin, Laura Viani, and Richard Reilly. «Virtual reality sound localization testing in cochlear implant users». In: (Aug. 2017). DOI: 10.1109/NER.2017.8008369 (cit. on p. 8).

[8] Andrea Gulli, Federico Fontana, Eva Orzan, Alessandro Aruffo, and Enrico Muzzi. «Spontaneous head movements support accurate horizontal auditory localization in a virtual visual environment». In: *PLOS ONE* 17.12 (Dec. 2022), pp. 1–17. DOI: 10.1371/journal.pone.0278705. URL: https://doi.org/10.1371/journal.pone.0278705 (cit. on p. 9).

[9] J. Cubick and Torsten Dau. «Validation of a Virtual Sound Environment System for Testing Hearing Aids». In: *Acta Acustica united with Acustica* 102 (May 2016), pp. 547–557. DOI: 10.3813/AAA.918972 (cit. on pp. 12, 13).

[10] Fargeot, Simon, Vidal, Adrien, Aramaki, Mitsuko, and Kronland-Martinet, Richard. «Perceptual evaluation of an ambisonic auralization system of measured 3D acoustics». In: *Acta Acust.* 7 (2023), p. 56. DOI: 10.1051/aacus/2023052. URL: https://doi.org/10.1051/aacus/2023052 (cit. on pp. 12–14).

[11] Wonyoung Yang and Murray Hodgson. «Validation of the Auralization Technique: Comparative Speech-Intelligibility Tests in Real and Virtual Classrooms». In: *Acta Acustica united with Acustica* 93 (Nov. 2007), pp. 991–999 (cit. on pp. 12, 15).

[12] Sylvain Emmanuel Favrot and Jörg Buchholz. «LoRA: A Loudspeaker-Based Room Auralization System». English. In: *Acta Acustica united with Acustica* 96.2 (2010), pp. 364–375. ISSN: 1610-1928. DOI: 10.3813/AAA.918285 (cit. on pp. 12, 13, 16).

[13] Murray Hodgson, Natalie York, Wonyoung Yang, and Mark Bliss. «Comparison of Predicted, Measured and Auralized Sound Fields with Respect to Speech Intelligibility in Classrooms Using CATT-Acoustic and ODEON». In: *Acta Acustica united with Acustica* 94 (Nov. 2008). DOI: 10.3813/AAA.918106 (cit. on pp. 12, 13, 17).

[14] Sylvain Emmanuel Favrot and Jörg Buchholz. «Validation of a loudspeaker-based room auralization system using speech intelligibility measures». English. In: Preprint 7763 (2009). 126th Audio Engineering Society Convention, AES126 ; Conference date: 07-05-2009 Through 10-05-2009, p. 7763. URL: http://www.aes.org/events/126/ (cit. on pp. 12, 17).

[15] Monika Rychtáriková, Tim van den Bogaert, Gerrit Vermeir, and Jan Wouters. «Perceptual validation of virtual room acoustics: Sound localisation and speech understanding». In: *Applied Acoustics* 72.4 (2011), pp. 196–204. ISSN: 0003-682X. DOI: https://doi.org/10.1016/j.apacoust.2010.11.012. URL: https://www.sciencedirect.com/science/article/pii/S0003682X10002677 (cit. on pp. 12, 18).

[16] Sylvain Favrot and Jörg M. Buchholz. «Distance perception in loudspeaker-based room auralization». English. In: 2 (2009). 127th Audio Engineering Society Convention - 2009 ; Conference date: 09-10-2009 Through 12-10-2009, pp. 859–866 (cit. on pp. 12, 20).

[17] Michael Schoeffler, Jan Gernert, Maximilian Neumayer, Susanne Westphal, and Juergen Herre. «On the validity of virtual reality-based auditory experiments: a case study about ratings of the overall listening experience». In: *Virtual Reality* 19 (Aug. 2015). DOI: 10.1007/s10055-015-0270-8 (cit. on pp. 12, 13, 21).

[18] G Parsehian, L Gandemer, C Bourdin, and Richard Kronland-Martinet. «Design and perceptual evaluation of a fully immersive three-dimensional sound spatialization system». In: (Sept. 2015). URL: https://hal.science/hal-01306631 (cit. on pp. 12, 22).

[19] Pavel Zahorik, Douglas Brungart, and Adelbert Bronkhorst. «Auditory distance perception in humans: A summary of past and present research». In: *Acta Acustica united with Acustica* 91 (May 2005), pp. 409–420 (cit. on pp. 12, 22, 23).

[20] Andrew J Kolarik, Brian C J Moore, Pavel Zahorik, Silvia Cirstea, and Shahina Pardhan. «Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss». en. In: 78.2 (Feb. 2016), pp. 373–395 (cit. on pp. 12, 23, 24).

[21] Monika Rychtarikova, Tim Bogaert, Gerrit Vermeir, and Jan Wouters. «Binaural Sound Source Localization in Real and Virtual Rooms». In: *J Audio Eng Soc* 4 (Jan. 2012) (cit. on pp. 12, 25).

[22] Antje Ihlefeld and Barbara Shinn-Cunningham. «Effect of source spectrum on sound localization in an everyday reverberant room». In: *The Journal of the Acoustical Society of America* 130 (July 2011), pp. 324–33. DOI: 10.1121/1.3596476 (cit. on pp. 12, 26).

[23] Stéphanie Bertet, Jérôme Daniel, Etienne Parizet, and O. Warusfel. «Investigation on Localisation Accuracy for First and Higher Order Ambisonics Reproduced Sound Sources». In: *Acta Acustica united with Acustica* 99 (2013), pp. 642–657. URL: https://hal.science/hal-00848764 (cit. on pp. 12, 27).

[24] Thirsa Huisman, Axel Ahrens, and Ewen MacDonald. «Sound source localization in virtual reality with ambisonics sound reproduction». In: (Feb. 2021). DOI: 10.31234/osf.io/5sef6 (cit. on pp. 12, 27).

[25] Axel Ahrens, Kasper Duemose Lund, Marton Marschall, and Torsten Dau. «Sound source localization with varying amount of visual information in virtual reality». In: 14.3 (2019), e0214603 (cit. on pp. 12, 28).

[26] Juan C. Gil-Carvajal, Jens Cubick, Sébastien Santurette, and Torsten Dau. «Spatial Hearing with Incongruent Visual or Auditory Room Cues». In: *Scientific Reports* 6.1 (Nov. 2016), p. 37342. ISSN: 2045-2322. DOI: 10.1038/srep37342. URL: https://doi.org/10.1038/srep37342 (cit. on pp. 12, 29).

[27] Jörg M. Buchholz and Virginia Best. «Speech detection and localization in a reverberant multitalker environment by normal-hearing and hearing-impaired listeners». In: *The Journal of the Acoustical Society of America* 147.3 (Mar. 2020), pp. 1469–1477. ISSN: 0001-4966. DOI: 10.1121/10.0000844. eprint: https://pubs.aip.org/asa/jasa/article-pdf/147/3/1469/15339667/1469_1_online.pdf. URL: https://doi.org/10.1121/10.0000844 (cit. on pp. 12, 30).

[28] Victor Benichoux, Marc Rébillat, and Romain Brette. «On the variation of interaural time differences with frequency». In: *The Journal of the Acoustical Society of America* 139.4 (2016), pp. 1810–1821 (cit. on p. 31).

[29] George F. Kuhn. «Model for the interaural time differences in the azimuthal plane». In: *The Journal of the Acoustical Society of America* 62.1 (July 1977), pp. 157–167. ISSN: 0001-4966. DOI: 10.1121/1.381498. eprint: https://pubs.aip.org/asa/jasa/article-pdf/62/1/157/11465711/157_1_online.pdf. URL: https://doi.org/10.1121/1.381498 (cit. on p. 31).

[30] Tingli Cai, Brad Rakerd, and William M Hartmann. «Computing interaural differences through finite element modeling of idealized human heads». en. In: *J Acoust Soc Am* 138.3 (Sept. 2015), pp. 1549–1560 (cit. on p. 31).

[31] Parvaneh Parhizkari. «Binaural HearingHuman Ability of Sound Source Localization». In: *Blekinge Institute of Technology* (2008). URL: https://www.diva-portal.org/smash/get/diva2:830971/FULLTEXT01.pdf (cit. on p. 33).

[32] Goldstein. *Localizzazione e Organizzazione Acustica.* URL: https://moodle2.units.it/pluginfile.php/379238/mod_resource/content/0/Cap_12_Ita_Localizzazione%20e%20Organizzazione%20Acustica_Goldstein.pdf (cit. on p. 33).

[33] Nti-Audio. *Reverberation Time.* URL: https://www.nti-audio.com/en/applications/room-building-acoustics/reverberation-time (cit. on p. 34).

[34] Ecophon. *Chiarezza del discorso.* URL: https://www.ecophon.com/it/about-ecophon/acoustic-knowledge/room-acoustic-descriptors/speech-clarity (cit. on p. 35).

[35] A. Guastamacchia et al. «Set up and preliminary validation of a small spatial sound reproduction system for clinical purposes». In: (Jan. 2022), pp. 4991–4998. DOI: 10.61782/fa.2023.0698 (cit. on pp. 38, 40, 75).

[36] Diego Murillo Gomez, Filippo Fazi, and Mincheol Shin. «Evaluation of ambisonics decoding methods with experimental measurements». In: (Apr. 2014). DOI: 10.14279/depositonce-6 (cit. on pp. 62, 76).