Leon Merkel, Bsc BA

Perceptual evaluation of professional line source and point source loudspeakers for immersive sound reinforcement

MASTER'S THESIS
to achieve the university degree of
Diplom-Ingenieur
Master's degree program: Electrical and Audio Engineering

submitted to
University of Music and Performing Arts Graz

Supervisor

Prof. Dr. Franz Zotter
Institute for Electronic Music and Acoustics (IEM)

Graz, February 25, 2025

AFFIDAVIT

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to KUGonline is identical to the present master's thesis.

Date, Signature

Acknowledgment

I want to express my honest appreciation to all those who have contributed to the success of this project. In particular, I am deeply thankful to my supervisor, Franz Zotter, whose support and guidance were instrumental—from refining the research question to the seamless implementation of the experiment and structure and writing of this thesis. His comprehensive help at any time has been substantial for this work.

I also wish to thank my co-supervisors, Phil Coleman and Frederik Roskam from L'Acoustics in London, for their expert assistance in designing the experiment and engaging in scientific discussions that considerably enriched the project.

I sincerely appreciate Sherif el Barbari, who enabled me to access the lab space at L'Acoustics UK and provided the required research resources. In addition, I am grateful to Hagai Pippich for his practical support in setting up the loudspeakers and amplifiers and transporting equipment to the lab.

Special thanks are due to Rapolas Daugintis from the Audio Experience Design group at Imperial College London, who borrowed me the Kemar dummy and Eigenmic Ambisonics microphone for my measurements and helped me recruit participants for the listening experiment.

I also appreciate the contributions of Lukas Gölles and Stefan Riedel. They shared their previous research and engaged in in-depth discussions, which helped me create a meaningful experimental design and supported me in evaluating it.

I am grateful to Thomas Röck for sharing his expertise with the Typst writing environment and always being ready to assist.

Finally, I want to thank my parents, Karin and Christoph, and my sister, Insa, for their support and trust throughout my university career.

Abstract

Immersive playback systems aim to create a balanced perception of sounds from different directions and establishing an impression of envelopment over an extended audience area. Current perceptual and simulation-based research showed that coverage designs of constant decay (0 dB per distance doubling) preserve the object level balance. In contrast, a decay of −3 dB generates a uniform sensation of envelopment for off-center listening positions. However, point-source loudspeakers remain widely used for immersive sound reinforcement systems in mid-sized venues. Coverage of any point-source loudspeaker inherently decays by −6 dB per distance doubling, so using them can conflict with the design goals outlined above. This thesis investigates perceived differences of eight surrounding line source loudspeakers in direct comparison with eight surrounding point source loudspeakers in a suitable closely spaced setup. The perceptual qualities, object level balance, spatial definition, and envelopment were compared in a MUSHRA listening experiment. Acoustic measurements of the experimental setup were carried out to gather omnidirectional and binaural room impulse responses (BRIRs). The BRIRs were used to check whether the ratings of the listening experiment could be reproduced on headphones. The ratings obtained from the loudspeaker and headphone-based experiments are highly correlated, confirming the transferability. Based on the acoustic measurements, linear regression models are devised and show high correlations of instrumented metrics with the perceptual results. The results confirm that flown line sources, exhibiting a decay of −2 dB per distance doubling, help preserve object-level balance, increase spatial definition, and provide a uniform envelopment experience throughout the audience area when compared to point source loudspeakers.

Kurzfassung

Immersive Beschallungssysteme streben eine ausgewogene Wahrnehmung von Direktschallobjekten aus verschiedenen Richtungen und den Eindruck einer gleichmäßig umgebenden Einhüllung durch Diffusschall für den gesamten Zuhörerbereich an. Aktuelle Untersuchungen mittels Hörversuchen und Simulationen haben gezeigt, dass eine konstante Direktschallpegelverteilung ohne entfernungsbedingten Pegelabfall (0 dB pro Entfernungsverdopplung) die Pegelverhältnisse von Schallobjekten aus verschiedenen Richtungen beibehält. Im Gegensatz dazu erzeugt ein entfernungsbedingter Pegelabfall von −3 dB pro Entfernungsverdopplung einen Eindruck gleichmäßiger Einhüllung an dezentralen Hörpositionen. Dennoch werden Punktschallquellen nach wie vor häufig für immersive Beschallungssysteme in mittelgroßen Lautsprechersituationen verwendet. Diese Lautsprecher weisen allerdings einen Direktschall-Pegelabfall von −6 dB pro Entfernungsverdopplung auf, wodurch sie den oben genannten optimalen Eigenschaften für immersive Beschallungsszenarien nur bedingt genügen. Diese Arbeit untersucht wahrnehmbare Unterschiede zwischen acht Linienschallquellen in einer Surroundkonfiguration und acht Punktschallquellen in derselben Anordnung im direkten Vergleich. In einem MUSHRA-Hörversuch wurden beiden Lautsprechertypen in der immersiven Beschallungsaufstellung hinsichtlich der Wahrnehmungskriterien, Ausgewogenheit der richtungsabhängigen Objektlautstärken, räumliche Definition und gleichmäßiger Einhüllung, verglichen. Akustische Messungen des Versuchsaufbaus wurden durchgeführt, um omnidirektionale und binaurale Raumimpulsantworten (BRIRs) zu erhalten. Die BRIRs wurden verwendet, um zu überprüfen, ob die Bewertungen des Hörversuchs mit Lautsprechern im Raum mithilfe eines zweiten Hörversuchs über Kopfhörer reproduziert werden können. Die Experimente mit Lautsprechern und Kopfhörern zeigen eine starke Korrelation, was die Übertragbarkeit bestätigt. Auf Grundlage der akustischen Messungen werden geeignete akustische Messgrößen eingeführt und mittels linearer Regression an die Bewertungen der Hörversuche angepasst, wobei deutliche Korrelationen zu den wahrgenommenen Bewertungen sichtbar werden. Die Ergebnisse bestätigen, dass Linienschallquellen, die einen Pegelabfall von −2 dB pro Entfernungsverdopplung aufweisen, im Vergleich zu Punktquellen-Lautsprechern dazu beitragen, die Pegelverhältnisse richtungsabhängiger Direktschallobjekte zu erhalten, die räumliche Definition zu erhöhen und einen Eindruck gleichmäßiger Einhüllung im gesamten Zuschauerbereich zu erzeugen.

Contents

1. Introduction

1.1. Motivation

With growing audiences in the live event industry, acoustical amplification of instruments and human voices with loudspeakers became essential [1, chp. 10]. Heil et al. invented the line source principle [2], [3] for loudspeaker systems, which enable high-quality sound reinforcement for large and distant audiences.

Before that, the limited coverage capabilities of single loudspeaker systems inevitably lead to the need for distributed multi-mono systems [4, chp. 7] with careful time alignment in overlapping transition areas [5, chp. 8]. True stereo playback for all listeners was often practically impossible with left and right speaker systems. Both speaker systems should cover the whole audience, which implies a large transition area of cross-firing loudspeakers, resulting in intense interference patterns with significant phase cancellations for correlated signals [6, chp. 3]. As frequency homogeneity throughout the audience is one of the main goals of sound reinforcement, most large-scale sound systems are traditionally dual-mono [7, chp. 12], implying no panning algorithms and phantom sources.

In the late 2010s, several loudspeaker manufacturers released immersive mixing technologies, such as L-ISA[1] (L'Acoustics, 2016)[2], Soundscape[3] (d&b, 2018)[4] and Spacemap Go[5] (Meyer Sound, 2020)[6], Flechter-Machine (Adamson, 2022)[7] with dedicated signal processor units to distribute sound objects on loudspeakers and add artificial reverberation for augmented immersion. The underlying object-based panning algorithms are based on either vector-base or multi-direction amplitude panning (VBAP/MDAP [8]). Differences lie in implementation defaults, for example, L-ISA's low-frequency build-up compensation and source-width manipulation using decorrelation methods [9, chp. 2]. All algorithms require similar loudspeaker deployments of at least five full-range loudspeakers equally distributed above the stage to widen the sweet area of accurate frontal localization and improve audio-visual consistency [9, chp. 2]. Additional side and rear loudspeakers equally spaced around the audience are recommended to be fed with artificial reflections and reverberation to create a realistic space enveloping the listener [10]. Concerning system design goals, all frontal speakers must provide coverage, defined in [11, p. 2] as "direct sound in an acceptable frequency response variation" over the whole spatialized area to facilitate object-based panning.

In [12], Zotter et al. found a maximum acceptable mixing imbalance of ± 3 dB from the intended instrument balance for sound objects from different directions. More accurately, a constant A-weighted [13] direct sound pressure within the audience should be aimed to preserve the object level balance when listening to a spatial mix [14]. Gölles et al. [15] confirmed this design goal of 0 dB per distance doubling by evaluating simulations and conducting perceptual listening experiments with prototype miniature variable curvature line arrays [16]. For uncorrelated signals such as additional reverberation in the surrounding loudspeakers, Riedel et al. [12], [17], [18] found an aiming decay goal of -3 dB per distance doubling based on simulations of surround loudspeaker deployments with different angular resolutions and direct level distributions to

[1] https://l-isa.l-acoustics.com

[2] https://l-isa.l-acoustics.com/innovations-l-acoustics-l-isa-studio-immersive-audio-software/

[3] https://www.dbsoundscape.com/global/en/

[4] https://www.dbaudio.com/global/en/about-db/press/newsroom/15022018-introducing-the-db-soundscape-the-evolution-of-the-listening-experience/

[5] https://meyersound.com/product/spacemap-go/

[6] https://meyersound.com/news/spacemap-go

[7] https://adamson-fletcher-machine.com

create the most uniform sensation of envelopment throughout the audience, which Gölles et al. perceptually confirmed with prototype miniature line source arrays [14].

To reach both design goals for immersive sound reinforcement systems, a strict dual-target design, accomplishing simultaneous playback of two, a -3 dB/dod and 0 dB/dod, signal buses, as presented in [19], is not purposeful today. It requires a high channel count for individual array-element processing, amplification and multiple mix bus structures for direct and diffuse sounds. To match both requirements, a compromise decay of -1 to -2 dB per distance doubling is suggested in [20].

Previous studies compared prototype miniature line sources adjusted to distinct decay designs for the given listening environment. By contrast, this work targets a comparison of real professional coaxial point sources and compact fixed curvature line sources to verify the hypothesis that line sources provide superior success over true point sources in spatial consistency and envelopment for immersive sound reinforcement. Perceptual comparison investigates an immersive sound system consisting of eight compact line sources directly next to eight point source loudspeakers. In particular, L'Acoustics Syva[8] compact fixed-curvature line sources fulfill the suggested direct sound decay goal by achieving -2 dB per distance doubling at a suspension height of 1.28 m, while X8[9] loudspeakers from the same company are used as point sources.

Two listening experiments evaluate the spatial qualities, adapted from the Spatial Audio Quality Inventory [21], which are most affected by the loudspeaker type. The first listening experiment utilized loudspeakers in a laboratory of L'Acoustics in London. The experiment's transferability to headphones is investigated by measuring binaural room impulse responses in the laboratory and repeating it the experiment with headphones in Graz. Additionally, standard acoustics metrics to process binaural room impulse responses and measurements across the audience area are adopted to find objective relations to the subjective ratings from the experiments.

1.2. Structure of the thesis

Section 2 describes the fundamentals, particularly different loudspeaker types in Section 2.2, array system designs in Section 2.3, immersive sound reinforcement principles in Section 2.4, and relevant acoustic metrics are introduced in Section 2.5. In Section 3.1, the design of stimuli, trials, rating criteria, and repetitions for the listening experiment is demonstrated. Section 3.2 depicts the setup and configuration of the onsite and headphone experiment.

The Section 4 are individually presents the results from both experiments, in Section 4.1 and Section 4.2. and their correlation is analyzed in Section 4.3. Further, in Section 4.4, the relationship between subjective ratings and introduced acoustic metrics is investigated.

Section 5 evaluates the results and discusses them. In Section 6, the investigation is concluded in its scientific context, including suggestions for future research related to the thesis outcomes.

[8]https://www.l-acoustics.com/products/syva/
[9]https://www.l-acoustics.com/products/x8/

2. Fundamentals

This section examines the fundamentals and immersive sound reinforcement principles, including loudspeaker types, system design goals, and objectively assessable metrics.

2.1. Loudspeaker directivity

The radiation behaviour of a loudspeaker (array) is one of the most basic characteristics in system design and deployment decisions for the demanded purposes, as it allows accurate predictions of direct sound pressure coverage. Designing the directional characteristics permits the control where sound from a loudspeaker radiates to. Measured directivity data facilitates the calculations of meaningful metrics such as the directivity index (DI).

Directivity data, stored with additional metadata in the SOFA-format[10], contains directional complex-valued transfer functions $H_{\varphi,\vartheta}[\omega]$ or real-valued impulse responses $h_{\varphi,\vartheta}[n]$, characterizing the radiated sound from the source, where φ is the azimuth angle in the range of [−180°, +180°], ϑ the elevation angle in range of [−90°, +90°], $\omega = 2\pi f$ the angular frequency with f as the frequency in $\frac{1}{s}$ (Hz) and n the time index in samples. This data is commonly visualized as single-frequency (-band) two-dimensional (2D) polar plots, which are cross-sections of the three-dimensional (3D) balloon plots, or as isobaric contour plots, showing the frequency-dependent intensity over the horizontal or vertical radiation angle.

The frequency-dependent directivity factor Q(ω) estimates the single-direction sound energy concentration emitted from a loudspeaker and is defined as the ratio of the spectral energy density (squared magnitude) in the preferred direction $[\varphi_0, \vartheta_0]$ and the mean energy spectral density of all directions as

$$Q(\omega) = \frac{\left|H_{\varphi_0,\vartheta_0}[\omega]\right|^2}{\frac{\sum_{i=0}^{N_\varphi-1}\sum_{j=0}^{N_\vartheta-1}\left|\sin\vartheta_j\right|\left|H_{\varphi_i,\vartheta_j}[\omega]\right|^2}{\sum_{i=0}^{N_\varphi-1}\sum_{j=0}^{N_\vartheta-1}\left|\sin\vartheta_j\right|}} \ , \tag{2.1}$$

where $\left|\sin\vartheta_j\right|$ are the surface weights, N_φ and N_ϑ the number of sampled azimuth and elevation angles. The logarithmic representation of $Q(\omega)$ is the directivity index DI(ω) in decibels (dB) reading as

$$\mathrm{DI}(\omega) = 10\log_{10}(Q(\omega)) \ . \tag{2.2}$$

The higher those parameters are for a loudspeaker, the more capable it is of focusing the sound energy into the desired direction. Controlled directivity beneficially increases sound energy in intended directions (e.g., the audience area) while decreasing sound energy in unwanted directions (e.g., reflective boundary surfaces).

2.2. Loudspeaker types and principles

Point and line sources are the two most commonly utilized source types for sound reinforcement. This section briefly examines the principles of both loudspeaker types, focusing on their coverage and direct sound decay over distance.

[10] https://www.sofaconventions.org/mediawiki/index.php/SOFA_(Spatially_Oriented_Format_for_Acoustics)

Point source

A frequency-independent DI of 0dB ($Q = 1$) indicates an omnidirectional radiation behaviour with no explicit beam direction—an ideal point source. The sound pressure decay over the distance of the source can be derived by evaluating the acoustic intensity I (power per surface area). The emitted acoustic power by any source remains constant on the expanding surface. The surface area ($A_{\text{sphere}} = 4\pi r^2$) of a spherical source with radius r increases proportional to the squared distance, such as the intensity $I \propto \frac{1}{r^2}$. Knowing the intensity is proportional to the squared pressure ($I \propto p^2$), the sound pressure loss is directly proportional to the distance r as $p \propto \frac{1}{r}$ [22, chp. 5]. Therefore, sound pressure decay per doubling of the distance (dod) of an ideal infinitely small point source from any distance r to twice the distance $2r$ yields

$$20\log_{10}\left(\frac{p_{2r}}{p_r}\right) = 20\log_{10}\left(\frac{\frac{1}{2r}}{\frac{1}{r}}\right) = 20\log_{10}\left(\frac{1}{2}\right) = -6\,\text{dB} \;, \tag{2.3}$$

per doubling of the distance (dod). This corresponds to the spherical wave propagation in the far field, which is also called **Fraunhofer region**.

According to piston radiator theory [22, chp. 7], a loudspeaker diaphragm can be modelled as a piston radiator whose surface is effectively partitioned into many small oscillating elements. At lower frequencies—where the wavelength is long compared to the diaphragm dimensions—these elements combine to produce an almost omnidirectional sound field. However, as the frequency increases and the wavelength decreases, the interference among these individual elements becomes more pronounced. This interference results in a more focused main lobe in the radiation pattern, with regions of destructive interference appearing outside the main beam. Consequently, the directivity increases with frequency, leading to a narrower beam of sound radiation.

It has been established that conically horn-loaded point sources can be used to control and increase directivity, matching the actual demands. Those geometric shapes of horns have already been extensively investigated [23], [24], [25]. Further, so-called waveguides have been invented to carefully time-align sound paths from a single exciting source to enlarged outlet regions to control the broad-band directivity precisely. By appropriately arraying multiple loudspeakers, their acoustical interaction can also be beneficial in creating a particular radiation pattern.

Line source

One of the most common array types used to control vertical directivity is the vertical alignment of loudspeakers to a line source array. The ideal continuous line source radiates sound cylindrical in the near field, such that the expanding surface area is defined with $A_{\text{cylinder}} = 2\pi r S$, where S is the fixed height of the source. The expanding area and so the intensity is proportional to r ($I \propto \frac{1}{r}$), and the sound pressure $p \propto \frac{1}{\sqrt{r}}$ results in a reduced decay for a distance increase from any r to $2r$ of

$$20\log_{10}\left(\frac{p_{2r}}{p_r}\right) = 20\log_{10}\left(\frac{\frac{1}{\sqrt{2r}}}{\frac{1}{\sqrt{r}}}\right) = 20\log_{10}\left(\frac{1}{\sqrt{2}}\right) = -3\,\text{dB} \;, \tag{2.4}$$

per dod. The decreased decay behaviour indicates the presence of a complex acoustic near-field or **Fresnel region** with extensive interference of sound contributions from different sound emission positions on the extended source.

The border radius r_b describes the distance at which the Fresnel zone transits into the Fraunhofer zone, implying a decay of -6 dB/dod, illustrated in Figure 2.1. The border distance r_b can be expressed as a function of line length S and frequency f. It varies depending on whether the observation point is located at the middle height of the array ($\tilde{r}_b$), cf. [26] (as shown in

Figure 2.2) or at its end ($\hat{r}_b$), [27] (indicating a flown array, where the audience is underneath the line source). The two definitions differ by a factor of 4 and read as

$$\tilde{r}_b = \frac{fS^2}{2c} \ , \qquad \text{or} \qquad \hat{r}_b = \frac{2fS^2}{c} \ , \tag{2.5}$$

where c is the speed of sound in $\frac{\text{m}}{\text{s}}$.

Examining real line sources with finite length, the Fresnel Zone analysis illustrates the principle of operation [3].

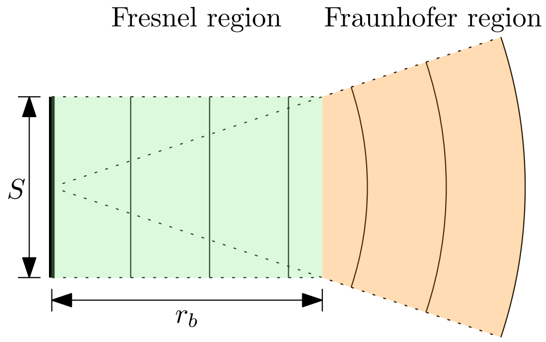


Figure 2.1: Sketch of the transition of the Fresnel region (green) to the Fraunhofer region (orange) of an ideal straight line source of length S indicating the border distance with r_b (adapted from Fig. 1 from [3]).

Consider a continuous straight line source of length S and an observer at a perpendicular distance r at its middle, as sketched in Figure 2.2. From the first zone, incident sound arrives within a quarter wavelength $(r + \frac{\lambda}{4})$ path difference, and only constructive interference is observed, while sounds arriving between $r + \frac{\lambda}{4}$ and $r + \frac{3\lambda}{4}$ additional path lengths, interfere destructively with the first zone. The third zone again adds constructively, as it arrives within $r + \frac{3\lambda}{4}$ and $r + \frac{5\lambda}{4}$. This pattern evolves in the same manner until the total line length S expires.

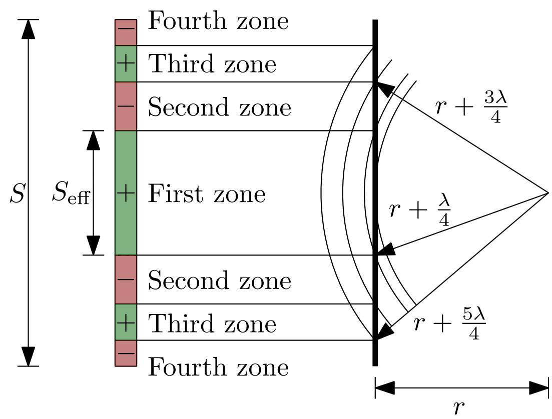


Figure 2.2: Sketch of a straight line source of length S in relation to an observer in perpendicular distance r and the corresponding Fresnel zones and effective line length S_{eff} (adapted from Fig. 3 from [3]).

From the second zone on, the alternating constructive and subtractive interferences of all zones respectively cancel, leaving the first zone as the only contributing part of sound energy to the listener. This defines the effective line length S_{eff}. For a given line length S, distance r, and frequency f, S_{eff} can be approximated by evaluating the curvature of the phase (second derivative) at the point of stationary phase [27]. As long as $S_{\text{eff}} < S$, the size of the Fresnel region grows with distance, and a part of the distance-dependent level decay gets compensated by the increase of S_{eff}. When S_{eff} reaches S, the Fresnel region transits smoothly into the Fraunhofer region. Those region allocations are highly frequency and line-length-dependent, reflected in a noticeable gradient in sound focusing capabilities from low to high frequencies.

Progressively curved line sources

In order to control the SPL coverage of line sources, the most rational approach is to adjust the binary Fresnel Zone for different frequencies at the covered audience area, which results in a progressive curving of flown line sources [28]. When arraying discrete loudspeaker elements to a line source with discontinuities, the wavefront sculpture technology (WST) by Urban et al. [3] defines requiring criteria. To create a wavefront that is as free from destructive interference as possible

a) a maximum center-to-center chassis spacing of $d_c \leq \frac{\lambda}{2}$ for the highest operating frequency to prevent spatial aliasing,
b) a maximum deviation of radiation curvature $u \leq \frac{\lambda}{4}$ from radiating a flat wavefront of the individual loudspeakers,
c) a constant product of each loudspeaker inclination angle α_i and aiming distance $r_{a,i}$ throughout the array, and
d) a maximum individual enclosure inclination angle of $\alpha_i \leq \frac{3°}{d_{c,i}}$,

should be adhered to.

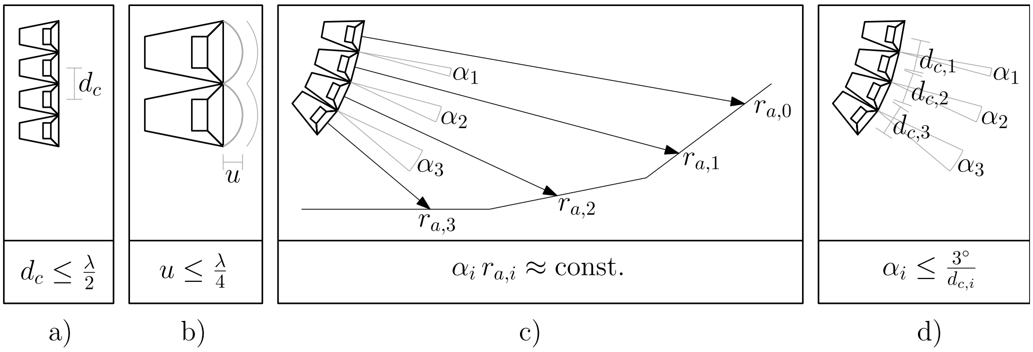


Figure 2.3: Illustration of the WST criteria from [3]. It shows the maximum center-to-center distance of each chassis in a), the maximum radiation curvature deviation from a flat wavefront in b), the constant product of aiming distance and inclination angle in c), and the maximum individual inclination angle in d).

These core constraints in fundamental line source theory are illustrated in Figure 2.3 and help to ensure that high frequencies are not compromised by spatial aliasing, requiring sufficiently small spacing of adjacent loudspeaker elements. At the same time, the waveguide outlets must be designed according to the necessary tilt angles so that deviations from the curved wavefront remain within a quarter of the wavelength, resulting in a smoothly curved wavefront. By maintaining a constant product of aiming distance and inclination and not exceeding the maximum tilt angle per enclosure, the array progressively curves along its length, which improves coherence and minimizes destructive interference.

Based on direct-sound-simulations, iterative algorithms are commonly provided as software by loudspeaker companies to determine the individual splay angles for a line array deployment. These algorithms consider a simplified audience area geometry and corresponding directivity data of used loudspeakers. Gölles and Zotter [27] presented a solution to find a continuous line contour of ideal point sources providing a selected level decay over distance. From this contour, a discretization can be obtained for inclination angles of any array loudspeaker type of a given height.

2.3. Array system design

The main parts of designing a frontal array system are the deployment of loudspeakers and their appropriate signal processing. Moreover, room acoustical treatment and scalability are considered for fixed installations. The typical demand for such systems is to provide a similar listening experience throughout the audience, which implies consistent direct-sound coverage with a homogenous frequency response [29, chp. 3].

Direct sound coverage

The uncorrupted emitted sound from a loudspeaker without any reflections from surrounding boundary surfaces and late reverberation is defined as its direct sound. The choice of the loudspeaker type plays a crucial role in achieving uniform direct sound coverage for different audience sizes.

Considering direct sound pressure level (SPL) between 1 kHz and 10 kHz along the audience depth of different suspended loudspeaker types gives insights into the application fields for different source types. Simulations of ideal point and progressively curved and tilted line sources accurately demonstrate the different throw capabilities. Figure 2.4 shows a comparison of the SPL decay over the distance between 1 kHz and 10 kHz of a progressively curved line source of 1 m length and an ideal point source normalized at 1 m audience depth. The suspension height is set to 2 m above ear height (1.7 m), giving a height of 3.7 m to fly both sources.

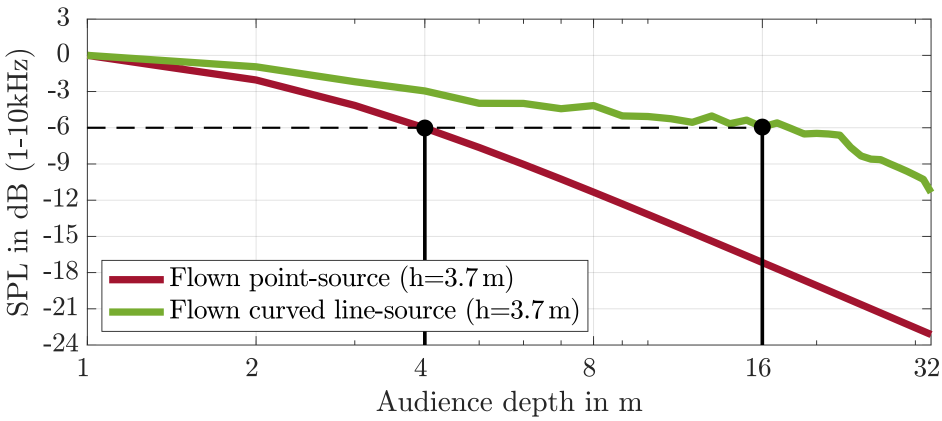


Figure 2.4: Exemplary comparison of distant dependent SPL ($1-10$ kHz) decay of ideal flown point and progressively curved and tilted line source.

This line source exhibits a direct SPL decay of -1.5 dB per distance doubling, while the point source decays at -6 dB/dod for distances from 4 m on. At 4 m, the point source SPL ($1-10$ kHz) already drops to -6 dB, whereas the line source only reaches this level at 16 m. At this 16 m distance, the point source has further decayed to -17 dB SPL ($1-10$ kHz). Beyond 22 m, the effective length of the line source approximates its actual physical length for most frequencies, marking the transition from the Fresnel to the Fraunhofer region. In this region, the SPL ($1-10$ kHz) decays at -6 dB/dod, similarly as a point source, see Section 2.2.

In a frontal sound reinforcement system, the suspension height, tilt, and rotation angle of the point source are the only parameters that can be used to rough-adjust the covered region. By contrast, the individual splay angles of line source arrays allow precise manipulation of the direct sound decay, even for complex auditorium geometries. So-called delay lines (additional delayed speakers to match the propagating wavefront), either for point or line sources, are set up further inside the audience to supplement if the direct sound needs to cover greater distances.

Frequency homogeneity

The equality of the frequency response across the audience is a second key quality for sound system design, directly impacting the preservation of tonal balance. Individual splay angles result in the curvature that defines the SPL coverage, which often requires a trade-off to frequency homogeneity. Figure 2.5 illustrates two line source designs: one for maximum direct coverage and the other for consistent frequency homogeneity. The direct sound SPL ($1 - 10\,\mathrm{kHz}$) is simulated by summing up impulses from a set of Green's functions that represent a curved line source at various positions. The right part of Figure 2.5 shows the direct sound SPL ($1 - 10\,\mathrm{kHz}$), for which the green array performs better as its SPL remains constant for up to $20\,\mathrm{m}$. The red and green array designs deliver a $-12\,\mathrm{dB}$ and a $-3\,\mathrm{dB}$ attenuation from $2\,\mathrm{m}$ to $16\,\mathrm{m}$ audience depth, respectively.

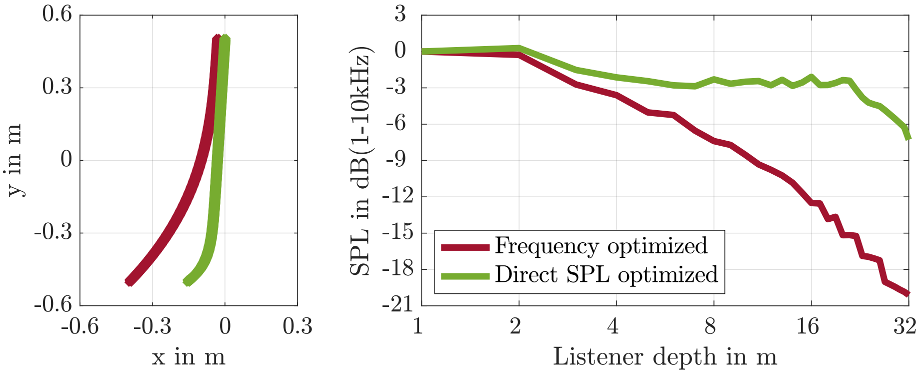


Figure 2.5: Two complementary line source designs, one is aiming for even direct sound coverage (green) while the other is optimized for consistent frequency responses over distance (red).

While the red design delivers a weaker direct sound coverage in Figure 2.5, its frequency responses at different distances ($2\,\mathrm{m}$, $8\,\mathrm{m}$, $32\,\mathrm{m}$) in Figure 2.6 prove to be outperforming the green design. While the green high-pass magnitude responses, showing the array optimized for coverage, differ noticeably in their cut-on frequency at either $125\,\mathrm{Hz}$, $500\,\mathrm{Hz}$, $4\,\mathrm{kHz}$, the high-pass characteristics of the red array remain similar at all audience distances; The slope of the high-pass is approximately 3dB per octave. A suitable compromise between delivering constant direct sound and achieving frequency stability should be targeted concerning the present requirements.

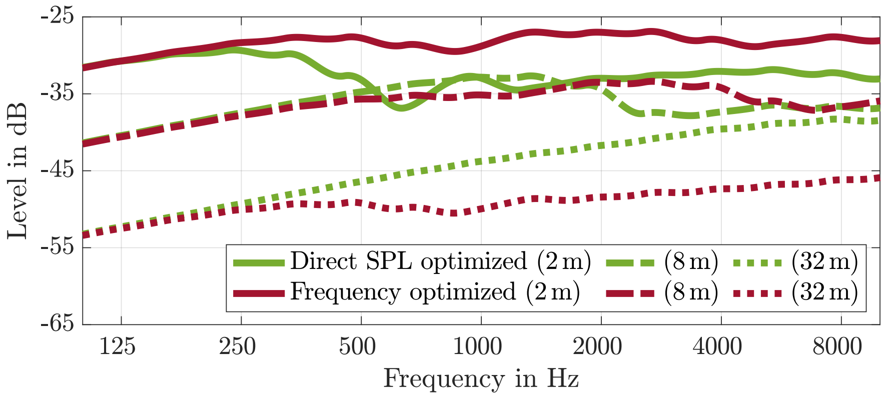


Figure 2.6: Magnitude responses from two line array designs at different distances. One is optimized for direct sound coverage (green), and the other for consistent frequency responses over distance (red).

In practice, the attenuation of high frequencies due to air dissipation over greater distances results in additional inhomogeneities. Carefully complying with those atmospheric influences, such as designing variable curvatures of the line source, can compensate for this attenuation and lead to a more balanced frequency response over distance. Additional tuning/equalizing of separate zones within the line array, covering different aiming distances, can improve the homogeneity.

2.4. Immersive sound reinforcement

The phrase “immersive sound reinforcement” first appeared in 2004 in a publication by Brandenburg et al. [30] about wave field synthesis (WFS) loudspeaker systems for cinemas and larger concerts, which concluded with a prediction about an increasing relevance of this technique in the future, which not has happened yet, at the time. However, their fundamental definition of immersive sound reinforcement as multiple sound object reproduction from several surrounding directions for larger audiences has remained almost the same until today. By dividing the term immersive sound reinforcement into its key elements, Frisch defines sound reinforcement as: *“transforming electrical signals [...] into acoustic signals by means of loudspeakers and reproducing them in a quality demanded by the respective purpose”* [31, chp. 13, p. 335] and Herre and Quackenbush describe immersion in the context of audio with *“an experience that provides to the listener the sensation of being fully immersed or present in a sound scene .[...] via different presentation modes, such as surround sound [...], 3D audio [...], and binaural audio to headphones”* [32, p. 1].

Summarizing both definitions, immersive sound reinforcement is the electro-acoustical reproduction of multidirectional audio content on two or three-dimensional loudspeaker layouts for an extended listener area. Besides reconsidering content and mixing techniques, the low latency spatialization algorithms and the comprehensive system design, including the physical deployment and loudspeaker characteristics, are the most crucial adjustable variables when creating successful audio immersions [9, chp. 2].

An immersive sound reinforcement system is designed to deliver an enveloping auditory experience with multiple loudspeaker subsystems with distinct purposes, as the

- **frontal** system supplies the direct sound and is responsible for accurately reproducing stage imaging of the performer,
- **surround** loudspeakers are positioned around the audience to enhance the sense of immersion, primarily responsible for artificial reverberation and effects,
- **height** speakers above the audience sometimes complement the system to enhance the vertical dimension and contribute to the realism of the spatial experience, and
- **fill** systems supplement audience parts that are not fully covered with direct sound from the main frontal systems, such as the first rows near the stage or balconies.

The main frontal system should consist of at least five sources, each uniformly covering the whole audience with direct sound to ensure stable localization of sound objects for all listeners [9]. Ideally, the surround loudspeakers are evenly distributed at the side and rear audience boundaries to facilitate realistic reflection and reverberation characteristics of a closed space or creative effects to envelop the listeners uniformly. Similar demands hold for height speakers that recreate artificial acoustic ceiling properties or fly-over effects [33]. Spatial fill systems for shaded audience areas require at least a similar horizontal resolution as the frontal system to comply with its direct sound object panning [34]. An exemplary system is depicted in Figure 2.7.

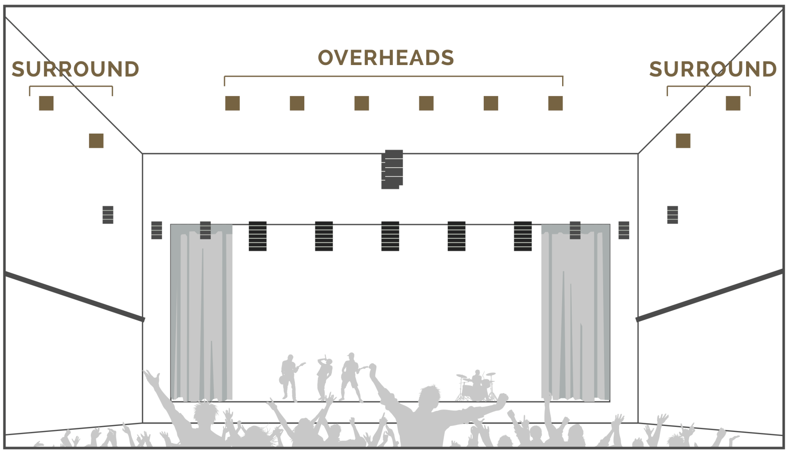


Figure 2.7: Exemplarily immersive sound reinforcement system L-ISA [35, p. 9].

2.5. Acoustic metrics

Three acoustic metrics are introduced to predict the subjective ratings from the listening experiment. These metrics—calculated from BRIR and RIR measurements at the listening positions—relate the subjective results to objective measures, verify their plausibility, and provide additional measures to evaluate immersive system designs. They are adapted from existing acoustic measures in such a way that they can approximate the subjective ratings as closely as possible. Finally, a position-independent linear regression of these (logarithmic) metrics against the actual ratings is performed to examine how well they correlate across different locations.

2.5.1. Front-to-surround ratio

We introduce the front-to-surround (FS) direct level ratio to assess object level balance. It compares the minimum A-weighted direct sound level from the three frontal loudspeakers to the maximum A-weighted direct sound level from all loudspeakers, providing a comprehensive measure of how the direct sound from the front (where most objects will be placed) contrasts with sound arriving from side and rear sources.

In order to avoid overemphasizing high frequencies when integrating over the entire audible range—and to better align with the human ear's roughly logarithmic frequency resolution—a $\frac{1}{\omega}$ weighting term is included. This simulates a quasi-third-octave integration of energy, preventing higher frequencies from dominating the calculations. All measurements are carried out using an omnidirectional measurement microphone, ensuring consistent capture of direct and diffuse sound contributions. Mathematically, the FS ratio is calculated as

$$\mathrm{FS} = 10 \log_{10} \left(\frac{\min\left(\sum^{\omega} \frac{1}{\omega} \, |H_A[\omega] \, H_{\mathrm{front}}^{\mathrm{direct}}[\omega]|^2\right)}{\max\left(\sum^{\omega} \frac{1}{\omega} |H_A[\omega] \, H_{\mathrm{all}}^{\mathrm{direct}}[\omega]|^2\right)} \right) , \tag{2.6}$$

where $H_A[\omega]$ denotes the A-weighting filter and $H_{\mathrm{front}}^{\mathrm{direct}}[\omega]$, $H_{\mathrm{all}}^{\mathrm{direct}}[\omega]$ are the direct sound transfer functions for the front and all loudspeakers, respectively.

2.5.2. Direct-to-diffuse ratio

To complement the spatial definition, the typical measure of the direct-to-reverberation level ratio was slightly adapted to the direct-to-diffuse (DD) ratio, suiting the loudspeaker layout described in Section 3.2. The DD ratio compares the mean A-weighted direct sound level of the three frontal speakers to the mean A-weighted diffuse sound level of the surrounding loudspeakers using measurements taken with an omnidirectional microphone at the individual listening locations. It is defined with

$$\mathrm{DD} = 10\log_{10}\left(\frac{\frac{1}{N_{\mathrm{ls,front}}}\sum^{\mathrm{ls,front}}\sum^{\omega}\frac{1}{\omega}\;|H_A[\omega]\;H^{\mathrm{direct}}_{\mathrm{front}}[\omega]|^2}{\frac{1}{N_{\mathrm{ls,surround}}}\sum^{\mathrm{ls,surround}}\sum^{\omega}\frac{1}{\omega}\;|H_A[\omega]\;H^{\mathrm{diffuse}}_{\mathrm{surround}}[\omega]|^2}\right)\;, \tag{2.7}$$

where $N_{\mathrm{ls,front}}$ is the number of front loudspeakers, $N_{\mathrm{ls,surround}}$ is the number of ride and rear loudspeaker, and $H^{\mathrm{diffuse}}_{\mathrm{surround}}[\omega]$ is the transfer function of the diffuse part (excluding the prominent direct sound peak of the impulse response) for the side and rear loudspeakers at the observation point.

2.5.3. Interaural level difference

The interaural level difference (ILD) describes the level difference between the left and right ear, giving a reasonably good measure of the overall uniformity of the surrounding sound field. We perform measurements using a KEMAR dummy head to determine this metric, capturing BRIRs for each loudspeaker. The mean A-weighted level (BRIR without truncation) at the left ear is compared to the mean A-weighted level at the right ear, yielding:

$$\mathrm{ILD} = 10\log_{10}\left(\frac{\sum^{\mathrm{ls}}\sum^{\omega}\frac{1}{\omega}\;|H_A[\omega]\;H^{\mathrm{all}}_{\mathrm{left\ ear}}[\omega]|^2}{\sum^{\mathrm{ls}}\sum^{\omega}\frac{1}{\omega}\;|H_A[\omega]\;H^{\mathrm{all}}_{\mathrm{right\ ear}}[\omega]|^2}\right)\;. \tag{2.8}$$

2.5.4. Linear regression of metrics to experiment ratings

To obtain reasonable comparisons between the logarithmic relative metric values in decibels and the unitless median ratings from the experiment, a linear regression[11] is performed, adjusting the metrics to the experimental rating scale from 0 to 100. An ordinary least squares approach optimizes the linear regression parameters α and β to giving the best agreement of the median experiments ratings y_i with regressed metrics $\tilde{y}_i$, such

$$\tilde{y}_i = \alpha x_i + \beta, \quad \text{and in matrix notation:} \quad \tilde{\mathbf{y}} = \mathbf{X}^T\begin{bmatrix}\beta\\ \alpha\end{bmatrix}\;, \tag{2.9}$$

where x_i are the provided predictor variables (acoustic metric) for each configuration i from a number N_c investigated configurations. The matrix $\mathbf{X}$ is composed of a vector of ones, and the metric values for each configuration as

$$\mathbf{X} = \begin{bmatrix}1 & x_0\\ \vdots & \vdots\\ 1 & x_{N_c}\end{bmatrix}\;. \tag{2.10}$$

In the estimation process of the optimal regression parameters α_{opt} and β_{opt} the sum of the squared residuals are minimized as,

$$\min_{\{\alpha,\beta\}}\sum_{i=1}^{N_c}(y_i - \tilde{y}_i)^2, \tag{2.11}$$

with N_c as the number of configurations investigated in the experiment. In matrix notation the pseudo-inverse of the metric values $\mathbf{X}^\dagger$ multiplied with the experiment rating vector $\mathbf{y}$,

$$\begin{bmatrix}\beta_{\mathrm{opt}}\\ \alpha_{\mathrm{opt}}\end{bmatrix} = \mathbf{X}^\dagger\mathbf{y} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}\;, \tag{2.12}$$

[11] https://de.mathworks.com/help/stats/fitlm.html

yielding the optimal regression parameters β_{opt} and α_{opt} and matching the metric results with median experiment ratings, while matrix $\mathbf{X}$ is compounded by the results from listening positions 1 and 2, which are independent. As the participants cannot compare both positions directly, the regression is implemented separately, resulting in regression parameters for both positions individually.

3. Methods & experimental setup

This section presents a carefully conceived experimental framework to evaluate the spatial qualities of modern immersive sound reinforcement systems for extended audience areas. The conclusions of perceptual listening experiences and acoustic metrics should help to find system design goals for loudspeakers used for multidirectional audio content.

3.1. Experimental design

To investigate the proposed hypothesis, a listening experiment is designed to compare the perceptual spatial qualities of line and point source loudspeakers for mid-sized immersive sound reinforcement. The design of the experiment mainly aims to investigate the performance of both loudspeaker types delivering sound at unequal distant off-center listening locations, as they are representative of most of the audience. The testing procedure "MUlti Stimulus test with Hidden Reference and Anchor" (MUSHRA) method was chosen as the International Telecommunication Union (ITU) confirms its suitability for the "subjective assessment of intermediate quality level of audio systems" [36, p. 3].

MUSHRA method

MUSHRA particularly aims to evaluate minor impairments that may appear in high-quality multichannel sound systems, making its procedure consistent with the proposed needs. The advantage of the MUSHRA method is that it gains representative and reliable results by having only a few (< 20) experienced participants. The MUSHRA procedure begins with a training phase where participants familiarize themselves with the test setup, the range of impairments, and the test signals. During the test, listeners are presented with a known reference signal, a hidden reference, a hidden anchor, and the stimuli under test. They can independently switch between the stimuli using an interactive system to compare them directly regarding the queried criteria. The individual stimuli are continuously graded from 0 to 100 based on their similarity to the given reference [36]. Several trials are used to evaluate the tested system for different signal types, investigating program-dependent ratings. Multiple repetitions of the trials can be performed to consolidate the ratings and investigate different general conditions. Typically, the order of investigated stimuli, presented trials, rated criteria, and conducted repetitions is randomized to eliminate order or fatigue effects, ensure valid perceptual judgments, and prevent systematic influences from affecting results.

Immersive loudspeaker layout

The design goal of this experiment is to evaluate the effect of the loudspeaker type on the quality of spatial reproduction of immersive content. An exclusively two-dimensional surrounding loudspeaker layout with eight sources in equal angular resolution of 45° (sketched in Figure 3.1), bounding a rectangular shape of $10\,\mathrm{m} \times 7\,\mathrm{m}$, was chosen for the listening experiment. To permit comparisons at each speaker location, a line and point source were positioned as closely to each other as possible to minimize localization disparities and allow unprejudiced comparability. As line sources with the suggested decay of $\approx -2\,\mathrm{dB/dod}$ [19], *L'Acoustics* "Syva" speakers are utilized and suspended to a bottom height of $1.28\,\mathrm{m}$ to match this decay goal. The coaxial "X8" loudspeakers from the same company were chosen as point sources to compare with. The exact geometry, system tuning, and signal processing are described in Section 3.2.

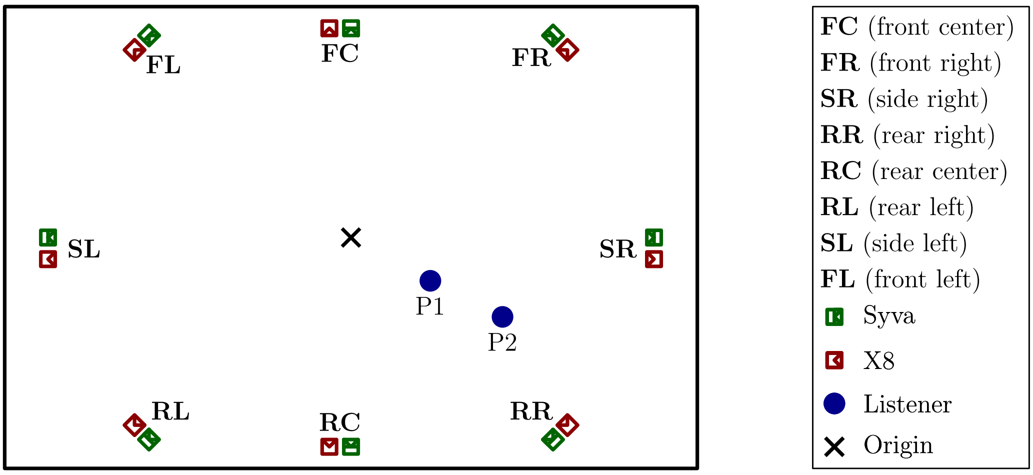


Figure 3.1: Sketch of the utilized surround loudspeaker layout.

3.1.1. Stimuli (loudspeaker configurations)

This experiment investigates different loudspeaker configurations on the fundamental layout, shown in Figure 3.1, consisting of line, point, and representative mix of both speaker types. Seven stimuli (loudspeaker configurations) are evaluated, including three hidden idealized references, three realistic configurations, and a hidden anchor. All configurations consist of eight sources at the sketched locations. The reference configurations are optimized to the specific position of the listener, which includes level and magnitude compensation for the exact listening position, such that all sources produce the same direct sound level and frequency response. These idealized references are impractical in the real world, as multiple simultaneous position-specific corrections can not be satisfied without compromise. The individual configurations (**C1-C7**) are:

- **C1(reference): Syva (position EQ)** The hidden and actual reference configuration utilized only Syva loudspeakers at all eight source locations, with additional individual-level compensation and equalization toward a flat magnitude between 125Hz and 16kHz at the listening position.
- **C2: Syva & X8 (position EQ)** A second reference-like configuration emulates a typical immersive sound reinforcement design with three frontal Syva line sources, supplemented by X8 point sources at the two side and three rear locations, and uses the idealized position-specific EQ and level corrections.
- **C3: X8 (position EQ)** To investigate the influence of the room acoustical excitation, a configuration with idealized position-specific magnitude and level correction consisting solely of X8 loudspeaker at all source locations is added as the third idealized configuration.

Idealized

- **C4: Syva** The first realistic configuration consists of eight Syva loudspeakers at all source locations with level matching and inter-loudspeaker equalization in the origin.
- **C5: Syva & X8** Three Syva and five X8 loudspeakers realize the typical configuration of line sources in the front and point sources in the side and rear positions, which are also tuned to be similar in origin.
- **C6: X8** The complementary stimuli to C4 is the configuration only of X8 point sources at all eight locations with level compensation and equalization for the origin.

Realistic

- **C7(anchor): X8 C-9dB** The hidden anchor is deliberately given poorer properties to distinguish it from the other configurations. Therefore, a configuration of eight X8 point sources with equalization and level adjustments in the origin is utilized at all locations. To create an imbalance in the intended object level balance, blur the spatial definition, and disrupt the sense of enclosure, the front center loudspeaker is attenuated by 9dB. Anchor

3.1.2. Criteria

To evaluate the criteria for rating different loudspeakers used in immersive sound reinforcement, it is crucial to identify the spatial qualities that are most affected by the type of loudspeaker. In Section 2.1, it was clarified that the directivity and near-field behaviour of line and point sources are the most characteristic distinctions affecting direct sound coverage and acoustic room excitation.

Object level balance

For immersive system design, direct sound coverage from the frontal main system is crucial for localization accuracy, clarity and preservation of the intended mix balance. Situated at an off-center location in a surrounding loudspeaker layout, different distances to the frontal loudspeakers determine direct sound level. The maximum tolerable imbalance of sound-object levels at off-center locations is determined with ± 3 dB by Zotter et al. [12]. Besides the horizontal dispersion, the distance decay can be seen as a key feature in providing uniform direct sound coverage. As shown in Section 2.1, the decay behaviour of the different speaker types varies substantially, implying a direct influence on the sound object level balance, which was chosen as the first rating criterion.

"The object level balance describes the perceptual coherence of the level balance between the sound objects in a spatial mix. It emphasizes preserving level balance, ensuring no object dominates or fades into the background while maintaining a consistent sound image."

Spatial definition

The loudspeaker directivity influences how strongly the reverberant sound field is, if a room is excited. Highly controlled directional sources only target the audience area and prevent intense early reflections or excitation of the reverberation compared to their direct sound. The ratio between the direct sound and reverberation is a meaningful measure of the perceived distance and clarity. Evaluating the influence of the loudspeaker type on reverberation, distance impression, and source width, some qualities from the "Spatial Audio Quality Inventory" (SAQI) by Lindau et al. [21] were assumed to be correlated and summarized to the second rating criterion.

"Spatial definition refers to the spatial clarity and distinctiveness of the sound objects to reach a similarly high level of presence and directness. High definition implies a balanced ratio of direct and reverberant sounds and allows individual voices to be clearly perceivable without being masked."

Spatial envelopment

As stated in [33], immersive systems are supposed to take playback reproduction to a new level by placing the listener inside a sound scene. Riedel et al. [18] found that -3 dB/dod is the optimal individual loudspeaker decay for the sensation of being equally surrounded by sound. So, the third rating criterion aims to investigate the effect of the loudspeaker type on the sensation of being uniformly enveloped by sound from all directions.

“Spatial envelopment refers to the sensation of being equally surrounded by sound from all directions. High envelopment occurs when the listener cannot identify a dominant direction, and the sound feels cohesive and uniform.”

3.1.3. Trials

Representative audio scenes have to be arranged to obtain a meaningful evaluation of the criteria. Three trials are used to playback different kinds of audio material. Two musical excerpts are utilized to rate the object level balance and spatial definition, and a deterministic uncorrelated noise signal provides ideal characteristics to grade the spatial envelopment.

Jazz-Pop

The first playback scene is a live recording of a four-piece jazz-pop band, giving distinct sound object differentiation, localizability, and transient components. The seven-second (7s) looped section includes a rhodes-like[12] sounding electrical piano playing the harmonies with voicings, an electrical bass complementing the fundamental tones, an acoustical drum set as rhythm element and female vocals as the lead voice of the piece. The drums are captured with spot microphones for the bass drum, snare drum, hi-hat, and a pair of main overhead microphones. The bass and piano are directly recorded, and a handheld dynamic microphone is utilized to record the vocals. Recreating a typical stage situation, the individual direct sounds of the channels are mixed and distributed on the frontal three sources in the layout (see Figure 3.1). The piano is placed on the left source (FL), the vocals and the bass are located in the center (FC), and the drum channels are sent to the right (FR). To avoid spatial artefacts, such as phantom image instability and localization blur, which depend on the position of the listener, direct sound objects are not amplitude panned between loudspeakers but are assigned directly to individual loudspeakers instead. Artificial reverberation is added to all surrounding sources to enhance the sensation of being in a realistic concert room. Creating realistic uncorrelated reflections and reverberation, impulse responses (IR) from the Kirishima Concert Hall[13] are utilized, measured using a dodecahedron loudspeaker on the stage and stereo pair of Neumann SKM-140 at different positions in the audience and on stage. Using the plugin IR1 from Waves[14], the IRs were altered towards a mean reverberation time of 0.9 s, typical for small venues. Unique RIRs from distinctive positions in the concert hall are allocated to the individual source locations to generate directional uncorrelated reflections. The allocation of RIRs to the individual source locations are:

- FC: conductors stand, mono-sum, 4 m from source.
- FR: row 3 in the stalls, right channel, 8 m from source.
- SR: row 10 in the stalls, right channel, 16 m from source.
- RR: row 19 in the stalls, right channel, 25 m from source.
- RC: row 19 in the stalls, mono-sum channel, 25 m from source.
- RL: row 19 in the stalls, left channel, 25 m from source.
- SL: row 10 in the stalls, left channel, 16 m from source.
- FL: row 3 in the stalls, left channel, 8 m from source.

The direct objects are convolved with all eight impulse responses with individual-level ratios depending on a personal mixing decision.

[12]https://rhodesmusic.com
[13]https://www.maki-and-associates.co.jp/projects/KRM?lang=en
[14]https://www.waves.com/plugins/ir1-convolution-reverb

String orchestra

To complement the modern pop scene, a recording of a chamber string orchestra performing Vivaldi's symphony in G major (RV149) gives rich harmonic content, wide spatial panning options, and high dynamic range typically staged in acoustically rich environments, enabling detailed evaluation of spatial definition, object level balance, and spatial envelopment. The direct sound captured with spot microphones is spatially mixed in the three front loudspeakers, placing the first and second violins at FL, the violas at FC, and the celli and double-basses at FR, recreating the American orchestra seating arrangement [37, chp. 7.1]. The RIRs used for the Jazz-Pop trial, which had a mean reverberation time of 1.9 s, proved suitable for the chamber string orchestra in the same deployment on the sources as described before.

Uncorrelated pink noise

In order to simplify rating the sensation of uniform envelopment, uncorrelated sources are advantageous. Independent, uncorrelated pink noise sources ensure a uniform spatial energy distribution, minimizing phase interference and enhancing the perception of diffuse sound fields. This approach aligns with the principles of diffuse field generation, as it avoids constructive or destructive interference patterns and provides consistent energy across all frequencies, which is essential for accurately evaluating envelopment [38, chp. 6], [39].

In the third trial, solely used for rating the envelopment, all source locations are fed with pink noise (1/f energy distribution—equal energy per octave) generated by the "mcfx_gain_delay" plugin from Matthias Kronlachner's plugin suite[15]. To prevent correlation, each source was fed by an independent instance of the plugin.

3.1.4. Repetitions

Typically, in surround sound reinforcement, the ideal sweet spot is at the origin of the loudspeaker layout, such that the distance and angles to all sources are similar. For immersive surround sound, the goal is to achieve a sweet area, supplying an entire audience area, which can be remarkably challenging at off-center locations. This experiment is repeated at two listening positions on a line from the origin, aiming to target the middle between the side right and rear right loudspeakers. The line is angled by a tangent ratio of 2:1, and the first listening location **P1** lies at a distance of 1.5 m from the origin and the second **P2**, at 3 m distance. The choice of positions represents reasonable locations for the many parts of the audience area as described in [40] for representative measurement locations within the audience. The exact locations are shown in the geometry sketch in Figure 3.3 in Section 3.2.1. The listening position to start the first trial was alternated between the sessions of the participants to prevent systematic biases.

[15]https://github.com/kronihias/mcfx

3.2. Experimental setup

The first listening experiment is conducted onsite in a room using a loudspeaker setup. The setup is virtualized and repeated binaurally on headphones to investigate the transferability of such an experiment to headphones.

3.2.1. Onsite experimental setup

The company L'Acoustics UK provided their laboratory (LAUK2 Lab, see Figure 3.2) in Kentish Town (London), all loudspeakers, amplification, connections, and interfaces for the onsite experiment and pursuing measurement.

Figure 3.2: L'Acoustics UK, immersive L-ISA experimental laboratory (LAUK2 Lab).

The existing loudspeaker layout does not match the required dimension, so a layout with custom positions corresponding to Figure 3.1 was installed. In total eight L'Acoustics Syva[16] line sources and eight L'Acoustics X8[17] point sources are used in the experiment, individually amplified by four L'Acoustics LA4X[18] and fed from a RME Digiface AVB[19] via the Audio/Video Bridging (AVB) network communication protocol [41].

The Syva line source consists of six 5-inch low-frequency (LF) and three 1.75-inch high-frequency (HF) drivers, internally crossover filters, and vertically arranged (from bottom to top: 4x LF, 3x HF, 2x LF). The X8 point source contains a coaxial arrangement of an 8-inch LF driver and a 1.5-inch HF driver, including a passive crossover filter.

Geometry

Giving an overview of the dimensions of the loudspeaker layout of the experimental setup, Figure 3.3 sketches the geometry from a top view and an exemplary side view, demonstrating the installation conditions in height and the distance gap of both loudspeaker types. The sources bound an area of $10\,\mathrm{m} \times 7\,\mathrm{m}$ and covered a listening area of approximately $9\,\mathrm{m} \times 6\,\mathrm{m}$.

[16] https://www.l-acoustics.com/products/syva/
[17] https://www.l-acoustics.com/products/x8/
[18] https://www.l-acoustics.com/products/la4x/
[19] https://rme-audio.de/de_digiface-avb.html

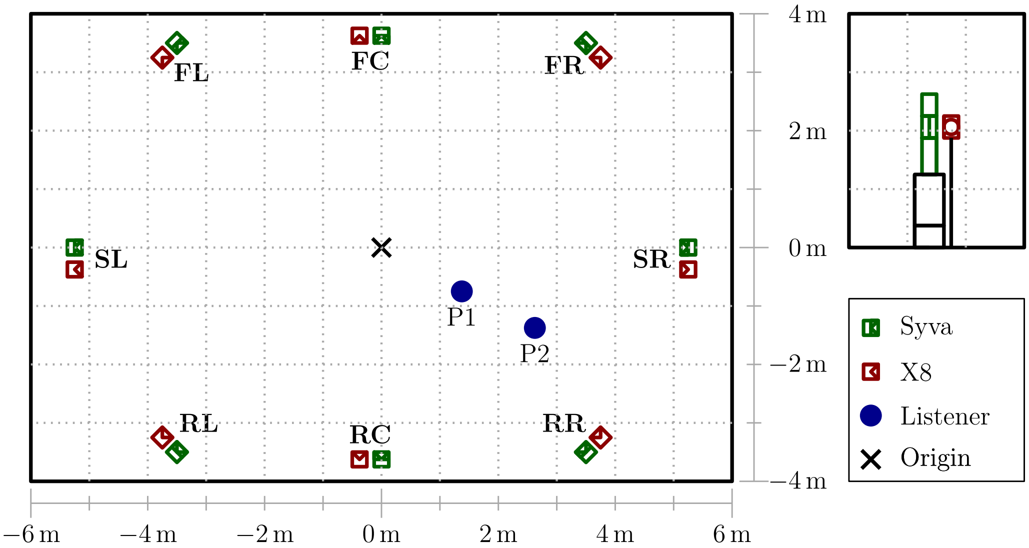


Figure 3.3: Sketch of the experimental setups geometry. Top view on the left, and exemplary side view of a Syva and X8 loudspeaker pair on the right.

Room acoustics

This study focuses on comparing different loudspeaker types and examines their influence on the room's reverberant excitation. The acoustical behaviour of the experimental space has a crucial impact on the perceived sound experience in a reverberant environment.

The room consists of a union of two main volumes (gable roof V_{gable} and ground level V_{room}) with an approximate overall volume of

$$\begin{aligned} V = V_{\text{room}} + V_{\text{roof}} &= (l \times w \times h) + \left(\frac{w \times h_r}{2} \times l_r \right) \\ &= 17\,\text{m} \times 8.8\,\text{m} \times 4.2\,\text{m} + \frac{8.8\,\text{m} \times 3\,\text{m}}{2} \times 20\,\text{m} = 892\,\text{m}^3 \ , \end{aligned} \tag{3.1}$$

where l is the length, w is the width, and h is the heigh of the room. h_r denotes the height of the roof and l_r its length. The roof is partly acoustically reflective. It consists of a glass ceiling, and the rest is covered by melamine foam. The boundary surfaces of the cuboid volume are walls, largely made of sound-hard concrete, and a wooden parquet floor.

As one of the most representative properties to describe room acoustics, the reverberation time (RT60) provides a valuable estimation of the overall acoustic behaviour of the room. The room has a mean reverberation time of approximately RT60 $\sim 0.81\,\text{s}$, which stays relatively uniform between 400 Hz and 4 kHz, as illustrated in Figure 3.4.

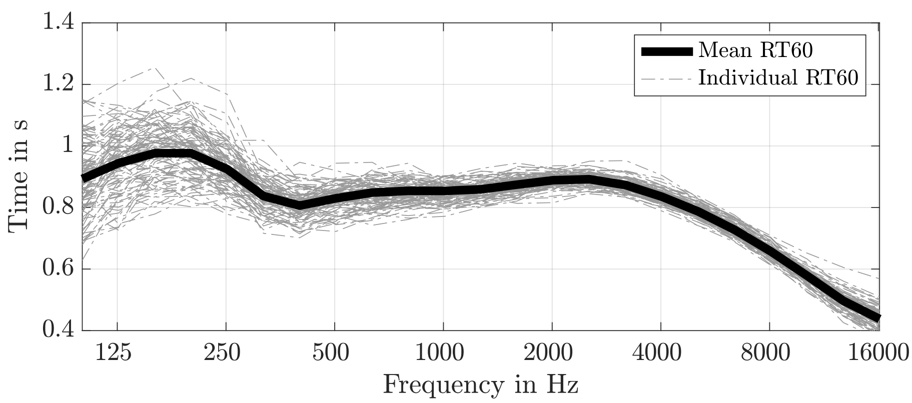


Figure 3.4: Mean reverberation time over frequency of the utilized onsite experiments room. The fine dashed lines indicate the RT60s from the individual source-receiver transfer paths.

Above 4 kHz, absorption (maybe including air absorption) causes a linear decrease to 0.44 s at 16 kHz. Below 400 Hz, the mean RT60 rises to 1 s at 160 Hz, from where it drops slightly again. For frequencies below 100 Hz, the RT60 is expected to increase again, but this was not analyzed further. The spread of the individually measured RT60 values towards frequencies below 250 Hz, demonstrates position-dependent behaviour due to the eigenfrequencies of the room.

The radiation characteristics of the loudspeakers in the room strongly interact with the room acoustics. The excitation of the reverberant field and intensity of early reflections depend highly on the directivity of the loudspeaker. In reverberant environments, the acoustic clarity measure (C50) is often used to evaluate acoustic intelligibility and direct sound supply capabilities in the audience. It is defined as the ratio of sound energy arriving within the first 50 ms compared to the sound energy of the reverberation after 50 ms [29, chp. 2] as

$$\mathrm{C50} = 10\log_{10}\left(\frac{\int_0^{50\,\mathrm{ms}} h^2(t)\,\mathrm{d}t}{\int_{50\,\mathrm{ms}}^{\infty} h^2(t)\,\mathrm{d}t}\right) \quad , \quad \text{and} \quad \mathrm{C50}[\omega] = 10\log_{10}\left(\frac{\left|\mathcal{F}\{h^{<50\,\mathrm{ms}}[n]\}\right|^2}{\left|\mathcal{F}\{h^{>50\,\mathrm{ms}}[n]\}\right|^2}\right) \quad , \tag{3.2}$$

where $h(t)$ refers to the continous-time RIR and $h[n]$ to its discrete-time representation. Comparing the frequency-dependent C50 of the two loudspeaker types, Syva and X8, gives insights into their influence on the perceived clarity and excitation of the reverberant field.

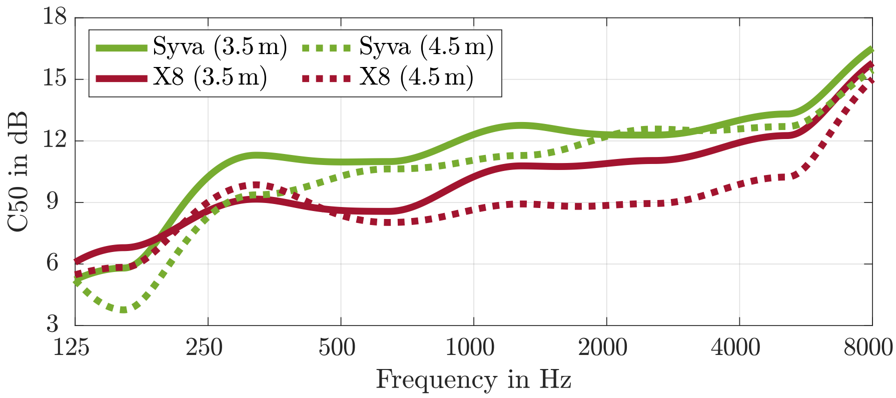


Figure 3.5: Comparison of the clarity measure (C50) over frequency of a central Syva and X8 loudspeaker in 3.5 m and 4.5 m distance.

For example, the RIRs of both loudspeaker types in the front center (FC) are measured at 3.5 m and 4.5 m distance and 1.2 m height, and their resulting C50 is displayed in Figure 3.5. Above 250 Hz, it is visible that the Syva line source shows C50 values of +2 dB compared to the X8 point source with slightly decreasing differences towards higher frequency for the closer source-receiver distance of 3.5 m. At the larger distance of 4.5 m, Syva's superior values begin at a higher frequency at 400 Hz but reach higher differences between 1 kHz and 5 kHz. In contrast to the reasonably consistent C50 between 500 Hz and 5 kHz of Syva for both distances, the X8 loudspeaker shows 2 dB of clarity loss between 1 kHz and 5 kHz at the more distant observation point. Unlike Syva, below 400 Hz, the X8's C50 remains consistent between the two distances. In the essential frequency range from 200 Hz to 8 kHz, the Syva line source can be observed to excite the reverberant field less and is able to emphasize direct and early sound more than the X8 point source. Below 200 Hz, both speaker types perform similarly except for some apparent interference effects. The higher reverberant field excitation of X8 is evident, as its directivity index (DI) would be approximately 3 dB lower than that of Syva. This difference arises from the spherical wave propagation typical of point sources, in contrast to the cylindrical propagation of line sources.

SPL decay

To achieve the targeted distance dependency of an A-weighted [13] direct sound decay of 0 to -3 dB/dod, the L'Acoustics simulation software Soundvision[20] recalled an optimal installation height for the Syva of 1.3 m for a seated audience with an ear height of 1.2 m. Placing the Syva on the associated Syva Low[21] (0.85 m) and an additional plinth (0.435 m) resulted in a height of 1.28 m, which is assumed to be sufficiently close to the targeted height.

The targeted decay is verified with on-axis ground-plane RIR measurements in 0.5 m intervals up to 12 m distance with loudspeakers mounted 1.28 m lower. The measured RIRs are available in [42]. Aligning the center of the coaxial driver of the X8 with the middle of the HF section of the Syva allowed us to observe the differences in on-axis distance, depending on the decay of both types of loudspeakers. Figure 3.6 shows the distant dependent A-weighted direct SPL of Syva in green and X8 in red. The vertical dashed black lines indicate the bounds of the listening locations of the experiment and the whole audience area.

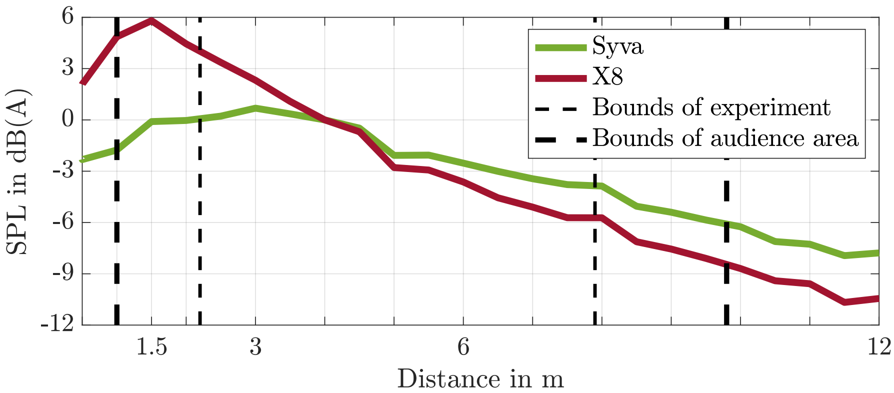


Figure 3.6: A-weighted SPL over a distance of Syva and X8.

The Syvas SPL variance of 6 dB over the audience is significantly smaller than the 16 dB accomplished by X8s. We observe a decay of 0 dB between 2 m and 4 m distance and a decay of 4 dB between 4 m and 8 m for Syva, which results in the goal of an average decay of -2 dB/dod per two doublings of the distance. In the same distance range, between 2 m and 8 m, X8 exhibits a drop in level by 10 dB, corresponding to an average decay of -5 dB/dod.

SPL coverage

To examine SPL distributions and acoustic metrics within the whole audience area, sweep measurements were taken with the omnidirectional ISEMcon EMX7150[22] measurement microphone on a 8 m × 6 m grid with a spatial resolution of 1 m and measured at a height of 1.2 m, matching the ear height of a seated person. A total of 9 (width) × 7 (depth) × 16 (loudspeaker) = 1008 impulse responses were obtained from the measurements (explained in Equation (3.3)). A repository[23] of all measured impulse responses is available at [42]. Figure 3.7 illustrates the dimensions of the measurement setup.

[20] https://www.l-acoustics.com/products/soundvision/
[21] https://www.l-acoustics.com/products/syva-low/
[22] https://www.isemcon.com/datasheets/EMX7150-US-r04.pdf
[23] https://phaidra.kug.ac.at/view/o:135786

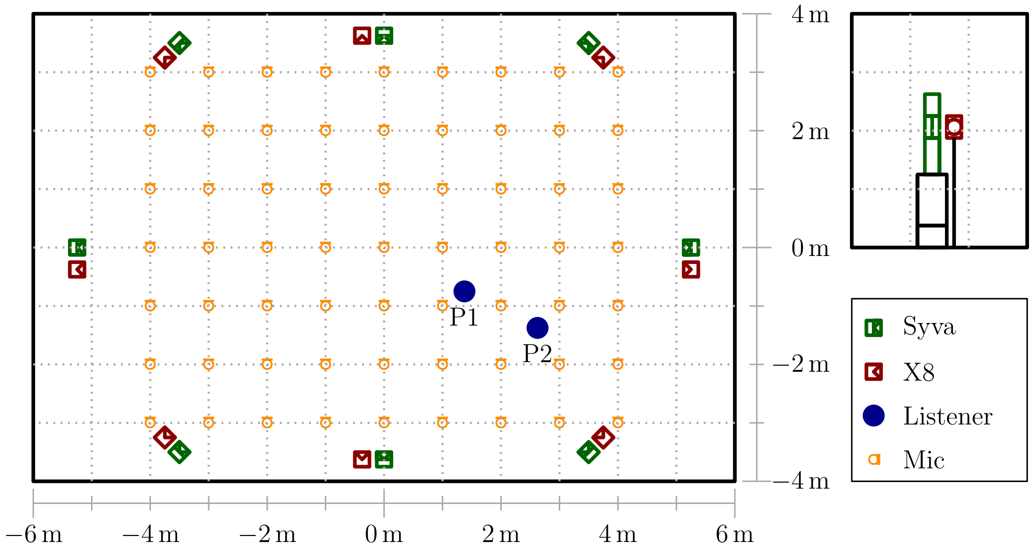
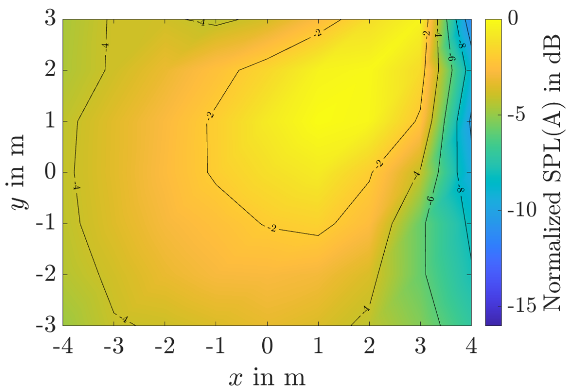


Figure 3.7: Sketch of the measurement locations in the experimental setup.

The coverage and advantages of Syva line source are even more substantial when plotting the SPL distribution across the whole audience for a single loudspeaker. Figure 3.8 displays the corresponding normalized A-weighted direct sound SPL map.



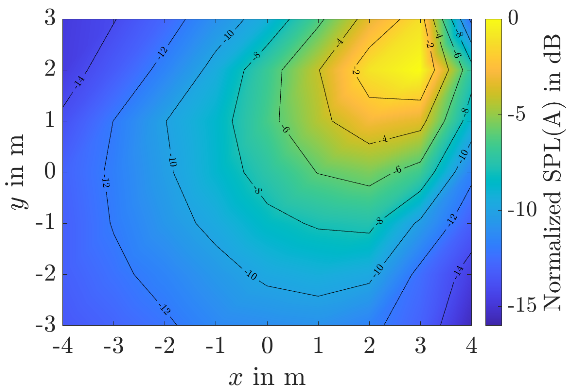


Figure 3.8: Normalized direct SPL(A) distribution over the audience area, measured with 1m spatial resolution. The left plot shows the SPL coverage of a single Syva line source at position FR, and the right plot shows the coverage of one X8 point source at the same position.

The direct sound coverage variability of a single Syva exceeds 6 dB only at the audience boundaries, while the X8 reaches this border after approximately 3 m. At the boundaries, the direct SPL of X8 decreases by 12 dB to 15 dB.

Loudspeaker tuning

To minimize the chance of identifying the loudspeaker type either by slight localization differences or frequency response, both speaker's HF sections are placed as close as possible to each other (see Figure 3.9), and their magnitudes are equalized using mixed-phase filters towards a flat magnitude between 125 Hz and 16 kHz. Those finite impulse response (FIR) filters show linear-phase behaviour for the high frequencies and minium-phase filtering for the lower frequencies, giving both low processing latency and accurate high-frequency filtering without altering the phase response of the loudspeaker.

Figure 3.9: Closest possible distance (0.22 m) between HF-sections of Syva and X8, due to the pole mounting of the X8.

To match the frequency response of both speaker types, all 16 loudspeakers (indexed with ls) are measured by recording a two-second exponential sweep $x_{\mathrm{sweep}}[n]$ played back from each loudspeaker individually and captured at nine microphone locations $y_{\mathrm{ls,mic}}[n]$ (indexed with mic) around the origin in 1.2 m head height with the ISEMcon EMX7150 omnidirectional microphone. All measured RIR are available[24] in [42]. The exact measurement grid of the receiver location is shown in Figure 3.10.

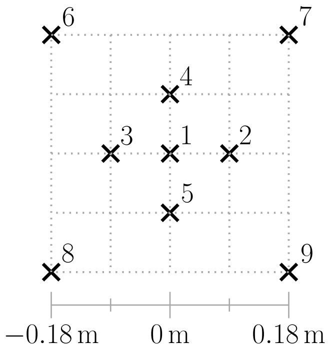


Figure 3.10: Measurement microphone grid for loudspeaker equalization at an area around the desired hearing position, enabling further averaging and weighting.

The IRs $h_{\mathrm{ls,mic}}[n]$ are obtained by dividing the Fourier transformation ($\mathcal{F}\{\}$) of the recorded sweep $y_{\mathrm{ls,mic}}[n]$ by the initial sweep $\mathcal{F}\left\{x_{\mathrm{sweep}}[n]\right\}$ and then calculate its inverse Fourier transformation($\mathcal{F}^{-1}\{\}$) as

$$h_{\mathrm{ls,mic}}[n] = \mathcal{F}^{-1}\left\{\frac{\mathcal{F}\left\{y_{\mathrm{ls,mic}}[n]\right\}}{\mathcal{F}\left\{x_{\mathrm{sweep}}[n]\right\}}\right\} . \tag{3.3}$$

The truncation of the resulting IR $h_{\mathrm{ls,mic}}[n]$ at 200 ms (9600 samples at 48 kHz sampling frequency) after the initial peak at N_i removes late reverberation, which does not affect the direct sound frequency response. For smooth decay of the IR, a half Hann window $h_{\mathrm{Hann}}[n]$ of 200 ms is applied to the truncated $h_{\mathrm{ls,mic}}[n]$ to define $\hat{h}_{\mathrm{ls,mic}}[n]$ as

$$\hat{h}_{\mathrm{ls,mic}}[n] = h_{\mathrm{Hann}}[n]\ h_{\mathrm{ls,mic}}[N_i < n < N_i + 9600] . \tag{3.4}$$

The loudspeaker frequency response is corrected by dividing the measured magnitudes $\left|\hat{H}_{\mathrm{ls,mic}}[\omega]\right|$ with the absolute magnitude response $|H_{\mathrm{correction}}[\omega]|$ of the measurement microphone (provided by the manufacturer), as

[24] https://phaidra.kug.ac.at/view/o:135786

$$\left|H_{\mathrm{ls,mic}}[\omega]\right| = \frac{\left|\hat{H}_{\mathrm{ls,mic}}[\omega]\right|}{\left|H_{\mathrm{correction}}[\omega]\right|} \ , \tag{3.5}$$

obtaining the corrected magnitude $\left|H_{\mathrm{ls,mic}}[\omega]\right|$ for each microphone-loudspeaker path with ω as the cycle frequency in $\frac{1}{\mathrm{s}}$. To gain individual loudspeaker correction filters for a reasonable listening region, the nine microphone locations are averaged and weighted by g_{mic}, emphasizing the central position,

$$H_{\mathrm{ls}}[\omega] = \sqrt{\frac{1}{N_{\mathrm{mic}}} \sum_{\mathrm{mic}=0}^{N_{\mathrm{mic}}} \left|g_{\mathrm{mic}} H_{\mathrm{ls,mic}}[\omega]\right|^2} \ , \tag{3.6}$$

where $H_{\mathrm{ls}}[\omega]$ is the magnitude response of the loudspeaker, $N_{\mathrm{mic}} = 9$ the number of measured microphone positions, and $g_{\mathrm{mic}} = [1, 0.8, 0.8, 0.8, 0.8, 0.6, 0.6, 0.6, 0.6]$ the set of weights. To adapt equalization to the spectral resolution of human hearing, the magnitude responses $H_{\mathrm{ls}}[\omega]$ were smoothed by decomposition into third-octave bands with a center cycle frequencies ω_c and fractional octave parameter $n_{\mathrm{oct}} = 3$. A normalized $\cos^2$ filter bank $W_{\omega_c}[\omega]$ is created,

$$W_{\omega_c}[\omega] = \frac{\widetilde{W}_{\omega_c}[\omega]}{\sqrt{\sum^{\omega} \widetilde{W}^2_{\omega_c}[\omega]}} \quad \text{with} \quad \widetilde{W}_{\omega_c}[\omega] = \cos^2\left[\frac{\pi}{2}\ \mathrm{Clip}\left\{\log_2\left(\frac{\omega}{\omega_c}\right) n_{\mathrm{oct}}, -1, 1\right\}\right] \ . \tag{3.7}$$

The smoothing applies the filter bank to $H_{\mathrm{ls}}[\omega]$ and retrieves the back the smoothed result by projecting the band levels back to the uniformly resolved Fourier transform variable ω

$$\tilde{H}_{\mathrm{ls}}[\omega] = \sum^{\omega_c} W_{\omega_c}[\omega] \sum^{\omega} \widetilde{W}_{\omega_c}[\omega] H_{\mathrm{ls}}[\omega] \ . \tag{3.8}$$

The resulting smoothed and position-averaged magnitudes are shown in decibels as $20\log_{10}\left(|\tilde{H}_{\mathrm{ls}}[\omega]|\right)$ in Figure 3.11.

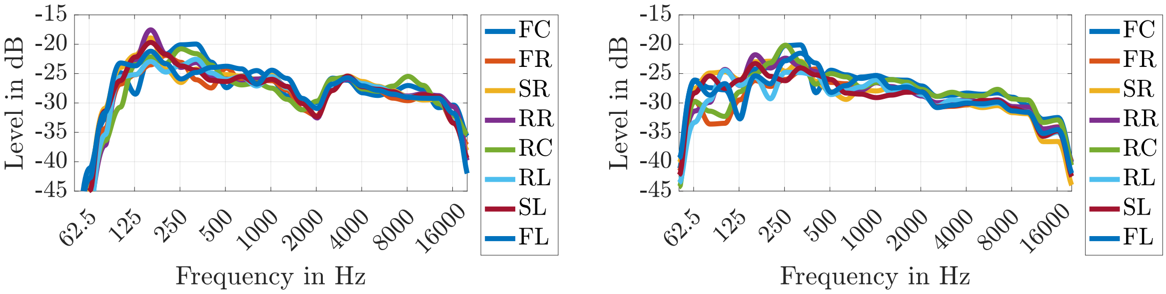


Figure 3.11: Smoothed and averaged magnitudes of all eight Syva loudspeakers (left) and X8 speakers (right) in the origin of the layout.

The minimum phase IR $h_{\mathrm{ls}}^{\mathrm{min}}[n]$ of $\tilde{H}_{\mathrm{ls}}[\omega]$ is calculated by employing the real valued cepstrum as

$$\begin{aligned}
h_{\mathrm{ls}}[n] &= \mathcal{F}^{-1}\left\{\ln\left(\tilde{H}_{\mathrm{ls}}[\omega]\right)\right\} \\
h_{\mathrm{ls}}^{c}[n] &= \begin{cases} h_{\mathrm{ls}}[n] & \text{for} \quad n = 0 \\ 2h_{\mathrm{ls}}[n] & \text{for} \quad 1 \leq n \leq \frac{N_{h_{\mathrm{ls}}}}{2} - 1 \\ h_{\mathrm{ls}}[n] & \text{for} \quad n = \frac{N_{h_{\mathrm{ls}}}}{2} \\ 0 & \text{for} \quad \frac{N_{h_{\mathrm{ls}}}}{2} + 1 \leq n \leq N_{h_{\mathrm{ls}}} \end{cases} \\
h_{\mathrm{ls}}^{\mathrm{min}}[n] &= \Re\left\{\mathcal{F}^{-1}\left\{e^{\left(\mathcal{F}\left\{h_{\mathrm{ls}}^{c}[n]\right\}\right\}}\right\} \ ,
\end{aligned} \tag{3.9}$$

where $N_{h_{\mathrm{ls}}}$ is the number of samples of h_{ls} and depends on the size (NFFT) of the Fourier transformation. $\Re\{\}$ denotes the real part of the resulting complex number. Given a target IR $h_t[n]$ that has band-pass characteristics from 125 Hz to 16 kHz, the equalizing linear-phase IR

$h_{\text{ls}}^{\text{lin}}[n]$ can then be obtained by the division of the target response $h_t[n]$ of $h_{\text{ls}}^{\text{min}}[n]$ in the Fourier transform domain, as

$$h_{\text{ls}}^{\text{lin}}[n] = \mathcal{F}^{-1}\left\{ \left| \frac{\mathcal{F}\{h_t[n]\}}{\mathcal{F}\{h_{\text{ls}}^{\text{min}}[n]\}} \right| \right\} . \tag{3.10}$$

A circular shift is helpful to see a continuous image of the main peak of the symmetric impulse response

$$h_{\text{ls}}^{\text{lin}}[n] = \begin{cases} h_{\text{ls}}^{\text{lin}}[n] & \text{for} \quad \frac{N_{h_{\text{ls}}^{\text{lin}}}}{2} \leq n \leq N_{h_{\text{ls}}^{\text{lin}}} - 1 \\ h_{\text{ls}}^{\text{lin}}[n] & \text{for} \quad 0 \leq n \leq \frac{N_{h_{\text{ls}}^{\text{lin}}}}{2} - 1 \end{cases} . \tag{3.11}$$

This response has linear phase properties. Symmetrical truncation to $N_{\text{lin}} = 257$ samples around the main peak reduces latency to $N = 127$ samples or $t = 2.65\,\text{ms}$ at $48\,\text{kHz}$, and frequency resolution to $\Delta f_{\text{lin}} = \frac{2f_s}{N_{\text{lin}}} = 375\,\text{Hz}$. As this filter is only practical for filtering in third-octave resolution above $1\,\text{kHz}$, a separate minimum-phase filter is calculated for the low-frequency range. The linear phase equalization $H_{\text{ls}}^{\text{lin}}[\omega]$ applied on the smoothed and averaged measurement $\tilde{H}_{\text{ls}}[\omega]$, yields a partially equalized magnitude

$$\hat{H}_{\text{ls}}[\omega] = \tilde{H}_{\text{ls}}[\omega] H_{\text{ls}}^{\text{lin}}[\omega] \ , \tag{3.12}$$

to be refined with a subsequent minimum-phase filter. This filter is found from the real-valued cepstrum Equation (3.9) of $\left|\frac{H_t[\omega]}{\hat{H}_{\text{ls}}[\omega]}\right|$ and results in $\hat{h}_{\text{ls}}^{\text{eq}}[n]$. The truncation to $N_{\text{min}} = 4096$ samples results in a frequency resolution of $\Delta f_{\text{min}} = \frac{f_s}{N_{\text{min}}} = 11.72\,\text{Hz}$ providing third octave frequency resolution down to $50\,\text{Hz}$ at $f_s = 48\,\text{kHz}$.

Equalizing each loudspeaker is done by filtering with the corresponding 4096 tap equalization FIR-filter $\hat{h}_{\text{ls}}^{\text{eq}}[n]$. After filtering, the root mean square (RMS) values RMS_{ls} of each loudspeaker at the origin are calculated to verify equalization using

$$\text{RMS}_{\text{ls}} = 20\log_{10}\left(\sqrt{\frac{1}{N}\sum^{N} \left(\hat{h}_{\text{ls}}^{\text{eq}}[n] * h_{\text{ls}}[n]\right)^2} \right) \ , \tag{3.13}$$

where N is the number of samples after convolution. Dividing the minimum of all RMS_{ls} values by each RMS_{ls} yields the gains $g_{\text{ls}}^{\text{origin}}$ to compensate for the level differences of the loudspeakers

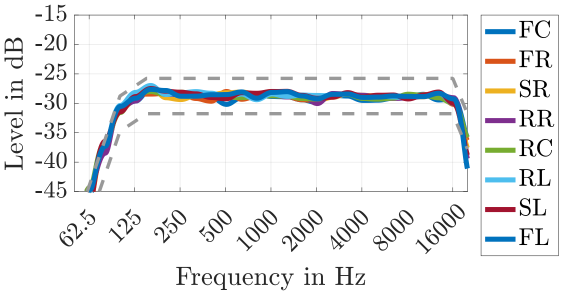
$$g_{\text{ls,origin}} = \min_{\text{ls}}(\text{RMS}_{\text{ls}}) \frac{1}{\text{RMS}_{\text{ls}}} \ , \tag{3.14}$$

with a maximum gain of 1, where unequal source distances to the origin are regarded. The decibel values of calculated gains $g_{\text{ls,origin}}$ for all loudspeakers are collected in Table 3.1. The higher values for X8 loudspeakers result from an additional offset to match the level with the Syva loudspeakers.

$20\log_{10}(g_{\text{ls,origin}})$ in dB	FC	FR	SR	RR	RC	RL	SL	FL
Syva	-0.90	0.00	-0.35	-0.70	-1.20	-0.44	-0.29	-0.43
X8	0.50	1.96	2.69	1.54	0.60	2.12	2.15	1.98

Table 3.1: Decibel values of compensation gains $g_{\text{ls,origin}}$ for all loudspeakers in the origin.

The filtered and level-compensated magnitude responses show nearly identical flat responses and similar levels for both speaker types, illustrated in Figure 3.12.



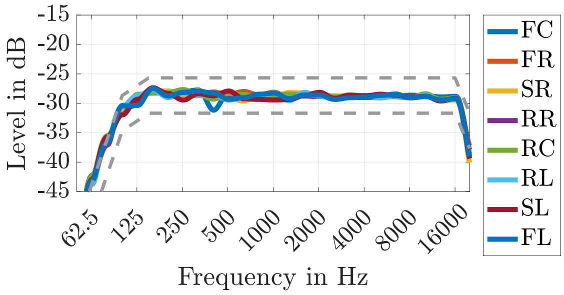


Figure 3.12: Magnitudes of all eight Syva loudspeakers (left) and X8 speakers (right) in the origin of the layout after filtering with equalizing FIR-filter, dashed lines represent ±3 dB limits.

Using the software Smaart v8.0 from Rational Acoustics[25], the equality of magnitude and level for all loudspeakers is verified in the origin. For the realistic application scenarios, **C4**, **C5**, and **C6**, the tuning procedure is complete.

To create the ideal configurations **C1**, **C2**, and **C3**, the same tuning process is repeated but measured with the grid from Figure 3.10 at the actual listening locations **P1** and **P2**. By doing so, filters and gains are calculated to match the directional level and magnitude at those explicit listening locations. This creates idealized conditions at off-center locations, which are used as references in the experiment. The measured magnitudes, calculated gains and resulting frequency responses are attached in Section A.1.

Additional level adjustments references of spatial envelopment

For the spatial envelopment criterion, the optimal properties for each loudspeaker should be a decay of −3 dB/dod, as demonstrated in [17]. Assuming a surround configuration delivering similar sound pressure from each source at the origin, a source-dependent gain factor $g_{\mathrm{ls,pos}}^{\text{-3dB}}$ must be applied for each distinct listening position to provide an idealized reference. The target gain value can be calculated for each loudspeaker with the square root distance ratio

$$g_{\mathrm{ls,pos}}^{\text{-3dB}} = \sqrt{\frac{d_{\mathrm{ls,origin}}}{d_{\mathrm{ls,pos}}}}\,, \tag{3.15}$$

where $d_{\mathrm{ls,origin}}$ is the distance from each loudspeaker (tweeter-section), indicated with ls, to the origin at head height (1.2 m). $d_{\mathrm{ls,pos}}$ is the distance from each loudspeaker to the listening position (also 1.2 m head height). This incremental diffuseness gain for −3 dB decay $g_{\mathrm{ls,pos}}^{\text{-3dB}}$ presupposes an equal level from each source at the distinct location, using the gains $g_{\mathrm{ls}}^{\mathrm{P1}}$ and $g_{\mathrm{ls}}^{\mathrm{P2}}$ from Table A.1 and Table A.2. The diffuseness gain values for all loudspeakers and both listening positions are shown in Table 3.2 and Table 3.3.

$20\log_{10}\left(g_{\mathrm{ls,P1}}^{\text{-3dB}}\right)$ in dB	FC	FR	SR	RR	RC	RL	SL	FL
Syva	-0.92	0.22	1.21	1.41	0.45	-0.54	-1.01	-1.11
X8	-0.98	0.29	1.26	1.45	0.33	-0.59	-1.00	-1.12

Table 3.2: Decibel values of −3 dB compensation gains $g_{\mathrm{ls,P1}}^{\text{-3dB}}$ for all loudspeakers at position P1.

$20\log_{10}\left(g_{\mathrm{ls,P2}}^{\text{-3dB}}\right)$ in dB	FC	FR	SR	RR	RC	RL	SL	FL
Syva	-1.89	0.01	2.51	3.17	0.08	-1.22	-1.86	-2.01
X8	-1.97	0.13	2.69	3.33	-0.11	-1.29	-1.84	-2.02

Table 3.3: Decibel values of −3 dB compensation gains $g_{\mathrm{ls,P2}}^{\text{-3dB}}$ for all loudspeakers at position P2.

[25] https://www.rationalacoustics.com

3.2.2. Headphone experimental setup

Transferring the listening experiment virtually to headphones requires accurately capturing the room acoustical behaviour, including the loudspeaker characteristics. The exact loudspeaker configuration used for the onsite experiment needs to be measured regarding binaural room impulse responses (BRIR). A repository[26] of all measured BRIR is available at [42]. As the final playback format is binaural, dummy head measurements provide a straightforward choice for virtualization. Inserting the measured BRIRs into the REAPER[27] and MUSHRA environment[28] of the onsite experiment, enabling to replicate the experiment on headphones.

Kemar dummy head

The G.R.A.S "45BB KEMAR" dummy head and torso, including the pinna simulator and ear canal extension [43], provide an accurate reproduction of the acoustic characteristics of human hearing, which made it the reasonable choice for the virtualization measurements. Individual ear calibration, using the G.R.A.S "Pisonphone Typ 42AA" emitting a 1/3-octave band noise with center frequency 250 Hz at 10 Pascal (114 dB SPL) when mounted to the ear, ensures compensation unequal sensitivities of the two ear microphones. For the measurement, the large ears of type KB0065 are employed. The impulse responses from all 16 loudspeakers to both ears are measured at the two listening positions and in the origin for a diffuse field equalization. The ear height of the dummy head is 1.2 m to imitate a seated person.

Diffuse field equalization

The ear-canal simulator of the KEMAR dummy head alters the frequency response of the measurements and is not suitable for a realistic binaural playback, which should only involve outer ears. Diffuse field equalization is, therefore, an important post-processing step for dummy head recordings, as it helps compensate for unwanted magnitude responses. The frequency characteristics of a dummy head vary significantly depending on the incident angle. Therefore, diffuse field equalization adjusts the average magnitude response from all directions to a target frequency response.

Let $h_{\mathrm{ls}}^{\mathrm{omni}}[n]$ be the measured RIR at the origin and $h_{\mathrm{ls,ear}}^{\mathrm{head}}[n]$ the BRIR at one ear, then

$$|H_{\mathrm{diff,eq}}[\omega]| = \sqrt{\frac{\sum^{\mathrm{ls}} |H_{\mathrm{ls}}^{\mathrm{omni}}[\omega]|^2}{\sum^{\mathrm{ls}} \sum^{\mathrm{ear}} |H_{\mathrm{ls,ear}}^{\mathrm{head}}[\omega]|^2}}\ , \tag{3.16}$$

denotes the magnitude response of the diffuse field equalization as shown in Figure 3.13.

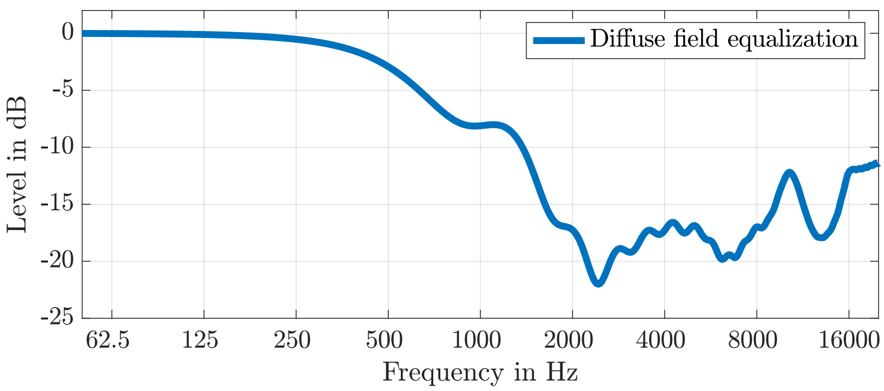


Figure 3.13: Magnitude response of the diffuse field equalization filter.

[26] https://phaidra.kug.ac.at/view/o:135786
[27] https://www.reaper.fm
[28] https://git.iem.at/rudrich/MUSHRA

Headphones

The choice of headphones can play an important role in virtualization and binaural playback. Binaural playback typically aims to reproduce real sound scenes where the sound objects are located outside the head. Satongar et al. [44] analyzed the measurable error of the interaural level difference (ILD) for different headphones and concluded that the more open the headphones are, the less the spectral distortion. Another spectral error analysis was done by Schneiderwind et al. [45], which concluded that extra-aural headphones introduce the least perceptive audible distortions. As headphones for binaural playback in the second experiment, the 3D-printed ultralight circumaural open headphones shown in Figure 3.14 from [46] were chosen to ensure a high degree of externalization for realistic binaural sound field reproduction.

Figure 3.14: Utilized ultralight circumaural open headphones from [46].

Due to the open construction, extra filtering is necessary to flatten the magnitude of the headphones. The applied correction magnitude response is shown in Figure 3.15.

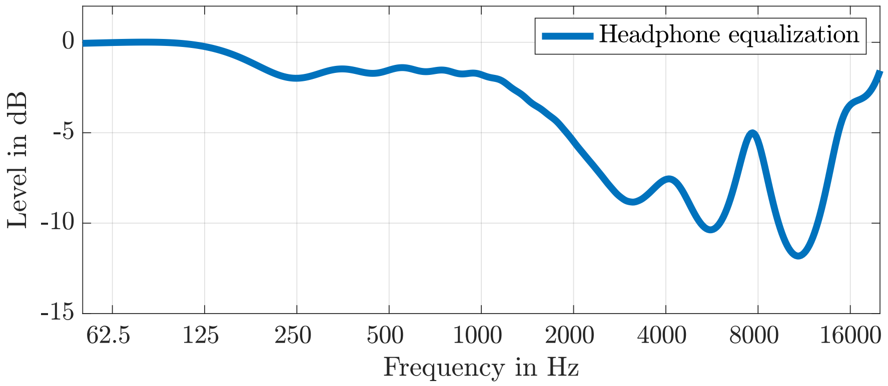


Figure 3.15: Magnitude response of the headphone equalization filter.

4. Results

This section of the thesis presents and compares the results of both listening experiments and relates them to the introduced acoustic metrics. All anonymized ratings are available[29] in [42]. The plots shown in this section provide a statistical overview of the listening experiment results and display the individual ratings for each criterion x_p of each participant p, the median, and the 95% confidence interval.

The median is the value in the middle of the data set, determined by sorting all data and searching for the middle value, splitting the data set into two same-size parts. The median is robust against outliers and representative of the typical ratings, enabling meaningful conclusions to be drawn for the central rating tendency.

The 95% confidence interval (CI) represents the range within the median that is expected to be located with a 95% certainty. It reflects the variability of the data set and is a meaningful complement to the median as it quantifies the reliability of the results. The utilized bootstrap procedure[30] employed resamples of the original data set X with replacement $N_b = 10000$ times, resulting in N_b new datasets X_b of the same size as X containing the samples drawn (possibly multiple times the same sample as the drawing process is with replacement) [47]. The median of each dataset X_b is collected in the bootstrap vector B, and its confidence interval is determined by calculating the 2.5^{th}- and 97.5^{th}-percentiles.

If the location of the median of one of the conditions tested lies outside of the CI of another condition under test, it roughly indicates that there could be significant differences. The Bonferroni corrected [48] p-value from a two-sided Wilcoxon signed rank test[31] [49] gives further insights into the significance between two conditions. p-values below 0.05 are denoted as significant. In this section, only an exclusive selection of essential p-values contributing the investigation are presented. The p-values for all configuration and position pairs of both listening experiments are attached in Section A.2.

In addition to the p-values from the significance tests described above, a robust version of Cohen's d effect size was calculated[32] to quantify the importance of the differences between the positions and configurations. Typically, Cohen's d is calculated by dividing the difference between condition means by the pooled standard deviation [50, chp. 2], [51]. Effect sizes between two conditions of $d \approx 0.2$ indicate small effects, $d \approx 0.5$ medium effects, and $d \approx 0.8$ large effects. However, since these effect size values seem to not provide additional detailed statistical insight beyond the p-values in this case, they are only collected in the Section A.3 rather than discussed Section 4.1 and Section 4.2.

4.1. Onsite experiment

In total, 17 listeners (students and employees from L'Acoustics) with relevant experience in immersive audio playback systems participated in the onsite experiment. Two assessors were excluded from the results as they rated the hidden reference less than 90 in more than 15% of the trials, cf. [36]. The age of one female and 16 male participants varied between 23 and 53 years, on average 34 years. The average duration for the experiment was 40 minutes.

[29] https://phaidra.kug.ac.at/view/o:135786
[30] https://de.mathworks.com/help/stats/bootci.html
[31] https://de.mathworks.com/help/stats/signrank.html
[32] https://de.mathworks.com/help/stats/meaneffectsize.html

4.1.1. Familiarization

At the start of the onsite listening experiment, participants were introduced to the musical excerpts and the playback setup. A simplified application of the repertory grid technique for rating multichannel sound [52] was employed, focusing solely on the assessment stage, omitting the elicitation process, and presenting pre-selected rating criteria in advance. Participants were asked to rank six spatial qualities they perceive to change the most when an immersive mix is played on different loudspeaker configurations. Most of the definitions for these pre-selected qualities were drawn from the SAQI [21]:

Localizability
"If localizability is low, spatial extent and location of a sound source are difficult to estimate, or appear diffuse, resp. If localizability is high, a sound source is clearly delimited. Low/high localizability is often associated with high/low perceived extent of a sound source. Examples: sound sources in highly diffuse sound field are poorly localizable." [21, p. 6]

Loudness
"Perceived loudness of a sound source. Disappearance of a sound source can be stated by a loudness equaling zero. Example of a loudness contrast: whispering vs. screaming." [21, p. 6]

Clarity
"Clarity/clearness with respect to any characteristic of elements of a sound scene. Impression of how clearly different elements in a scene can be distinguished from each other, how well various properties of individual scene elements can be detected. The term is thus to be understood much broader than the in realm of room acoustics, where clarity is used to predict the impression of declining transparency with increasing reverberation." [21, p. 7]

Level of reverberation (direct/reverberation-ratio)
"Perception of a strong reverberant sound field, caused by a high ratio of reflected to direct sound energy. Leads to the impression of high diffusivity in case of stationary excitation (in the sense of a low D/R-ratio). Example: The perceived intensity of reverberation differs significantly between rather small and very large spaces, such as living rooms and churches." [21, p. 6]

Further, the general **timbre** was asked for, referring to tonal and frequency-dependent changes between the different layouts. The **object level balance** describes the loudness balance of different directional objects in the spatial mix.

During the familiarization phase, an audio loop is played in which loudspeaker configurations switch automatically, with the starting position randomized for each listener. The participant is asked to rate:

Which spatial qualities are perceptually most affected by the chosen loudspeaker type?

By analyzing the median ratings, it emerges that clarity, timbre, D/R-ratio, and object level balance appear to be most influenced by the loudspeaker setup. Localizability yields fewer discernible changes, and participants detect almost no differences in loudness. The sorted ranking with corresponding median values (P1 and P2 for different listening locations) shows a slight tendency rather than a strong effect, as shown in Table 4.1.

Quality	P1	P2	Mean
Clarity	27.1	18.1	22.5
Timbre	23.3	11.0	17.1
D/R-ratio	19.4	14.2	16.8
Object level balance	19.5	13.3	16.4
Localizability	12.4	10.1	11.3
Loudness	9.9	1.4	5.7

Table 4.1: Median results from the familiarization procedure, ranked by its mean values of both listening locations.

These observations align reasonably well with our evaluation criteria suggested in Section 3.1.2. Nevertheless, this initial familiarization procedure did not offer extensive scientific insights but merely reinforced that focusing on qualities such as clarity, D/R-ratio, and object level balance is meaningful. The clarity and D/R-ratio are summarized in one queried quality **spatial definition**, and the **object level balance** is asked anyway as a rating criterion.

4.1.2. Object level balance (onsite)

In Figure 4.1 the median and 95% CI are illustrated for the object level balance for all idealized references, realistic conditions, and anchor configuration, and they are plotted separately for both listening locations. Position 1 (P1) is illustrated in the darker red, while position 2 (P2) is brighter. The three configurations in the left section are the idealized references layouts (C1-C3) with position compensated level and magnitude, while the first reference configuration (C1) shows the hidden reference ratings. The central three configurations are the realistic application scenarios (C4-C6), while the hidden anchor (C7) is shown in a separate section on the right. The grey cross markers indicate the individual ratings.

Table 4.2 shows the p-value of a two-sided Wilcoxon signed rank test with Bonferroni correction comparing the object level balance ratings for distinct configuration pairs for positions P1 and P2. The hidden reference is compared to all other deployments except for the hidden anchor, and the realistic layouts are compared among each other. Values below 0.05 are denoted as significant [49] and highlighted in bold.

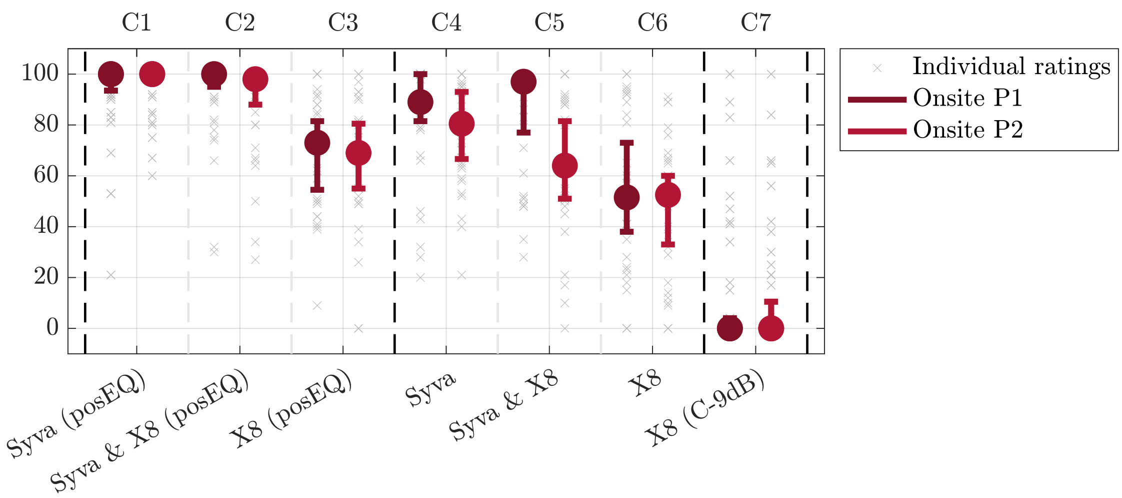


Figure 4.1: Onsite object level balance ratings for the two listening locations. The dots indicate medians; vertical lines are the 95% CI for both listening positions. On the left are the three idealized references C1-C3, in the middle three realistic configurations C4-C6, and on the right is the anchor C7.

p	C1|C2	C1|C3	C1|C4	C2|C5	C3|C6	C4|C5	C4|C6	C5|C6
P1	1.039	**0.005**	0.565	0.544	0.069	0.966	**0.005**	**0.002**
P2	0.223	$<$ **0.001**	**0.005**	**0.006**	**0.031**	0.202	**0.001**	**0.025**

Table 4.2: Bonferroni-corrected p-values from a two-sided Wilcoxon signed rank test for pairwise comparisons of the loudspeaker layouts for the object level balance ratings. Values below 0.05 are denoted as significant.

p	C1(P1|P2)	C2(P1|P2)	C3(P1|P2)	C4(P1|P2)	C5(P1|P2)	C6(P1|P2)	C7(P1|P2)
	0.276	0.733	0.918	0.223	**0.002**	0.289	0.679

Table 4.3: Bonferroni-corrected p-values for pairwise comparisons of the object level balance ratings of the onsite listening experiment for the listening positions, P1 and P2. Values below 0.05 are denoted as significant.

The median values show that most participants identified the hidden reference C1 with a median of 100 and hidden anchor C7 with a median of 0, confirming the validity of the experimental design for the object level balance. Configurations C1-C3 do not show position-dependent mean assessments, meeting the prior expectations. The median ratings for references C1-C2 containing Syva loudspeakers in the front responsible for direct sound objects show values above 98 and $p > 0.2$ when comparing C1 with C2, indicating no significant differences, and the preservation of object level balance seems to be unaffected by the side and rear loudspeaker type for those layouts. The X8 reference C3 is rated significantly lower ($p < 0.01$, see Table 4.2) compared to the first two C1 and C2, which could be explained by the stronger acoustical excitation of the room due to the directivity of X8. The directionality decreases towards lower frequencies, leading to higher reflection and reverberation levels (overall room gain). The direct sound of the objects containing low-frequency content was emphasized, diminishing the intended level balance, e.g., the shells from the drum kit, the celli, and double basses are perceived louder, which implies an object level change for C3.

The high median ratings of > 89 at P1 for the realistic application scenarios C4 and C5 consisting of Syva loudspeakers in the front exemplify their benefit for the level balance, particularly for the direct sound objects, due to their better coverage. Compared with the idealized reference scenarios C1 and C2, no significant differences are observed at P1 ($p > 0.05$). Regarding P2, significant differences are determined between the idealized and realistic scenarios ($p < 0.05$). As Syva does not exhibit constant SPL over the listening area, position dependencies are observable in C4 and C5, as P1 is rated better in each case. The location variations, with a median variability between 64 and 97, are more prominent for the combined configuration C5 of Syva in the front and X8 in the side and rear positions. Nevertheless, C5 shows the highest median rating of 97 at P1 for the realistic configurations, even higher than the configuration C4, which consists only of Syva line sources and is less position-dependent. Besides the hidden anchor C7, configuration C6, which consists of only X8 point sources, has the lowest median ratings, slightly above 51, for both P1 and P2. It is particularly striking that the median of the ratings shows no position dependencies. However, the 95% CI of a length of 35 rating points at P1 indicates a strong variability of the ratings. The dominant sources at FC and FR could explain why their level relation does not change between P1 and P2, and FL is distant anyway.

4.1.3. Spatial definition (onsite)

The median rating, 95% CI, and individual ratings of the spatial definition from the onsite experiment are displayed in Figure 4.2. To prove the significance between ratings, Bonferroni-corrected p-values from a Wilcoxon signed rank test for key-layout pairs are displayed in Table 4.4. Moreover, the position dependence is tested individually with p-values comparing P1 and P2 for each layout, as shown in Table 4.5. Most participants found the hidden reference and anchor in the median, indicating a meaningful experimental design.

The median spatial definition ratings of > 95 for the Syva reference C1 and the mixed reference C2 are significantly higher ($p < 0.05$, see Table 4.4) than the median rating of the X8 reference C3, showing median values between 59 and 75. Between C1 and C2, both with Syva loudspeakers in the front, no significant difference ($p > 0.8$) is observed. By comparing the positions P1 and P2 for the C1-C3 references unexceptionally, a slight position dependency is noticeable, which is barely not significant for C3 ($p = 0.074$). Even with directional equality of level and magnitude, the increasing distance to the front sources at position 2 could deteriorate the ratings slightly, as the perception of spatial definition also implies the distance impression. The acoustical excitation of the room additionally affects the spatial definition at P2.

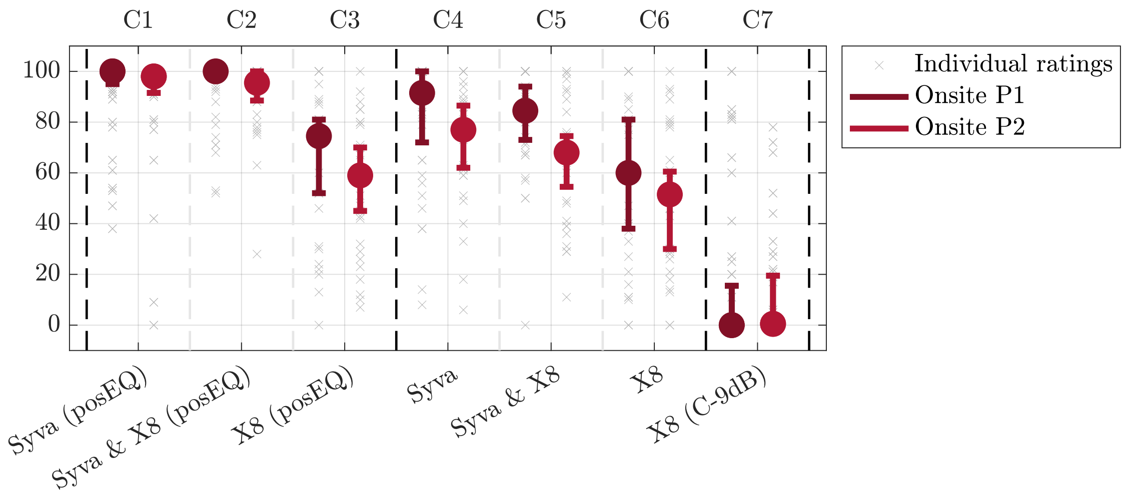


Figure 4.2: Onsite spatial definition ratings for the two listening locations. The dots indicate medians; vertical lines are the 95% CI for both listening locations. On the left are the three idealized references C1-C3, in the middle three realistic configurations C4-C6, and on the right is the anchor C7.

p	C1|C2	C1|C3	C1|C4	C2|C5	C3|C6	C4|C5	C4|C6	C5|C6
P1	0.823	**0.006**	0.407	0.369	0.375	0.861	**0.004**	**0.025**
P2	0.858	**0.001**	**0.031**	**< 0.001**	0.540	0.548	**0.016**	**0.049**

Table 4.4: Bonferroni-corrected p-values from a two-sided Wilcoxon signed rank test for pairwise comparisons of the loudspeaker layouts for the spatial definition ratings. Values below 0.05 are denoted as significant.

p	C1(P1|P2)	C2(P1|P2)	C3(P1|P2)	C4(P1|P2)	C5(P1|P2)	C6(P1|P2)	C7(P1|P2)
	0.689	0.891	0.074	0.156	**0.007**	0.482	0.055

Table 4.5: Bonferroni-corrected p-values for pairwise comparisons of the spatial definition ratings of the onsite listening experiment for the listening positions, P1 and P2. Values below 0.05 are denoted as significant.

Comparing the spatial definition median ratings of the realistic configurations C4 and C5 with Syva loudspeakers in front, a slight but not significant ($p > 0.5$) preference for the layout C4, that has also line sources in the side and rear is observed at both P1 and P2. The Syva deployment C4 shows median values of 77 for P2 and 91.5 at P1, while the combined configuration C5 is rated by median values between 68 and 84.5. The combined configuration C5 with point sources on the side and rear might produce a more blurred and reverberant impression due to their directivity behaviour, explaining the slightly poorer assessments. The layout C6 consisting only of X8 has a median rating of 51 for P2 and 60 for P1, which is significantly worse in spatial definition ($p < 0.05$). Generally, all realistic conditions C4-C6 show lower ratings for P2, which leads to the conclusion that spatial definition degrades when having non-constant decay over distance. Besides a stronger loss in direct sound level over distance, the excitation of the reverberant field might also degrade the ratings. In Table 4.5 only the combined Syva & X8 configuration C5 shows significant position differences.

4.1.4. Spatial envelopment (onsite)

The median spatial envelopment ratings for the hidden reference and anchor, shown in Figure 4.3, also confirm the validity of the experimental design for this criterion. Additionally, the p-values of configuration layout pairs and location dependencies are calculated and displayed in Table 4.6 and Table 4.7. For this attribute, the reference conditions C1-C3 employ position-specific correction (see Table 3.2 and Table 3.3 in Section 3.2.1) to an ideal -3 dB/dod decay.

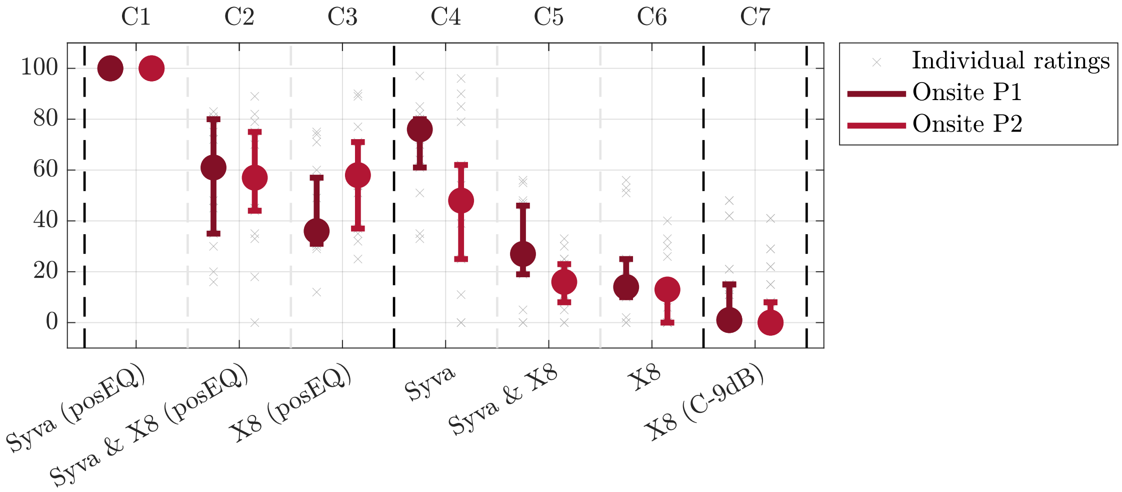


Figure 4.3: Onsite spatial envelopment ratings for the two listening locations. The dots indicate medians; vertical lines are the 95% CI for both listening locations. On the left are the three idealized references C1-C3, in the middle three realistic configurations C4-C6, and on the right is the anchor C7.

p	C1|C2	C2|C3	C1|C4	C2|C5	C3|C6	C4|C5	C4|C6	C5|C6
P1	**0.001**	**0.032**	**0.001**	**0.001**	**0.003**	**0.002**	**0.001**	**0.014**
P2	**0.001**	0.872	**0.001**	**0.001**	**0.001**	**0.021**	**0.001**	0.527

Table 4.6: Bonferroni-corrected p-values from a two-sided Wilcoxon signed rank test for pairwise comparisons of the loudspeaker layouts on spatial envelopment ratings. Values below 0.05 are denoted as significant.

p	C1(P1|P2)	C2(P1|P2)	C3(P1|P2)	C4(P1|P2)	C5(P1|P2)	C6(P1|P2)	C7(P1|P2)
	1.000	0.552	**0.027**	**0.035**	**0.004**	0.141	0.383

Table 4.7: Bonferroni-corrected p-values for pairwise comparisons of the spatial envelopment ratings of the onsite listening experiment for the listening positions, P1 and P2. Values below 0.05 are denoted as significant.

In contrast to the median ratings for the object level balance and spatial definition, it stands out that the hidden reference layout C1 is rated exceptionally high for spatial envelopment with significant differences to all other layouts (see Table A.9 and Table A.10 in Section A.2). C1 shows a median rating of 100. In contrast, the medians of the other reference, C2 and C3, lie between 36 and 61, showing significant differences against the reference at both P1 and P2. Additionally, a significant difference between C2 and C3 is only observed for P1 ($p = 0.032$). Due to the importance of all surrounding loudspeakers for the spatial envelopment, conditions C2 and C3 with point sources in the rear and side locations introduce higher and non-uniform acoustical excitation, as the distances to back-walls are different, causing a more sensitive decrease of envelopment even if they are compensated to exhibit $-3\,\mathrm{dB/dod}$ at the specific position. The distinct tonal colouration, which made detecting the hidden reference easier, also resulted in a higher variability for the other configurations, as the assessing focus appeared to have varied between the tonal similarity to the reference and the actual sensation of uniform envelopment. Position-related differences are only significant for the idealized reference X8 configuration C3 (p=0.027).

When evaluating the results of the realistic configurations, condition C4 with Syva line sources in the side and rear shows significantly ($p < 0.02$) better ratings for the envelopment, especially at P1. The long 95% CI for the C4 layout of 37 rating points at P2 can be explained by the difficulty of solely rating the envelopment against a reference and ignoring timbral changes between the configurations. Except for the X8 layout C6 ratings, all realistic conditions C4-C6 show ratings that significantly prefer P1 ($p < 0.04$). Relative to the Syva condition C4, significant disadvantages ($p < 0.03$) in achieving uniform envelopment arise when point sources are used for the side and rear positions in C5 and C6, as reflected by median ratings below 28 points. Due to the SPL loss over distance for the X8 point source ($-5\,\mathrm{dB/dod}$), uncorrelated pink noise playback is likely localized from the direction of the closest loudspeaker. The non-significant position dependence ($p(\mathrm{C6}(\mathrm{P1}|\mathrm{P2})) = 0.414$ in Table 4.7) of the X8 configuration C6 could explained by the anyhow low ratings, and the longer 95% CI at P2 indicates rating uncertainties.

4.2. Experiment repeated on headphones

In the headphone experiment, 18 listeners participated (students and academic staff from the IEM), who had relevant experience in immersive audio playback systems. Two assessors were excluded from the results as they rated the hidden reference less than 90 in more than 15% of the trials, cf. [36]. The ages of three female and 15 male participants varied between 22 and 44, with an average age of 29. The average duration for the experiment was 42 minutes.

4.2.1. Object level balance (headphones)

The object level balance median and 95% CI ratings of the headphone experiment are displayed in Figure 4.4. Position 1 (P1) is illustrated with a darker blue, while a brighter colour was used for position 2 (P2). The three conditions on the left are idealized references with position-specific gains and equalization. In the middle, C4-C6 describe the realistic application scenarios. The right section shows the ratings for the hidden anchor C7. The significance of p-values of exclusive configuration and position pairs are collected in Table 4.8 and Table 4.9, where significant levels are highlighted in bold font.

The median ratings of the object level balance indicate that most participants found the hidden reference (median > 92) and anchor (median < 4), with slight uncertainties at P2. However, the applicability of the experimental design is undoubtedly approved. The median ratings of the object level balance concerning the other reference configurations C2 and C3 are significantly ($p < 0.05$) lower than the reference layout C1, which consists only of Syva loudspeakers. Respectively, both C2 and C3 show higher ratings at P2, however not significantly, as shown in Table 4.9.

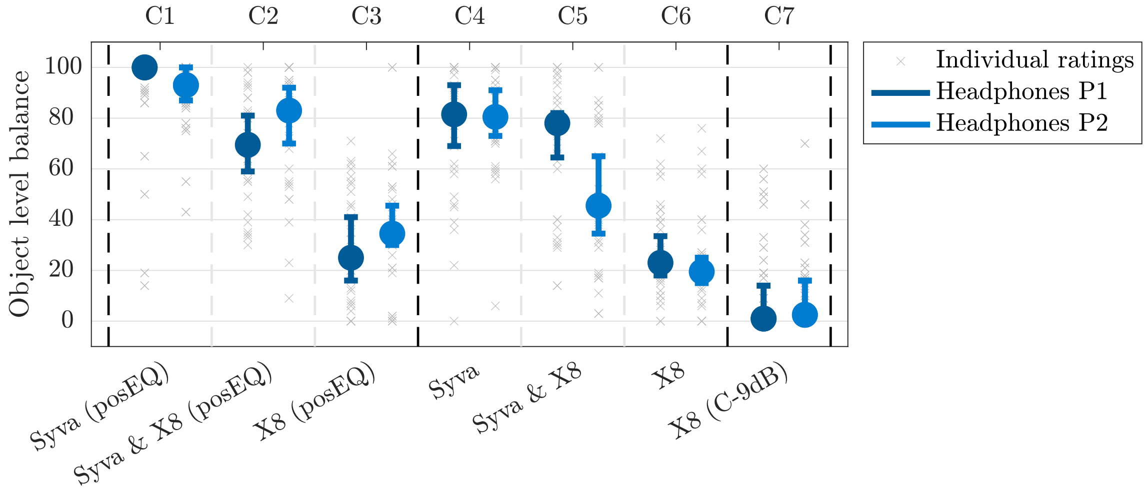


Figure 4.4: Object level balance ratings for the two listening locations of the headphone experiment. The dots indicate that medians and vertical lines are the 95% CI for both listening locations. On the left are the three idealized references C1-C3, in the middle three realistic configurations C4-C6, and on the right is the anchor C7.

p	C1|C2	C1|C3	C1|C4	C2|C5	C3|C6	C4|C5	C4|C6	C5|C6
P1	**0.003**	$<$ **0.001**	0.063	0.881	1.422	0.583	$<$ **0.001**	$<$ **0.001**
P2	0.055	$<$ **0.001**	0.050	**0.003**	**0.043**	$<$ **0.001**	$<$ **0.001**	**0.003**

Table 4.8: Bonferroni-corrected p-values from a two-sided Wilcoxon signed rank test for pairwise comparisons of the headphone experiment loudspeaker layouts on object level balance ratings. Values below 0.05 are denoted as significant.

p	C1(P1|P2)	C2(P1|P2)	C3(P1|P2)	C4(P1|P2)	C5(P1|P2)	C6(P1|P2)	C7(P1|P2)
	0.409	0.102	0.282	0.787	$<$ **0.001**	0.229	0.756

Table 4.9: Bonferroni-corrected p-values for pairwise comparisons of the object level balance ratings of the headphone listening experiment for the listening positions, P1 and P2. Values below 0.05 are denoted as significant.

The median ratings for the object level balance of the realistic configurations show the highest values of about 81 for the layout C4 consisting only of Syva loudspeakers. Additionally, no significant ($p = 0.787$) position dependence is observed for C4. At P1, the combined configuration C5 of Syva and X8 is rated comparable (no significant difference, $p = 0.583$) to C4 with a median of 78. A significant difference ($p < 0.001$) between C4 and C5 is observed at P2, rated with a median of 45.5, indicating a disadvantageous influence for the object level balance of point sources in the side and rear. Besides the hidden anchor C7, the X8 layout C6 scored the worst, with median ratings of 19.5 at P2 and 23 at P1. The short 95% CI for both locations indicates a high certainty for the results. Except for C5 ($p = 0.014$), no configuration show significant position differences ($p > 0.22$)

4.2.2. Spatial definition (headphones)

The median ratings with their 95% CI for the spatial definition of the headphone-based repetition of the listening experiment are displayed in Figure 4.5. Table 4.10 and Table 4.11 show a choice of meaningful p-values for configuration and location pairs, with significant levels highlighted in bold font.

As expected, all idealized reference layouts on the left side of the plot show no significant differences ($p > 0.38$) in the median ratings between the two listening positions, P1 and P2. The hidden reference C1 is rated best with a median of 99.5 and 100 at P1 and P2, respectively, and a 95% CI of shorter 10 rating points. The combined reference configuration C2 achieves medians of 78.5 and 78. The 95% CI of C2 is 12 at P1 and 21.5 at P2, indicating a slightly higher uncertainty at the second location. The X8 reference configuration C3 is rated worst, having median results of 38.5 and 40 and 95% CI of 33 at P1 and 24.5 at P2. The significant differences between most of those layouts ($p < 0.053$) are unexpected in the first place. However, they are attributable to the acoustical room excitation, which is more strongly perceived in the headphone conditions of this experiment.

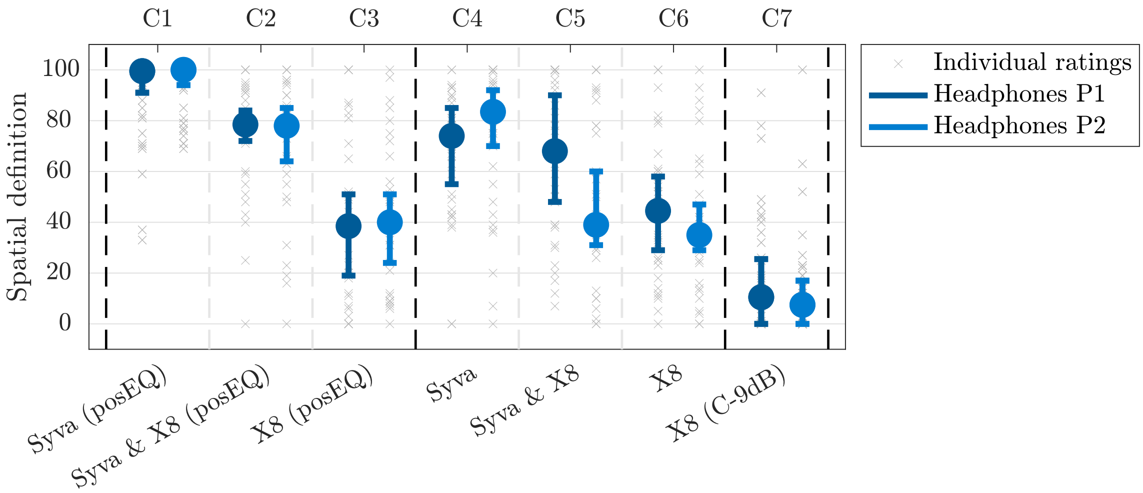


Figure 4.5: Spatial definition ratings for the two listening locations of the headphone experiment. The dots indicate medians; vertical lines are the 95% CI for both listening locations. On the left are the three idealized references C1-C3, in the middle three realistic configurations C4-C6, and on the right is the anchor C7.

p	C1|C2	C1|C3	C1|C4	C2|C5	C3|C6	C4|C5	C4|C6	C5|C6
P1	0.052	< **0.001**	**0.004**	0.743	0.483	0.817	**0.035**	0.115
P2	**0.003**	< **0.001**	**0.012**	0.127	0.963	**0.007**	**0.001**	1.177

Table 4.10: Bonferroni-corrected p-values from a two-sided Wilcoxon signed rank test for pairwise comparisons of the loudspeaker layouts on the headphone experiments spatial definition ratings. Values below 0.05 are denoted as significant.

p	C1(P1|P2)	C2(P1|P2)	C3(P1|P2)	C4(P1|P2)	C5(P1|P2)	C6(P1|P2)	C7(P1|P2)
	0.387	0.737	0.853	0.280	**0.014**	0.308	0.224

Table 4.11: Bonferroni-corrected p-values for pairwise comparisons of the spatial definition ratings of the headphone listening experiment for the listening positions, P1 and P2. Values below 0.05 are denoted as significant.

The realistic application configurations C4-C6 in the middle of the plot show fewer but still observable differences between the loudspeakers used. Significant differences are determined between the C4 Syva layout and both deployments, C5 and C6, having X8 loudspeakers in the side and rear positions ($p < 0.01$) at P1 and P2. The differences are significant when comparing the C4 with C6. At P1, the ratings for C4 and C5 do not show significant differences ($p = 0.817$). The configuration C4, that consists only of Syva line sources, has the highest ratings for the spatial definition, with a median of 83.5 (32 rating points 95% CI) at P2 and 74 (20 rating points 95% CI) at P1. The headphone result for the combined configuration C5 is slightly worse and uncertain, with a 95% CI of 41 rating points and a median of 68 at P1. The same layout at P2 performs even worse, with a median of 39 and a 95% CI of 30 rating points, indicating that the side and rear X8 loudspeakers degrade the spatial definition at further off-center locations. The ratings of C6 with only X8 loudspeakers show median ratings of 44.5 at P1 and 35 at P2, with a 95% CI of 28.5 and 15.5 rating points. The hidden anchor C7 was rated with median values of 10.5 and 7.5, higher than expected, but also showed higher uncertainties due to a 95% CI of 26 and 17 rating points at P1 and P2.

4.2.3. Spatial envelopment (headphones)

Figure 4.6 illustrates the median and 95% CI for the spatial envelopment ratings of the headphone experiment. On the far left of the plot, median ratings of 100 for the idealized reference configuration C1, which consists of only Syva loudspeakers, demonstrate that most participants found the hidden reference. The median ratings for the hidden anchor C7 on the right are similar to the realistic application scenario configurations C5 and C6, indicating difficulty detecting the anchor on headphones, which is confirmed by the p-values above $p > 0.25$ (see Table A.18 and Table A.19). A selection of p-values of configuration pairs and between the listening positions are shown in Table 4.12 and Table 4.13, where values of $p < 0.05$ are denoted as significant and highlighted in bold font.

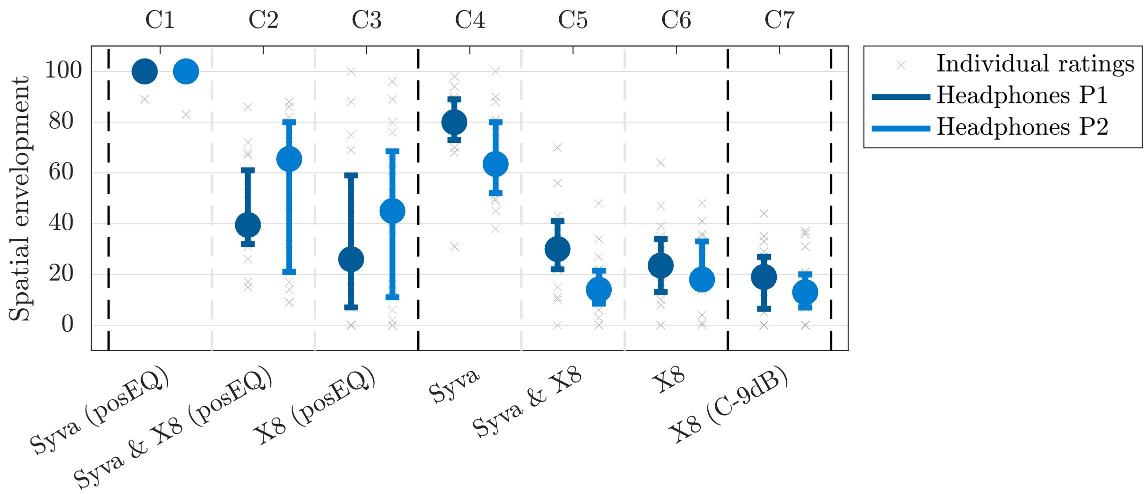


Figure 4.6: Spatial envelopment ratings for the two listening locations of the headphone experiment. The dots indicate that medians and vertical lines are the 95% CI for both listening locations. On the left are the three idealized references C1-C3, in the middle three realistic configurations C4-C6, and on the right is the anchor C7.

p	C1|C2	C2|C3	C1|C4	C2|C5	C3|C6	C4|C5	C4|C6	C5|C6
P1	**0.006**	0.435	**0.008**	0.697	1.554	**0.008**	**0.007**	0.619
P2	**0.008**	0.070	**0.010**	**0.045**	0.265	**0.007**	**0.007**	0.578

Table 4.12: Bonferroni-corrected p-values from a two-sided Wilcoxon signed rank test for pairwise comparisons of the loudspeaker layouts on the headphone experiments spatial envelopment ratings. Values below 0.05 are denoted as significant.

p	C1(P1|P2)	C2(P1|P2)	C3(P1|P2)	C4(P1|P2)	C5(P1|P2)	C6(P1|P2)	C7(P1|P2)
	1.000	0.162	0.203	**0.046**	**0.004**	0.660	0.511

Table 4.13: Bonferroni-corrected p-values for pairwise comparisons of the spatial envelopment ratings of the headphone listening experiment for the listening positions, P1 and P2. Values below 0.05 are denoted as significant.

The hidden reference C1 is rated significantly higher than the other idealized reference configurations C2 and C3 ($p < 0.01$) at both positions, P1 and P2. Besides C1, the configurations C2 and C3 have a high variability in the ratings, clarified by the long 95% CIs of up to 59 rating points. The timbral disparities and individually conditioned binaural playback could explain the high variance of those ratings, while the hidden reference C1 was easily found compared to the given reference. However, the combined configuration C2 was rated slightly better in the median than the X8 configuration C3. No significant differences are observed between C2 and C3 at positions P1 and P2 ($p > 0.05$). P2 was rated better for both layouts, with a median of 65.5 for the combined C2 configuration and 45 for the C3 configuration that consists of only X8 loudspeakers. The C2 configuration shows a median of 39.5 at P1, and C3 has a median rating of 26 at P1. The p-values above $p > 0.15$ indicate no significant difference between the positions P1 and P2 for all idealized reference configurations C1-C3.

The realistic application scenario configurations C4-C6 in the middle of the plot generally show a higher certainty for the spatial envelopment ratings with a maximal length of 95% CI with 28 rating points. The configuration C4, which consists only of Syva, shows the highest median ratings for the spatial envelopment, with 80 at P1 and 63.5 at P2, and significant differences towards the other realistic configurations C5 and C6 are observed with p-values below $p < 0.01$. Configurations C5 and C6, with X8 point sources in the side and rear, have very low and similar median ratings with no significant differences, indicated by p-values of $p > 0.25$. The combined configuration C5 of Syva and X8 loudspeakers show median ratings of 30 at P1 and 14 at P2. The X8 configuration C6 has a median of 23.5 at P1 and 18 at P2, with 95% CIs of 21 and 19 rating points, respectively. As previously mentioned, the hidden anchor C7 shows similar median ratings of 19 and 13 at P1 and P2, respectively. Significant position-related differences within a configuration are only observed for the configurations C4 and C5 with $p < 0.05$.

4.3. Comparison of onsite and headphone experiment

This section compares the results of the onsite loudspeaker experiment from London with the headphone experiment from Graz. All three rating criteria, object level balance, spatial definition, and spatial envelopment, are contrasted separately in conjunct plots for simplified visibility. The plot structure and colouration are the same as in previous illustrations. The red median and 95% CI represent the ratings from the onsite experiment, while the blue versions are the results from the headphone experiment. The darker colour indicates listening position P1, and the brighter one indicates P2. Two correlation coefficients R and $\hat{R}$ give the coherence between the two experiments, while R contains the correlation of all stimuli and $\hat{R}$ illustrates the correlation of the hidden reference C1, three realistic configurations C4-C6 and the hidden anchor C7.

4.3.1. Comparison of object level balance

In Figure 4.7, the object level balance ratings (median and 95% CI) from both experiments are illustrated. The overall trend of the median ratings resembles both experiments reasonably well. Overall, the headphone experiment ratings are mostly lower than the onsite experiment, except for the hidden reference C1 and anchor C7 and the realistic application scenario configuration C4, that consists of only Syva. The position differences of the ratings reproduce well in the headphone experiment, except for the idealized configurations C2 and C3. The inequalities concerning position differences and lower ratings of those positional equalized of the headphone experiment could explained by the successively applied filters (position correction-, diffuse-field-, and headphone-equalization-filters), which modifies the audible frequency response and possibly affect the natural perception.

The lower ratings of the binaural headphone experiment could be attributable to the higher amount of perceived reverberation in static binaural playback [53, chp. 16]. Sound reflections from room boundary surfaces are perceived more explicitly when omitting head rotations. Configurations with point sources C2, C3, C5, and C6 show stronger room excitation (reflections and reverberation), possibly changing the direct sound balance, resulting in higher deviations of the reference level balance.

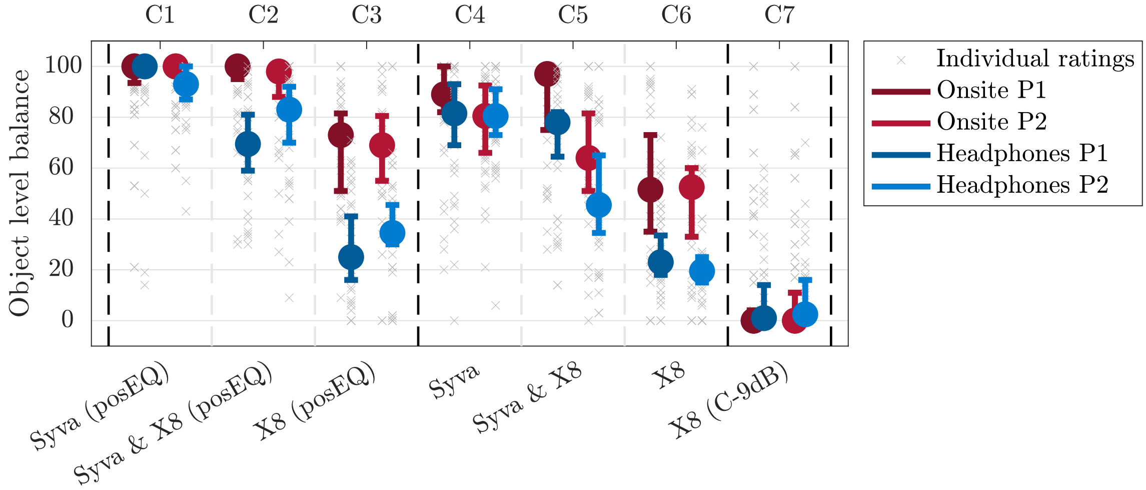


Figure 4.7: Object level balance median ratings with 95% CI from the onsite experiment (red) and headphone experiment (blue).

The overall correlation coefficients of the median ratings for the object level balance between the two experiments of all configurations C1-C7 are collected in Table 4.14. The R-values for both positions P1 and P2 combined and for the individual positions are displayed, once excluding the idealized reference configurations C2 and C3 indicated by the $\hat{R}$ and secondly calculated by considering all configurations C1-C7.

$\hat{R}_{\text{OLB}}$	$\hat{R}^{\text{P1}}_{\text{OLB}}$	$\hat{R}^{\text{P2}}_{\text{OLB}}$	R_{OLB}	$R^{\text{P1}}_{\text{OLB}}$	$R^{\text{P2}}_{\text{OLB}}$
0.94	0.95	0.93	0.89	0.88	0.91

Table 4.14: Correlation coefficients between both experiments for the object level balance. $\hat{R}$-values consider only the hidden reference, anchor, and the three uncompensated layouts, while R considers all layouts. Positional correlations are indicated with superscripts.

Excluding the previously mentioned questionable idealized configurations C1-C3 for binaural playback, the overall correlation of $\hat{R}_{\text{OLB}} = 0.94$ shows a strong correlation between the two experiments for C1 and the realistic configurations C4-C7. Observing the individual listening location, P1 correlates slightly stronger with $\hat{R}^{\text{P1}}_{\text{OLB}} = 0.95$ compared to $\hat{R}^{\text{P2}}_{\text{OLB}} = 0.93$ for P2. Including all configurations C1-C7, the overall correlation is less but still strong with $R_{\text{OLB}} = 0.89$.

4.3.2. Comparison of spatial definition

To compare the median and 95% CI of the spatial definition results from the onsite and headphone experiment, both are contrasted in Figure 4.8. A substantial resemblance of both experiments can be observed when considering only the median values. Similar to the object level balance, the ratings from the headphone experiments are lower on average. The position differences for the individual configurations are reproduced in the headphone experiment except for idealized reference configuration C3, which consists solely of X8 and the realistic Syva configuration C4. The most visible differences are in idealized configurations C2 and C3, which could be explained by the extensive filtering procedure of those configurations. The hidden reference C1 is rated high in the headphone experiment, as most participants indicated this hidden reference, possibly also regarding tonal similarities.

The realistic application scenario configurations C5 and C6, which have X8 in the rear and side, generally perform worse in the headphone experiment, which could also be attributable to the static binaural playback. The higher median ratings of the hidden anchor C7 in the headphone experiment compared to the onsite experiment are particularly noticeable.

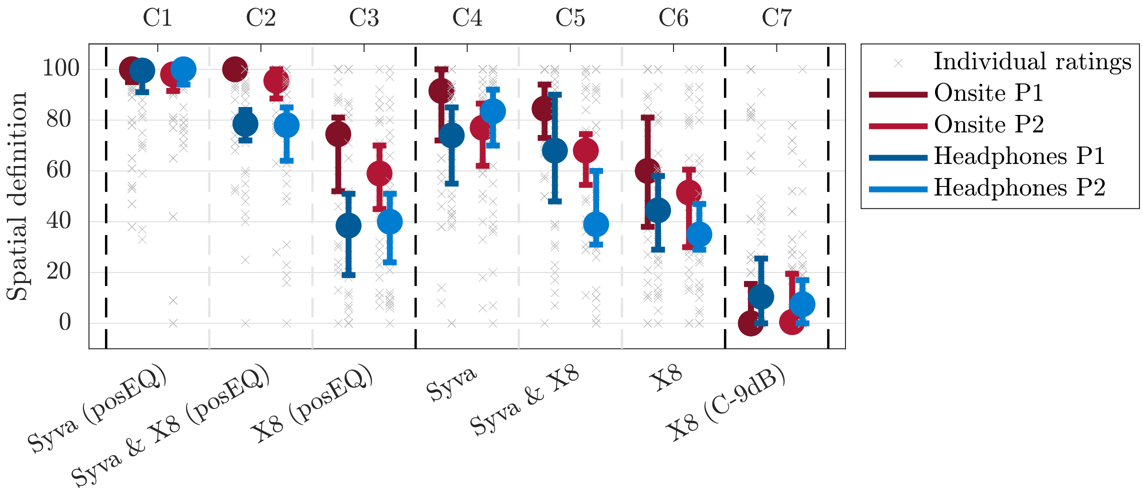


Figure 4.8: Spatial definition median ratings with 95% CI from the onsite experiment (red) and headphone experiment (blue).

The correlation coefficients for the spatial definition are gathered in Table 4.15. The $\hat{R}_{\text{SD}} = 0.93$ reveals a strong correlation for the spatial definition between both experiments, only considering the configurations C1 and C4-C7. Listening position P1 shows a higher correlation of $\hat{R}_{\text{SD}}^{\text{P1}} = 0.96$ compared to P2 with $\hat{R}_{\text{SD}}^{\text{P2}} = 0.91$. Considering all configurations C1-C7, the correlation shows a $R_{\text{SD}} = 0.90$, while the individual positions P1 and P2 have almost indistinguishable values.

$\hat{R}_{\text{SD}}$	$\hat{R}_{\text{SD}}^{\text{P1}}$	$\hat{R}_{\text{SD}}^{\text{P2}}$	R_{SD}	$R_{\text{SD}}^{\text{P1}}$	$R_{\text{SD}}^{\text{P2}}$
0.93	0.96	0.91	0.90	0.91	0.90

Table 4.15: Correlation coefficients between both experiments for the spatial definition. $\hat{R}$-values consider only the hidden reference, anchor, and the three uncompensated layouts, while R considers all layouts. Positional correlations are indicated with superscripts.

4.3.3. Comparison of spatial envelopment

The median spatial envelopment ratings with 95% CI of the onsite and headphone experiment are compared in Figure 4.9. The similarity of the ratings of both experiments is visible, with a minor deviation of the idealized reference configurations C2 and C3 and the hidden anchor C7. The realistic application scenario configurations C4-C6 show slightly higher ratings in the headphone experiment compared to the onsite experiment. Perceiving more reverb in static binaural playback could explain the more uniform sensation of envelopment in the headphone experiment. For the realistic configurations C4-C6, the position differences in the headphone experiment match the median ratings of the onsite experiment.

Due to excessive filtering operations, the long 95% CI for the idealized configurations C2 and C3 underlines previous observations of uncertain ratings for those layouts. However, the overall trend of those layouts reproduces the onsite ratings reasonably well. A noticeable difference between both experiments concerning the hidden anchor C7 is observed. The higher ratings for C7 in the headphone experiment could be attributable to the more inaccurate discoverability of the attenuated center source due to higher perceived reverb and non-individual binaural reproduction.

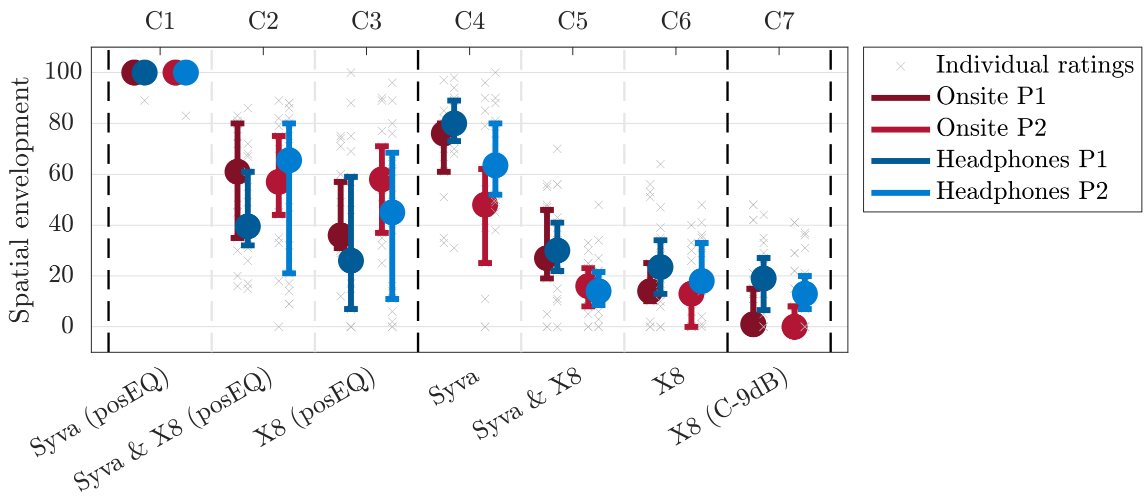


Figure 4.9: Spatial envelopment median ratings with 95% CI from the onsite experiment (red) and headphone experiment (blue).

The correlation coefficients between the onsite and headphone experiment, gathered in Table 4.16, show exceptional values of > 0.93. Especially when omitting the two idealized reference configurations C2 and C3, the consolidated coefficient indicated a robust correlation of $\hat{R}_{\mathrm{SE}} = 0.99$. Almost no differences in the correlation coefficient are visible between the two listening positions, P1 and P2. Including all configurations in the calculation of R, a correlation between both experiments of $R_{\mathrm{SE}} = 0.94$ is achieved, while the second listening position P2 correlates stronger with $R_{\mathrm{SE}}^{\mathrm{P2}} = 0.96$ compared to P1 with $R_{\mathrm{SE}}^{\mathrm{P1}} = 0.93$.

$\hat{R}_{\mathrm{SE}}$	$\hat{R}_{\mathrm{SE}}^{\mathrm{P1}}$	$\hat{R}_{\mathrm{SE}}^{\mathrm{P2}}$	R_{SE}	$R_{\mathrm{SE}}^{\mathrm{P1}}$	$R_{\mathrm{SE}}^{\mathrm{P2}}$
0.99	0.99	0.98	0.94	0.93	0.96

Table 4.16: Correlation coefficients between both experiments for the spatial envelopment. $\hat{R}$-values consider only the hidden reference, anchor, and the three uncompensated layouts, while R considers all layouts. Positional correlations are indicated with superscripts.

4.4. Ratings relation to acoustic metrics

The acoustic metrics presented in Section 2.5 aim to reproduce the ratings for the queried criteria of the experiment. Under the assumption of the previously presented similarity of both experiments, only the metric relation to the onsite experiment is examined. The regression results for the headphone experiment are collected in Section A.5. The linear regression of the calculated metrics towards the subject ratings facilitates reasonable comparisons.

4.4.1. Relation of object level balance to front-to-surround ratio

The front-to-surround (FS) ratio compares the minimum A-weighted direct sound level from the three front loudspeakers to the maximum A-weighted level from all deployed loudspeakers, giving a measure of how the sound delivered from front sources contrasts with sound arriving from the rear and side directions. The linear regression $\widetilde{\mathrm{FS}} = \alpha_{\mathrm{FS}}\ \mathrm{FS} + \beta_{\mathrm{FS}}$ map the FS ratio values towards the experiment median ratings for the object level balance for each position individually, as no direct relation of both ratings is assumed. The parameters for positions P1 and P2 are displayed in Equation (4.1)

$$\begin{aligned} \widetilde{\mathrm{FS}}^{\mathrm{P1}} &= \alpha_{\mathrm{FS}}^{\mathrm{P1}}\ \mathrm{FS} + \beta_{\mathrm{FS}}^{\mathrm{P1}} = -8.86\ \mathrm{FS} + 90.26\ , \\ \widetilde{\mathrm{FS}}^{\mathrm{P2}} &= \alpha_{\mathrm{FS}}^{\mathrm{P2}}\ \mathrm{FS} + \beta_{\mathrm{FS}}^{\mathrm{P2}} = -6.73\ \mathrm{FS} + 74.36\ , \end{aligned} \tag{4.1}$$

where $\widetilde{\mathrm{FS}}$ are the regressed FS ratio, α_{FS} and β_{FS} the corresponding regression parameters. As the idealized configuration negatively impacts the correlation of $\widetilde{\mathrm{FS}}$ with OLB ratings, the plotted regression is restricted to the realistic application scenario configurations C4-C7.

In Figure 4.10 the regression results are contrasted to the onsite ratings. The overall trend of $\widetilde{\mathrm{FS}}$ shows similar behavior as the onsite ratings for the object level balance. The position differences of the metric are more distinct and indicate a higher dependence on the distance from the origin, except for the hidden anchor configuration C7. The metric values for the X8 configuration C6 are slightly higher compared to the median rating from the onsite experiment but still lie within the CIs.

The correlation coefficients between the metrics and ratings are collected in Table 4.17. Considering a regression with all configurations, the correlation is $R_{\mathrm{FS}} = 0.94$. Position P1 shows a stronger correlation of $R^{\mathrm{P1}}_{\widetilde{\mathrm{FS}}} = 0.97$ compared to P2 ($R^{\mathrm{P2}}_{\widetilde{\mathrm{FS}}} = 0.92$). When observing only the plotted regression values of realistic configurations, the correlation is $\hat{R}_{\widetilde{\mathrm{FS}}} = 0.97$. The individual correlations for the two listening positions, P1 and P2, show the values as the combined correlation.

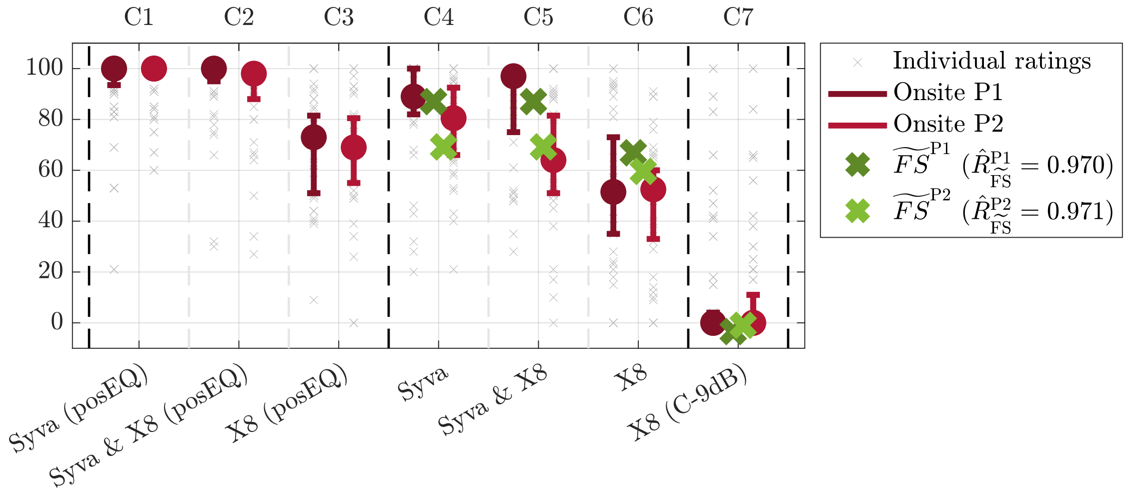


Figure 4.10: Comparison of the linear regression results of the FS ratio for configurations C4-C7 with the object level balance ratings from the onsite experiment. The darker green cross indicates P1, and the brighter cross marks P2.

$\hat{R}_{\widetilde{\mathrm{FS}}}$	$\hat{R}^{\mathrm{P1}}_{\widetilde{\mathrm{FS}}}$	$\hat{R}^{\mathrm{P2}}_{\widetilde{\mathrm{FS}}}$	$R_{\widetilde{\mathrm{FS}}}$	$R^{\mathrm{P1}}_{\widetilde{\mathrm{FS}}}$	$R^{\mathrm{P2}}_{\widetilde{\mathrm{FS}}}$
0.97	0.97	0.97	0.94	0.97	0.92

Table 4.17: Correlation coefficients between the onsite experiment results and the linear regression of the DD ratio for the object level balance. $\hat{R}$-values consider configurations C4-C7, while R considers C1-C7.

Examining the acoustic metric within the audience area, RIRs are measured in a 8 m × 6 m grid with 1 m discretization (illustrated in Figure 3.7 in Section 3.2.1), and the FS ratio is calculated at each grid point. Figure 4.11 shows the normalized FS ratio for the realistic configuration C4, which consists solely of Syva on the left and configuration C6 on the right.

The normalized FS ratio of the C4 configuration shows a large area within the FS ratio that does not exceed −3 dB. In the far back, this −3 dB value is not even exceeded throughout the entire width. About half of the audience lies within −3 dB of this normalized ratio, indicating a stable balance between panned direct sound objects in the front. The ratio decreases towards a maximum of −9 dB in the frontal left and right corner sections. In this region, the dominance of the opposing left or right frontal speaker degrades the ratio, as its direct sound contribution is less than that of the nearby source on the side.

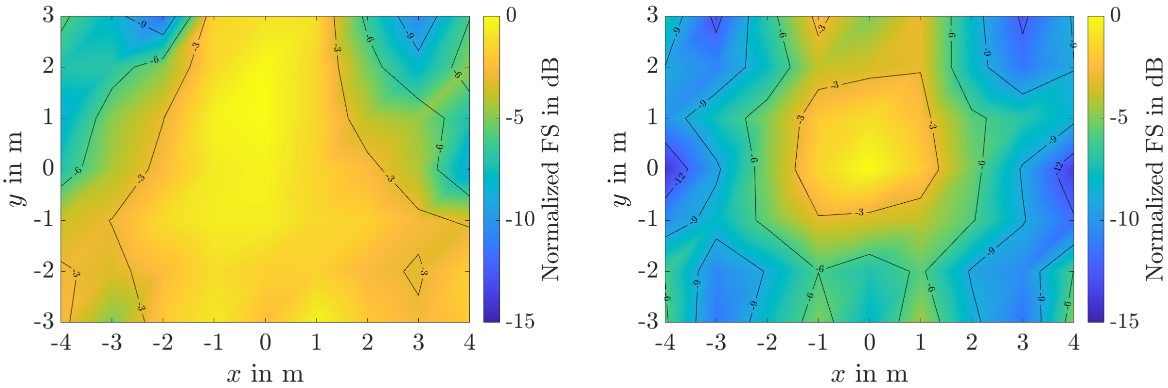


Figure 4.11: Normalized FS ratio for the majority of the audience area (8 m × 6 m) for the Syva configuration C4 (left) and X8 configuration C6 (right).

In contrast, deploying only X8 point sources, as realized in configuration C6, the area of less than $-3\,\mathrm{dB}$ normalized FS ratio is minimized to a region spanning in a radius of 1 m around the sweet spot in the origin compared to configuration C4. The FS ratio for this deployment deteriorates in all directions, indicating an unambiguous level change of frontally panned direct sound objects. This directional direct sound level variance at the off-center positions confirms the degradation of the perceived object level balance when utilizing point sources in such deployment.

4.4.2. Relation of spatial definition to direct-to-diffuse ratio

The direct-to-diffuse (DD) ratio relates the mean A-weighted direct sound level of the three frontal speakers to the mean A-weighted diffuse sound level of the side and rear loudspeakers of the given layout, giving a measure of how much direct sound is delivered from the front loudspeaker compared to the amount of diffuse sound level produced by the side and rear loudspeakers. Comparable to the FS ratio, the DD ratio is adjusted to the spatial definition rating scale using a linear regression for each listening position separately. The corresponding regression parameters α_{DD} and β_{DD} are

$$
\begin{aligned}
\widetilde{\mathrm{DD}}^{\mathrm{P1}} &= \alpha_{\mathrm{DD}}^{\mathrm{P1}}\ \mathrm{DD} + \beta_{\mathrm{DD}}^{\mathrm{P1}} = 37.61\ \mathrm{DD} + 264.19\ , \\
\widetilde{\mathrm{DD}}^{\mathrm{P2}} &= \alpha_{\mathrm{DD}}^{\mathrm{P2}}\ \mathrm{DD} + \beta_{\mathrm{DD}}^{\mathrm{P2}} = 18.45\ \mathrm{DD} + 174.61\ ,
\end{aligned}
\tag{4.2}
$$

resulting in the values $\widetilde{\mathrm{DD}}^{\mathrm{P1}}$ and $\widetilde{\mathrm{DD}}^{\mathrm{P2}}$ for the particular DD ratio. The idealized configurations are likewise excluded from the regression calculation.

The comparison of $\widetilde{\mathrm{DD}}$ with the onsite spatial definition ratings is depicted in Figure 4.12, where a similar tendency is visible. All regressed metric values are located within the 95% CIs, and the position differences are represented well, especially for configurations C4 and C5, having Syva loudspeaker in the front. Except for position P2 of the X8 configuration C6 and anchor C7, the metric matches even the median of the ratings very well. The higher value of the $\widetilde{\mathrm{DD}}$ for C7 at P2 is attributable to the higher impact on the DD ratio for more central locations as the attenuated center degrades the ratio in this region, and outer regions are less influenced.

Considering all configurations C1-C7, the strong correlation of $R_{\mathrm{DD}} = 0.97$ signifies the successful reproduction of the experiment rating with the presented metric. Excluding the idealized configuration ratings from the regression results in a slightly lower correlation of $\hat{R}_{\widetilde{\mathrm{DD}}} = 0.96$. It was observed that position P1 shows better replica of the ratings with $R_{\widetilde{\mathrm{DD}}}^{\mathrm{P1}} = 0.99$ and $\hat{R}_{\widetilde{\mathrm{DD}}}^{\mathrm{P1}} = 0.99$ compared to P2 ($R_{\widetilde{\mathrm{DD}}}^{\mathrm{P2}} = 0.96$ and $\hat{R}_{\widetilde{\mathrm{DD}}}^{\mathrm{P2}} = 0.93$) independent from the considered set of configurations.

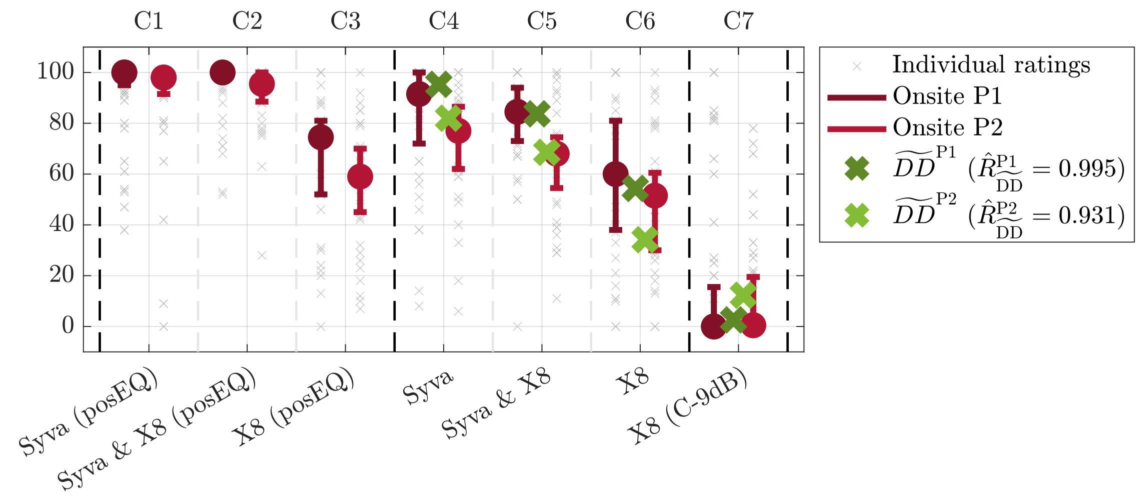


Figure 4.12: Comparison of the linear regression results of the DD ratio for configurations C4-C7 with the spatial definition ratings from the onsite experiment. The darker green cross indicates P1, and the brighter cross marks P2.

$\hat{R}_{\widetilde{DD}}$	$\hat{R}^{P1}_{\widetilde{DD}}$	$\hat{R}^{P2}_{\widetilde{DD}}$	$R_{\widetilde{DD}}$	$R^{P1}_{\widetilde{DD}}$	$R^{P2}_{\widetilde{DD}}$
0.96	0.99	0.93	0.97	0.99	0.96

Table 4.18: Correlation coefficients between the onsite experiment results and the linear regression of the DD ratio for the spatial definition. $\hat{R}$-values consider configurations C4-C7, while R considers C1-C7.

The map data of the normalized DD ratio for the configurations C4, consisting only of Syva (left), and C6, consisting solely of X8 (right), are depicted in Figure 4.13. Conceding the Syva configuration C4, the DD ratio stays within −2 dB in most of the audience area, showing only values below −2 dB at the left and right boundaries of the audience area.

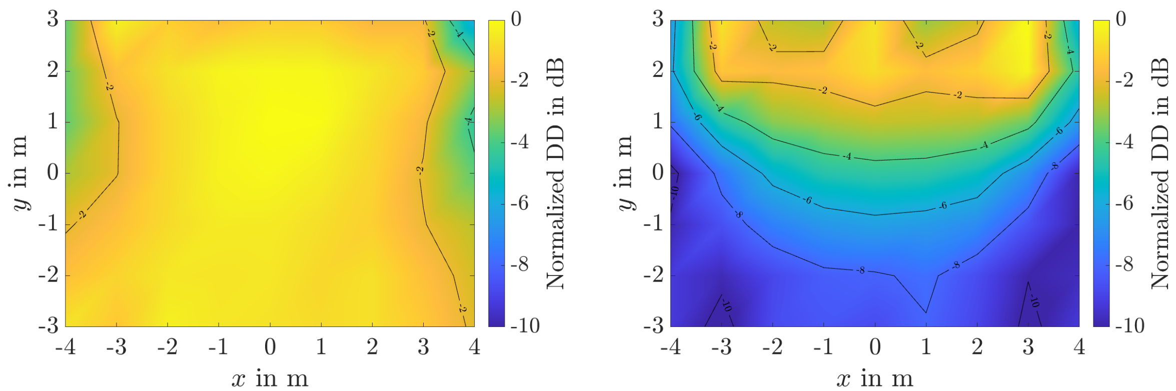


Figure 4.13: Normalized DD ratio for the majority of the audience area (8 m × 6 m) for the Syva configuration C4 (left) and X8 configuration C6 (right).

Comparing the DD ratio values in the audience, when utilizing the C6 configuration—deployment of only X8—, a gradient from the front to the back of the audience area can be observed, with values decreasing from −2 dB to −10 dB. The direct sound decay of the point source is −5 dB per distance doubling, which seems to affect the DD ratio and the diffuse sound component increases already at very short distances to the front loudspeakers. Compared to the Syva C4 configuration, the region within the normalized DD ratio remains −2 dB and does not extend further than 1.5 dB from the frontal sources.

4.4.3. Relation of spatial envelopment to the interaural level difference

In previous studies [18], the interaural level difference (ILD) was utilized as an objective evaluation measure to measure the sensation of uniform envelopment. In this work, the adjusted ILD describes the difference in the A-weighted level between the right and the left ear. Performing a linear regression of the ILD towards the spatial envelopment ratings using the positional independent regression parameters $\alpha_{\text{ILD}}^{\text{P1}}$, $\beta_{\text{ILD}}^{\text{P1}}$, $\alpha_{\text{ILD}}^{\text{P2}}$, $\beta_{\text{ILD}}^{\text{P2}}$ as

$$\begin{aligned} \widetilde{\text{ILD}}^{\text{P1}} &= \alpha_{\text{ILD}}^{\text{P1}}\ \text{ILD} + \beta_{\text{ILD}}^{\text{P1}} = -95.02\ \text{ILD} + 93.72\ , \\ \widetilde{\text{ILD}}^{\text{P2}} &= \alpha_{\text{ILD}}^{\text{P2}}\ \text{ILD} + \beta_{\text{ILD}}^{\text{P2}} = -36.23\ \text{ILD} + 50.04\ , \end{aligned} \tag{4.3}$$

giving $\widetilde{\text{ILD}}^{\text{P1}}$ and $\widetilde{\text{ILD}}^{\text{P2}}$ for further comparisons. As previously mentioned, the regression is only performed for the realistic configurations C4-C7 due to extensive deviations from the idealized configurations. The $\widetilde{\text{ILD}}$ results are compared with the onsite ratings of the spatial envelopment in Figure 4.14.

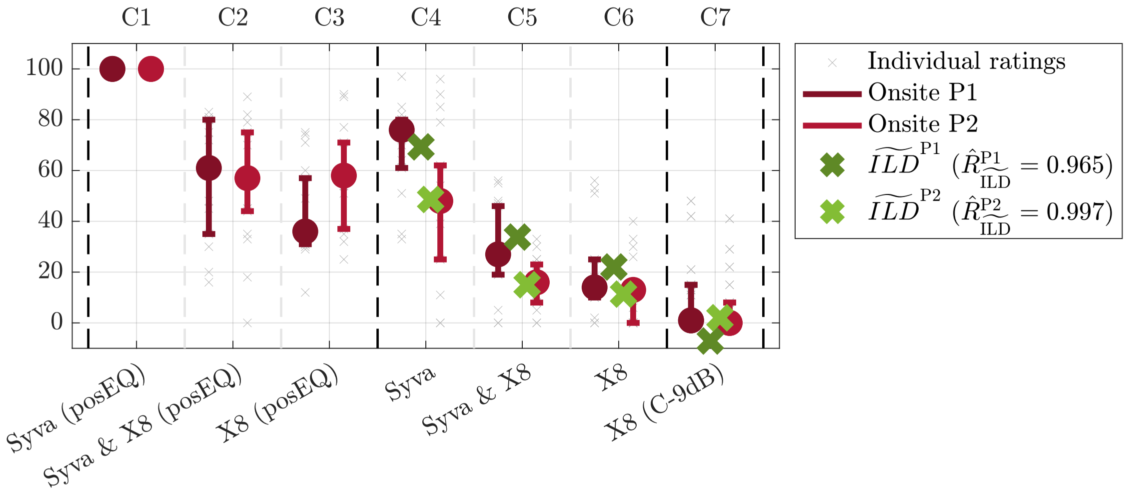


Figure 4.14: Comparison of the linear regression results of the ILD for configurations C4-C7 with the spatial envelopment ratings from the onsite experiment. The darker green cross indicates P1, and the brighter cross marks P2.

Except for P1 in the configuration C6 and C7, all $\widetilde{\text{ILD}}$ values are within the 95% CI of the onsite ratings. The position differences for the configurations C4 and C5 are reflected in the acoustic metrics after the regression. The low values of the metric at P1 for the anchor configuration C7 are explainable by the distinct influence on the uniform directional sound level at P1 due to the attenuated level of the center speaker.

By the exclusion of the idealized configurations for the linear regression, the consolidated correlation between the onsite ratings and adapted ILD metric clarifies the strong correlation with a value $\hat{R}_{\widetilde{\text{ILD}}} = 0.98$. The separate position correlation indicates a slightly better reproduction for P2 with $\hat{R}_{\widetilde{\text{ILD}}}^{\text{P2}} = 0.99$ compared to P1 with a value of $\hat{R}_{\widetilde{\text{ILD}}}^{\text{P1}} = 0.97$. When including the idealized configurations in the regression calculation, the correlation coefficient $R_{\widetilde{\text{ILD}}} = 0.425$ indicates a distinctive degradation of the reproducibility of the ratings.

$\hat{R}_{\widetilde{\text{ILD}}}$	$\hat{R}_{\widetilde{\text{ILD}}}^{\text{P1}}$	$\hat{R}_{\widetilde{\text{ILD}}}^{\text{P2}}$	$R_{\widetilde{\text{ILD}}}$	$R_{\widetilde{\text{ILD}}}^{\text{P1}}$	$R_{\widetilde{\text{ILD}}}^{\text{P2}}$
0.98	0.97	0.99	0.425	0.21	0.64

Table 4.19: Correlation coefficients between the onsite experiment results and the linear regression of the ILD for the spatial envelopment. $\hat{R}$-values consider configurations C4-C7, while R considers C1-C7.

5. Discussion

The discussion of the results of this thesis is divided into parts, corresponding to the three criteria evaluating the ratings from both experiments with the invented acoustic metrics. In the last part, the consolidated ratings give an overview of the performance of the different loudspeaker layouts for immersive surround sound reinforcement.

Object level balance

The experiment ratings of the onsite experiments (red ratings in Figure 5.1) express that loudspeaker layouts containing frontal line sources are beneficial for preserving the direct sound object level balance in a larger audience area. This holds for both idealized and realistic configurations. Comparing the realistic application scenario configuration C4 (solely Syva) with the combined configuration C5 does not significantly benefit the object level balance for positions P1 and P2 ($p > 0.2$, see Table 4.2). Slight position dependencies in the realistic configurations (only significant for the C5 configuration, $p = 0.002$) result from non-constant direct sound decays, explaining more significant differences with point sources in the side and rear positions as those sources' distant dependent direct sound decays are higher compared to line sources. The configurations consisting of point sources do not show any position-dependent differences because the capability of preserving the object level balance at both positions is very low, affected mainly by the left front source FL, and position differences are even less perceivable. Significant deterioration of the object level balance ratings of the idealized reference X8 configuration C3 compared to the C1 Syva deployment are determined with $p < 0.05$ for both P1 and P2. This can be explained by their stronger excitation of the reverberant field (especially for low frequencies), which interferes with the direct sounds and alters the intended mix balance.

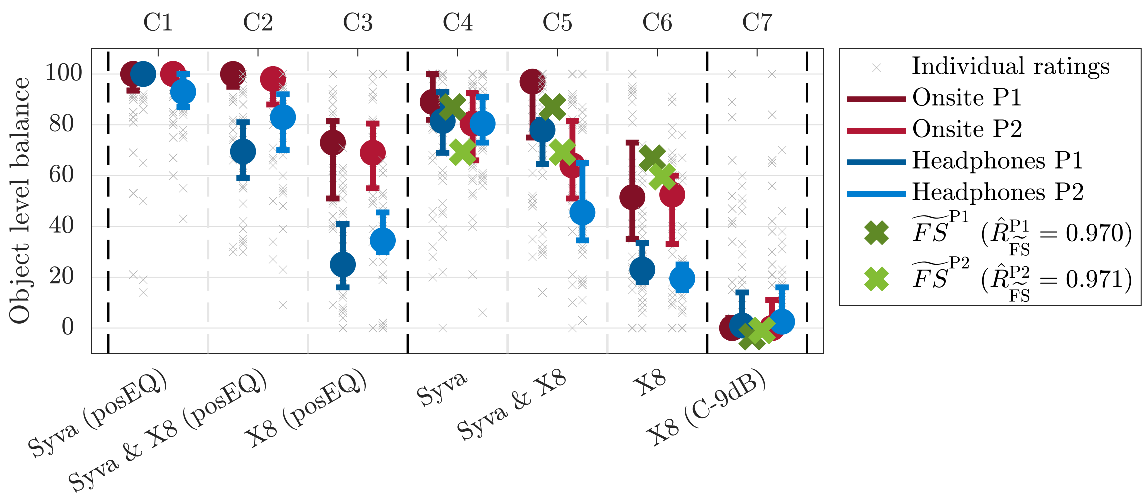


Figure 5.1: Compiled results of ratings from the onsite and virtual listening experiment for the criterion object level balance, related to the introduced metric FS ratio of the direct sound which correlates with $\hat{R}_{\widetilde{\mathrm{FS}}} = 0.97$ with the onsite ratings.

By comparing the object level balance of the onsite experiment (red) with the headphone experiment (blue), a proper agreement of the ratings with a correlation of $R_{\mathrm{OLB}} = 0.89$ demonstrated the transferability of the onsite experiment to a binaural headphone version of the experiment. Due to the reproduction of headphones with static and non-individualized BRIRs, perceived reverberation increases, resulting in lower ratings of the object level balance on average. This phenomenon is more noticeable for configurations containing X8 point sources (C2, C3, C5, and C6), as their excitation of the reverberant field is anyway stronger.

By utilizing the linear regression of the presented FS ratio towards the object level balance results from the onsite experiment, a strong correlation of $\hat{R}_{\widetilde{\mathrm{FS}}} \geq 0.97$ for both listening locations indicates the ability to reproduce the median of the ratings of uncompensated configurations meaningfully. Position dependence also becomes apparent in the linear regression of the FS ratio but does not match the ratings ideally. The invented metric could be utilized in future system design decisions, determining where the direct sounds object level balance is preserved.

The experiment ratings for the mixing balance, as evaluated in [14], can be compared well with the object level balance examined. Utilizing professional line sources and real point sources, the observations coincide with those from the comparable experiment that used miniature line arrays [14]. Loudspeakers with a minor direct sound decay over distance, Syva (-2 dB/dod) in this study and prototyped variable curvature array with a design producing 0 dB/dod in [14], show in both cases a better preservation of object level balance or mixing balance. Even though the ratings of both experiments show explicit similarities, a direct comparison is not meaningful due to the differences in the experimental design.

Spatial definition

The assessments of spatial definition in the onsite experiment (red ratings in Figure 5.2) show a significant benefit of frontal line sources compared to frontal point sources. Slight preferences in spatial definition ratings are observed for the realistic configuration consisting only of Syva loudspeaker C4 compared to the combined C5 configuration but do not appear significant ($p > 0.5$). The consistently lower median ratings at listening position P2 indicate, for all realistic configurations C4, C5 and C6, a fundamental decrease in spatial definition towards positions further away from the origin. Due to the significantly lower ratings of X8 point sources, regardless of position compensation ($p < 0.05$), the hypothesis can be formulated that point sources degrade the perceived spatial definition. This can be explained by their wider vertical dispersion, which causes a stronger excitation of the reverberant field, decreased clarity, and a more distant impression of direct sounds.

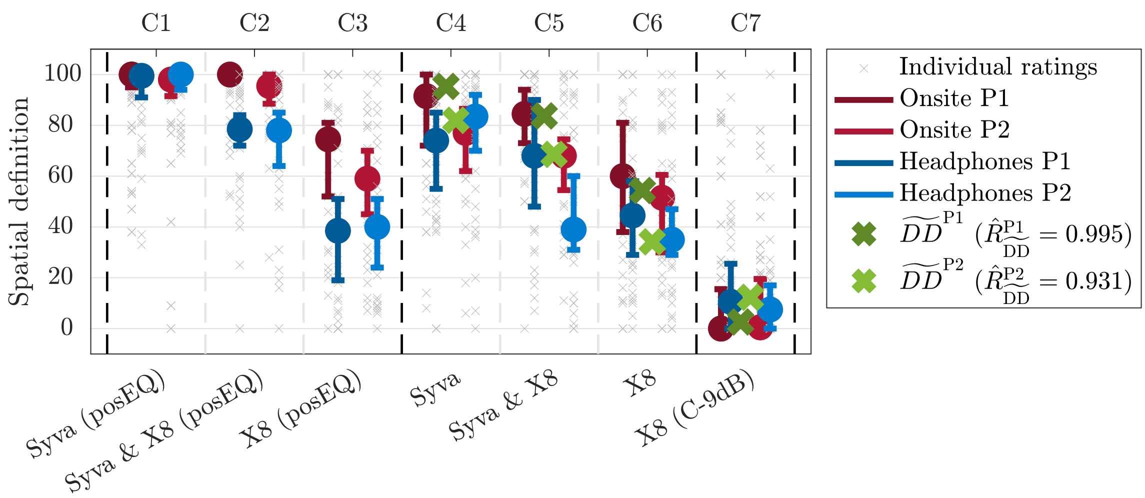


Figure 5.2: Compiled results of ratings from the onsite and virtual listening experiment for the criterion spatial definition, related to the introduced metric DD ratio which correlates with $\hat{R}_{\widetilde{\mathrm{DD}}} = 0.96$ with the onsite ratings.

The comparison of the spatial definition ratings of the onsite experiment (red) with the headphone experiments ratings (blue) shows a similar distribution tendency with an overall correlation of $R_{\mathrm{SD}} = 0.94$, verifying the transferability to a binaural headphone experiment, also for this criterion. Except for the Syva configuration C4 median ratings at position P2 and the C7 anchor ratings, the headphone experiment ratings are predominantly slightly lower than the onsite experiment. As previously mentioned, the amount of diffuse reverberation affects the impression of spatial definition, which could explain the lower ratings of the headphone experiment as a higher amount of reverberation is perceived in static and non-individualized binaural headphone reproduction. This reproduction issue primarily concerns the configurations containing point sources, as their excitation of the reverberant field is generally more substantial. The insignificant ($p = 0.28$) but unexceptional higher median spatial definition rating of the realistic Syva configuration C4 at listening position P2 compared to P1 in the headphone experiment could be attributed to timbre similarities with the reference. Furthermore, in a direct comparison of the Syva with the combined configuration of Syva and X8, the impression of spatial definition at position P1 is more similar ($p = 0.052$). The difference at the second listening location is more significant ($p = 0.003$) and could result in higher assessments of the Syva configuration, as positions P1 and P2 are not directly comparable during the experiment.

The linear regression of the presented DD ratio towards the spatial definition ratings of the onsite experiment shows a strong overall correlation of $\hat{R}_{\widetilde{\mathrm{DD}}} = 0.96$. Especially for position P1, the regression results achieved an outstanding match of the subjective results of $\hat{R}^{\mathrm{P1}}_{\widetilde{\mathrm{DD}}} = 0.995$. Presumably, the ratings of the realistic configuration C6, that consists only of X8, do not show significant position differences ($p = 0.482$), degrade the correlation at position P2, as the calculated DD ratio demonstrates higher position differences. However, sufficient reproduction of the results of the perceptual experiments is given for the configurations of a realistic application scenario. For the design of immersive systems, the invented DD ratio could evaluate the deployment decisions regarding spatial definition.

Spatial envelopment

The median onsite spatial envelopment ratings (red in Figure 5.3) of the idealized and realistic layouts consisting of only Syva line sources, C1 and C4, significantly ($p < 0.02$) improve the sensation of uniform envelopment at both tested listening positions P1 and P2, compared to deployments containing X8 loudspeakers (C2, C3, C5, C6, and C7). While the idealized configuration of solely Syva loudspeakers C1 does not show significant position differences ($p = 1.00$) in the spatial envelopment ratings, the realistic Syva configuration C4 induces significant position differences ($p = 0.035$) due to their deviation (−2 dB/dod) to the optimal direct sound decay of −3 dB/dod for ideal uniform envelopment [17].

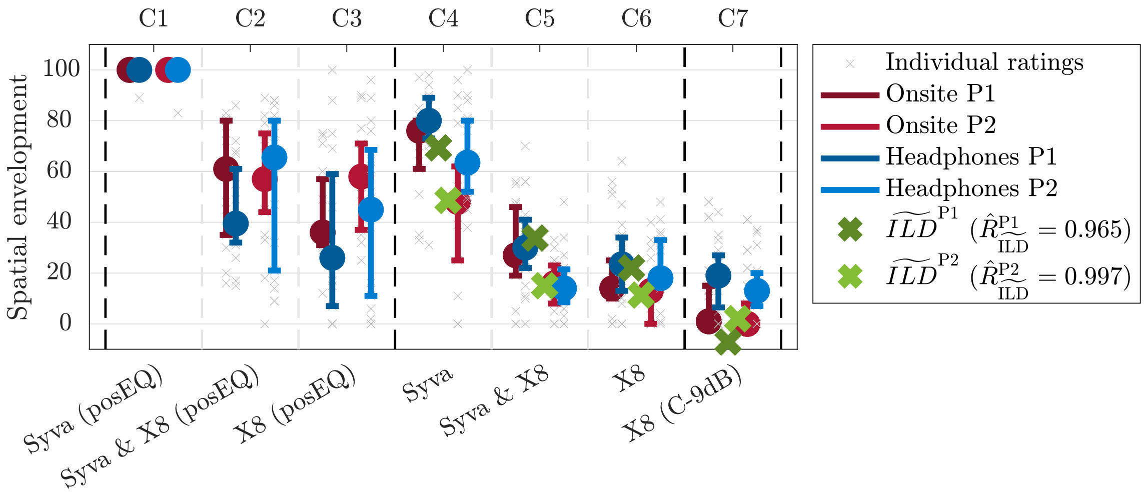


Figure 5.3: Compiled results of ratings from the onsite and virtual listening experiment for the criterion spatial envelopment, related to the introduced metric ILD of the direct sound which correlates with $\hat{R}_{\widetilde{\mathrm{ILD}}} > 0.98$ with the onsite ratings.

Regarding the median ratings of the idealized configurations C2 and C3, an influence of the excitation of the reverberant field to the sensation of spatial envelopment can be assumed, as they produce the same level and magnitude as the Syva reference C1 at the listener position.

Human hearing is highly susceptible to tonal difference during pink noise playback such that sound colourations caused by reflections and reverberation are explicitly perceived, altering the similarities to the reference.

Besides the direct sound coverage of the tested configurations, the colouration due to acoustical effects on indoor sound propagation is assumed to have a particular influence on the envelopment assessments. This also explains the significantly ($p < 0.03$) worse spatial envelopment ratings of the realistic configurations containing X8 loudspeaker (C5 and C6) compared to Syva configuration C4.

When comparing the spatial envelopment ratings of the onsite experiment (red) with the headphone experiment ratings (blue), the correlation coefficient of $R_{\mathrm{SE}} = 0.98$ reflects a distinctive agreement of both. The longer 95% CI in the headphone experiment ratings for the idealized C2 and C3 configurations point to a higher level of uncertainty, possibly caused by focusing on the tonal similarities compared to the reference instead of the sensation of envelopment. The static binaural playback, utilizing no customized head-related-transfer functions (HRTF), leads to the individual conditioned performance of the binaural reproduction. The quality of binaural playback affects the sensation of uniform envelopment, whereby the individual poor perception of spatial cues could significantly degrade the immersion. Additionally, the higher perceived reverb in static and non-individualized binaural reproduction could be attributable to the slightly higher ratings in deploying realistic application scenarios.

The adapted logarithmic metric ILD obtained from dummy head measurements of the experimental setup at the listening positions is adjusted to the onsite experiments rating scale via linear regression. The summarized correlation coefficient of both positions shows a strong correlation of $R_{\widetilde{\mathrm{ILD}}} = 0.98$, indicating a meaningful reproduction of the subjective onsite ratings for the spatial envelopment. The regressed $\widetilde{\mathrm{ILD}}$ show a slightly stronger correlation towards the experiments ratings for position 2 with $R^{\mathrm{P2}}_{\widetilde{\mathrm{ILD}}} = 0.997$ compared to position 1, $R^{\mathrm{P1}}_{\widetilde{\mathrm{ILD}}} = 0.966$.

The experiment, conducted in [14], likewise evaluates the sensation of envelopment for three different decay conditions. Comparing the rating trends of the -3 dB/dod condition from [14] with the Syva configuration (-2 dB/dod) used in this study and the 0 dB/dod condition [14] with the X8 point source configuration (-5 dB/dod) from this work, superior envelopment ratings for the configurations closer to -3 dB/dod are observed in both experiments.

Overall assessment of loudspeaker layouts

The ratings of the selected criteria, both listening locations, three trials, and two performed experiments are consolidated to summarize the tested loudspeaker layouts for immersive sound reinforcement. $10 \times 31 = 310$ ratings per layout are collected, and the median and 95% CI interval are calculated. The results are displayed in Figure 5.4 but do not have an implicit scientific impact because different ratings collected might not be comparable. Nevertheless, the trend observed in the previous sections becomes apparent.

Idealized configurations, consisting solely of Syva line sources C1, are rated best, and the other idealized configurations containing point sources (C2 and C3) are rated unambiguously worse as they tend to have a higher impact on the excitation of the reverberant field due to their directivity. The more point sources deployed in the configuration, the lower the consolidated ratings.

For the realistic configuration, a similar trend is observable. The highest ratings have the configuration of solely Syva loudspeakers C4, followed by the combined C5 configuration. The lowest ratings appear for configuration C6, consisting only of X8 point sources. The overall ratings are slightly lower, caused by the natural influence of the acoustical transfer path altering the level and magnitude arriving at the listener’s ears.

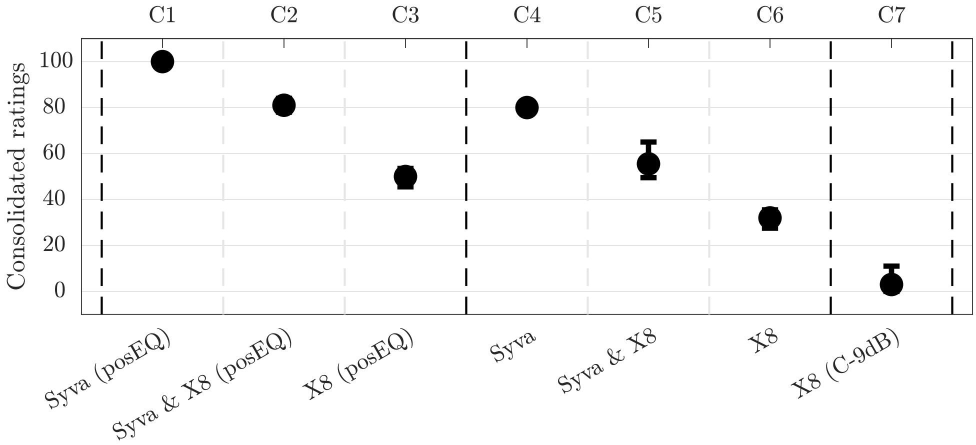


Figure 5.4: Consolidation of the results for all trials, criteria, repetitions, and both experiments to summarize an overall assessment of the investigated loudspeaker configurations.

6. Conclusion

This thesis presents a novel methodology with two MUSHRA-based listening experiments using multichannel musical excerpts to compare the perceived spatial qualities of professional coaxial point sources and professional compact fixed-curvature line sources in immersive sound reinforcement, focusing on off-center positions representative for extended audience areas. In-situ and binaural measurements yielded a set of acoustic metrics linking measured sound-field parameters to observed perceptual outcomes, illustrating how different surround loudspeaker layouts influence immersive reproduction.

The ratings from onsite and headphone listening experiments show that the direct sound decay of elevated, compact fixed-curvature line sources (-2 dB/distance-doubling) extends the sweet area for preserving object level balance, thereby validating earlier prototype-based findings [14] with professional loudspeakers. Similarly, frontally deployed line sources improve spatial definition by reducing room excitation, whereas point source directivity introduces more reverberant energy and diminishes clarity. Surround line sources further enhance uniform envelopment at off-center positions, confirming earlier perceptual and analytic studies that propose a -3 dB/ distance-doubling target for enveloping fields [14],[17].

A direct comparison of the onsite and headphone experiment results shows strong correlations above 0.87 across all three criteria, validating the transferability of the onsite methodology to static, non-individualized binaural playback. In headphone reproduction, participants perceive stronger reverberation, which primarily lowers ratings for object level balance and spatial definition in point source deployments that excite the reverberant field more intensely. It can be speculated that the use of non-individualized HRTFs reduces externalization and introduces confusion about specific spatial cues and colouration artifacts, but it did not seem to alter the overall results for the configurations under test. Consequently, the headphone results closely reproduce the onsite experiment ratings, affirming the viability of virtual listening experiments for immersive sound reinforcement research.

The newly introduced FS ratio, DD ratio, and ILD measures, which target a numerical measurement-based estimation of object level balance, spatial definition, and spatial envelopment, respectively, achieve high correlations exceeding 0.94 with the ratings of the onsite listening experiment in linear regression. These close correspondences underscore the predictive power of the proposed metrics for system design and layout decisions in immersive surround sound reinforcement, enabling objective characterization of loudspeaker deployments' perceptual attributes from measured or simulated data.

Outlook

Future research could concern a binaural headphone experiment incorporating individualized HRTFs and dynamic rendering, including head rotations, using high-order Ambisonics measurements of large-scale line source systems. By measuring each driver individually, researchers might be able to subsequently refine magnitude and distance decay settings, facilitating a comprehensive examination of loudspeaker deployments and their tuning. Such an experiment could deepen our understanding of the perceived spatial qualities and provide suggestions and validations for optimal immersive layout design.

Another important research focus involves extending and validating the proposed acoustic metrics across various loudspeaker types and immersive reinforcement layouts. Their predictive capability could lead to the adoption of standard tools for accurately characterizing system performance and perceptual outcomes even without linear regression to subjective ground truth.

Declaration of utilized resources

The software *MATLAB*[33], along with all its toolboxes, was used to implement the algorithms and create all the plots.

The sketches and vector graphics were created using *IPE*[34].

To verify equalization, the software *Smaart v8*[35] was employed.

For conducting MUSHRA experiments, the application *MUSHRA*[36] was utilized.

Audio playback during the listening experiments was carried out using *REAPER*[37].

To enhance the writing, the translator *DeepL.com*[38], the text correction tool *Grammarly*[39], and the large language model *ChatGPT-4o*[40] by OpenAI were used.

[33]https://de.mathworks.com/products/matlab.html
[34]https://ipe.otfried.org
[35]https://www.rationalacoustics.com/pages/smaart-home
[36]https://git.iem.at/rudrich/mushra
[37]https://www.reaper.fm
[38]https://www.deepl.com/de/translator
[39]https://app.grammarly.com
[40]https://chatgpt.com/

Appendix

A.1 Layouts with idealized magnitude and level

This section illustrates the measured and smoothed magnitudes, the calculated gains $g_{\mathrm{ls,P1}}$ and $g_{\mathrm{ls,P1}}$, and the idealized and level-matched magnitudes for listening position P1 and P2.

Position 1

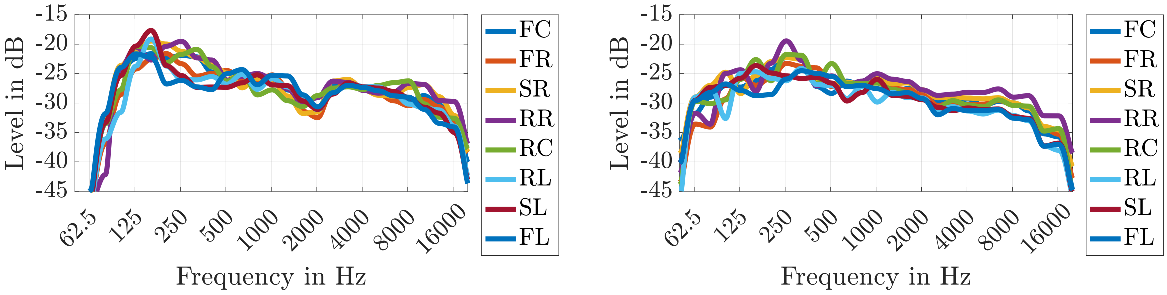


Figure A.1: Smoothed and averaged magnitudes of all eight Syva loudspeakers (left) and X8 speakers (right) at position P1 of the layout.

$20\log_{10}(g_{\mathrm{ls,P1}})$ in dB	FC	FR	SR	RR	RC	RL	SL	FL
Syva	-0.69	-0.36	-1.43	-2.04	-1.16	-0.21	-0.27	0.00
X8	1.01	0.99	0.55	-0.70	0.87	2.70	2.36	2.49

Table A.1: Decibel values of compensation gains $g_{\mathrm{ls,P1}}$ for all loudspeakers for position 1.

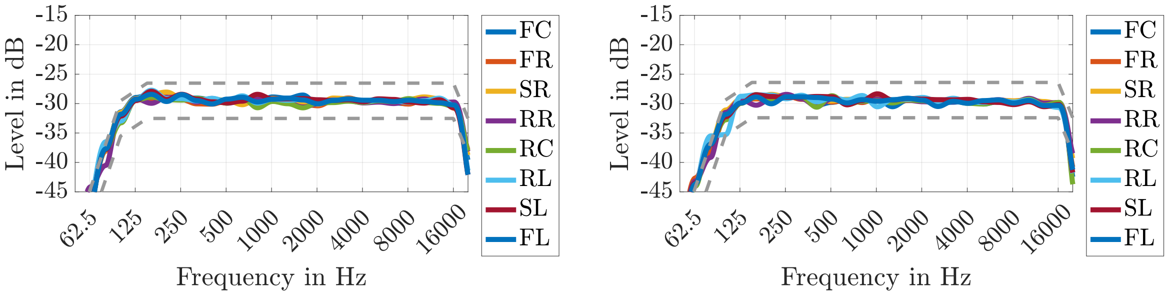


Figure A.2: Magnitudes of all eight Syva loudspeakers (left) and X8 speakers (right) at position P1 of the layout after filtering with equalizing FIR-filter, dashed lines represent ± 3 dB limits.

Position 2

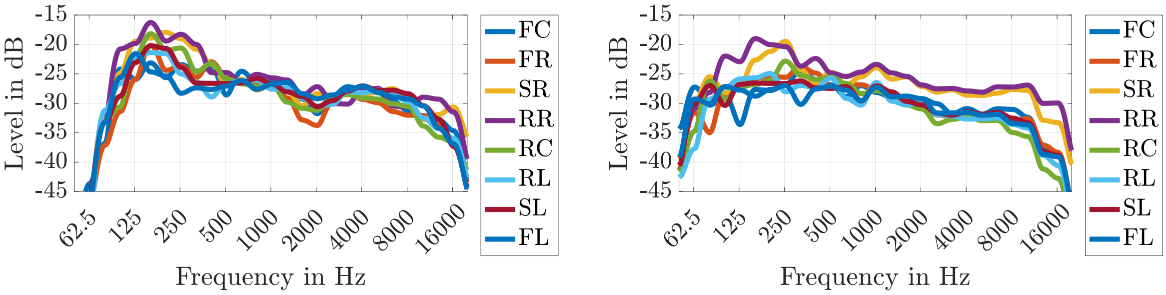


Figure A.3: Smoothed and averaged magnitudes of all eight Syva loudspeakers (left) and X8 speakers (right) at position P2 of the layout.

$20\log_{10}(g_{\mathrm{ls,P2}})$ in dB	FC	FR	SR	RR	RC	RL	SL	FL
Syva	-1.08	0.00	-2.27	-2.56	-0.38	-0.92	-1.69	-1.36
X8	0.82	1.32	-2.29	-3.61	2.83	2.20	1.41	1.74

Table A.2: Decibel values of compensation gains $g_{\mathrm{ls,P2}}$ for all loudspeakers for position 2.

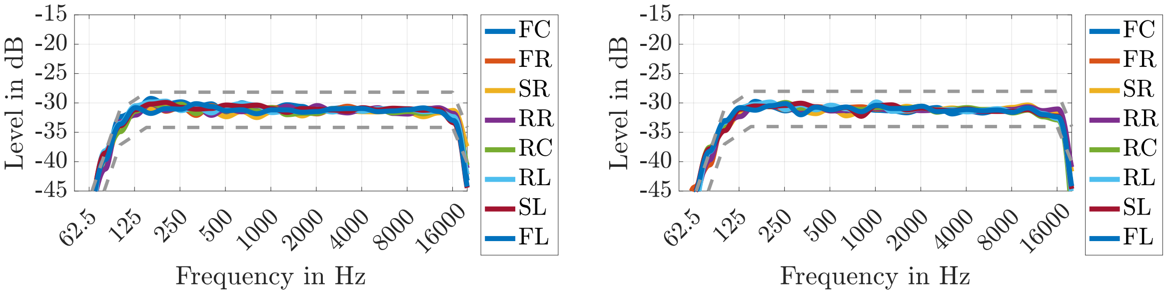


Figure A.4: Magnitudes of all eight Syva loudspeakers (left) and X8 speakers (right) at position P2 of the layout after filtering with equalizing FIR-filter, dashed lines represent $\pm 3\,$dB limits.

A.2 Bonferroni-corrected p-values

This section shows Bonferroni-corrected p-values for pairwise comparisons between all configurations and between the listening positions for each configuration, respectively.

A.2.1 Object level balance p-values (onsite experiment)

p(P1)	C1	C2	C3	C5	C5	C6	C7
C1	0.000	1.039	0.005	0.565	0.681	0.000	0.000
C2	0.000	0.000	0.000	0.630	0.544	0.000	0.000
C3	0.000	0.000	0.000	0.179	0.155	0.069	0.000
C5	0.000	0.000	0.000	0.000	0.966	0.005	0.000
C5	0.000	0.000	0.000	0.000	0.000	0.002	0.000
C6	0.000	0.000	0.000	0.000	0.000	0.000	0.000
C7	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table A.3: Bonferroni-corrected p-values for pairwise comparisons between all configurations of the object level balance ratings of the onsite listening experiment for the listening position P1. Values below 0.05 are denoted as significant.

p(P2)	C1	C2	C3	C5	C5	C6	C7
C1	0.000	0.052	0.000	0.004	0.030	0.000	0.000
C2	0.000	0.000	0.000	0.643	0.743	0.001	0.000
C3	0.000	0.000	0.000	0.028	0.054	0.483	0.106
C5	0.000	0.000	0.000	0.000	0.817	0.035	0.000
C5	0.000	0.000	0.000	0.000	0.000	0.115	0.000
C6	0.000	0.000	0.000	0.000	0.000	0.000	0.006
C7	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table A.4: Bonferroni-corrected p-values for pairwise comparisons between all configurations of the object level balance ratings of the onsite listening experiment for the listening position P2. Values below 0.05 are denoted as significant.

p	C1(P1|P2)	C2(P1|P2)	C3(P1|P2)	C4(P1|P2)	C5(P1|P2)	C6(P1|P2)	C7(P1|P2)
	0.276	0.733	0.918	0.223	0.002	0.289	0.679

Table A.5: Bonferroni-corrected p-values for pairwise comparisons of the object level balance ratings of the onsite listening experiment for the listening positions, P1 and P2. Values below 0.05 are denoted as significant.

A.2.2 Spatial definition p-values (onsite experiment)

p(P1)	C1	C2	C3	C5	C5	C6	C7
C1	0.000	0.823	0.006	0.407	0.364	0.001	0.000
C2	0.000	0.000	0.002	0.429	0.369	0.001	0.000
C3	0.000	0.000	0.000	0.293	0.078	0.375	0.001
C5	0.000	0.000	0.000	0.000	0.861	0.004	0.000
C5	0.000	0.000	0.000	0.000	0.000	0.025	0.000
C6	0.000	0.000	0.000	0.000	0.000	0.000	0.000
C7	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table A.6: Bonferroni-corrected p-values for pairwise comparisons between all configurations of the spatial definition ratings of the onsite listening experiment for the listening position P1. Values below 0.05 are denoted as significant.

p(P2)	C1	C2	C3	C5	C5	C6	C7
C1	0.000	0.858	0.001	0.031	0.018	0.001	0.000
C2	0.000	0.000	0.000	0.010	0.000	0.000	0.000
C3	0.000	0.000	0.000	0.031	0.417	0.540	0.000
C5	0.000	0.000	0.000	0.000	0.548	0.016	0.000
C5	0.000	0.000	0.000	0.000	0.000	0.049	0.000
C6	0.000	0.000	0.000	0.000	0.000	0.000	0.001
C7	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table A.7: Bonferroni-corrected p-values for pairwise comparisons between all configurations of the spatial definition ratings of the onsite listening experiment for the listening position P2. Values below 0.05 are denoted as significant.

p	C1(P1|P2)	C2(P1|P2)	C3(P1|P2)	C4(P1|P2)	C5(P1|P2)	C6(P1|P2)	C7(P1|P2)
	0.689	0.891	0.074	0.156	0.007	0.482	0.055

Table A.8: Bonferroni-corrected p-values for pairwise comparisons of the spatial definition ratings of the onsite listening experiment for the listening positions, P1 and P2. Values below 0.05 are denoted as significant.

A.2.3 Spatial envelopment p-values (onsite experiment)

p(P1)	C1	C2	C3	C5	C5	C6	C7
C1	0.000	0.001	0.001	0.001	0.001	0.001	0.001
C2	0.000	0.000	0.032	0.208	0.001	0.001	0.001
C3	0.000	0.000	0.000	0.021	0.012	0.003	0.001
C5	0.000	0.000	0.000	0.000	0.002	0.001	0.001
C5	0.000	0.000	0.000	0.000	0.000	0.014	0.002
C6	0.000	0.000	0.000	0.000	0.000	0.000	0.002
C7	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table A.9: Bonferroni-corrected p-values for pairwise comparisons between all configurations of the spatial envelopment ratings of the onsite listening experiment for the listening position P1. Values below 0.05 are denoted as significant.

p(P2)	C1	C2	C3	C5	C5	C6	C7
C1	0.000	0.001	0.001	0.001	0.001	0.001	0.001
C2	0.000	0.000	0.872	0.890	0.001	0.001	0.001
C3	0.000	0.000	0.000	0.794	0.001	0.001	0.001
C5	0.000	0.000	0.000	0.000	0.021	0.007	0.012
C5	0.000	0.000	0.000	0.000	0.000	0.527	0.038
C6	0.000	0.000	0.000	0.000	0.000	0.000	0.484
C7	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table A.10: Bonferroni-corrected p-values for pairwise comparisons between all configurations of the spatial envelopment ratings of the onsite listening experiment for the listening position P2. Values below 0.05 are denoted as significant.

p	C1(P1|P2)	C2(P1|P2)	C3(P1|P2)	C4(P1|P2)	C5(P1|P2)	C6(P1|P2)	C7(P1|P2)
	1.000	0.552	0.027	0.035	0.004	0.141	0.383

Table A.11: Bonferroni-corrected p-values for pairwise comparisons of the spatial envelopment ratings of the onsite listening experiment for the listening positions, P1 and P2. Values below 0.05 are denoted as significant.

A.2.4 Object level balance *p*-values (headphone experiment)

p(P1)	C1	C2	C3	C5	C5	C6	C7
C1	0.000	0.003	0.000	0.063	0.021	0.000	0.000
C2	0.000	0.000	0.000	0.600	0.881	0.000	0.000
C3	0.000	0.000	0.000	0.000	0.000	1.422	0.019
C5	0.000	0.000	0.000	0.000	0.583	0.000	0.000
C5	0.000	0.000	0.000	0.000	0.000	0.000	0.000
C6	0.000	0.000	0.000	0.000	0.000	0.000	0.007
C7	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table A.12: Bonferroni-corrected *p*-values for pairwise comparisons between all configurations of the object level balance ratings of the headphone listening experiment for the listening position P1. Values below 0.05 are denoted as significant.

p(P2)	C1	C2	C3	C5	C5	C6	C7
C1	0.000	0.055	0.000	0.050	0.000	0.000	0.000
C2	0.000	0.000	0.000	0.576	0.003	0.000	0.000
C3	0.000	0.000	0.000	0.000	0.126	0.043	0.002
C5	0.000	0.000	0.000	0.000	0.000	0.000	0.000
C5	0.000	0.000	0.000	0.000	0.000	0.003	0.000
C6	0.000	0.000	0.000	0.000	0.000	0.000	0.043
C7	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table A.13: Bonferroni-corrected *p*-values for pairwise comparisons between all configurations of the object level balance ratings of the headphone listening experiment for the listening position P2. Values below 0.05 are denoted as significant.

p	C1(P1|P2)	C2(P1|P2)	C3(P1|P2)	C4(P1|P2)	C5(P1|P2)	C6(P1|P2)	C7(P1|P2)
	0.409	0.102	0.282	0.787	< 0.001	0.229	0.756

Table A.14: Bonferroni-corrected *p*-values for pairwise comparisons of the object level balance ratings of the headphone listening experiment for the listening positions, P1 and P2. Values below 0.05 are denoted as significant.

A.2.5 Spatial definition p-values (headphone experiment)

p(P1)	C1	C2	C3	C5	C5	C6	C7
C1	0.000	0.052	0.000	0.004	0.030	0.000	0.000
C2	0.000	0.000	0.000	0.643	0.743	0.001	0.000
C3	0.000	0.000	0.000	0.028	0.054	0.483	0.106
C5	0.000	0.000	0.000	0.000	0.817	0.035	0.000
C5	0.000	0.000	0.000	0.000	0.000	0.115	0.000
C6	0.000	0.000	0.000	0.000	0.000	0.000	0.006
C7	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table A.15: Bonferroni-corrected p-values for pairwise comparisons between all configurations of the spatial definition ratings of the headphone listening experiment for the listening position P1. Values below 0.05 are denoted as significant.

p(P2)	C1	C2	C3	C5	C5	C6	C7
C1	0.000	0.003	0.000	0.012	0.000	0.000	0.000
C2	0.000	0.000	0.013	1.705	0.127	0.006	0.000
C3	0.000	0.000	0.000	0.008	1.538	0.963	0.001
C5	0.000	0.000	0.000	0.000	0.007	0.001	0.000
C5	0.000	0.000	0.000	0.000	0.000	1.177	0.000
C6	0.000	0.000	0.000	0.000	0.000	0.000	0.006
C7	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table A.16: Bonferroni-corrected p-values for pairwise comparisons between all configurations of the spatial definition ratings of the headphone listening experiment for the listening position P2. Values below 0.05 are denoted as significant.

p	C1(P1|P2)	C2(P1|P2)	C3(P1|P2)	C4(P1|P2)	C5(P1|P2)	C6(P1|P2)	C7(P1|P2)
	0.387	0.737	0.853	0.280	0.014	0.308	0.224

Table A.17: Bonferroni-corrected p-values for pairwise comparisons of the spatial definition ratings of the headphone listening experiment for the listening positions, P1 and P2. Values below 0.05 are denoted as significant.

A.2.6 Spatial envelopment *p*-values (headphone experiment)

p(P1)	C1	C2	C3	C5	C5	C6	C7
C1	0.000	0.006	0.001	0.008	0.008	0.007	0.009
C2	0.000	0.000	0.435	0.012	0.697	0.037	0.009
C3	0.000	0.000	0.000	0.025	0.717	1.554	0.588
C5	0.000	0.000	0.000	0.000	0.008	0.007	0.007
C5	0.000	0.000	0.000	0.000	0.000	0.619	0.271
C6	0.000	0.000	0.000	0.000	0.000	0.000	1.069
C7	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table A.18: Bonferroni-corrected *p*-values for pairwise comparisons between all configurations of the spatial envelopment ratings of the headphone listening experiment for the listening position P1. Values below 0.05 are denoted as significant.

p(P2)	C1	C2	C3	C5	C5	C6	C7
C1	0.000	0.008	0.006	0.010	0.009	0.007	0.009
C2	0.000	0.000	0.070	0.766	0.045	0.052	0.042
C3	0.000	0.000	0.000	0.238	0.248	0.265	0.139
C5	0.000	0.000	0.000	0.000	0.007	0.008	0.007
C5	0.000	0.000	0.000	0.000	0.000	0.578	0.352
C6	0.000	0.000	0.000	0.000	0.000	0.000	0.286
C7	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table A.19: Bonferroni-corrected *p*-values for pairwise comparisons between all configurations of the spatial envelopment ratings of the headphone listening experiment for the listening position P2. Values below 0.05 are denoted as significant.

p	C1(P1|P2)	C2(P1|P2)	C3(P1|P2)	C4(P1|P2)	C5(P1|P2)	C6(P1|P2)	C7(P1|P2)
	1.000	0.162	0.203	0.046	0.004	0.660	0.511

Table A.20: Bonferroni-corrected *p*-values for pairwise comparisons of the spatial envelopment ratings of the headphone listening experiment for the listening positions, P1 and P2. Values below 0.05 are denoted as significant.

A.3 Cohen's d effect sizes

This section shows Cohen's d effect sizes for pairwise comparisons between all configurations and between the listening position for each configuration, respectively.

A.3.1 Object level balance d effect sizes (onsite experiment)

d(P1)	C1	C2	C3	C5	C5	C6	C7
C1	0.000	0.042	1.162	0.448	0.384	1.206	3.407
C2	−0.042	0.000	1.101	0.397	0.355	1.173	3.296
C3	−1.162	−1.101	0.000	−0.687	−0.478	0.395	1.887
C4	−0.448	−0.397	0.687	0.000	0.074	0.905	2.717
C5	−0.384	−0.355	0.478	−0.074	0.000	0.743	2.154
C6	−1.206	−1.173	−0.395	−0.905	−0.743	0.000	1.064
C7	−3.407	−3.296	−1.887	−2.717	−2.154	−1.064	0.000

Table A.21: Cohen's d effect sizes for pairwise comparisons between all configurations of the object level balance ratings of the onsite listening experiment for the listening position P1. Effect sizes of $d \approx 0.2$ indicate small effects, $d \approx 0.5$ medium effects, and $d \approx 0.8$ large effects.

d(P2)	C1	C2	C3	C5	C5	C6	C7
C1	0.000	0.346	1.403	1.010	1.434	1.949	3.984
C2	−0.346	0.000	0.951	0.583	1.014	1.559	3.188
C3	−1.403	−0.951	0.000	−0.375	0.105	0.722	2.047
C4	−1.010	−0.583	0.375	0.000	0.466	1.059	2.501
C5	−1.434	−1.014	−0.105	−0.466	0.000	0.608	1.857
C6	−1.949	−1.559	−0.722	−1.059	−0.608	0.000	1.039
C7	−3.984	−3.188	−2.047	−2.501	−1.857	−1.039	0.000

Table A.22: Cohen's d effect sizes for pairwise comparisons between all configurations of the object level balance ratings of the onsite listening experiment for the listening position P2. Effect sizes of $d \approx 0.2$ indicate small effects, $d \approx 0.5$ medium effects, and $d \approx 0.8$ large effects.

d	C1(P1|P2)	C2(P1|P2)	C3(P1|P2)	C4(P1|P2)	C5(P1|P2)	C6(P1|P2)	C7(P1|P2)
	−0.220	0.152	0.034	0.375	0.591	0.213	0.006

Table A.23: Cohen's d effect sizes for pairwise comparisons of the object level balance ratings of the onsite listening experiment for the listening positions, P1 and P2. Effect sizes of $d \approx 0.2$ indicate small effects, $d \approx 0.5$ medium effects, and $d \approx 0.8$ large effects.

A.3.2 Spatial definition d effect sizes (onsite experiment)

d(P1)	C1	C2	C3	C5	C5	C6	C7
C1	0.000	−0.114	0.770	0.317	0.408	0.922	1.878
C2	0.114	0.000	0.859	0.408	0.539	0.994	1.965
C3	−0.770	−0.859	0.000	−0.422	−0.479	0.248	1.066
C4	−0.317	−0.408	0.422	0.000	−0.002	0.624	1.488
C5	−0.408	−0.539	0.479	0.002	0.000	0.683	1.623
C6	−0.922	−0.994	−0.248	−0.624	−0.683	0.000	0.739
C7	−1.878	−1.965	−1.066	−1.488	−1.623	−0.739	0.000

Table A.24: Cohen's d effect sizes for pairwise comparisons between all configurations of the spatial definition ratings of the onsite listening experiment for the listening position P1. Effect sizes of $d \approx 0.2$ indicate small effects, $d \approx 0.5$ medium effects, and $d \approx 0.8$ large effects.

d(P2)	C1	C2	C3	C5	C5	C6	C7
C1	0.000	0.069	1.535	1.003	1.165	1.671	4.627
C2	−0.069	0.000	1.479	0.935	1.112	1.617	4.462
C3	−1.535	−1.479	0.000	−0.658	−0.311	0.181	1.596
C4	−1.003	−0.935	0.658	0.000	0.315	0.830	2.762
C5	−1.165	−1.112	0.311	−0.315	0.000	0.483	2.002
C6	−1.671	−1.617	−0.181	−0.830	−0.483	0.000	1.303
C7	−4.627	−4.462	−1.596	−2.762	−2.002	−1.303	0.000

Table A.25: Cohen's d effect sizes for pairwise comparisons between all configurations of the spatial definition ratings of the onsite listening experiment for the listening position P2. Effect sizes of $d \approx 0.2$ indicate small effects, $d \approx 0.5$ medium effects, and $d \approx 0.8$ large effects.

d	C1(P1|P2)	C2(P1|P2)	C3(P1|P2)	C4(P1|P2)	C5(P1|P2)	C6(P1|P2)	C7(P1|P2)
	−0.131	0.053	0.363	0.315	0.655	0.193	0.263

Table A.26: Cohen's d effect sizes for pairwise comparisons of the spatial definition ratings of the onsite listening experiment for the listening positions, P1 and P2. Effect sizes of $d \approx 0.2$ indicate small effects, $d \approx 0.5$ medium effects, and $d \approx 0.8$ large effects.

A.3.3 Spatial envelopment d effect sizes (onsite experiment)

d(P1)	C1	C2	C3	C5	C5	C6	C7
C1	0.000	1.751	3.248	2.494	4.424	6.037	9.783
C2	−1.751	0.000	0.519	−0.431	1.052	1.505	1.994
C3	−3.248	−0.519	0.000	−1.260	0.625	1.166	1.790
C4	−2.494	0.431	1.260	0.000	2.063	2.887	4.074
C5	−4.424	−1.052	−0.625	−2.063	0.000	0.518	1.102
C6	−6.037	−1.505	−1.166	−2.887	−0.518	0.000	0.582
C7	−9.783	−1.994	−1.790	−4.074	−1.102	−0.582	0.000

Table A.27: Cohen's d effect sizes for pairwise comparisons between all configurations of the spatial envelopment ratings of the onsite listening experiment for the listening position P1. Effect sizes of $d \approx 0.2$ indicate small effects, $d \approx 0.5$ medium effects, and $d \approx 0.8$ large effects.

d(P2)	C1	C2	C3	C5	C5	C6	C7
C1	0.000	2.312	2.613	1.839	9.756	6.550	10.067
C2	−2.312	0.000	−0.115	0.135	1.716	1.742	2.171
C3	−2.613	0.115	0.000	0.236	2.164	2.109	2.679
C4	−1.839	−0.135	−0.236	0.000	1.130	1.216	1.484
C5	−9.756	−1.716	−2.164	−1.130	0.000	0.293	0.836
C6	−6.550	−1.742	−2.109	−1.216	−0.293	0.000	0.361
C7	−10.067	−2.171	−2.679	−1.484	−0.836	−0.361	0.000

Table A.28: Cohen's d effect sizes for pairwise comparisons between all configurations of the object spatial envelopment of the onsite listening experiment for the listening position P2. Effect sizes of $d \approx 0.2$ indicate small effects, $d \approx 0.5$ medium effects, and $d \approx 0.8$ large effects.

d	C1(P1|P2)	C2(P1|P2)	C3(P1|P2)	C4(P1|P2)	C5(P1|P2)	C6(P1|P2)	C7(P1|P2)
	0.000	0.175	−0.531	0.707	0.717	0.359	0.245

Table A.29: Cohen's d effect sizes for pairwise comparisons of the spatial envelopment ratings of the onsite listening experiment for the listening positions, P1 and P2. Effect sizes of $d \approx 0.2$ indicate small effects, $d \approx 0.5$ medium effects, and $d \approx 0.8$ large effects.

A.3.4 Object level balance d effect sizes (headphone experiment)

d(P1)	C1	C2	C3	C5	C5	C6	C7
C1	0.000	1.454	3.114	0.721	1.011	4.543	5.678
C2	−1.454	0.000	1.479	−0.344	−0.034	1.868	2.613
C3	−3.114	−1.479	0.000	−1.627	−1.274	0.073	0.735
C4	−0.721	0.344	1.627	0.000	0.263	1.942	2.568
C5	−1.011	0.034	1.274	−0.263	0.000	1.509	2.094
C6	−4.543	−1.868	−0.073	−1.942	−1.509	0.000	0.835
C7	−5.678	−2.613	−0.735	−2.568	−2.094	−0.835	0.000

Table A.30: Cohen's d effect sizes for pairwise comparisons between all configurations of the object level balance ratings of the headphone listening experiment for the listening position P1. Effect sizes of $d \approx 0.2$ indicate small effects, $d \approx 0.5$ medium effects, and $d \approx 0.8$ large effects.

d(P2)	C1	C2	C3	C5	C5	C6	C7
C1	0.000	0.670	3.246	0.709	1.532	5.001	6.475
C2	−0.670	0.000	1.763	−0.117	0.904	2.575	3.247
C3	−3.246	−1.763	0.000	−2.227	−0.417	0.715	1.403
C4	−0.709	0.117	2.227	0.000	1.074	3.356	4.291
C5	−1.532	−0.904	0.417	−1.074	0.000	0.913	1.351
C6	−5.001	−2.575	−0.715	−3.356	−0.913	0.000	0.779
C7	−6.475	−3.247	−1.403	−4.291	−1.351	−0.779	0.000

Table A.31: Cohen's d effect sizes for pairwise comparisons between all configurations of the object level balance ratings of the headphone listening experiment for the listening position P2. Effect sizes of $d \approx 0.2$ indicate small effects, $d \approx 0.5$ medium effects, and $d \approx 0.8$ large effects.

d	C1(P1|P2)	C2(P1|P2)	C3(P1|P2)	C4(P1|P2)	C5(P1|P2)	C6(P1|P2)	C7(P1|P2)
	0.336	−0.383	−0.254	−0.103	0.576	0.310	0.088

Table A.32: Cohen's d effect sizes for pairwise comparisons of the object level balance ratings of the headphone listening experiment for the listening positions, P1 and P2. Effect sizes of $d \approx 0.2$ indicate small effects, $d \approx 0.5$ medium effects, and $d \approx 0.8$ large effects.

A.3.5 Spatial definition d effect sizes (headphone experiment)

d(P1)	C1	C2	C3	C5	C5	C6	C7
C1	0.000	0.803	1.560	0.871	0.765	1.918	3.237
C2	−0.803	0.000	1.038	0.232	0.269	1.173	2.311
C3	−1.560	−1.038	0.000	−0.778	−0.621	−0.130	0.590
C4	−0.871	−0.232	0.778	0.000	0.076	0.801	1.733
C5	−0.765	−0.269	0.621	−0.076	0.000	0.593	1.350
C6	−1.918	−1.173	0.130	−0.801	−0.593	0.000	0.916
C7	−3.237	−2.311	−0.590	−1.733	−1.350	−0.916	0.000

Table A.33: Cohen's d effect sizes for pairwise comparisons between all configurations of the spatial definition ratings of the headphone listening experiment for the listening position P1. Effect sizes of $d \approx 0.2$ indicate small effects, $d \approx 0.5$ medium effects, and $d \approx 0.8$ large effects.

d(P2)	C1	C2	C3	C5	C5	C6	C7
C1	0.000	0.945	1.824	0.685	1.367	2.409	5.438
C2	−0.945	0.000	0.975	−0.096	0.683	1.246	2.811
C3	−1.824	−0.975	0.000	−0.980	−0.161	0.079	1.017
C4	−0.685	0.096	0.980	0.000	0.713	1.209	2.488
C5	−1.367	−0.683	0.161	−0.713	0.000	0.248	1.075
C6	−2.409	−1.246	−0.079	−1.209	−0.248	0.000	1.160
C7	−5.438	−2.811	−1.017	−2.488	−1.075	−1.160	0.000

Table A.34: Cohen's d effect sizes for pairwise comparisons between all configurations of the spatial definition ratings of the headphone listening experiment for the listening position P2. Effect sizes of $d \approx 0.2$ indicate small effects, $d \approx 0.5$ medium effects, and $d \approx 0.8$ large effects.

d	C1(P1|P2)	C2(P1|P2)	C3(P1|P2)	C4(P1|P2)	C5(P1|P2)	C6(P1|P2)	C7(P1|P2)
	−0.072	0.134	−0.040	−0.184	0.433	0.201	0.244

Table A.35: Cohen's d effect sizes for pairwise comparisons of the spatial definition ratings of the headphone listening experiment for the listening positions, P1 and P2. Effect sizes of $d \approx 0.2$ indicate small effects, $d \approx 0.5$ medium effects, and $d \approx 0.8$ large effects.

A.3.6 Spatial envelopment d effect sizes (headphone experiment)

d(P1)	C1	C2	C3	C5	C5	C6	C7
C1	0.000	3.045	1.977	2.261	4.701	6.055	5.966
C2	−3.045	0.000	0.318	−1.822	0.628	0.991	1.241
C3	−1.977	−0.318	0.000	−1.384	0.066	0.266	0.438
C4	−2.261	1.822	1.384	0.000	2.996	3.830	3.973
C5	−4.701	−0.628	−0.066	−2.996	0.000	0.369	0.671
C6	−6.055	−0.991	−0.266	−3.830	−0.369	0.000	0.342
C7	−5.966	−1.241	−0.438	−3.973	−0.671	−0.342	0.000

Table A.36: Cohen's d effect sizes for pairwise comparisons between all configurations of the spatial envelopment ratings of the headphone listening experiment for the listening position P1. Effect sizes of $d \approx 0.2$ indicate small effects, $d \approx 0.5$ medium effects, and $d \approx 0.8$ large effects.

d(P2)	C1	C2	C3	C5	C5	C6	C7
C1	0.000	1.362	1.669	2.180	10.069	6.661	8.351
C2	−1.362	0.000	0.292	−0.306	1.151	0.924	1.197
C3	−1.669	−0.292	0.000	−0.654	0.688	0.488	0.739
C4	−2.180	0.306	0.654	0.000	2.858	2.245	2.820
C5	−10.069	−1.151	−0.688	−2.858	0.000	−0.480	0.168
C6	−6.661	−0.924	−0.488	−2.245	0.480	0.000	0.585
C7	−8.351	−1.197	−0.739	−2.820	−0.168	−0.585	0.000

Table A.37: Cohen's d effect sizes for pairwise comparisons between all configurations of the spatial envelopment ratings of the headphone listening experiment for the listening position P2. Effect sizes of $d \approx 0.2$ indicate small effects, $d \approx 0.5$ medium effects, and $d \approx 0.8$ large effects.

d	C1(P1|P2)	C2(P1|P2)	C3(P1|P2)	C4(P1|P2)	C5(P1|P2)	C6(P1|P2)	C7(P1|P2)
	−0.594	−0.249	−0.149	0.867	0.913	0.082	0.240

Table A.38: Cohen's d effect sizes for pairwise comparisons of the spatial envelopment ratings of the headphone listening experiment for the listening positions, P1 and P2. Effect sizes of $d \approx 0.2$ indicate small effects, $d \approx 0.5$ medium effects, and $d \approx 0.8$ large effects.

A.4 Comparison of ratings regarding different musical excerpt

In this section, the ratings of the object level balance and spatial definition for the different musical excerpts are contrasted.

Onsite experiment ratings (musical excerpt)

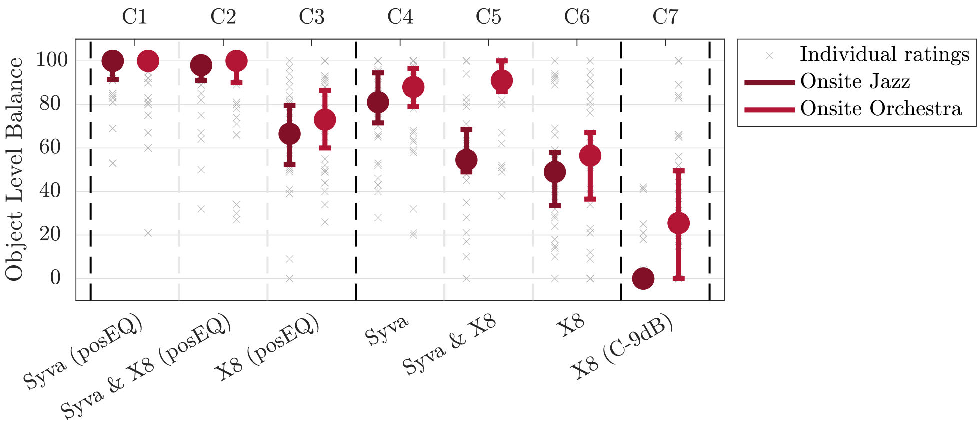


Figure A.5: Object level balance ratings for the two musical excerpts of the onsite experiment. The dots indicate that medians and vertical lines are the 95% CI for both listening locations. On the left are the three idealized references C1-C3, in the middle three realistic configurations C4-C6, and on the right is the anchor C7.

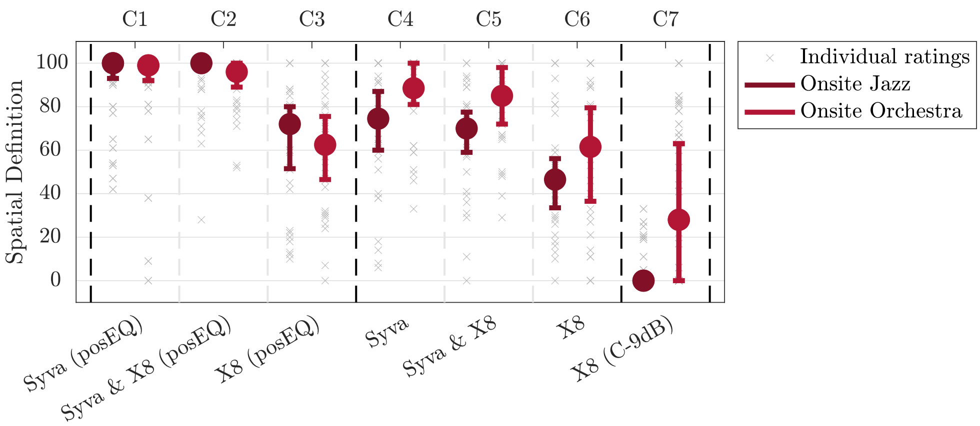


Figure A.6: Spatial definition ratings for the two musical excerpts of the onsite experiment. The dots indicate that medians and vertical lines are the 95% CI for both listening locations. On the left are the three idealized references C1-C3, in the middle three realistic configurations C4-C6, and on the right is the anchor C7.

Headphone experiment ratings (musical excerpt)

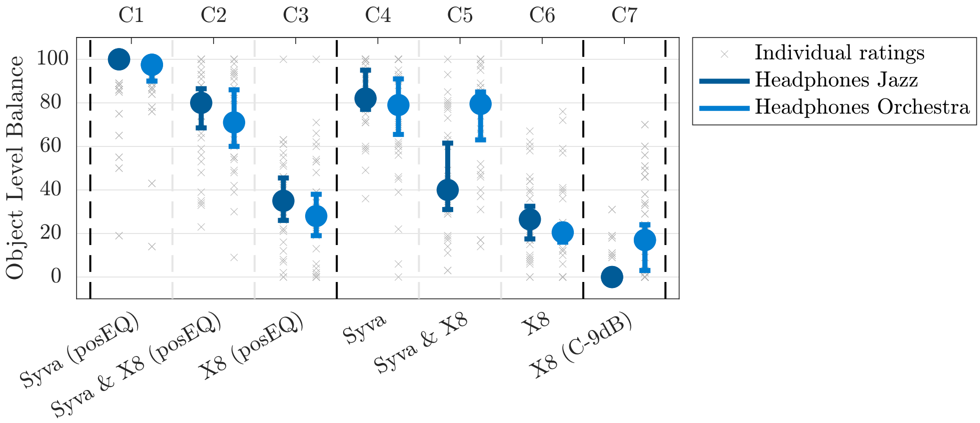


Figure A.7: Object level balance ratings for the two musical excerpts of the headphone experiment. The dots indicate that medians and vertical lines are the 95% CI for both listening locations. On the left are the three idealized references C1-C3, in the middle three realistic configurations C4-C6, and on the right is the anchor C7.

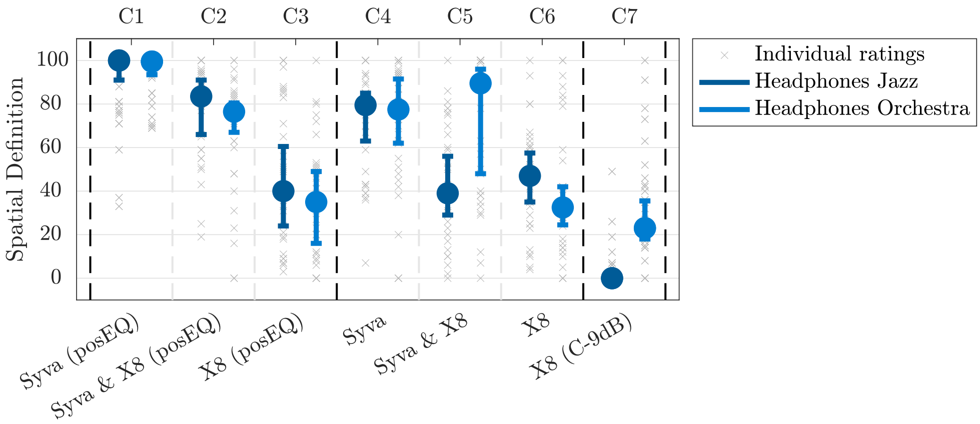


Figure A.8: Spatial definition ratings for the two musical excerpts of the onsite experiment. The dots indicate that medians and vertical lines are the 95% CI for both listening locations. On the left are the three idealized references C1-C3, in the middle three realistic configurations C4-C6, and on the right is the anchor C7.

A.5 Ratings relation to acoustic metrics (headphone experiment)

This section displays the linear regression of the FS ratio, DD ratio and ILD to the headphone experiment ratings of the object level balance, spatial definition and spatial envelopment.

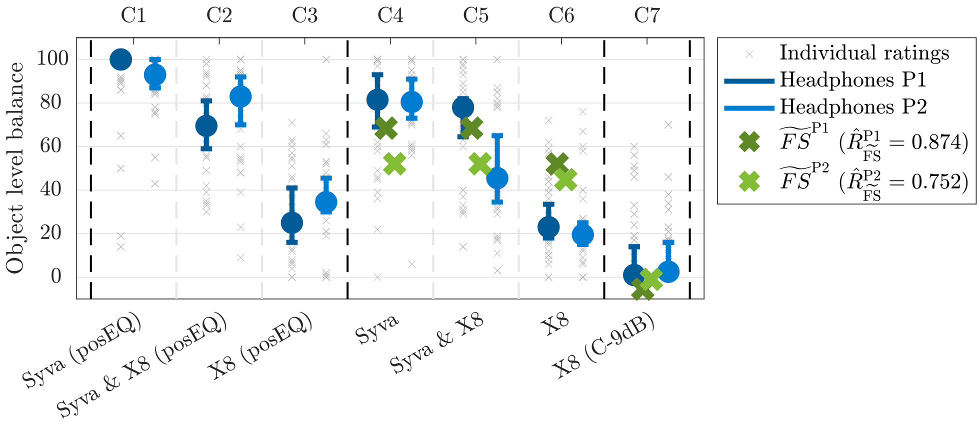


Figure A.9: Comparison of the linear regression results of the FS ratio for configurations C4-C7 with the object level balance ratings from the headphone experiment. The darker green cross indicates P1, and the brighter cross marks P2.

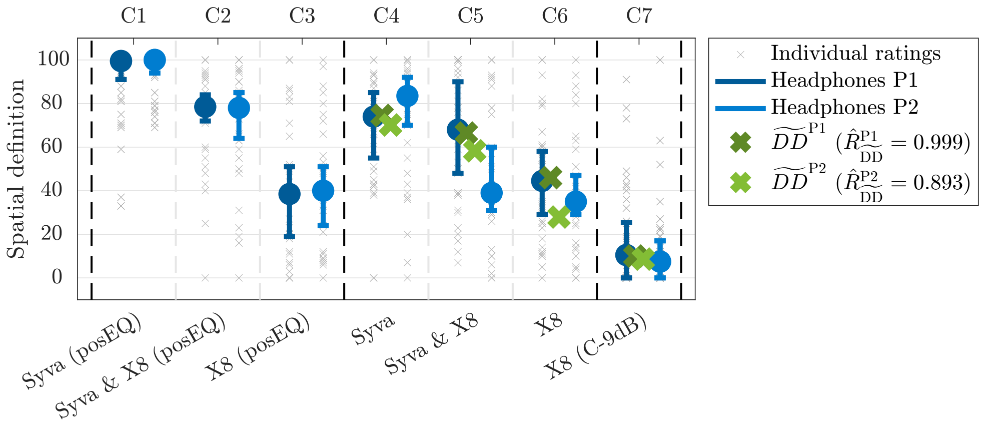


Figure A.10: Comparison of the linear regression results of the DD ratio for configurations C4-C7 with the spatial definition ratings from the headphone experiment. The darker green cross indicates P1, and the brighter cross marks P2.

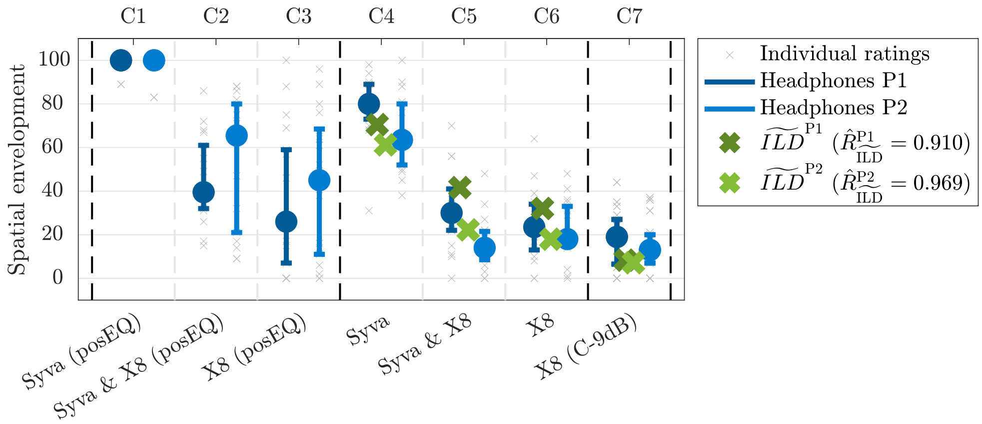


Figure A.11: Comparison of the linear regression results of the ILD for configurations C4-C7 with the spatial envelopment ratings from the headphone experiment. The darker green cross indicates P1, and the brighter cross marks P2.

Bibliography

[1] D. Howard and J. Angus, *Acoustics and Psychoacoustics*. Routledge, 2013.

[2] C. Heil and M. Urban, “Sound fields radiated by multiple sound sources arrays,” in *Audio Engineering Society Convention 92*, 1992.

[3] M. Urban, C. Heil, and P. Bauman, “Wavefront Sculpture Technology,” *Journal of the audio engineering society*, vol. 51, no. 10, pp. 912–932, Oct. 2003.

[4] D. Davis and E. Patronis, *Sound System Engineering*. Routledge, 2012. doi: 10.4324/9780080959603.

[5] B. McCarthy, *Sound Systems: Design and Optimization: Modern Techniques and Tools for Sound System Design and Alignment*. Routledge, 2012. doi: 10.4324/9780080467849.

[6] J. Blauert, *Spatial hearing: the psychophysics of human sound localization*. MIT press, 1996. doi: 10.7551/6391.001.0001.

[7] P. Newell and K. Holland, *Loudspeakers: For Music Recording and Reproduction*. Routledge, 2018. doi: 10.4324/9781315149202.

[8] V. Pulkki, “Virtual Sound Source Positioning using Vector Base Amplitude Panning,” *Journal of the Audio Engineering Society*, vol. 45, no. 6, pp. 456–466, Jun. 1997.

[9] J. Paterson and H. Lee, *3D Audio*. Routledge, 2021.

[10] P. Coleman, A. Franck, P. Jackson, R. Hughes, L. Remaggi, and F. Melchior, “Object-Based Reverberation for Spatial Audio,” *Journal of the Audio Engineering Society*, vol. 65, no. 1/2, pp. 66–77, Feb. 2017, doi: 10.17743/jaes.2016.0059.

[11] L-Acoustics, “L-ISA Hyperreal Sound: Why Amplitude-Based Algorithms Are Best Suited to Frontal Loudspeaker Configurations,” 2020.

[12] F. Zotter, S. Riedel, L. Gölles, and M. Frank, “Acceptable imbalance of sound-object levels for off-center listeners in immersive sound reinforcement,” *Fortschritte der Akustik, DAGA, Hamburg*, 2023.

[13] “Electroacoustics – Sound level meters – Part 2: Pattern evaluation tests.” [Online]. Available: https://webstore.iec.ch/publication/3525

[14] L. Gölles, F. M., and Z. F., “Improving Surround Sound Reinforcement at Off-Center Listening Positions with Miniature Line Arrays,” 2024.

[15] L. Gölles, M. Frank, and F. Zotter, “Simulating the Sweet Area of Immersive Sound Reinforcement with Surrounding Mini Line Arrays,” 2024.

[16] L. Gölles, F. Zotter, and L. Merkel, “Miniature line array for immersive sound reinforcement,” in *Audio Engineering Society Conference: AES 2023 International Conference on Spatial and Immersive Audio*, 2023.

[17] S. Riedel and F. Zotter, “Surrounding line sources optimally reproduce diffuse envelopment at off-center listening positions,” *JASA Express Letters*, vol. 2, no. 9, Sep. 2022, doi: 10.1121/10.0014168.

[18] S. Riedel, M. Frank, F. Zotter, and R. Sazdov, “A Study on Loudspeaker SPL Decays for Envelopment and Engulfment across an Extended Audience,” *Journal of the Audio Engineering Society*, no. 6, Jan. 2024.

[19] L. Gölles and F. Zotter, “Dual-Target Design for Large-Scale Sound Reinforcement: Simulation and Evaluation,” 2024.

[20] L. Gölles, M. Frank, and F. Zotter, “Evaluating a Dual-Target Line-Array Design for Medium-Scale Surround Sound Reinforcement,” 2024.

[21] A. Lindau, V. Erbes, S. Lepa, H.-J. Maempel, F. Brinkmann, and S. Weinzierl, “A Spatial Audio Quality Inventory (SAQI).” Technische Universität Berlin, 2014. doi: 10.14279/DEPOSITONCE-15275.

[22] L. Kinsler, A. Frey, A. Coppens, and J. Sanders, *Fundamentals of Acoustics*, 4th ed. New York: Wiley, 2000.

[23] D. B. Keele Jr., “What's so Sacred about Exponential Horns?,” *Journal of the Audio Engineering Society*, no. 1038, 1975.

[24] E. Geddes and D. Clark, “Computer Simulation of Horn-Loaded Compression Drivers,” *Journal of the Audio Engineering Society*, vol. 35, no. 7/8, pp. 556–566, 1987.

[25] P. D. Bauman, A. B. Adamson, and E. R. Geddes, “Acoustic Waveguides - In Practice,” *Journal of the Audio Engineering Society*, vol. 41, no. 6, pp. 462–470, 1993.

[26] M. Ureda, “Line Arrays: Theory and Applications,” in *Proceedings of the 110th Audio Engineering Society Convention*, Amsterdam, Netherlands, 2001.

[27] L. Gölles and F. Zotter, “Theory of continuously curved and phased line sources for sound reinforcement,” *Acta Acustica*, vol. 7, p. 52, 2023, doi: 10.1051/aacus/2023045.

[28] M. Ureda, “"J" and "Spiral" Line Arrays,” *Journal of the Audio Engineering Society*, no. 5485, Nov. 2001.

[29] W. Ahnert and D. Noy, *Sound Reinforcement for Audio Engineers*. Focal Press, 2023. doi: 10.4324/9781003220268.

[30] K. Brandenburg, S. Brix, and T. Sporer, “Wave field synthesis: New possibilities for large-scale immersive sound reinforcement,” *Fraunhofer IDMT & Ilmenau Tech University, Mo5 E*, vol. 1, pp. 1–507, 2004.

[31] H. Frisch, “Sound Reinforcement,” in *Handbook of Engineering Acoustics*, Springer Berlin Heidelberg, 2013, pp. 335–351. doi: 10.1007/978-3-540-69460-1_13.

[32] J. Herre and S. R. Quackenbush, “MPEG-H 3D Audio: Immersive Audio Coding,” *Acoustical Science and Technology*, vol. 43, no. 2, pp. 143–148, Mar. 2022, doi: 10.1250/ast.43.143.

[33] L-Acoustics, “Immersive Hyperreal Sound System Design.” Accessed: Jan. 18, 2025. [Online]. Available: https://l-isa.l-acoustics.com/technology/system-design/immersive-hyperreal-sound/

[34] L-Acoustics, “L-ISA Hyperreal Sound: Optimizing Coverage and Spatial Rendering for Fill Systems,” 2021.

[35] L-Acoustics, “Extraordinary Sound Spaces: Product Catalog.” 2019. Accessed: Jan. 16, 2025. [Online]. Available: https://l-isa.l-acoustics.com/wp-content/uploads/2019/05/L-Acoustics-Creations_Extraordinary-Sound-Space-Brochure_Web.pdf

[36] International Telecommunication Union, Radiocommunication Sector (ITU-R), “Recommendation ITU-R BS.1534-3: Method for the subjective assessment of intermediate quality level of audio systems (MUSHRA),” Geneva, Switzerland, 2015. [Online]. Available: https://www.itu.int/rec/R-REC-BS.1534-3-201510-I/en

[37] J. Meyer, *Akustik und musikalische Aufführungspraxis*, 4th Edition. in Fachbuchreihe Das Musikinstrument. Kassel, Germany: Bosse Verlag, 1989.

[38] T. Holman, *Surround Sound: Up and Running*, 2nd ed. Boston: Focal Press, 2008. [Online]. Available: https://www.routledge.com/Surround-Sound-Up-and-running/Holman/p/book/9780240808291

[39] D. Griesinger, “The Theory and Practice of Perceptual Modeling – How to Use Electronic Reverberation to Add Depth and Envelopment Without Reducing Clarity,” in *Proceedings of the Tonmeister Conference*, Hannover, Germany, 2000.

[40] L-Acoustics, “Optimum Measurement Locations.” Accessed: Jan. 17, 2025. [Online]. Available: https://www.l-acoustics.com/wp-content/uploads/2021/07/Optimum_measurement_locations.pdf

[41] J. Imtiaz, J. Jasperneite, and L. Han, “A performance study of Ethernet Audio Video Bridging (AVB) for Industrial real-time communication,” in *2009 IEEE Conference on Emerging Technologies & Factory Automation*, IEEE, Sep. 2009, pp. 1–8. doi: 10.1109/etfa.2009.5347126.

[42] L. Merkel, “Repository of measured RIR and BRIR of the experimental setup, loudspeaker equalization filters, and anonymized ratings of both conducted listening experiments.” [Online]. Available: https://phaidra.kug.ac.at/view/o:135786

[43] GRAS Sound & Vibration, “KEMAR - Knowles Electronics Manikin for Acoustic Research.” Accessed: Sep. 14, 2024. [Online]. Available: https://www.grasacoustics.com/products/kemar

[44] D. Satongar, C. Pike, Y. W. Lam, and A. I. Tew, “The Influence of Headphones on the Localization of External Loudspeaker Sources,” *Journal of the Audio Engineering Society*, vol. 61, no. 1/2, pp. 31–40, 2013.

[45] C. Schneiderwind, A. Neidhardt, and D. Meyer, “Comparing the Effect of Different Open Headphone Models on the Perception of a Real Sound Source,” in *Proceedings of the 150th Audio Engineering Society Convention*, Online, 2021.

[46] A. Mülleder and F. Zotter, “Ultralight Circumaural Open Headphones,” in *Proceedings of the 154th Audio Engineering Society Convention*, Espoo, Finland, 2023.

[47] T. J. DiCiccio and B. Efron, “Bootstrap confidence intervals,” *Statistical Science*, vol. 11, no. 3, Sep. 1996, doi: 10.1214/ss/1032280214.

[48] H. Abdi, “The Bonferonni and Šidák Corrections for Multiple Comparisons,” *Encyclopedia of measurement and statistics*, vol. 3, 2007.

[49] J. D. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference*. Chapman, Hall/CRC, 2010. doi: 10.1201/9781439896129.

[50] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. New York: Lawrence Erlbaum Associates, 1988.

[51] D. Lakens, “Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs,” *Frontiers in Psychology*, vol. 4, 2013, doi: 10.3389/fpsyg.2013.00863.

[52] C. Sylvain and W. Florian, “Extraction of auditory features and elicitation of attributes for the assessment of multichannel reproduced sound,” *Journal of the Audio Engineering Society*, vol. 54, no. 9, pp. 815–826, Sep. 2006.

[53] M. Vorländer, *Auralization - Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. Springer Berlin Heidelberg, 2008. doi: 10.1007/978-3-540-48830-9.