Audio Engineering Society

Conference Paper 6

Presented at the International Conference on Acoustics & Sound Reinforcement
2024 January 23–26, Le Mans, France

A Study on Loudspeaker SPL Decays for Envelopment and Engulfment across an Extended Audience

Stefan Riedel[1], Matthias Frank[1], Franz Zotter[1], and Robert Sazdov[2]

[1]*Institute of Electcronic Music and Acoustics, University of Music and Performing Arts, Graz, Austria*
[2]*School of Communication, Arts and Social Sciences, University of Technology, Sydney, Australia*

Correspondence should be addressed to Stefan Riedel (riedel@iem.at)

ABSTRACT

Listener envelopment and listener engulfment refer to the sensations of being 'surrounded by sound' and 'being covered by sound', respectively. In multichannel loudspeaker arrangements, listeners at off-center seats typically experience a reduced sensation of envelopment and engulfment due to a directional imbalance towards nearby loudspeakers. The experiment presented in this study investigates the effect of different loudspeaker sound pressure level (SPL) decay profiles on the off-center distance limit, at which envelopment or engulfment break down. Three different profiles are considered: 0, -3, and -6 dB SPL decay per doubling of distance, simulated by controlling the levels of point-source loudspeakers based on the listener position. The experiment results indicate a significant expansion of the off-center limit of envelopment when horizontally surrounding loudspeakers exhibit a -3 dB SPL decay. Regarding engulfment, the experiment shows that the off-center limit is expanded by a wide distribution of height loudspeakers that covers the entire audience area. A computational model confirms that the optimal loudspeaker SPL decay for envelopment is the one that minimizes the interaural level difference (ILD) and interaural coherence (IC) over an extended area. An interesting finding from simulations is that purely lateral multichannel arrangements can benefit from a 0 dB rather than -3 dB SPL decay per doubling of distance.

1 Introduction

Listener envelopment (LEV) and listener engulfment (LEG) are perceptual attributes that characterize a listener's spatial impression [1, 2, 3, 4]. The attributes are commonly defined as [2, 3, 5, 6]:

- Envelopment: being surrounded by sound,
- Engulfment: being covered by sound.

Previous experiments have demonstrated how envelopment [7, 8, 5] and engulfment [3, 6] are affected by the number of sound sources, their directional distribution, and the source signal characteristics. For uncorrelated, 1.8 kHz lowpass-filtered noise signals, and listeners centered in the loudspeaker arrangement, as few as four horizontal sound sources are sufficient to create an enveloping impression that is similar to the impression rendered by a dense circular or hemispherical setup, [6, 7]. For broadband noise signals, which are localized more precisely by human listeners due to additional high-frequency cues [9], 12 circularly distributed sound sources are required to yield perceptual similarity to a surrounding 24-loudspeaker reference

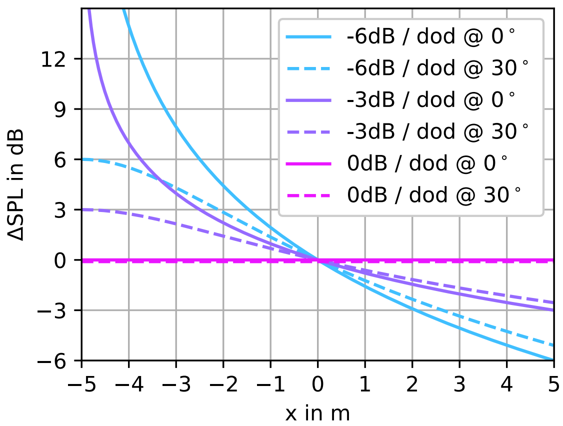


Fig. 1: Effect of source SPL decay on ΔSPL coverage relative to the audience center for an omnidirectional source at 0° versus 30° elevation. The corresponding locations are $\boldsymbol{r}_0 = [-5,0,0]$ m and $\boldsymbol{r}_0 = [-5,0,2.9]$ m relative to the center.

setup [5, 6, 7, 10]. When broadband reverberant signals are reproduced over a 3D, spherical arrangement of sources, at least 12 reverberant sources are required to yield perceptual responses that cannot be discriminated from a 64-source reference [11].

In sound reinforcement applications, such as concert venues and movie theaters, many listeners in the audience are located off-center, where perceived envelopment can degrade due to directional imbalances [5, 8, 12, 13]. Point-source loudspeakers at ear-height level lead to directional imbalances at off-center listening positions, which reduces listener envelopment [5, 13]. The sound pressure level (SPL) decay of the individual loudspeakers is an important factor in this matter, and it is -6 dB per doubling of distance (dod) for the typical point sources. A recent simulation study showed that the degradation in envelopment can be reduced by placing a point-source surround layer at 30° elevation [14], which is often seen in movie theaters to avoid directional imbalances for the outer seats. Figure 1 shows that the relative level change for an elevated point source (-6 dB/dod @ 30°) is indeed similar to a line source at ear height (-3 dB/dod @ 0°). The increased loudspeaker height represents a trade-off because broadband signals will be perceived as elevated, however this elevation percept would not be as obvious for natural signals or lowpass-signals [6, 15].

This paper is concerned with the possible improvements in diffuse-field rendering using surrounding line-source loudspeakers. Line source arrays can be designed to meet various sound pressure level (SPL) coverage targets based on their curvature [16, 17, 18], making them highly suitable for immersive sound reinforcement applications [5, 12]. In large-scale sound reinforcement, line arrays are the de facto standard to achieve a consistently flat direct SPL across the audience area [16, 19]. The proposed research question can be formulated as: Which kind of SPL decay, 0 dB, -3 dB, or -6 dB per doubling of distance (dod) optimally renders envelopment / engulfment across a large audience area with loudspeakers at different elevation layers (0°, 30°, or 60°)?

To address the proposed research question, we simulate three variants of direct SPL decay in our experiment, namely 0 dB, -3 dB, or -6 dB per doubling of the distance (dod). This is achieved via a real-time gain compensation of the individual loudspeaker signals, controlled by the listener's position within the loudspeaker array, assuming a point-source SPL decay (-6 dB/dod) of the installed loudspeakers (d&B 12S-D and 8S).

2 Listening Experiment

A listening experiment was conducted in which participants were asked to walk from the center towards the loudspeaker direction at $-70°$, while maintaining a frontal look direction, and they had to log the location at which they perceived that envelopment/engulfment was noticeably impaired in comparison to the central listening position.

2.1 Experiment Setup

The listening experiment was conducted at the IEM CUBE ($T_{60} = 0.5\,\mathrm{s}$), a large academic reproduction studio with a size of approximately $A = 120\,\mathrm{m}^2$. The installed loudspeaker hemisphere is composed of 12 d&B 12S-D at ear-height level (layer L1), eight d&B 8S at 30° elevation (L2), and five d&B 8S at $\geq 60°$ elevation (L3), cf. Fig. 2. The loudspeakers were individually equalized in third-octave bands by minimum-phase FIR filters, to compensate for spectral differences and level differences at the central listening position. The SPL decay of the installed point-source loudspeakers can be assumed to be -6 dB per doubling of distance.

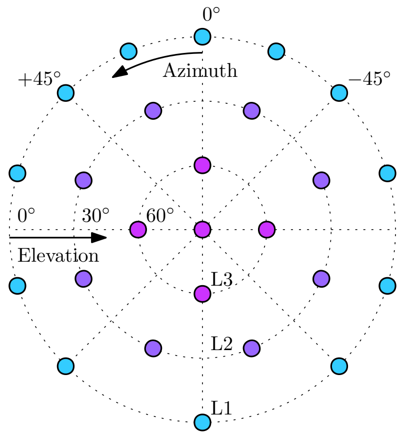


Fig. 2: Loudspeaker direction set used in this study. Elevation layers are defined as L1 (0°), L2 (30°), and L3 ($\geq$ 60°).

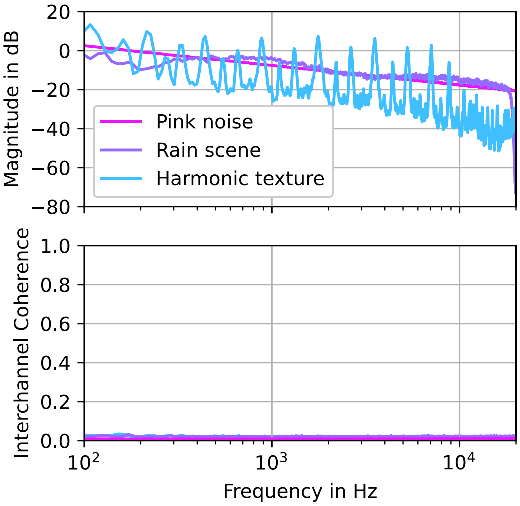


Fig. 3: Average magnitude spectra (top) and average interchannel magnitude-squared coherence (bottom) of the multichannel stimuli.

During the experiment, the participants wore a cap fitted with motion-capture markers, which enabled position-tracking via an OptiTrack system. Tracking was necessary to render different level compensations in real time [12]. Additionally, it allowed to electronically capture the position where participants would indicate the sensation of envelopment/engulfment to be impaired. The motion-capture cap was fitted tightly and did not occlude the ears, such that the head-related transfer functions of the participants were not affected.

2.2 Experiment Design

Level compensations to simulate different SPL decays using point-source loudspeakers were implemented. The signal gains σ_i allow to render a 0 dB, -3 dB, or -6 dB direct SPL decay per doubling of distance:

$$\sigma_i(\boldsymbol{r}) = \left(\frac{\|\boldsymbol{r} - \boldsymbol{r}_i\|}{\|\boldsymbol{r}_i\|} \right)^{1-\beta}, \qquad (1)$$

where $\boldsymbol{r} \in \mathbb{R}^3$ is the position of the listener, $\boldsymbol{r}_i$ is the position of the i-th loudspeaker, and $1-\beta$ is the decay exponent. For $\beta = 1$ no compensation is applied (-6 dB SPL decay), for $\beta = 0.5$ a 'half' compensation is applied (-3 dB SPL decay), and for $\beta = 0$ a 'full' level compensation is applied (0 dB SPL decay). Note that for a centered listener ($\boldsymbol{r} = \boldsymbol{0}$) the expression evaluates to $\sigma_i(\boldsymbol{0}) = 1$ for any value of β. The position-interactive compensation was implemented in the *PureData* (PD) real-time processing language.

Three different multichannel sound stimuli were pre-rendered for the experiment: (1) uncorrelated pink noise signals, (2) a diffuse rain scene, and (3) a harmonic texture stimulus. The multichannel pink noise signals were generated as independent noises in *Python*. The multichannel rain scene and harmonic stimulus were created using the IEM GranularEncoder [6], rendering uncorrelated full-sphere 5^{th}-order Ambisonic signals from stereo sound files, which could then be decoded to the 25-channel loudspeaker setup with the AllRAD (Allround Ambisonic Decoding) plug-in [20]. The rain scene and the harmonic scene were included as ecologically valid stimuli in the context of spatial audio, both in terms of their spectral and spatio-temporal characteristics (cf. audio files online [21]). Figure 3 shows channel-averaged magnitude spectra and the average interchannel magnitude-squared coherence of the multichannel stimuli. The calculation employed Welch's method [22] to average the cross- and autospectra of the signal channels over 128 signal blocks (4096 samples block length, Hann-windowed, 50% overlap, sampling rate of 44.1 kHz).

Different loudspeaker layers are considered to reproduce the multichannel stimuli. Combinations of the available loudspeaker layers (L1, L2, and L3) are denoted by concatenated abbreviations, e.g. L1L2, when they are unions of two layers. Two sets were rated per attribute: L1 and L1L2 for envelopment, and L3 and L2L3 for engulfment. The sets L1 and L1L2 were chosen for envelopment as they reproduce surrounding impressions, cf. Fig. 2, whereas L3 and L2L3 are capable of covering a centered listener by sound from above [6]. All stimulus signals were extracted from the 25-channel audio files and loudness-normalized by dividing the signals by the square root of the number of active loudspeakers (e.g. the gain would be $g = 1/\sqrt{12}$ in case of L1, or $g = 1/\sqrt{20}$ in case of L1L2).

2.3 Results

Twelve experienced listeners participated in the experiment ($N = 12$), who were either staff or students of the authors' institution. Figure 4 (top) shows the experimental results for envelopment (left) and engulfment (right). A three-way analysis of variance (ANOVA) was performed using MATLAB. Results indicate significant main effects of compensation/layer condition, sound stimulus, and participant for both envelopment and engulfment (all $p < 0.001$). No significant interaction between sound stimulus and compensation/layer is found for envelopment or engulfment ($p = 0.981$ and $p = 0.087$ respectively). The results of the ANOVA confirm that the level compensation/layer and the sound stimulus significantly affect the off-center limits for envelopment and engulfment.

In order to focus on the effects of loudspeaker layer and compensation of SPL decay, the data is pooled across the different sound stimuli, cf. Fig. 4 (bottom). Pooling of data across the sound stimuli seems valid here, as the ANOVA did not indicate a significant interaction between sound stimulus and the compensation/layer condition. Below, the statistical analysis proceeds non-parametrically to increase robustness against outliers within the limited sample size.

Regarding envelopment, the median off-center limit is largest for the half compensation conditions (H-L1 and H-L1L2), cf. Fig. 4 (bottom left). In fact, the off-center limit is roughly 50% larger than for the other compensations for both the L1 layer and the L1L2 layer combination, increasing from around 2 meters to around 3 meters. This finding is confirmed by highly significant differences in pairwise Wilcoxon signed-rank tests, see Tab. 1. Interestingly, for the L1L2 layer combination, applying no compensation is significantly better than the full compensation. However, the half-compensation scheme still gives the largest off-center limit.

Table 1: Bonferroni-Holm corrected p-values for pairwise Wilcoxon signed-rank tests between compensation conditions. Bold numbers indicate $p < 0.05$. Full (F), Half (H), and No (N) compensation refer to 0 dB, -3 dB and -6 dB SPL coverage per doubling of distance.

	Envelopment		
	No vs. Half	No vs. Full	Half vs. Full
L1	$<$ **0.001**	0.912	$<$ **0.001**
L1L2	**0.011**	**0.002**	$<$ **0.001**
	Engulfment		
	No vs. Half	No vs. Full	Half vs. Full
L3	1.000	1.000	1.000
L2L3	**0.001**	$<$ **0.001**	**0.008**

Table 2: Bonferroni-Holm corrected p-values for pairwise Wilcoxon signed-rank tests between layer conditions applying no loudspeaker level compensation (N). Bold numbers indicate $p < 0.05$.

	Envelopment L1 vs. L1L2	Engulfment L3 vs. L2L3
No (N)	**0.003**	$<$ **0.001**

Regarding engulfment, no significant differences between the compensations for the L3 layer are observed, as all compensations yield a rather small limit of around 1 meter, cf. Fig. 4 (bottom right). For the L2L3 layer combination, applying no compensation is significantly better than half or full compensation, and increases the median off-center limit to around 2.5 meters. When no compensation is applied, the conditions which include the L2 height layer lead to significant improvements. For example, the L1L2 condition results in a significantly larger off-center limit for envelopment compared to L1, cf. Tab. 2. Similarly, the L2L3 layer combination results in a significant improvement over L3 for engulfment when no level compensation is applied.

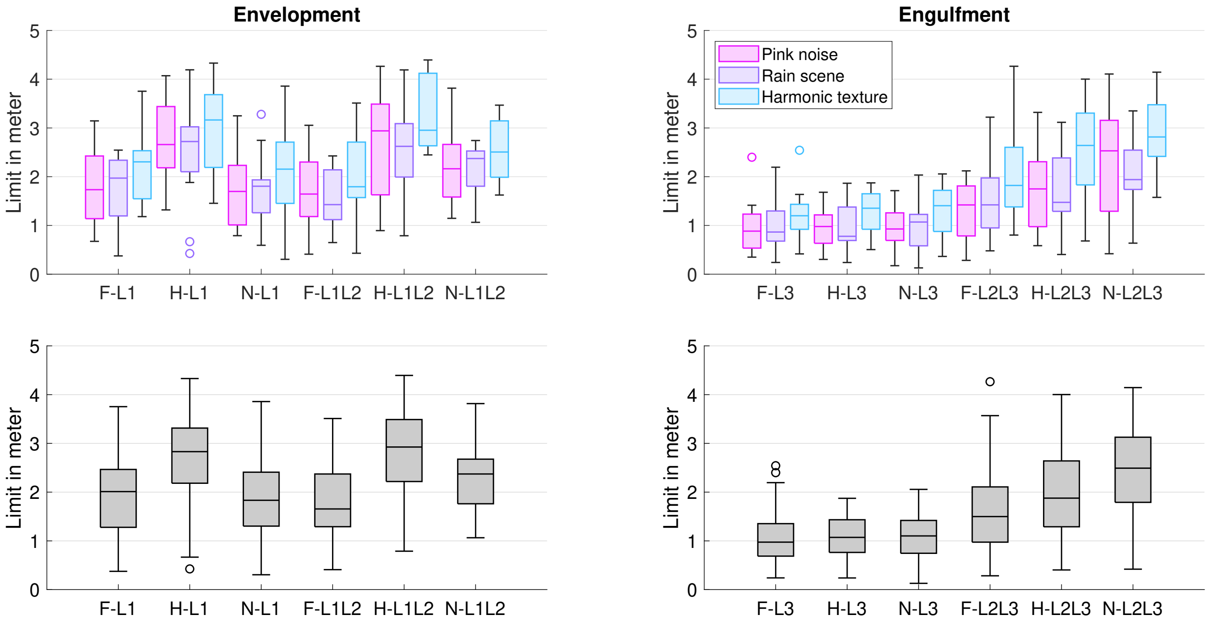


Fig. 4: Experimental off-center limits for envelopment (left) and engulfment (right) for $N = 12$ participants. Data is shown before (top) and after pooling across different sound stimuli (bottom). Full (F), Half (H), or No (N) level compensation simulate loudspeaker SPL decays of 0 dB, -3 dB, or -6 dB per doubling of distance. The loudspeaker layer abbreviations refer to elevation levels L1 (0°), L2 (30°), and L3 ($\geq$ 60°).

2.4 Discussion

The results of the experiment indicate that horizontally surrounding loudspeakers that exhibit a -3 dB SPL decay per doubling of distance optimally preserve envelopment over an extended audience in a large studio environment ($T_{60} = 0.5$ s). This result is in line with previous results obtained for a circular array under anechoic conditions [5]. The study presented in this paper extends to non-circular loudspeaker arrays, as e.g. the L1 layer is not an equiangular arrangement of sources. An interesting finding is that height layers contribute well to envelopment and engulfment when the loudspeakers at 30° elevation or higher exhibit a -6 dB SPL decay. This is likely due to the fact that elevated point-source loudspeakers are capable of an SPL coverage similar to horizontal line sources in the nearby part of the listening area, cf. Fig. 1 (dashed blue vs. purple line between -4 meters to 0 meters). This also explains the results for engulfment, which showed the largest off-center limit for the L2L3 layer combination, given that no level compensation to the inherent -6 dB SPL decay is applied.

While it is expected that the results of this experiment apply to a large extent to actual line-source loudspeakers that can accomplish the respective 0 dB/dod and -3 dB/dod coverage by their curvature design [16], the real-time interactive implementation with the d&B 12S-D and 8S point-source loudspeakers has subtle differences. In particular, line-source loudspeakers would neither excite the reverberant field interactively with the listener position, nor with the lower directivity of point-source loudspeakers. Nonetheless, the results are generally in line with the previous anechoic study [5], and it can be assumed that the diffuse reverberant energy of the room enhances envelopment for all loudspeaker SPL decay variations.

3 Model-based Evaluation

This section presents a sound field model to compute the interaural level differences (ILD) and interaural coherence (IC) of a listener in an arbitrary loudspeaker arrangement, which enables to predict the impairment of envelopment and engulfment at off-center listening positions [5, 23]. The model is evaluated on the data collected in the presented experiment, and subsequently allows to investigate other multichannel loudspeaker arrangements not included so far.

3.1 ILD and IC Computation

The computation of ILD and IC is based on a spherical harmonic (SH) sound field model previously presented by the authors [14]. The model description is adapted to account for the level compensation scheme applied to point-source loudspeakers in a room, as employed in the presented experiment. The model assumes source signals s_i, which arrive at the listener from directions $\boldsymbol{\Omega}_i = (\phi_i, \theta_i)$, where ϕ_i and θ_i refer to the azimuth and elevation of the i-th source with respect to the listener. The signal variances are determined by the level compensation $\sigma_i(\boldsymbol{r}) = (\|\boldsymbol{r}-\boldsymbol{r}_i\|/\|\boldsymbol{r}_i\|)^{1-\beta}$, see Eq. 1. To model the point-source SPL decay of the loudspeakers, the weights $1/\|\boldsymbol{r}-\boldsymbol{r}_i\|$ are applied, where $\|\boldsymbol{r}-\boldsymbol{r}_i\|$ is the distance between the i-th loudspeaker and the listener. Additionally, the far-field directivity $\Gamma(\boldsymbol{\Omega}_i)$ of the sources could be incorporated, which results in direct-sound weights $g_i = \Gamma(\boldsymbol{\Omega}_i)\cdot 1/\|\boldsymbol{r}-\boldsymbol{r}_i\|$.

The order-N Ambisonic sound field $\boldsymbol{\chi}(t) \in \mathbb{R}^{(\mathrm{N}+1)^2\times 1}$ is composed of the direct sound and an isotropic, diffuse component $\boldsymbol{n}(t) \in \mathbb{R}^{(\mathrm{N}+1)^2\times 1}$:

$$\boldsymbol{\chi}(t) = \sum_i g_i s_i(t)\boldsymbol{y}(\boldsymbol{\Omega}_i) + \boldsymbol{n}(t) \tag{2}$$

$$= \boldsymbol{Y}\mathrm{diag}\{\boldsymbol{g}\}\boldsymbol{s}(t) + \boldsymbol{n}(t)\,, \tag{3}$$

where $\boldsymbol{Y} = [\boldsymbol{y}(\boldsymbol{\Omega}_1)\dots\boldsymbol{y}(\boldsymbol{\Omega}_\mathrm{I})]^{(\mathrm{N}+1)^2\times \mathrm{I}}$ is a matrix of real-valued spherical harmonics (SH) evaluated at the source directions. Ear signals are obtained via convolution with binaural decoding filters $\boldsymbol{h}_l(\omega)$ and $\boldsymbol{h}_r(\omega) \in \mathbb{C}^{(\mathrm{N}+1)^2\times 1}$:

$$x_l(\omega) = \boldsymbol{h}_l^\mathsf{T}(\omega)\boldsymbol{\chi}(\omega) = \boldsymbol{\chi}^\mathsf{T}(\omega)\boldsymbol{h}_l(\omega)\,, \tag{4}$$

$$x_r(\omega) = \boldsymbol{h}_r^\mathsf{T}(\omega)\boldsymbol{\chi}(\omega)\,, \tag{5}$$

where ω denotes the radial frequency. The cross-spectrum of the ear signals is

$$S_{lr}(\omega) = E\{x_r^*(\omega)x_l(\omega)\} = \boldsymbol{h}_r^\mathsf{H}(\omega)E\{\overline{\boldsymbol{\chi}}\boldsymbol{\chi}^\mathsf{T}\}\boldsymbol{h}_l(\omega) \tag{6}$$

$$= \boldsymbol{h}_r^\mathsf{H}(\omega)\boldsymbol{C}_{xx}(\omega)\boldsymbol{h}_l(\omega)\,, \tag{7}$$

where $E\{\cdot\}$ denotes expectation over time, $(\cdot)^\mathsf{H}$ denotes Hermitian transposition, and $\overline{(\cdot)}$ denotes element-wise conjugation. The autospectra of the ear signals are obtained similarly as

$$S_{ll}(\omega) = \boldsymbol{h}_l^\mathsf{H}(\omega)\boldsymbol{C}_{xx}(\omega)\boldsymbol{h}_l(\omega)\,, \tag{8}$$

$$S_{rr}(\omega) = \boldsymbol{h}_r^\mathsf{H}(\omega)\boldsymbol{C}_{xx}(\omega)\boldsymbol{h}_r(\omega)\,. \tag{9}$$

We can compute interaural coherence and interaural level difference as:

$$\mathrm{IC}[b] = \frac{\max_\tau\left|\int_{-\infty}^{\infty} w_b^2(\omega)S_{lr}(\omega)e^{j\omega\tau}d\omega\right|}{\sqrt{\int_{-\infty}^{\infty} w_b^2(\omega)S_{ll}(\omega)d\omega\cdot\int_{-\infty}^{\infty} w_b^2(\omega)S_{rr}(\omega)d\omega}}\,, \tag{10}$$

$$\mathrm{ILD}[b] = 10\cdot\log_{10}\left(\frac{\int_{-\infty}^{\infty} w_b^2(\omega)S_{ll}(\omega)d\omega}{\int_{-\infty}^{\infty} w_b^2(\omega)S_{rr}(\omega)d\omega}\right)\,, \tag{11}$$

where b is the frequency band index and $w_b(\omega)$ denote bandpass windows, e.g. zero-phase gammatone windows. The range for the lag τ is limited to $-1\,\mathrm{ms} \le \tau \le 1\,\mathrm{ms}$. Assuming uncorrelated source signals $E\{\bar{\boldsymbol{s}}\boldsymbol{s}^\mathsf{T}\} = \boldsymbol{C}_{ss} = \mathrm{diag}\{[\sigma_i^2]\}$, the SH covariance matrix $\boldsymbol{C}_{xx}(\omega)$ is given by

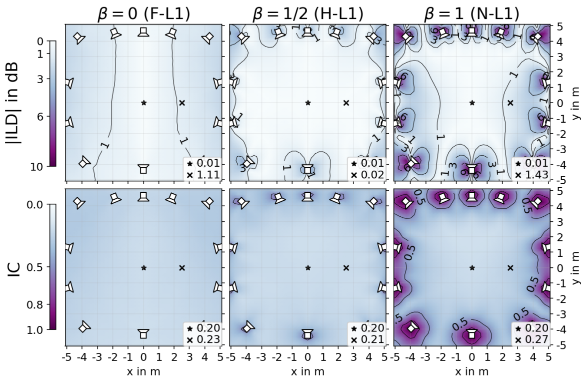
$$\boldsymbol{C}_{xx}(\omega) = \boldsymbol{Y}\mathrm{diag}\{\boldsymbol{g}\}^2\boldsymbol{C}_{ss}\boldsymbol{Y}^\mathsf{T} + \nu\boldsymbol{I}\,. \tag{12}$$

The diagonal loading models the diffuse field and is defined by acoustic parameters:

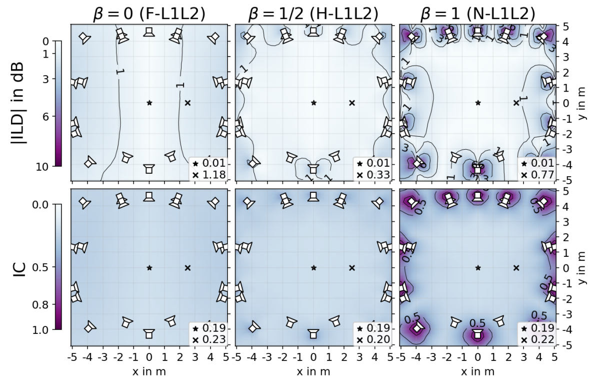
$$\nu = \sum_i \frac{\sigma_i^2}{4\pi r_{\mathrm{H},i}^2} = \sum_i \frac{\sigma_i^2\cdot T_{60}}{4\pi\cdot 0.057^2\cdot V\cdot\gamma_i}\,, \tag{13}$$

$$r_{\mathrm{H},i} = 0.057\cdot\sqrt{\frac{V}{T_{60}}}\cdot\sqrt{\gamma_i}\,, \tag{14}$$

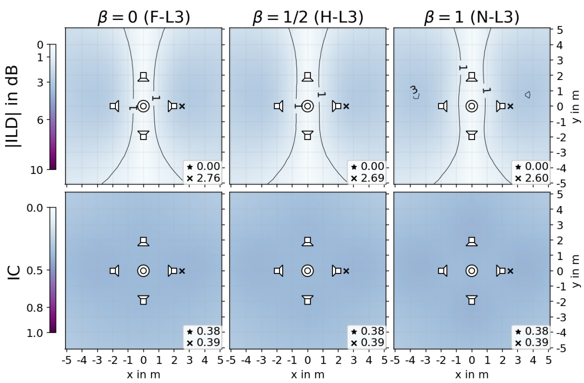
where V denotes the room volume, T_{60} denotes the reverberation time, and γ_i is the directivity factor of the i-th source. At the critical distance $r_\mathrm{H} = 0.057\cdot\sqrt{\frac{V}{T_{60}}}\cdot\sqrt{1}$ [24], the squared direct sound pressure caused by a single point source $p_1^2 = \mathrm{tr}\{\boldsymbol{Y}\mathrm{diag}\{\boldsymbol{g}\}^2\boldsymbol{C}_{ss}\boldsymbol{Y}^\mathsf{T}\} = \frac{(\mathrm{N}+1)^2}{4\pi}\cdot\frac{\sigma_1^2}{r_\mathrm{H}^2}$ is equal to the squared diffuse sound pressure modeled by uncorrelated plane waves $p_d^2 = \mathrm{tr}\{\nu\boldsymbol{I}\} = (\mathrm{N}+1)^2\cdot\frac{\sigma_1^2}{4\pi r_\mathrm{H}^2}$. The factor ν generally accounts for the diffuse sound field energy of all sources, hence the summation in Eq. 13. The directivity factor γ affects the amount of diffuse sound field energy, and the ratio between the direct and diffuse sound field energy will be further influenced by the direct-sound decay coefficient β. The far-field directivity $\Gamma(\phi,\theta)$ of the source is related to the directivity factor $\gamma = 4\pi/\int_0^{2\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\Gamma^2\cos\theta\mathrm{d}\theta\mathrm{d}\phi$ and the directivity index $\mathrm{DI} = 10\cdot\log_{10}(\gamma)$, respectively.



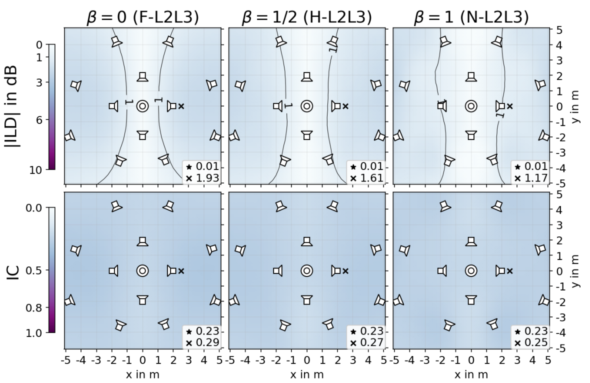
(a) L1



(b) L1L2



(c) L3



(d) L2L3

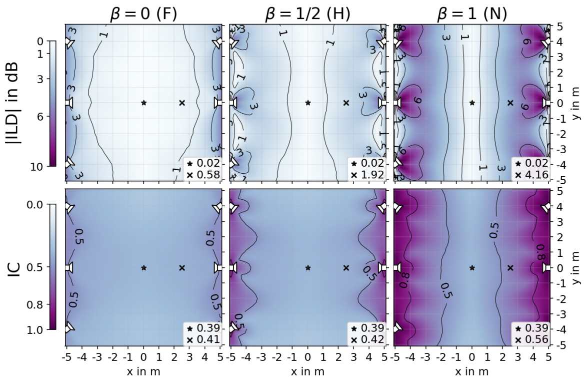
Fig. 5: Absolute interaural level difference (ILD) and interaural coherence (IC) evaluated across the listening area ($\theta = 0^\circ$, $z = 0$) for a frontal head orientation. The IEM CUBE has a $T_{60} = 0.5$ s, the directivity index of the d&B 12S-D and 8S was modeled as DI $= 10$ dB and DI $= 8$ dB. The number of loudspeakers per left-right hemifield varies at off-center positions, such that a loudspeaker SPL decay of -3 dB per doubling of distance is optimal for the layer L1, whereas the -6 dB SPL decay is optimal for the height layers.

3.2 Parameter Settings

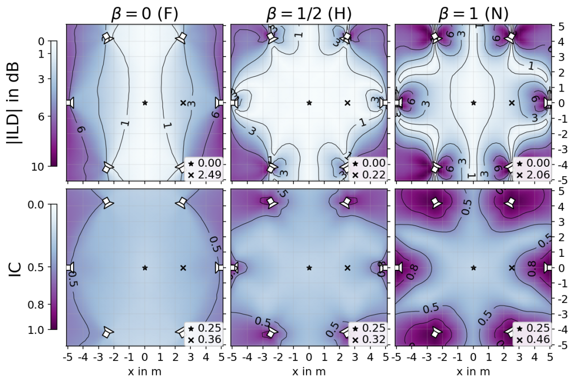
The implementation of the model uses the KU100 dummy head HRTFs [25] in a 5-th order Ambisonic representation (magnitude least-squares) [26]. The ILD and IC are computed in 40 gammatone frequency bands with center frequencies ranging from 42 Hz to 18 kHz, and are subsequently averaged in perceptually relevant frequency bands: 200 Hz to 12.8 kHz for ILD [27] and 200 Hz to 1.6 kHz for IC [28]. This yields a single, broadband value per simulated listener position and orientation. We consider only the frontal head orientation per position, to match approximately the orientation of the listener during the experiment task, and to focus on lateral imbalance. We estimate a directivity index (DI) of 10 dB for the d&B 12S-D loudspeaker and of 8 dB for the d&B 8S loudspeaker, based on the specified dispersion by the manufacturer ($110^\circ \times 55^\circ$ for 12S-D vs. $100^\circ \times 100^\circ$ for 8S). Note that the DI parameter accounts only for a correct diffuse sound energy level in our model, cf. Eqs. 12 to 14, while the direct-sound directivity Γ is an independent parameter ($\Gamma = 1$ is used as a simplification here). The reverberation time is set to $T_{60} = 0.5$ seconds as measured for the experiment room ($V = 500\,\mathrm{m}^3$).

3.3 Simulation Results & Discussion

Figure 5 shows the simulations of ILD and IC for the conditions presented in the listening experiment. Pre-



(a) 6-channel layout (lateral)



(b) 6-channel layout (circular)

Fig. 6: Absolute interaural level difference (ILD) and interaural coherence (IC) evaluated across the listening area ($\theta = 0°$, $z = 0$) for a frontal head orientation. Anechoic simulation ($T_{60} = 0$ s) comparing lateral versus equiangular arrangement of sources. Since the number of loudspeakers per left-right hemifield is maintained at off-center positions for the lateral setup, the 0 dB SPL decay is optimal in such scenarios.

vious work has established the following perceptual criteria for envelopment in loudspeaker-based reproduction: ILD ≤ 1 dB and IC ≤ 0.5 [5, 23]. The 1 dB ILD contours in Fig. 5 (a) and (b) seem to predict the off-center limit for the breakdown of envelopment reasonably well for the cases of no compensation and full SPL decay compensation (see F-L1 and N-L1 with perceptual limits measured and modeled around 2 meters). Interestingly, the model overestimates the limit for the half compensation, as the 1 dB ILD contour is around 4.5 meters off-center, cf. Fig. 5 (a) and (b), while most participants have marked their perceptual limit at around 3 meters. An explanation could be that the visual awareness of one's position in the loudspeaker array could influence the rating behaviour as a participant. It is proposed that listeners do not assume extreme off-center positions to be enveloping, even if relevant perceptual cues are present as shown by the model computations. Regarding the results for the layers L3 and L2L3, the 1 dB ILD contour predicted by the model accurately reflects the off-center limits measured in the experiment, cf. Fig. 5 (c) and (d). This demonstrates that lateral imbalances seem important cues for envelopment and engulfment, as the IC often remains below 0.5 even at off-center positions.

The model allows for the simulation of loudspeaker setups not considered in the experiment. For example, two variants of a 6-channel setup are simulated in Fig. 6. The lateral setup reveals that a 0 dB SPL decay might be preferable for lateral only setup, while for an equiangular setup, the -3 dB SPL decay is the optimum.

4 Conclusion

The findings of this study suggest that the listening area of envelopment can be increased by a direct sound pressure level (SPL) decay between -3 dB and 0 dB per doubling of distance along the audience [16]. The results suggest that -3 dB SPL decay is optimal to reproduce diffuse sound scenes on circular or rectangular surround layouts. If the loudspeaker setup is sparse and contains only few lateral surround channels to reproduce diffuse content (e.g. a 4-channel setup [7]), a 0 dB SPL decay achieves better spatial balance and envelopment at off-center positions. By means of a computational model, it is demonstrated that the optimal loudspeaker SPL decay is the one that minimizes the interaural level difference (ILD) and interaural coherence (IC) over an extended area. This model consistently explains the rating behaviour in the listening experiment above.

Regarding height loudspeaker layers the experiment showed no benefit of any level compensation for loudspeakers at elevations $\geq 30°$, which indicates that point-source loudspeakers can deliver a sufficiently consistent SPL in small- to medium-scale sound reinforcement when deployed at elevations $\geq 30°$. It could be shown that the distribution of the height loudspeakers should cover the entire audience to deliver a large listening area of diffuse engulfment.

A limitation of the presented study lies in using real-time level-compensated point-source loudspeakers, which does not accurately account for the increased vertical directivity of curved line-source loudspeakers. In particular, the increased directivity would reduce the diffuse response of the room [24], taking results closer to anechoic conditions, which has previously been studied [5]. Overall, the results of this experiment confirm findings seen under anechoic conditions [5], extending to loudspeaker arrays with height layers in an acoustically-treated listening room.

5 Data Availability

The authors provide open access to experiment data and model code [29] and to audio files of the experiment stimuli [21].

6 Acknowledgment

The authors gratefully received funding from the Austrian Science Fund (FWF): P 35254-N (Envelopment in Immersive Sound Reinforcement, EnImSo).

References

[1] Bradley, J. S. and Soulodre, G. A., "Objective measures of listener envelopment," *J. Acoust. Soc. Am.*, 98(5), pp. 2590–2597, 1995, doi: 10.1121/1.413225.

[2] Soulodre, G. A., Lavoie, M. C., and Norcross, S. G., "Objective measures of listener envelopment in multichannel surround systems," *J. Audio Eng. Soc.*, 51(9), pp. 826–840, 2003.

[3] Sazdov, R., Paine, G., and Stevens, K., "Perceptual Investigation into Envelopment, Spatial Clarity, and Engulfment in Reproduced Multi-Channel Audio," in *Proceedings of the 31st International AES Conference: New Directions in High Resolution Audio*, Audio Engineering Society, 2007.

[4] Sazdov, R., "Envelopment vs. Engulfment: Multidimensional scaling on the effect of spectral content and spatial dimension within a three-dimensional loudspeaker setup," in *Proceedings of the 3rd International Conference on Spatial Audio*, pp. 52–66, 2015.

[5] Riedel, S. and Zotter, F., "Surrounding line sources optimally reproduce diffuse envelopment at off-center listening positions," *JASA-EL*, 2(9), 2022, doi: 10.1121/10.0014168.

[6] Riedel, S., Frank, M., and Zotter, F., "The Effect of Temporal and Directional Density on Listener Envelopment," *J. Audio Eng. Soc.*, 71(7/8), pp. 455–467, 2023, doi: 10.17743/jaes.2022.0088.

[7] Hiyama, K., Komiyama, S., and Hamasaki, K., "The Minimum Number of Loudspeakers and its Arrangement for Reproducing the Spatial Impression of Diffuse Sound Field," in *Proceedings of the 113th AES Convention*, Los Angeles, 2002.

[8] Cousins, M. P., Fazi, F. M., Bleeck, S., and Melchior, F., "Subjective diffuseness in layer-based loudspeaker systems with height," in *Proceedings of the 139th AES Convention*, New York, 2015.

[9] Santala, O. and Pulkki, V., "Directional perception of distributed sound sources," *J. Acoust. Soc. Am.*, 129(3), pp. 1522–1530, 2011, doi: 10.1121/1.3533727.

[10] Sonke, J.-J., *Variable Acoustics by Wave Field Synthesis*, Ph.D. thesis, TU Delft, 2000.

[11] Kirsch, C., Poppitz, J., Wendt, T., van de Par, S., and Ewert, S. D., "Spatial resolution of late reverberation in virtual acoustic environments," *Trends in Hearing*, 25, 2021, doi: 10.1177/23312165211054924.

[12] Heidegger, P., Brands, B., Langgartner, L., and Frank, M., "Sweet Area using Ambisonics with Simulated Line Arrays," in *Fortschritte der Akustik, DAGA*, Vienna, 2020.

[13] Blochberger, M., Zotter, F., and Frank, M., "Sweet area size for the envelopment of a recursive and a non-recursive diffuseness rendering approach," in *Proceedings of the 5th International Conference on Spatial Audio*, pp. 151–157, Ilmenau, 2019.

[14] Riedel, S., Goelles, L., Frank, M., and Zotter, F., "Modeling the Listening Area of Envelopment," in *Fortschritte der Akustik (DAGA)*, pp. 1178–1181, Hamburg, 2023.

[15] Butler, R. A. and Humanski, R. A., "Localization of sound in the vertical plane with and without high-frequency spectral cues," *Perception & psychophysics*, 51(2), pp. 182–186, 1992, doi: 10.3758/BF03212242.

[16] Straube, F., Schultz, F., Bonillo, D. A., and Weinzierl, S., "An analytical approach for optimizing the curving of line source arrays," *J. Audio Eng. Soc.*, 66(1/2), pp. 4–20, 2018.

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

[18] Gölles, L., Zotter, F., and Merkel, L., "Miniature Line Array for Immersive Sound Reinforcement," in *Proceedings of the AES International Conference on Spatial and Immersive Audio*, Audio Engineering Society, 2023.

[19] Urban, M., Heil, C., and Bauman, P., "Wavefront sculpture technology," *J. Audio Eng. Soc.*, 51(10), pp. 912–932, 2003.

[20] Zotter, F. and Frank, M., "All-round ambisonic panning and decoding," *J. Audio Eng. Soc.*, 60(10), pp. 807–820, 2012.

[21] Riedel, S. et al., "Ambisonic stimulus files and binaural renders," `https://doi.org/10.5281/zenodo.10227483`, 2023.

[22] Welch, P., "The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms," *IEEE Transactions on audio and electroacoustics*, 15(2), pp. 70–73, 1967, doi: 10.1109/TAU.1967.1161901.

[23] Walther, A. and Faller, C., "Assessing diffuse sound field reproduction capabilities of multichannel playback systems," in *Proceedings of the 130th AES Convention*, London, 2011.

[24] Ahnert, W. and Steffen, F., *Sound reinforcement engineering: fundamentals and practice*, CRC Press, 2000.

[25] Bernschütz, B., "A spherical far field HRIR/HRTF compilation of the Neumann KU 100," in *Fortschritte der Akustik (AIA-DAGA)*, pp. 592–595, Merano, 2013, doi: 10.5281/zenodo.3928297.

[26] Schörkhuber, C., Zaunschirm, M., and Höldrich, R., "Binaural rendering of ambisonic signals via magnitude least squares," in *Fortschritte der Akustik (DAGA)*, pp. 339–342, München, 2018.

[27] Yost, W. A. and Dye Jr, R. H., "Discrimination of interaural differences of level as a function of frequency," *J. Acoust. Soc. Am.*, 83(5), pp. 1846–1851, 1988, doi: 10.1121/1.396520.

[28] Brughera, A., Dunai, L., and Hartmann, W. M., "Human interaural time difference thresholds for sine tones: The high-frequency limit," *J. Acoust. Soc. Am.*, 133(5), pp. 2839–2855, 2013, doi: 10.1121/1.4795778.

[29] Riedel, S. et al., "Supplementary data and code," `https://git.iem.at/enimso/aes-2024-lev-leg-area`, 2023.