Audio Engineering Society

Convention Paper 10691

Presented at the AES 156th Convention
2024 June 15–17, Madrid, Spain

This paper was peer-reviewed as a complete manuscript for presentation at this convention. This paper is available in the AES E-Library (http://www.aes.org/e-lib), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Improving Surround Sound Reinforcement at Off-center Listening Positions with Miniature Line Arrays

Lukas Gölles[1,2], Matthias Frank[1,2], and Franz Zotter[1,2]

[1]*Institute of Electronic Music and Acoustics, 8010 Graz, Austria*
[2]*University of Music and Performing Arts Graz, 8010 Graz, Austria*

Correspondence should be addressed to Lukas Gölles (goelles@iem.at)

ABSTRACT

Successful immersive sound reinforcement requires a large sweet area in which the audience is presented with a balanced sound. In smaller venues, state-of-the-art surround sound reinforcement employs point-source loudspeakers producing direct sounds that decay at a rate of 6 dB per distance doubling. In contrast, simulation studies and listening experiments using simulated line arrays have proven that a dual-target design should be pursued for the direct sound decay: 0 dB per distance doubling to preserve the mixing balance and -3 dB to preserve the envelopment at off-center listening positions. This contribution presents a listening experiment comparing different distance decay settings of eight surrounding miniature line arrays of 65 cm total length in a 10.3 m $\times$ 12 m room. The experimental results verify the findings of previous, more theoretical studies.

1 Introduction

Spatial audio, principally known from cinema and games, is gaining in importance, which also increases the interest in using such technology at live events. There are already several commercial solutions for large-scale sound reinforcement considering line arrays [1, Ch. 2, p. 24ff] which primarily concentrate on frontally positioned direct sound objects supplemented by surrounding effects to maintain the time-alignment of spatialized objects, particularly percussion elements, at off-center listening positions [2, 3].
However, direct sound objects can be positioned also around the listeners at venues with smaller listening areas for up to 250 people. These mid-scale surround sound applications still employ point-source loudspeakers in many cases. Experiments showed that a maximum mixing imbalance of ± 3 dB is acceptable for surround sound reproduction [4]. In these experiments, listeners had to balance a mix of two sound objects and then find an upper and lower level difference at which playback is still acceptable. Considering two opposing point-sources, a level difference of 3 dB is reached, when moving approximately 1/6 of the setup radius away from the central listening position, strongly limiting the sweet area for mixing balance. By contrast, it has already been shown that sound reinforcement using surrounding line-source arrays increases the sweet area [5]. Furthermore, listening experiments with simulated line arrays (done by a half or full compensation of the $1/r$ direct sound level decay of physical point sources) revealed similar results for an Ambisonic playback system [6]. Simulations using the A-weighted direct sound level of curved line-source loudspeakers predicted an enlarged listening area compared to a setup with conventional point-source loudspeakers [7]. Considering



Fig. 1: IEM CUBE with miniature line arrays.

the diffuse surrounding sound, experiments showed that a setup of point sources is also not able to reproduce a fully enveloping scene at off-center listening positions [8]. Simulations and listening experiments with simulated line arrays showed that a -3 dB decay per distance doubling (dod) has to be pursued for preserving the envelopment at off-center listening positions [6, 9, 10].

These simulated line arrays with point-source loudspeakers were able to compensate the level balances of the direct sound, but they neglected details that real line-source arrays exhibit: directivity and frequency-dependent radiation.

This contribution presents a listening experiment using a surrounding ring of eight real miniature line arrays applying the theory proposed in [11]. We use a dual-target line-array design, which is a combination of curving and phasing/beamforming of line arrays. The listening experiment is divided into two individual parts to find optimal array design targets: one for mixing balance preservation and the other for envelopment preservation. These aspects are evaluated at three off-center listening positions.

2 Preparation of the Experiment

2.1 Room and Source Setup

The listening experiment was conducted at the IEM CUBE, a 10.3 m $\times$ 12 m $\times$ 4.8 m studio with a reverberation time of 0.5 s. We used eight miniature line arrays with eight enclosures each in a circular layout, cf. Figure 1 and Figure 2. The 10 cm $\times$ 10 cm $\times$ 8.2 cm 3D printed enclosures of these line arrays are equipped with 2.5 inch broadband drivers, cf. [12]. In this experiment, three distance decay configurations were investigated: 0 dB/dod (constant source), -3 dB/dod (line source) and -6 dB/dod (point source). For the 0 dB/dod configuration, the position-dependent frequency responses of the small line arrays follow the desired target for frequencies above 1 kHz, while the radiation for lower frequencies tends towards -3 dB/dod. Furthermore, we added a special configuration which reduces the level of the center loudspeaker of the point-source setup by 9 dB. To preserve the signal components from the center, the signal was distributed to adjacent loudspeakers. This configuration was included to investigate the benefit of a center loudspeaker.

To enable seamless switching between the different configurations, we use mixed line array designs. These use a combination of curving and phasing/beamforming as described and evaluated in [11]. The curvature was designed to achieve 0 dB/dod when the line array elements were driven in phase. For the other configurations, delays were added to each enclosure to achieve -3 dB/dod and -6 dB/dod [13]. The nonlinear differential equation to design curving and phasing for a $-6 \cdot \beta$ dB direct sound level decay per distance doubling was presented in [11] and is implemented in an online solver [14]. This tool was used to calculate tilt angles and delays for the discrete line arrays, to which links are given in the footnote with the used parameters. Our setup required two different line array designs as the throw distances of odd-numbered loudspeakers [1,2] differ from the throw distances of even numbered loudspeakers [3,4]. We choose a rigging height of 2.3 m for which we assume seated audiences (ear height: 1.2 m). The additional delays change the frequency response of the loudspeakers for each listening position individually. For example, Figure 6 shows the frequency responses of loudspeaker 1, 4, 7 and 8 for all positions

[1] https://enimso.iem.sh/post/line-array-designer-two-target/?N=8&h=0.082&xr0=10&y0=1.1&g=0.31&beta=0&beta2=0.5&g2=0.693

[2] https://enimso.iem.sh/post/line-array-designer-two-target/?N=8&h=0.082&xr0=10&y0=1.1&g=0.31&beta=0&beta2=0.5&g2=0.693

[3] https://enimso.iem.sh/post/line-array-designer-two-target/?N=8&h=0.082&xr0=12&y0=1.1&g=0.283&beta=0&beta2=0.5&g2=0.668

[4] https://enimso.iem.sh/post/line-array-designer-two-target/?N=8&h=0.082&xr0=12&y0=1.1&g=0.283&beta=0&beta2=1&g2=1

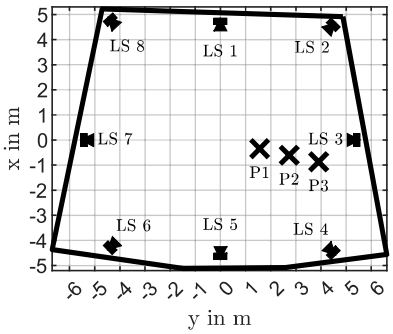


Fig. 2: Loudspeaker setup at the IEM CUBE with positions for the listening experiments marked as crosses.

with different distance decay settings. The graphs were separated for easier readability. As there are only slight differences in the responses, we omitted any position-dependent equalizing in detail, as it is not feasible in practice.
The three off-center listening positions were chosen to be on a straight line and that the level difference of two opposing sources that contain direct sound objects in the listening experiments, LS 3 and LS 7, is included. However, we have placed the line at a slight angle to the rear so that LS 3 does not take on an overriding role in the listening experiment because the results in the off-axis direction are representative of a larger number of listeners.
The listening experiment was divided into two parts to find the optimum distance decays for preservation of the mixing balance preservation of the envelopment at off-center listening positions.

2.2 Stimuli

In the first part, three music excerpts were played back in the following sequence with increasing complexity of the scene:
(i) A 12.09 s loop from Clara Berry and Wooldog's *Waltz For My Victims* from MUSDB18 [15] where the drums are panned to $\varphi = -90°$, the bass and the singing voice are placed to the frontal center, $\varphi = 0°$, and the piano is panned to $\varphi = 90°$. The spatial placement was chosen to enable clear separation of the instruments in the first track to be rated.
(ii) A 4.56 s loop from the live-concert recording *The Magical Music of Disney* played by Bergkapelle Leoben-Seegraben, a wind orchestra consisting of 63 musicians. We spatialized the recordings of 41 spot microphones in the frontal semicircle using the `MultiEncoder` of the IEM Plugin Suite[5].
(iii) A 13.28 s loop from the live-concert recording *Moment for Morricone* played by the same orchestra with equal spatial distribution of the spot microphones as (ii) but with an additional choir of 10 singers.
Ambisonic decoding was done using the AllRAD approach [16] with applied max-$\boldsymbol{r}_E$ weights [17]. At zenith and nadir, an imaginary loudspeaker was added with zero gain to maintain the directional resolution at the horizon [18]. Playback used third-order Ambisonics to achieve optimally smooth panning on the given ring with eight loudspeaker arrays [19]. We applied amplitude compensation of the different loudspeaker distances to the center but omitted delay compensation, as it can reduce the size of the sweet area [20].

In the second part, uncorrelated noise and recorded rain spatialized with the `GranularEncoder` of the IEM Plugin Suite are played back. The uncorrelated noise was routed directly to the corresponding loudspeaker arrays without Ambisonic processing, while the spatialized rain was decoded the same way as the scenes in the first part of the experiment. We added a low-pass filter with cut-off frequency at $f_c = 4\,\text{kHz}$ to reduce possible coloration caused by the phasing/beamforming and spatial aliasing due to the driver spacing.

3 Experimental procedure

Before the listening experiment started, we clarified the terms mixing balance and envelopment and performed demonstrations using different distance decay settings if necessary.
The experiment started at listening position P1, marked as cross in Figure 2, with the task of evaluating the mixing balance part followed by evaluation of the envelopment. Then, participants had to move to the second position P2, repeat both tasks, and continue at the third position P3 afterwards. To compare the different array configurations all at once, a Multi-Stimulus Test with Hidden Reference and Anchor (MUSHRA) [21] appeared to be a suitable test paradigm. The distance decay configurations were randomized within each trial, but participants were presented with the fixed order of audio scenes at each position as the tracks increase in complexity in this order.

[5] `https://plugins.iem.at`

3.1 Mixing Balance including consistency of directional images

In the first part, the participants were asked to switch between the mono reproduction played back from the center array as a reference and the spatial reproduction using the surrounding arrays. They had to address three questions within a single answer, where 0 denoted the worst and 100 the best configuration. Listeners were asked to use the whole range. If the participant did not use the entire range, the results were normalized to the desired range between 0 and 100. The two other configurations had to be positioned relatively to the best and worst:
(a) Is the mix balance preserved if you compare the spatial reproduction to the mono reproduction from the center array?
(b) Does the perceived distance change when comparing the sources positioned around the center of the spatial mix with the mono reproduction?
(c) Are the directional voices well distributed across the frontal semicircle or is there a hole around the center?

3.2 Envelopment

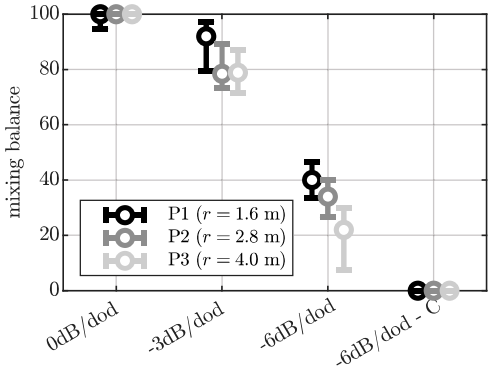
In the second part, the participants had to rate the envelopment, i.e. the feeling of being surrounded by sound. Here, 0 denoted not enveloping at all, which meant that the perceived sound collapses into a single direction. 100 denoted completely enveloping sound, i.e. there was no preferred perceived direction. The listeners were asked to rate a scene with highest score, when they had the feeling that the sound field had no preferred direction.

4 Results

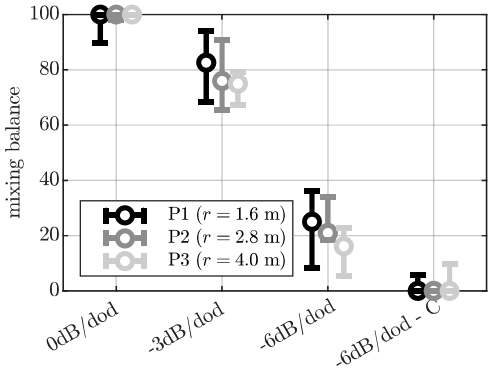
25 listeners (24 male, 1 female, 21 with experience in audio engineering) took part in the experiment with an average age of 33 years. The entire experiment took them on average 28 min 21 s.

4.1 Mixing balance including consistency of directional images

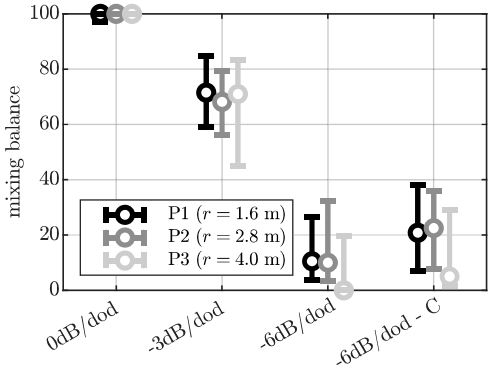
Figure 3 shows the result for the first part of the listening experiment regarding the mixing balance including consistency of directional images. Comparing the results for each scene, it becomes clear that



(a) Results for *Waltz For My Victims*



(b) Results for *The Magical Music of Disney*



(c) Results for *Moment for Morricone*

Fig. 3: Medians and corresponding 95% confidence intervals of mixing balance for different array configurations.

	0dB|−3dB	0dB|−6dB	−3dB|−6dB	−3dB|−6dB-C	−6dB|−6dB-C
P1	0.541	3.585	1.739	3.186	0.890
P2	1.041	4.197	1.635	3.106	0.846
P3	1.165	6.039	2.418	3.460	0.649

Table 1: Effect size for pairwise comparisons of pooled results of the mixing balance part for different array configurations; values greater 0.6 denote medium effect, values greater 0.8 denote large effect.

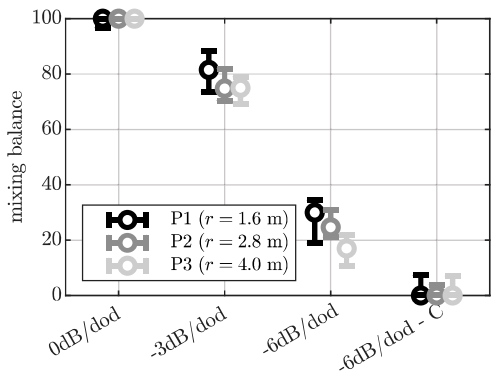


Fig. 4: Medians and corresponding 95% confidence intervals of mixing balance for different array configurations pooled over all audio scenes.

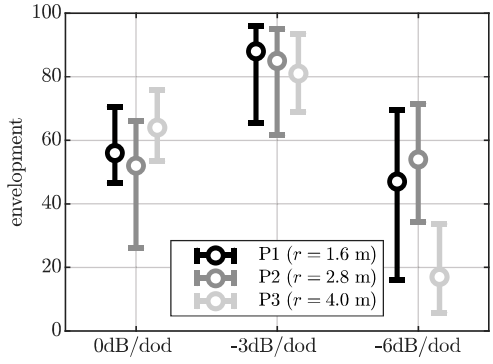
the participants rated the configuration with 0 dB/dod best. −3 dB/dod achieves slightly worse results. The ratings for the point-source configurations are clearly worse, especially for −6 dB/dod-C. The differences in the ratings across the audio scenes are significant, excepted for the most complex audio scene, *Moment for Morricone*, where a Wilcoxon signed-rank test with Bonferroni-Holm correction indicates $p > 0.5$. There is also weak significance ($p = 0.07$) for the first scene comparing 0 dB/dod with −3 dB/dod. We performed a correlation analysis between each scene at each position. The correlation coefficient is always above 0.85 and indicates strong correlation which allows the results of all scenes to be pooled, cf. Figure 4. It should be noted that the center loudspeaker increases clarity and stability (−6 dB/dod vs. −6 dB/dod - C) of a scene with a well defined voice in the center, cf. *Waltz For My Victims*. However, when considering a scene, in which instruments and a choir are distributed around the front semicircle without preference of a particular voice, cf. *Moment for Morricone*, the center loudspeaker becomes less important as the effects of stability and clarity are not as pronounced.

Comparing the pooled results of the two best configurations, 0 dB/dod with −3 dB/dod, the median differs about 20-30 points, cf. Figure 4. The signed-rank test reveals significant differences for all positions ($p < 0.001$). We also calculated the impact effect size [22] which indicates great effects for the two outermost positions, cf. Table 1. Although there are significant differences between these two configurations, it required a high level of attention to distinguish between the configurations which was reported by some participants after the experiment.

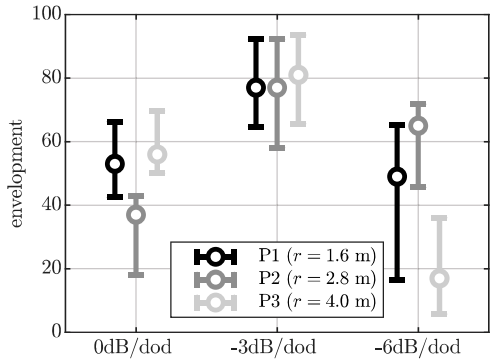
In terms of mixing balance, level decays 0 dB/dod and −3 dB/dod outperform the point-source configurations. Both, median values and effect sizes indicate that there are clear differences. Furthermore, −3 dB/dod and −6 dB/dod are worse for positions further out. The point-source configurations −6 dB/dod and −6 dB/dod-C were rated worst. The signed-rank test also indicates significant differences for all positions ($p < 0.01$) and the effect sizes from Table 1 show large effects. Furthermore, the participants noted that the voices are perceived more distantly, especially those placed around the center because the source directivity is higher for smaller distance decays cf. [11]. The effect sizes comparing 0 dB/dod with −6 dB/dod-C are not listed as the distribution of the results are clearly different which would result in extremely high values for the effect size.

4.2 Envelopment

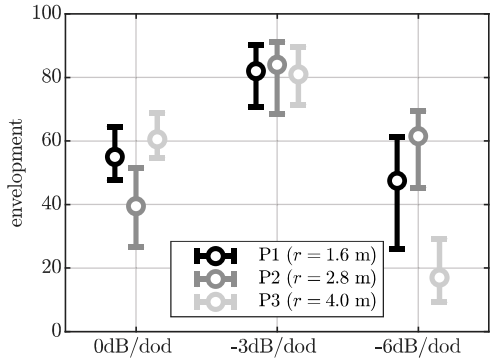
Array configurations with −3 dB/dod were rated best in terms of envelopment in the second part of the experiment, cf. Figure 5. The Wilcoxon signed-rank test with Bonferroni-Holm correction shows significant differences for all positions between the −3 dB/dod array and 0 dB/dod array as well as between the −3 dB/dod array and −6 dB/dod array ($p < 0.01$). For position P2, there is only a weak difference between −3 dB/dod and −6 dB/dod ($p = 0.0558$). Between 0 dB/dod and



(a) Results for decorrelated noise.



(b) Results for spatialized rain.



(c) Results pooled over all audio scenes.

Fig. 5: Medians and corresponding 95% confidence intervals of mixing envelopment for different array configurations.

	0dB|−3dB	0dB|−6dB	−3dB|−6dB
P1	0.554	0.348	0.617
P2	0.910	0.341	0.390
P3	0.444	1.170	2.143

Table 2: Effect size for pairwise comparisons of pooled results of the envelopment part for different array configurations; values greater 0.6 denote medium effect, values greater 0.8 denote large effect.

−6 dB/dod significant results are denoted only on position P3. In contrast to the results of the previous part about mixing balance, the effect sizes indicate smaller effects, cf. Table 2, as the spread is larger.
Furthermore, some participants noted that the 0 dB/dod configuration leads to an unnatural behavior as the perceived direction pulls to the left, even though they were sitting to the right of the central position. Considering a circular setup in the free field with an infinite number of loudspeakers with 0 dB/dod placed on the horizon, the energy vector [23, 24, 25] is $r_{\mathrm{E}} = 0$ in the center. Moving to the right, the direction of the energy vector is determined by more sound sources located on the left as the direct sound level of each loudspeaker is constant. This effect is also apparent when using a discrete set of loudspeaker directions.

5 Discussion

Our results clearly show that mixing balance is preserved best by the array configuration with 0 dB/dod, while envelopment is best for the configuration with −3 dB/dod. As a consequence, loudspeaker systems should optimally be able to realize these two targets at the same time. In practice, this would require two independent mix buses, to treat the direct sound objects and the enveloping parts individually. One drives line array in phase with curving designed for 0 dB/dod and the other one adds delays to each individual element of the line array to achieve −3 dB/dod. Optimal surround sound reinforcement using eight line arrays with eight elements each, has to consider a system of 64 amplifiers, a mixing console of two individual mix buses with eight channels each and a processor that distributes the outputs of the mixer to the arrays and adds delays to the bus containing the enveloping parts. These delays have to be added with high precision as small deviations lead to a deviation of the distance decay. Furthermore, immersive audio productions must separate the direct part

from the diffuse, enveloping part and provide both parts separately to enable separate processing at playback.

6 Conclusion

In this contribution, we compared mixing balance including consistency of directional images and envelopment for surround sound reinforcement using a circle of eight miniature line arrays with different direct SPL supply profiles at three off-center listening positions in a 10.3 m × 12 m room. The results of the listening experiment indicate that a dual-target design of the direct sound decay should be pursued for optimal surround sound reinforcement: 0 dB per distance doubling for preserving the mix balance at off-center listening positions and −3 dB per distance doubling for preserving the envelopment. Moreover, reducing the level of the center array while preserving its signal by distributing it to the neighboring arrays, decreases the mixing balance.
Further work could investigate whether such a dual-target design is necessary with both direct and enveloping/diffuse parts separately rendered, or whether a single target of 0 dB, −1.5 dB or −3 dB per distance doubling achieves acceptable playback. Moreover, it would be interesting to find out whether further improvements can be achieved with planar arrays that are not only able to control the SPL profile along a line into the audience, but across the audience area.

7 Acknowledgments

Our research was funded by the Austrian Science Fund (FWF): P 35254-N (Envelopment in Immersive Sound Reinforcement, EnImSo). We thank all listeners for their participation in the experiment, Bergkapelle Leoben-Seegraben for providing the concert recordings for the listening experiment and the company Herzkraft for cooperation during the recording.

8 Data Availability

To support reproducible research, the results of the listening experiments are made available under [26].

References

[1] Patterson, J. and Lee, H., editors, *3D Audio*, Perspective on Music Production, Routledge, United States, 1st edition, 2021, ISBN 9781138590069, doi: 10.4324/9780429491214.

[2] Mouterde, T., Epain, N., Moulin, S., and Corteel, E., "On the perception of musical groove in large-scale events with immersive sound," in *Audio Engineering Society Conference: AES 2023 International Conference on Spatial and Immersive Audio*, 2023.

[3] Mouterde, T., Epain, N., Moulin, S., and Corteel, E., "On the factors influencing groove fidelity in immersive live music events," in *Audio Engineering Society Conference: AES 2024 International Acoustics & Sound Reinforcement Conference*, 2024.

[4] Zotter, F., Riedel, S., Gölles, L., and Frank, M., "Acceptable Imbalance of Sound-Object Levels for Off-Center Listeners in Immersive Sound Reinforcement," in S. Lippert and O. von Estorff, editors, *Fortschritte der Akustik, DAGA*, Hamburg (Deutschland), 2023.

[5] Nettingsmeier, J. and Maier, P., "Scaling up: Making point-source multichannel material work for large listening areas," in *3rd International Conference on Spatial Audio (ICSA 2015)*, 2015.

[6] Heidegger, P., Brands, B., Langgartner, L., and Frank, M., "Sweet Area using Ambisonics with simulated Line Arrays," in *Fortschritte der Akustik, DAGA*, Wien (Österreich), 2021.

[7] Gölles, L. and Zotter, F., "Optimally Curved Arc Source for Sound Reinforcement," in *Fortschritte der Akustik, DAGA*, Wien (Österreich), 2021.

[8] Blochberger, M., Zotter, F., and Frank, M., "Sweet area size for the envelopment of arecursive and a non-recursive diffusenessrendering approach," in *Proceedings of 5th International Conference on Spatial Audio ICSA*, Ilmenau (Deutschland), 2019.

[9] Riedel, S. and Zotter, F., "Surrounding line sources optimally reproduce diffuse envelopment at off-center listening positions," *JASA Express Letters*, 2(9), p. 094404, 2022, doi: 10.1121/10.0014168.

[10] Riedel, S., Gölles, L., Zotter, F., and Frank, M., "Modeling the Listening Area of Envelopment," in *Fortschritte der Akustik, DAGA*, Hamburg, 2023.

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

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

[13] Gölles, L. and Zotter, F., "Dual-Target Design for Large-Scale Sound Reinforcement: Simulation and Evaluation," in *Audio Engineering Society Conference: AES 2024 International Conference on Acoustics & Sound Reinforcement*, 2024.

[14] Gölles, Lukas, "Line Array Designer considering Two Targets," https://enimso.iem.sh/post/line-array-designer-two-target/, 2023.

[15] Rafii, Z., Liutkus, A., Stöter, F.-R., Mimilakis, S. I., and Bittner, R., "The MUSDB18 corpus for music separation," 2017, doi: 10.5281/zenodo.1117372.

[16] Zotter, F. and Frank, M., "All-Round Ambisonic Panning and Decoding," *J. Audio Eng. Soc*, 60(10), pp. 807–820, 2012.

[17] Daniel, J., *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*, Ph.D. thesis, Université de Paris 6, 2000.

[18] Zotter, F., Riedel, S., and Frank, M., "All-Round Ambisonic Decoding: Spread and Correlation," in *Fortschritte der Akustik, DAGA*, Stuttgart, 2022.

[19] Frank, M., Gölles, L., Riedel, S., and Zotter, F., "Equalizing the coloration of different Ambisonic order weightings," in *Fortschritte der Akustik, DAGA*, Hamburg, 2023.

[20] Frank, M., "How to make Ambisonics sound good," in *Proceedings of Forum Acusticum*, Krakau, 2014.

[21] Union, I. T., "RECOMMENDATION ITU-R BS.1534-1 - Method for the subjective assessment of intermediate quality level of coding systems," *International Telecommunications Union, Geneva*, 2001.

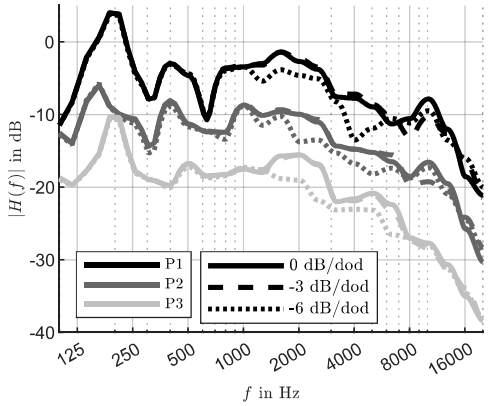
[22] Lötsch, J. and Ultsch, A., "A non-parametric effect-size measure capturing changes in central tendency and data distribution shape," *PLoS ONE*, 15, 2020, doi: 10.1371/journal.pone.0239623.

[23] Gerzon, M. A., "General metatheory of auditory localisation," in *Audio Engineering Society Convention 92*, Audio Engineering Society, 1992.

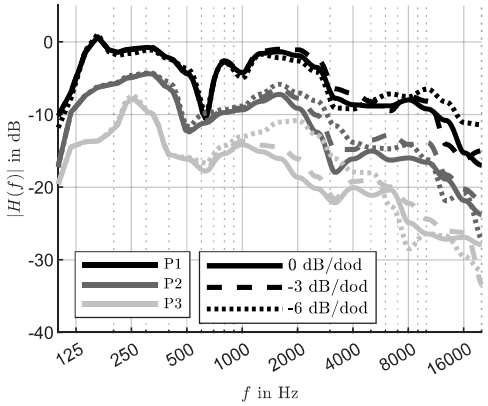
[24] Frank, M., "Localization using different amplitude-panning methods in the frontal horizontal plane," in *Proc. of the EAA Joint Symposium on Auralization and Ambisonics*, Berlin, 2014.

[25] Kurz, E. and Frank, M., "Prediction of the listening area based on the energy vector," in *Proc. of International Conferene on Spatial Audio, Graz*, Graz, 2017.

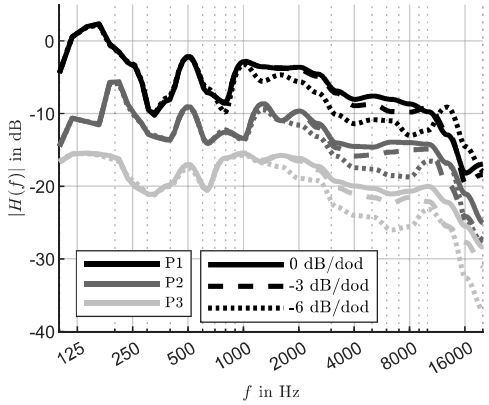
[26] Lukas Gölles, "Supplementary Data for Improving Surround Sound Reinforcement at Off-center Listening Positions with Mini Line Arrays," https://phaidra.kug.ac.at/o:133352, 2024.



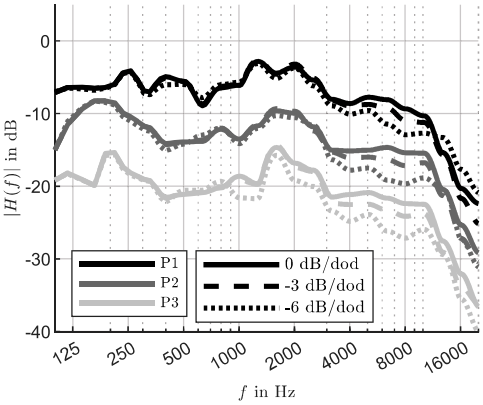
(a) Frequency responses of Loudspeaker 1



(b) Frequency responses of Loudspeaker 4



(c) Frequency responses of Loudspeaker 7



(d) Frequency responses of Loudspeaker 8

Fig. 6: Frequency responses of the first 2000 samples (41.67 ms at $f_s = 48\,\text{kHz}$) of the impulse respones of the miniature line arrays at different observation points with different distance decay settings; for easier readability we applied offsets in steps of 6 dB.