Simulating the Sweet Area of Immersive Sound Reinforcement with Surrounding Mini Line Arrays

Lukas Gölles[1], Matthias Frank[1], Franz Zotter[1]
[1] University of Music and Performing Arts, Institute of Electronic Music and Acoustics, 8010 Graz, Austria
Email: goelles@iem.at, frank@iem.at, zotter@iem.at

Introduction

The recent success in commercializing immersive music productions for home consumers boosts the urge to create surround audio experiences in concerts. Success requires a large sweet area in which the audience is presented with a balanced spatial audio reproduction. The sweet area needs to become larger than what point-source loudspeakers can achieve. Recent studies [1, 2, 3, 4] indicated that two design targets of the direct sound level should be pursued: 0 dB per doubling of the distance (dod) to preserve the mixing balance, and -3 dB/dod to preserve the envelopment at off-center listening positions.

The extended $\boldsymbol{r}_E$ vector model [5, 6, 7] has already been used to predict the position-dependent localization error of an Ambisonic playback system [8]. This contribution deals employs this model to predict the mean absolute localization error of an arrangement of eight surrounding miniature line arrays with different distance decay settings in a 10 m × 12 m room. In terms of direct-sound mixing balance, we use the maximum deviation from the median of all direct sound levels to estimate the area for plausible reproduction. To verify previous findings from listening experiments about envelopment [3, 4], we assume playback of decorrelated signals from all eight loudspeakers and use the length of the energy vector to predict diffuseness.

Line Array Curving and Phasing

To achieve an equalized on-axis direct sound level roll-off by $-6 \cdot \beta$ dB/dod, [9] presents a differential equation for a total inclination, composed linearly from a physical inclination / curvature and an electronic beamforming angle using delays. A mixed design implements two targets in a single system, e.g. the curvature is designed for 0 dB/dod and additional delays for each enclosure achieve -3 dB/dod. An online tool[1] solves this differential equation for a continuous source, discretizes the solution afterwards, and proposes tilt angles and delays for each enclosure in the chain of a line array. For practical reasons, we use mixed line array designs so that they may be compared to the results of listening experiments later. To support reproducible research, links to the online tool are provided with the parameters used.

Room and Source Geometry

The listening area is evaluated at the IEM CUBE, a 10 m × 12 m studio with a reverberation time of 500 ms. We use a setup of eight surrounding line arrays that are equally spaced in the azimuth and positioned as close to ear height (1.2 m for seated audience) as possible. The number of loudspeakers determines the optimal Ambisonic playback order $N = 3$ [10]. To simulate a real case scenario, we consider the dimensions of our miniature line array with eight enclosures. Each enclosure has a height of $h = 8.2$ cm [11]. For the sources positioned near the corners, we choose the farthest observation point of 12 m, the other sources were designed for 10 m. With these parameters, the online solver [2,3,4,5] is used to set up the tilt angles for a purely curved line source with $\beta = 0$ and to propose delays to achieve $\beta_2 = 0.5$ or $\beta_2 = 1$ with the $\beta_1 = 0$ curvature, cf. [9]. Note that the line array solver assumes that the listener's ears are located in the x-y plane ($z = 0$).

Simulating the sound pressure of a discrete line source

The continuous source contour is discretized into a polygon of frequency-dependent straight-line elements of 7.38 cm length with gaps in between [9]. The length of these segments gets shorter by the fifth power of the frequency, i.e. for 20 Hz the length of the straight line corresponds to 90% of the height $h = 8.2$ cm and for 20 kHz it is 60%. The sound pressure is evaluated on a grid of 5 cm space for a frequency range between 20 Hz and 20 kHz with a linear resolution of 128 points. The results were A-weighted and summarized for broadband analysis. Figure 1 shows the simulated A-weighted direct sound pressure levels for different distance decay settings of the center loudspeaker in the front, 0 dB/dod, -3 dB/dod, and -6 dB/dod. These results are used as damping weights $w_{\mathrm{d,l}}$ for calculating the energy vector below.

Energy vector

The normalized energy vector is calculated as the sum of the weighted unity vectors $\boldsymbol{\theta}_\mathrm{l}$ [5, 6, 7],

$$\boldsymbol{r}_E = \frac{\sum_L \left(w_{\tau,\mathrm{l}} \cdot w_{\mathrm{d,l}} \cdot g_\mathrm{l}\right)^2 \boldsymbol{\theta}_\mathrm{l}}{\sum_L \left(w_{\tau,\mathrm{l}} \cdot w_{\mathrm{d,l}} \cdot g_\mathrm{l}\right)^2}. \quad (1)$$

[1] https://enimso.iem.sh/post/line-array-designer-two-target/

[2] https://enimso.iem.sh/post/line-array-designer-two-target/?N=8&h=0.082&xr0=10&y0=1.1&g=0.312&beta=0&beta2=0.5&g2=0.683

[3] https://enimso.iem.sh/post/line-array-designer-two-target/?N=8&h=0.082&xr0=12&y0=1.1&g=0.282&beta=0&beta2=0.5&g2=0.657

[4] https://enimso.iem.sh/post/line-array-designer-two-target/?N=8&h=0.082&xr0=10&y0=1.1&g=0.312&beta=0&beta2=1&g2=1

[5] https://enimso.iem.sh/post/line-array-designer-two-target/?N=8&h=0.082&xr0=12&y0=1.1&g=0.282&beta=0&beta2=1&g2=1

$\boldsymbol{\theta}_\mathrm{l}$ denotes the unity vector pointing from the listening position to the l-th loudspeaker, g_l the gain weight of the l-th loudspeaker, $w_{\tau,\mathrm{l}}$ is the time weight and $w_{\mathrm{d},\mathrm{l}}$ denotes the damping weight that we get from direct sound level simulations. For N-th order Ambisonic encoding and decoding to L loudspeakers, these gains $\boldsymbol{g} = \begin{bmatrix} g_1 & g_2 & \dots & g_L \end{bmatrix}^\mathsf{T}$ are

$$\boldsymbol{g} = \boldsymbol{D}\,\mathrm{diag}\{\boldsymbol{a}\}\,\boldsymbol{y}_N(\boldsymbol{\theta}_s)\,. \tag{2}$$

$\boldsymbol{D}$ is the decoder matrix calculated by the AllRAD approach [12], the diagonal matrix $\mathrm{diag}\{\boldsymbol{a}\}$ is used to apply max-$\boldsymbol{r}_E$ weights [13] and $\boldsymbol{y}_N$ is the column vector containing the spherical harmonics up to order N evaluated at the desired normalized source direction $\boldsymbol{\theta}_s$. The time weights are $w_{\tau,\mathrm{l}} = 10^{\frac{w_\tau \cdot \tau_l}{1000}}$, where τ_l denotes the acoustic delay of the l-th loudspeaker to the listening position. $w_\tau = -1\,\frac{\mathrm{dB}}{\mathrm{ms}}$ is a multiplication factor that is known from the echo threshold and chosen as trade-off between transient and stationary signals [14, 15, 16, 17, 18].

Localization error

The localization error is the horizontal difference between the predicted direction of the energy vector and the desired panning direction relative to the listening position. Figure 2 shows the root mean square of the localization error for panning around the horizon in steps of $1°$ with applied max-$\boldsymbol{r}_\mathrm{E}$ weights for all three line array configurations, $0\,\mathrm{dB/dod}$, $-3\,\mathrm{dB/dod}$, and $-6\,\mathrm{dB/dod}$. For excellent localization, the error should stay below $10°$ [13]. The area that is enclosed by the $10°$ contour is $61\,\mathrm{m}^2$ for $0\,\mathrm{dB/dod}$, $58\,\mathrm{m}^2$ for $-3\,\mathrm{dB/dod}$ and $49\,\mathrm{m}^2$ for $-6\,\mathrm{dB/dod}$. The main difference between $0\,\mathrm{dB/dod}$ and $-3\,\mathrm{dB/dod}$ lies in the areas near the loudspeakers located at $\varphi \in \{0°, 90°, 180°, 270°\}$. Furthermore, the area for $0\,\mathrm{dB/dod}$ is elliptical while it is square to rectangular for the other configurations.

Mixing balance

While the localization results show only weak differences between the different distance decay settings, we should also consider the mixing balance for surround sound applications. Optimally, direct sound objects should reach all positions in the audience with the same level, regardless of their panning direction. It has been shown that listeners tolerate a mixing imbalance of $\pm 3\,\mathrm{dB}$ with regard to the optimal mix [19]. We define the maximum absolute deviation from the median $\tilde{w}_\mathrm{d}$ as mixing balance,

$$\mathrm{MB} = \max\left(|20\lg \boldsymbol{w}_\mathrm{d} - 20\lg \tilde{w}_\mathrm{d}|\right)\,. \tag{3}$$

Figure 3 shows the mixing balance for all three distance decay settings, $0\,\mathrm{dB/dod}$, $-3\,\mathrm{dB/dod}$, and $-6\,\mathrm{dB/dod}$. In contrast to the results of the localization error, the results are clearly different from each other because the localization error only takes the direction of the energy vector into account which neglects the actual level of the phantom source. The area that is enclosed by the $\pm 3\,\mathrm{dB}$ limit is $65\,\mathrm{m}^2$ for $0\,\mathrm{dB/dod}$, $19\,\mathrm{m}^2$ for $-3\,\mathrm{dB/dod}$, and $4\,\mathrm{m}^2$ for $-6\,\mathrm{dB/dod}$.

Envelopment

So far, our simulations took the localization and mixing balance into account and thus only considered the behavior of direct sound objects of an immersive mix. However, immersive sound reinforcement has also to face surround effects, such as reverberation. Riedel et al. [3, 4] showed that loudspeakers, which direct sound level rolls off by $-3\,\mathrm{dB/dod}$, would be optimal to preserve the envelopment at off-center listening positions for a horizontal loudspeaker arrangement. Here, we use the length of the energy vector to predict the diffuseness, $1 - ||\boldsymbol{r}_\mathrm{E}||$. The diffuseness has to stay above 80 % to achieve plausible reproduction [20]. Figure 4 shows the results for eight surrounding line arrays with different distance decays assuming uncorrelated noise with unit variance from all arrays. It can be clearly seen that $-3\,\mathrm{dB/dod}$ yields the largest area of $78\,\mathrm{m}^2$ that is enclosed by the 80 % limit. By contrast, the areas of $0\,\mathrm{dB/dod}$ and $-6\,\mathrm{dB/dod}$ are smaller with $12\,\mathrm{m}^2$ and $4\,\mathrm{m}^2$, respectively.

Furthermore, the directions of the energy vectors for $0\,\mathrm{dB/dod}$ show an interesting behavior, cf. Figure 5. They point in the opposite direction when moving from the central listening position. At off-center positions, the number of closer loudspeakers decreases while the number of farther loudspeakers increases. As there are more distant loudspeakers of (nearly) equal level, the direction of the energy vector is dominated by the more distant loudspeakers. We may compensate that effect for many listening positions when using $-3\,\mathrm{dB/dod}$. For $-6\,\mathrm{dB/dod}$, the closer loudspeakers dominate the direction of the energy vector.

Optimal Targets

Optimal surround sound reinforcement should consider $0\,\mathrm{dB/dod}$ for direct sound objects and $-3\,\mathrm{dB/dod}$ for the diffuse parts of an audio scene. To use both targets with a single system, mixed line array designs have to be pursued [9]. For a setup with eight surrounding line arrays with eight enclosures each, this requires a mixing desk with two individual eight-channel mix buses. The output of both buses must be fed into a processor that distributes the signal to 64 channels while adding delays to the envelopment bus for each enclosure individually. Furthermore, 64 amplifiers are necessary to drive the enclosures individually.

Conclusion

The simulations using the extended vector models could indicate that optimal immersive sound reinforcement must pursue two targets, $0\,\mathrm{dB/dod}$ for the direct sound and $-3\,\mathrm{dB/dod}$ for the diffuse part to create maximum sweet areas for mixing balance and envelopment. Further listening experiments have to be conducted to verify the results of the simulations in practice. Moreover, such experiments should investigate whether the hardware effort could be decreased by a single design target with $0\,\mathrm{dB/dod}$, $-1.5\,\mathrm{dB/dod}$ or $-3\,\mathrm{dB/dod}$ instead of the two targets while still providing plausible reproduction of audio scenes with both direct sound objects and enveloping diffuse sound.

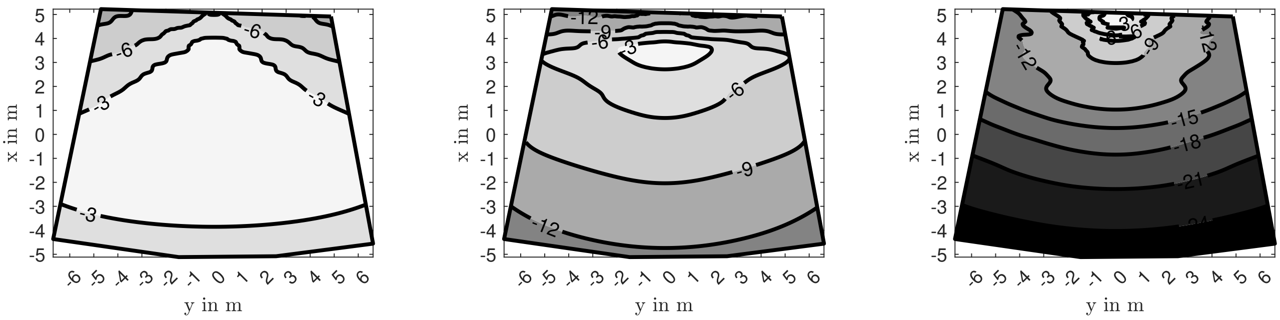

Figure 1: Simulated A-weighted direct sound level of the center line array with different distance decays: 0 dB/dod (left), −3 dB/dod (center) and −6 dB/dod (right).

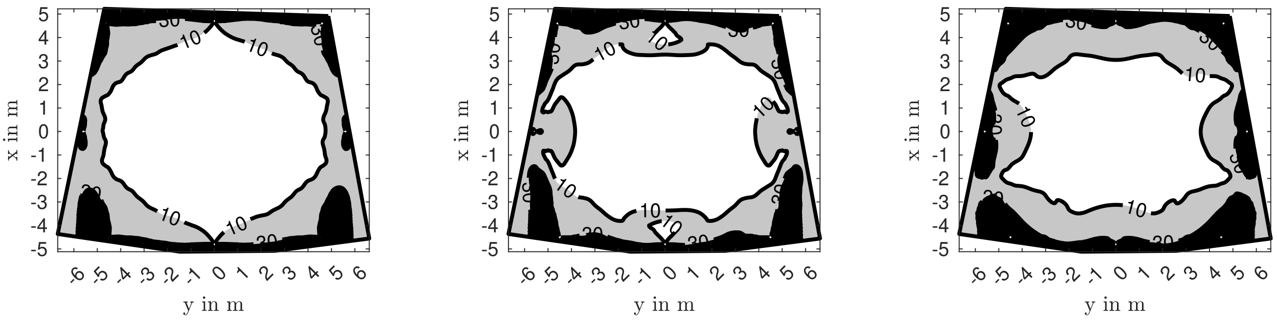

Figure 2: Simulations of the mean absolute localization error using the $\boldsymbol{r}_{\mathrm{E}}$ vector model for third order Ambisonic encoding and decoding for panning on the horizon with applied max $\boldsymbol{r}_{\mathrm{E}}$ weights: 0 dB/dod (left), −3 dB/dod (center) and −6 dB/dod (right).

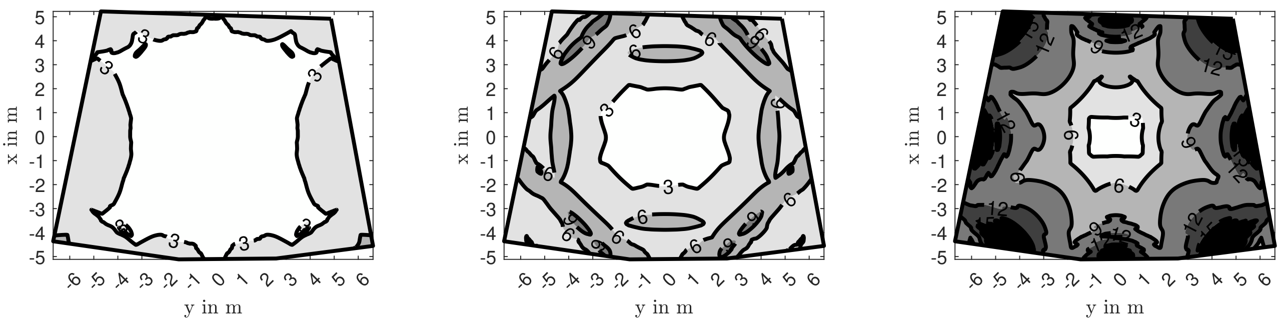

Figure 3: Simulated mixing balance for a setup of eight surrounding line arrays: 0 dB/dod (left), −3 dB/dod (center) and −6 dB/dod (right).

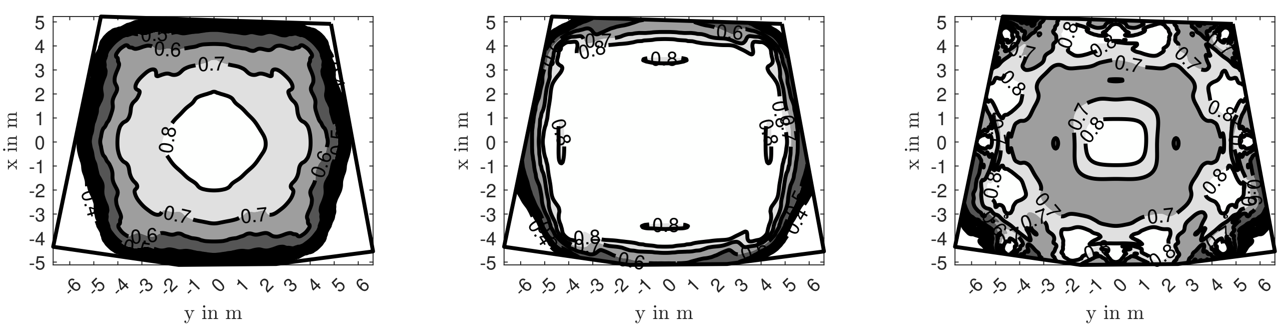

Figure 4: Simulated diffuseness of different line array settings for a setup of eight surrounding line arrays: 0 dB/dod (left), −3 dB/dod (center) and −6 dB/dod (right).

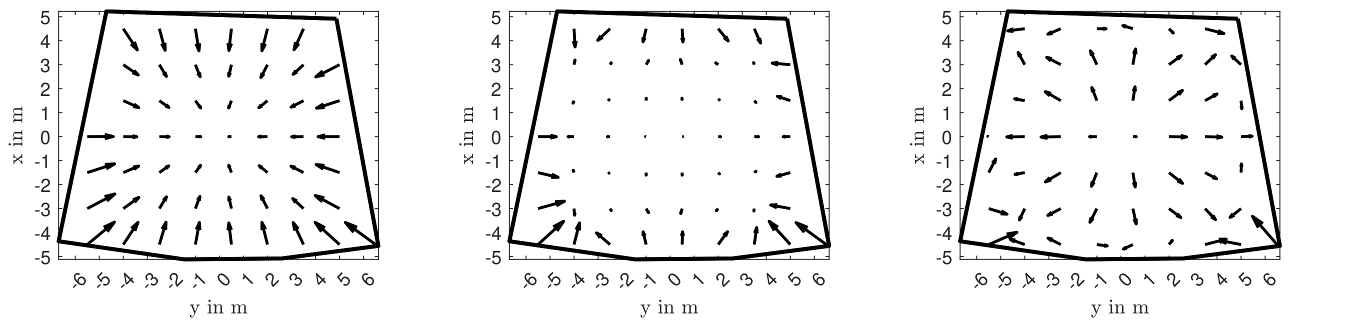

Figure 5: Simulated $\boldsymbol{r}_E$ vectors assuming decorrelated noise is played back by the line arrays: 0 dB/dod (left), −3 dB/dod (center) and −6 dB/dod (right).

Acknowledgment

Our research was funded by the Austrian Science Fund (FWF): P 35254-N, Envelopment in Immersive Sound Reinforcement (EnImSo).

Data Availability

In order to provide reproducible research, the direct sound level simulations are made available under [21].

References

[1] P. Heidegger, B. Brands, L. Langgartner, and M. Frank, "Sweet area using ambisonics with simulated line arrays," in *Fortschritte der Akustik, DAGA*, Wien, 08 2021.

[2] L. Gölles and F. Zotter, "Optimally curved arc source for sound reinforcement," in *Fortschritte der Akustik, DAGA*, Wien, 08 2021.

[3] S. Riedel and F. Zotter, "Surrounding line sources optimally reproduce diffuse envelopment at off-center listening positions," *JASA Express Letters*, vol. 2, no. 9, p. 094404, 09 2022.

[4] S. Riedel, L. Gölles, F. Zotter, and M. Frank, "Modeling the listening area of envelopment," in *Fortschritte der Akustik, DAGA*, 03 2023.

[5] M. A. Gerzon, "General metatheory of auditory localisation," in *Audio Engineering Society Convention 92*. Audio Engineering Society, 03 1992.

[6] P. Stitt, S. Bertet, and M. van Walstijn, "Extended energy vector prediction of ambisonically reproduced image direction at off-center listening positions," *J. Audio Eng. Soc*, vol. 64, no. 5, pp. 299–310, 2016.

[7] E. Kurz and M. Frank, "Prediction of the listening area based on the energy vector," in *International Conference on Spatial Audio*, 09 2017.

[8] M. Frank, "Localization using different amplitude-panning methods in the frontal horizontal plane," in *Proceedings of the EAA Joint Symposium on Auralization and Ambisonics*, 01 2014.

[9] L. Gölles and F. Zotter, "Theory of continuously curved and phased line sources for sound reinforcement," *Acta Acust.*, vol. 7, p. 52, 2023.

[10] M. Frank, L. Gölles, S. Riedel, and F. Zotter, "Equalizing the coloration of different ambisonic order weightings," in *Fortschritte der Akustik, DAGA*, Hamburg, 03 2023.

[11] 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*, Aug 2023.

[12] F. Zotter and M. Frank, "All-round ambisonic panning and decoding," *J. Audio Eng. Soc*, vol. 60, no. 10, pp. 807–820, 2012.

[13] ——, *Ambisonics*. Springer, 04 2019.

[14] H. Seraphim, "Über die Wahrnehmbarkeit mehrerer Rückwürfe von Sprachschall," *Acta Acustica united with Acustica*, vol. 11, no. 2, pp. 80–91(12), 1961.

[15] Rakerd, Hartmann, and Hsu, "Echo suppression in the horizontal and median sagittal planes," *The Journal of the Acoustical Society of America*, vol. 107 2, pp. 1061–4, 2000.

[16] P. Robinson, A. Walther, C. Faller, and J. Braasch, "Echo thresholds for reflections from acoustically diffusive architectural surfaces," *The Journal of the Acoustical Society of America*, vol. 134, pp. 2755–2764, 10 2013.

[17] E. Kurz, "Efficient prediction of the listening area for plausible reproduction," Master's thesis, University of Music and Performing Arts, Nov. 2018.

[18] M. Frank and J. Kristl, "Perceptual sweet area of 2-channel stereo playback," in *VDT International Convention*, 03 2024.

[19] F. Zotter, S. Riedel, L. Gölles, and M. Frank, "Acceptable imbalance of sound-object levels for off-center listeners in immersive sound reinforcement," in *Fortschritte der Akustik, DAGA*, 03 2023.

[20] ——, "Diffuse sound field synthesis," 2024.

[21] L. Gölles, "Simulated A weighted direct sound pressure levels of line arrays at the IEM CUBE," 2024. [Online]. Available: https://phaidra.kug.ac.at/o:133064