forum acusticum 2026
8-12 September · Graz, Austria

ESTIMATING SPATIAL SOUND FIELD FEATURE MAPS FROM A REDUCED SET OF MEASUREMENTS

Roman Kiyan[1], Rodi Younes[1], Stephan Preihs[1], Jürgen Peissig[1]

[1] *Institut für Kommunikationstechnik, Leibniz Universität Hannover*

Abstract

Acoustic features derived from binaural and spherical harmonic representations of a sound field can be used for objective analysis of sound field properties as well as for perceptual modeling. When a sound field with a particular spatial extent is considered, such as in multichannel loudspeaker reproduction, the position-dependent variation of sound field features may be used to define a sweet area where the reproduction is regarded as sufficiently similar to the loudspeaker setup's sweet spot. Because measuring spatially dense sound field feature maps requires considerable practical effort, reconstructing the spatial variation of sound field features from a reduced number of measurements is desirable. To explore the possibility of reconstructing spatial feature maps from spatially sparse measurements, sound field parameters were measured in a listening room using an automated rover carrying a dummy head or a spherical microphone array. Different strategies for spatial subsampling of the sound field feature maps obtained from the measurements are investigated to arrive at a minimal measurement setup required for the application of sweet area analysis in multichannel loudspeaker reproduction.

Keywords: *sound field analysis, sweet area, automated measurement, parameter estimation*

1. Introduction

In multichannel loudspeaker reproduction, the sweet area around a loudspeaker setup's sweet spot is the part of the listening area where particular properties of the sound field are considered to be sufficiently similar to those in the sweet spot itself. It is common to use perceptual properties such as localization [1], [2], envelopment [1], [3] or immersion [4] to define the sweet area. To assess the size of the sweet area using physical measurements, sound field features associated with these perceptual properties may be used. This study is concerned with the spatial resolution required in measurements for such investigations.

When reconstructing a sound pressure distribution across a listening area, sampling the sound field according to the Nyquist theorem is associated with substantial measurement effort [5]. Various approaches have been introduced to reconstruct sound field data from a reduced number of measurements, including neural networks [6], curve-fitting approaches [7], as well as compressed sensing (CS) algorithms [8]. In contrast to the goal of sound pressure reconstruction, acquisition of sound field features for perceptual models is more akin to the evaluation of room acoustic parameters over an extended area, which has been studied in the context of concert hall acoustics [9]–[11]. On the one hand, such spatial feature maps can be expected to be smoother across the listening area than the sound pressure distribution, allowing for coarser sampling. On the other hand, as these features are extremely dependent on the loudspeaker setup and signals [12], prior knowledge is required to determine an appropriate sampling grid. In this study, the effects of different sampling strategies are analyzed for noise and music stimuli with different loudspeaker setups.

2. Methodology

2.1. Measurement setup

Impulse response (IR) measurements were carried out in a listening room [13] largely compliant with ITU-R-BS-1116-3. A subset of loudspeakers comprising a 5.1.4 setup according to ITU-R-BS-2051-2 was used. IRs were measured with a Neumann KU 100 dummy head and an mh acoustics em32 Eigenmike attached to a microphone stand mounted on a four-wheeled rover vehicle. A motion capture system was used for positioning control. The measurement system is shown in Fig. 1. A target grid of $21 \times 18 = 378$ points with a spacing of 15 cm was selected, covering a rectangular area from $(-1.5\,\text{m}, -1.05\,\text{m})$ to $(1.5\,\text{m}, 1.5\,\text{m})$ around the sweet spot of the loudspeaker setup.

Corresponding author: *roman.kiyan@ikt.uni-hannover.de.*

Figure 1: Rover with the em32 in the listening room.

2.2. Sound field parameters

For sound field analysis, the measured IRs were convolved with the corresponding channels of eight music signals and two noise signals available in Mono, Stereo, 5.1 and 5.1.4 channel-based formats. The two white noise signals differ in their inter-channel correlation (ICC) with values of $\rho \approx 0$ and $\rho \approx 0.5$.
The selection of evaluated sound field features is based on previous studies on immersion in multichannel music reproduction as well as sweet spot analysis [4], [12], [14]. From the dummy head measurements, inter-aural level difference (ILD), inter-aural cross-correlation (IACC), and the inter-aural correlation coefficient (IACoef) were evaluated[1]. Sound field diffuseness [15] was computed from the Ambisonics representation obtained using the Eigenmike. The sound field parameters ψ were computed from 100 ms time windows with 50 % overlap in the four frequency bands given in Table 1 as well as from the broadband signals. The parameter time series were then averaged, yielding the mean values ψ_{mean}. To be able to compare the effects of sampling in (x, y) on the different parameters, the $\psi_{\text{mean}}(x, y)$ were rescaled to the value range $[0, 1]$ for each signal in each format, yielding the normalized sound field feature maps $\phi(x, y)$.

2.3. Reconstruction of sound field feature maps

2.3.1. Interpolation

Three main approaches were used for the reconstruction of sound field feature maps from coarse measurements. The first two approaches are based on sampling the sound field at uniform grid resolutions with greater sampling distances than the originally measured grid. To simulate measurements at different resolutions, interpolation was applied to the sound field feature maps. For this study, natural neighbor interpolation [16] was used for its improved smoothness over linear interpolation and lack of oscillations compared to cubic interpolation.
In the interpolation-based approaches, the sound field features ϕ are interpolated from the (slightly noisy) acquired coordinates to a grid strictly following the 15 cm spacing. These feature maps are interpolated to a coarser grid $(\tilde{x}, \tilde{y})$ which can be written as

$$\tilde{\phi}(\tilde{x}, \tilde{y}) = \underset{(x,y)\to(\tilde{x},\tilde{y})}{\text{interp}} \left\{\phi(x, y)\right\}. \tag{1}$$

The coarse sound field feature data $\tilde{\phi}$ is then re-interpolated to the fine grid, obtaining the estimated sound field feature map

$$\hat{\phi}(x, y) = \underset{(\tilde{x},\tilde{y})\to(x,y)}{\text{interp}} \left\{\tilde{\phi}(\tilde{x}, \tilde{y})\right\}. \tag{2}$$

The two interpolation-based methods differ in the coarse grid $(\tilde{x}, \tilde{y})$ being designed either with similar resolution in the x and y directions or trading a finer resolution in x for a coarser resolution in y based on observations detailed in Section 3.

2.3.2. Compressed sensing

A fundamentally different approach is based on CS [17], [18]. The idea behind CS is to reconstruct a signal from fewer measurements than required according to the Nyquist sampling theorem by leveraging a *sparse* representation in some basis. Typically, it is assumed that a signal's frequency (temporal) or wavenumber (spatial) domain representation is sparse (i. e. characterized by a limited number of prominent coefficients). It has been shown that *near-optimal* recovery can be achieved by taking a reduced number of samples at random locations – which results in an underdetermined system of equations when attempting to restore all samples – and finding the solution with the minimal L_1 norm in the frequency domain [17]. In the current study, a CS pipeline based on the discrete cosine transform (DCT) was implemented for reconstructing the sound field feature maps $\phi(x, y)$ using the `l1magic` MATLAB toolbox [19].

2.3.3. Reconstruction error

For all reconstruction methods, the reconstruction error is analyzed using the median absolute error

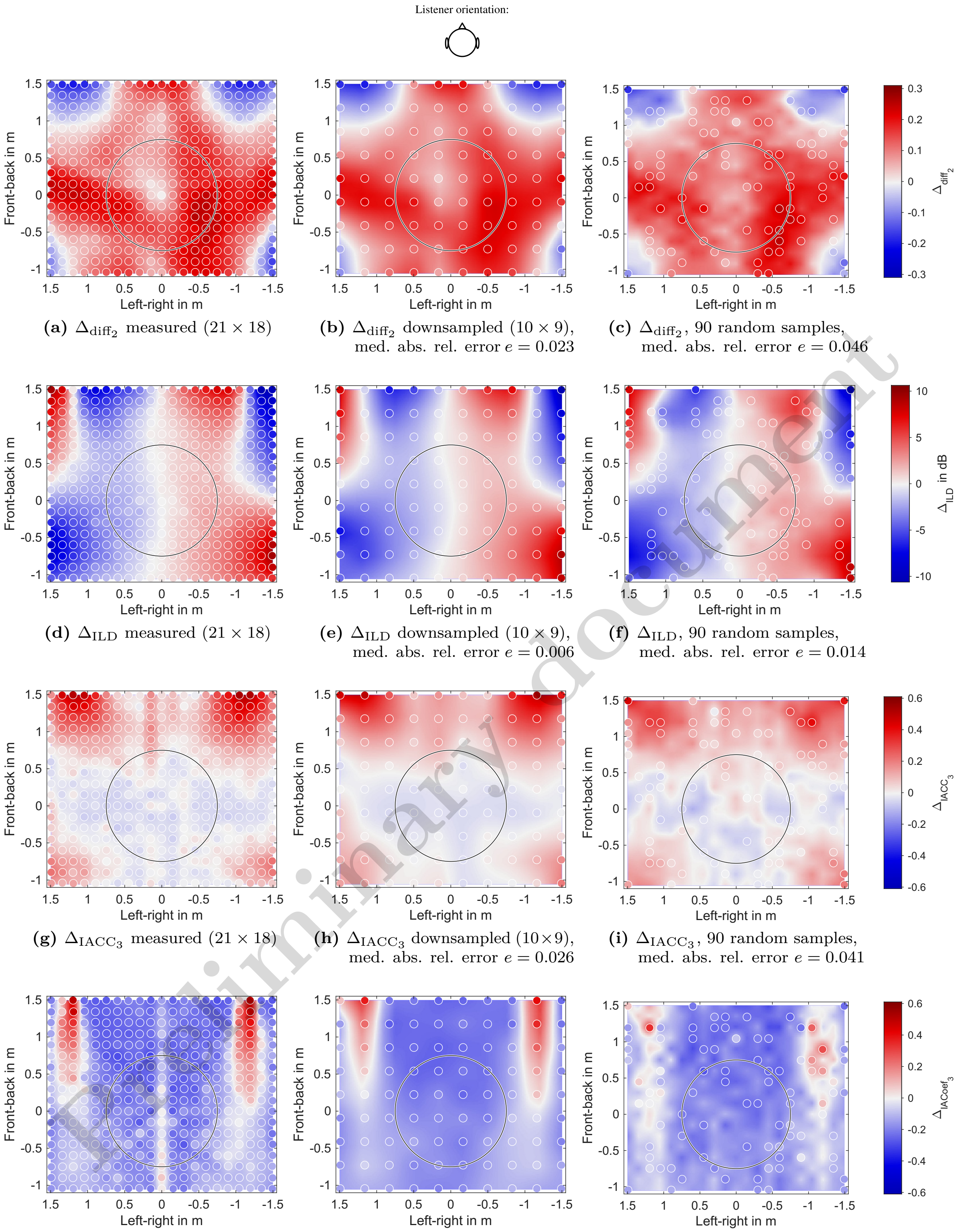
$$e = \underset{(x,y)}{\text{med}} \left|\hat{\phi}(x, y) - \phi(x, y)\right|, \tag{3}$$

which acts as a relative error metric due to the aforementioned feature normalization. The error e can be evaluated against the relative number of points in the reduced set of measurements compared to the original 378 points, assessing the error incurred due to the simulated reduction in measurement effort.

Table 1: Frequency bands used in sound field analysis

Band index	1	2	3	4
Freq. range in Hz	< 100	100... 500	500... 2.5k	> 2.5k

[1] IACoef is a measure of inter-aural correlation based on the time-domain inter-aural correlation function (IACF) at lag zero rather than the maximum of the magnitude of the IACF. For detailed definitions, see ref. [12].



(a) Δ_{diff_2} measured (21×18)

(b) Δ_{diff_2} downsampled (10×9), med. abs. rel. error $e = 0.023$

(c) Δ_{diff_2}, 90 random samples, med. abs. rel. error $e = 0.046$

(d) Δ_{ILD} measured (21×18)

(e) Δ_{ILD} downsampled (10×9), med. abs. rel. error $e = 0.006$

(f) Δ_{ILD}, 90 random samples, med. abs. rel. error $e = 0.014$

(g) Δ_{IACC_3} measured (21×18)

(h) Δ_{IACC_3} downsampled (10×9), med. abs. rel. error $e = 0.026$

(i) Δ_{IACC_3}, 90 random samples, med. abs. rel. error $e = 0.041$

(j) $\Delta_{\mathrm{IACoef}_3}$ measured (21×18)

(k) $\Delta_{\mathrm{IACoef}_3}$ downsampled (10×9), med. abs. rel. error $e = 0.027$

(l) $\Delta_{\mathrm{IACoef}_3}$, 90 random samples, med. abs. rel. error $e = 0.047$

Figure 2: Time-averaged sound field parameter deviations with respect to the listening position for semi-correlated noise reproduced over a 5.1 setup. Scatter points represent sampling positions, images are upsampled from data reconstructed to the (21×18) grid. Feature deviations in all columns are with respect to the sweet spot value (at the origin) in the left column. The circles show the measurement area from previous studies [4], [12].

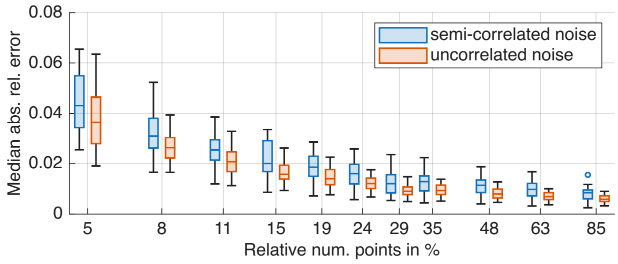
3. Results

3.1. Interpolation from uniform grids

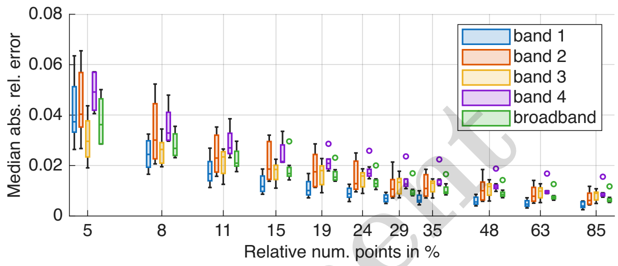
To simulate coarser measurements in uniform grids, logarithmically spaced numbers of points (rounded to integers) were chosen according to Table 2. To develop an intuition for the sound field features, grids and the error metric, exemplary heatmaps of the four selected sound field features are displayed in Fig. 2. These heatmaps show the non-normalized difference $\Delta_{\psi_{\mathrm{mean}}}$ to the sweet spot value in exemplary frequency bands. The results are displayed for semi-correlated noise reproduced using five loudspeakers. Results from the original measurement grid are displayed in the left column while interpolated results from a (10×9) grid (approx. 24 % of the original number of points) are shown in the central column.
For the given example, both visual assessment and the error e show that smooth feature maps such as the one for ILD are reproduced very accurately whereas fine structures such as the strong changes in diffuseness and IACoef around the sweet spot are not correctly reconstructed. In the exemplary heatmaps, the semi-correlated noise signal was specifically selected to illustrate strong variation in sound field features around the sweet spot [12]. For a more general overview, the dependencies of the error e on the relative number of grid points, signals and frequency bands are displayed for different sound field features in Fig. 3 and Fig. 4. Fig. 3 shows error results for the diffuseness maps for the noise signals by signal type and frequency band. In addition to the general trend of increasing error with the reduction of measurement points, some dependencies on the signal and on the frequency bands can be observed, but the errors for each given reduced grid are typically all in a similar range. Notably, even when evaluating high-frequency data, the resulting diffuseness maps are still comparable in smoothness to those computed for lower frequencies. When comparing the different sound field features in Fig. 4, it is clear that some feature maps can be reconstructed more precisely than others for the noise signals. This may be an artifact of the synthetic test signals, however, as the error distributions for the music signals behave more similarly across sound field features.

3.2. Higher lateral resolution

The feature maps in Fig. 2 exhibit lateral symmetries due to the symmetric loudspeaker setup and the front-facing head orientation in the binaural measurements. The fact that lateral variations such as the narrow band of IACoef values around the centerline of the loudspeaker setup are easily lost with coarser measurement grids suggests that different resolutions in the x and y directions may be appropriate. To evaluate this, grids retaining the full 21 points in the x direction with different numbers of points in y were compared to grids with the same number of points and approximately equal spacing in x and y. Error results for both types of grids are plotted against each other in Fig. 5a. Note that above approx. 67 % of

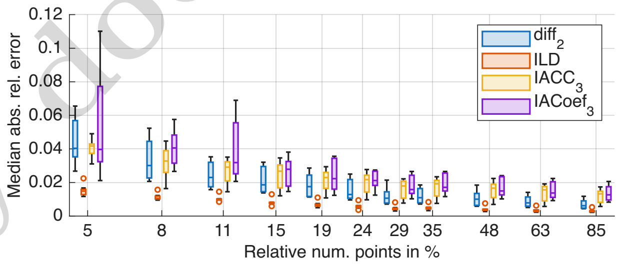


(a) Signal dependency (all formats, all frequency bands)

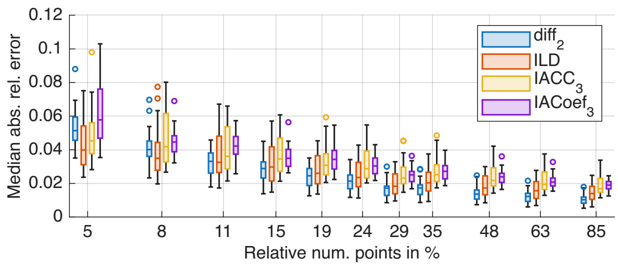


(b) Frequency dependency (both signals, all formats)

Figure 3: Diffuseness map reconstruction error for the noise signals. The plots show median and quartiles (boxes), maximum and minimum values within 1.5 times the interquartile range (whiskers) as well as outliers.



(a) Noise (both signals, all formats, all frequency bands)



(b) Music (eight signals, all formats, all frequency bands)

Figure 4: Reconstruction error by sound field feature.

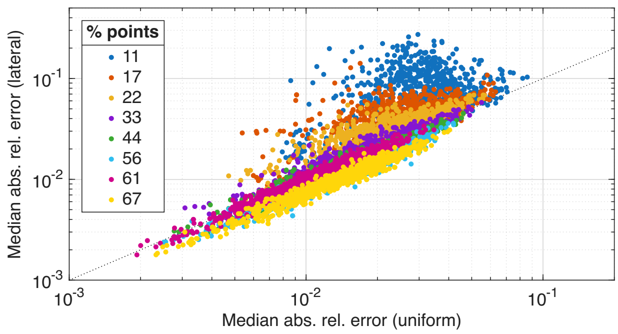
points, both approaches result in identical grids and are hence omitted. The error scatter plot shows that the grids with higher lateral resolution perform worse in terms of reconstruction error if just a few lines in y are used. Above 33 % of points, a slight improvement over the grids with uniform distances in x and y can be observed.

3.3. Reconstruction from random samples

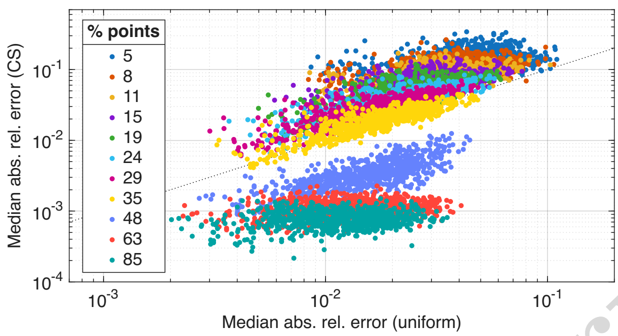
For the CS approach, a similar comparison to interpolation from uniform grids is displayed in Fig. 5b.

Table 2: Uniform grids used for downsampling. Configurations shown in Fig. 2 are highlighted in bold.

Num. points (x)	5	6	7	8	9	**10**	11	12	15	17	19	**21**
Num. points (y)	4	5	6	7	8	**9**	10	11	12	14	17	**18**
Rel. num points in %	5.3	7.9	11.1	14.8	19.0	**23.8**	29.1	34.9	47.6	62.9	85.4	**100.0**
Avg. spacing in cm	80	61.9	50.5	42.7	37.0	**32.6**	29.2	26.4	22.3	19.2	16.3	**15.0**



(a) Higher lateral resolution vs. uniform downsampling



(b) Random sampling with L_1-based reconstruction vs. uniform downsampling with interpolation

Figure 5: Median absolute relative errors e compared between reconstruction methods. The dotted lines indicate equal errors. Combined results for diffuseness, ILD, IACC, and IACoef in all frequency bands are displayed.

The error results show that for small reductions in the amount of data (retaining 63 % of points or more), CS is capable of practically perfect reconstruction of the sound field feature maps, which is not achieved by interpolation. The two methods perform equally well when keeping approx. 35 % of points whereas CS reconstruction fidelity substantially deteriorates with fewer samples. The right column of Fig. 2 shows CS reconstruction from approx. 24 % of points, showing that the overall structure of the heatmaps can be reconstructed, but with noticeable artifacts in the fine structure.

4. Discussion

As the results presented in Section 3 show, the evaluated interpolation-based methods are able to restore the overall structure of sound field feature maps from very few measurement points. This is highlighted by the median absolute relative error generally being below 5 % even when using as little as 10 %-15 % of the original number of measurement points. While exemplary features associated with immersive music experience were used here, the effects are expected to generalize to various acoustic features derived from binaural and Ambisonics signals. Despite good overall reconstruction, it was observed that the error metric averages out errors in the fine structure of the feature maps dependent on inter-aural time and phase relationships. While this fine structure consists of lateral variation in sound field features in the current dataset, this cannot be expected more generally for non-channel-based content or if head rotation is considered in binaural features. Thus the laterally dense grids tested here cannot be considered as a general solution.

The CS-based approach, while outperforming interpolation for minor reductions in measured data, cannot be generally recommended either due to the artifacts introduced in the reconstructed feature maps when limited data is available. This is not a shortcoming of the method itself but rather a consequence of the given data not being well-suited to CS. With just (21 × 18) sample points (and the same number of wavenumber domain coefficients) to begin with, the feature maps cannot be expected to be sparse enough in the wavenumber domain to profit appreciably from the CS approach.

Overall, the results suggest that prior knowledge on the expected sound field feature maps is required if fine structures such as the sharp variations in diffuseness and IACoef are to be properly captured. If, instead of the simple features used here, the sweet area is defined in terms of more sophisticated auditory models [2], this challenge can be expected to be exacerbated further. Therefore, for future measurement campaigns, it appears to be appropriate to first gather knowledge on the expected structure of the sound field feature maps to be measured (by small-scale measurements or by simulations) to define an optimal measurement procedure adapted to the task at hand.

5. Summary

In this study, the effect of different spatial sampling strategies on physical sound field features was investigated. Using different uniform grids as well as CS-based reconstruction from random samples, it was observed that measurements with an average spacing of up to approx. 45 cm preserve the overall structure of the spatial sound field feature distributions reasonably well. Fine structures require much denser spatial sampling to be adequately captured, though. In the case of the signals and loudspeaker setup used here, this can be achieved by sampling with a higher resolution in the lateral direction. For general sound fields, however, additional knowledge on the expected spatial feature maps is required. As for the application of CS, the limited sparsity of the sound field feature maps acquired in the studio-sized listening room does not justify the method, although it may be viable in larger larger listening rooms or concert hall environments. Based on the findings presented here, dense sampling of the sound field around the sweet spot can be recommended, whereas tradeoffs reducing the measurement effort by taking fewer samples can be made for the outskirts of the listening area.

Data availability

The IR data set used in the presented study is planned to be made available by the authors. The code producing the results of this paper is available at `https://gitlab.uni-hannover.de/roman.kiyan.jr/speakersweetspot`.

References

[1] M. Frank and F. Zotter, "Exploring the perceptual sweet area in ambisonics," presented at the AES Convention 142, Berlin, Germany, 2017.

[2] P. Lladó, A. Neidhardt, A. R. Souchaud, Z. Cvetković, and E. De Sena, "Perceptually-driven panning for an extended listening area," presented at the 2025 IEEE WASPAA, Tahoe City, CA, USA, 2025. DOI: 10.1109/WASPAA66052.2025.11230924.

[3] S. Riedel and F. Zotter, "Surrounding line sources optimally reproduce diffuse envelopment at off-center listening positions," *JASA Express Lett.*, vol. 2, no. 9, 2022, 094404. DOI: 10.1121/10.0014168.

[4] R. Kiyan, S. Preihs, and J. Peissig, "Determining the immersion sweet area in multichannel loudspeaker reproduction using spatial sound field features," presented at the AES Convention 156, Madrid, Spain, 2024.

[5] I. B. Witew, M. Vorländer, and N. Xiang, "Sampling the sound field in auditoria using large natural-scale array measurements," *J. Acoust. Soc. Am.*, vol. 141, no. 3, EL300–EL306, 2017. DOI: 10.1121/1.4978022.

[6] F. Lluís, P. Martínez-Nuevo, M. Bo Møller, and S. Ewan Shepstone, "Sound field reconstruction in rooms: Inpainting meets super-resolution," *J. Acoust. Soc. Am.*, vol. 148, no. 2, pp. 649–659, 2020. DOI: 10.1121/10.0001687.

[7] T. Pham Vu and H. Lissek, "Low frequency sound field reconstruction in a non-rectangular room using a small number of microphones," *Acta Acustica*, vol. 4, no. 2, p. 5, 2020. DOI: 10.1051/aacus/2020006.

[8] S. A. Verburg and E. Fernandez-Grande, "Reconstruction of the sound field in a room using compressive sensing," *J. Acoust. Soc. Am.*, vol. 143, no. 6, pp. 3770–3779, 2018. DOI: 10.1121/1.5042247.

[9] D. de Vries, E. M. Hulsebos, and J. Baan, "Spatial fluctuations in measures for spaciousness," *J. Acoust. Soc. Am.*, vol. 110, no. 2, pp. 947–954, 2001. DOI: 10.1121/1.1377634.

[10] X. Pelorson, J.-P. Vian, and J.-D. Polack, "On the variability of room acoustical parameters: Reproducibility and statistical validity," *Appl. Acoust.*, vol. 37, no. 3, pp. 175–198, 1992. DOI: https://doi.org/10.1016/0003-682X(92)90002-A.

[11] K. Fujii, T. Hotehama, K. Kato, *et al.*, "Spatial distribution of acoustical parameters in concert halls: Comparison of different scattered reflections," *J. of Temporal Design in Architect. and the Environ.*, vol. 4, no. 1, pp. 59–68, 2004.

[12] R. Kiyan, S. Preihs, and J. Peissig, "Signal Dependencies of the Immersion Sweet Area in Multichannel Loudspeaker Reproduction," *J. Audio Eng. Soc.*, vol. 73, no. 5, pp. 282–292, 2025. DOI: 10.17743/jaes.2022.0202.

[13] R. Hupke, M. Nophut, S. Li, R. Schlieper, S. Preihs, and J. Peissig, "The immersive media laboratory: Installation of a novel multichannel audio laboratory for immersive media applications," presented at the AES Convention 144, Milan, Italy, 2018.

[14] R. Kiyan, J. Bergner, S. Preihs, *et al.*, "Towards predicting immersion in surround sound music reproduction from sound field features," *Acta Acustica*, vol. 7, no. 45, 2023. DOI: 10.1051/aacus/2023040.

[15] V. Pulkki, "Spatial sound reproduction with directional audio coding," *J. Audio Eng. Soc.*, vol. 55, no. 6, pp. 503–516, 2007.

[16] R. Sibson, "A brief description of natural neighbour interpolation," in *Interpreting multivariate data*. John Wiley & Sons, 1981, pp. 21–36.

[17] D. Donoho, "Compressed sensing," *IEEE Trans. on Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006. DOI: 10.1109/TIT.2006.871582.

[18] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. on Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006. DOI: 10.1109/TIT.2005.862083.

[19] E. Candes and J. Romberg, *L1-magic: Recovery of sparse signals via convex programming*, 2005. [Online]. Available: https://candes.su.domains/software/l1magic/.