Meyer-Kahlen, Nils; Amengual, Sebastià V.; Lokki, Tapio

What the Spatial Decomposition Method can and cannot do

Published in:
ICA 2022 proceedings

Published: 01/01/2022

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Meyer-Kahlen, N., Amengual, S. V., & Lokki, T. (2022). What the Spatial Decomposition Method can and cannot do. In *ICA 2022 proceedings* (Proceedings of the ICA congress). Acoustical Society of Korea (ASK). https://ica2022korea.org/data/Proceedings_A12.pdf

ICA 2022
24th INTERNATIONAL CONGRESS ON ACOUSTICS
October 24(Mon) - 28(Fri), 2022 Gyeongju, Korea

PROCEEDINGS of the
24th International Congress on Acoustics

October 24 to 28, 2022 in Gyeongju, Korea

ABS-0825

What the spatial decomposition method can and cannot do

Nils MEYER-KAHLEN[(1)], Sebastià V. AMENGUAL GARÍ[(2)], Tapio LOKKI[(1)]

[(1)]Acoustics Lab, Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland, nils.meyer-kahlen@aalto.fi
[(2)]Reality Labs Research, Meta, Redmond, WA, USA

ABSTRACT
The spatial decomposition method (SDM) is a parametric approach for the processing of spatial room impulse responses. Although extensively used, there are some issues related to used microphone array, reproduction loudspeaker setup, and signal processing that are not widely understood. For example, there have been different observations with regards to its performance with rendering extremely short transient signals, where some authors have described issues such as “roughness” or “graininess”. Here, we wish to clarify what the limitations of the spatial decomposition method are and provide practical guidance to ensure the value of SDM as a tool for spatial room impulse response analysis and rendering in various contexts. Specifically, we discuss directional estimation performance, the ability to estimate two reflections at the time, estimation in the late tail of the response, and the roughness and whitening found in the quality of the final rendering. Finally, some post-processing techniques to compensate the possible audible artifacts are reviewed.

Keywords: Directional Room Impulse Responses, Measurement, Auralization

1 INTRODUCTION

The spatial decomposition method (SDM) is a parametric approach for processing spatial room impulse responses (SRIR) that was proposed almost 10 years ago [18]. Since then, it has been applied in a large number of different studies, regarding the evaluation of concert halls [8] and stage acoustics [3], smaller music venues [19], sound studios and movie theaters [15], and even cars [17]. In recent times, there have been divergent observations with regards to its performance, with some authors describing issues such as “graininess” with a few special signals. The first evaluation of SDM [18] found out to yield better results than than an earlier parametric room impulse response algorithm, SIRR [10], which is the basis for a more recent development called HO-SIRR [9]. In [9], SDM performed much worse than both. In a recent pilot study, the obtained perceptual results were often amongst the best of the methods tested, but depended on the used microphone array [13].

In this paper, we wish to clarify what the capabilities and limitations of the spatial decomposition method are, and show that while being aware of them, SDM can be a valuable tool for SRIR analysis and rendering. We first revisit the sound-field model underlying SDM in Section 2. In Section 3 we then describe the analysis and in Section 4 the rendering stage of the classical SDM. For each stage, we first show the algorithms available in the SDM toolbox and mention new additions to the SDM framework from more recent publications. We explain what the SDM can achieve when used correctly, and which limitations are expected in any case. Section 5, concludes the paper discussing the use of SDM and further directions for parametric SRIR methods.

2 THE UNDERLYING SOUND-FIELD MODEL

Every parametric audio method, may it be operating on running signals or on impulse responses, is based on a sound field model. The strength and simultaneously the most important limitation of SDM is the simplicity of the underlying model. It assumes that at any given sample at time t of a room impulse response, exactly one sound event occurs, originating from a direction of arrival (DoA) $\boldsymbol{\theta}(t)$, with sound pressure $p(t)$. It is assumed that when knowing the pressure and the DoA at each sample, the sound field is fully characterized.

Clearly, this parameterization is most reasonable for the direct sound and broad-band early reflections found in a room. The model is less valid in the late part of response, where the density of reflections is high. Yet, this does not mean that no directional information can be extracted at all, as also discussed below.

3 SDM ANALYSIS ...

3.1 ... can estimate the DoAs of individual reflections from microphone SRIRs or SH domain SRIRs

Now that the sound field model is established, algorithms for finding estimates $\hat{\boldsymbol{\theta}}(t)$ of the directions $\boldsymbol{\theta}(t)$ can be discussed. SDM was first implemented using estimation based on Time Difference of Arrival (TDoA) estimation. Later, also the broadband Pseudo Intensity Vector (PIV) was introduced to SDM [4, 22]. Note that both estimation principles are already described in pioneering work about room impulse response analysis [21].

The choice of analysis methods depends on the choice of microphone array used to capture the SRIR. If open arrays of omnidirectional capsules are employed [18], TDoA estimation is the best choice. PIV is preferred for spherical microphone arrays (SMA) known from capturing Ambisonics signals, such as tetrahedral arrays of cardioid capsules, or arrays with omnidirectional capsules on a rigid sphere. Yet it should be noted that open arrays can also be used for PIV estimation, and that directional microphones can yield satisfactory TDoA analysis results as well [19], albeit with degradations compared to omnidirectional capsules.

Time-Difference of Arrival Estimation For TDoA estimation, the responses of an open array of at least four microphones that are not in the same plane are used to obtain the direction within a block of size N, centered around each sample t. Typical choices of microphone arrays have been 3D intensity probes, consisting of six omnidirectional capsules at the centers of the faces of an imagined cube. First, the TDoAs between the microphone capsules are calculated with sub-sample accuracy using interpolated cross-correlation [16]. Then, the DoAs are obtained by finding the least squares solution for a plane wave arriving from $\hat{\boldsymbol{\theta}}$, as in

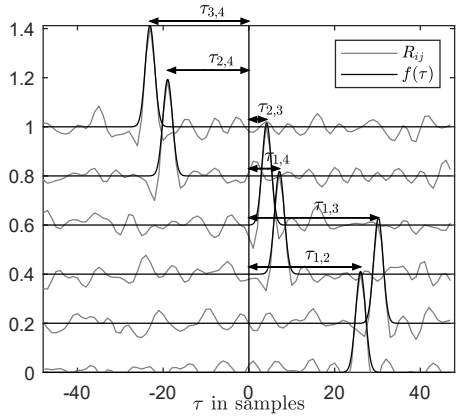


Figure 1. Cross-correlation functions with fitted Gaussian functions for one plane wave in isotropic noise captured with four microphones.

$$R_{i,j}(n) = \frac{1}{N}\sum_{t=1}^{N} h_i(t)h_j(t+n) \tag{1}$$

$$\tau_{i,j} = \arg\max_{n} \; R_{i,j}(n) \tag{2}$$

$$c_{i,j} = \frac{\ln R_{i,j}(\tau+1) - \ln R_{i,j}(\tau-1)}{4\ln R_{i,j}(\tau) - 2\ln R_{i,j}(\tau-1) - 2\ln R_{i,j}(\tau+1)} \tag{3}$$

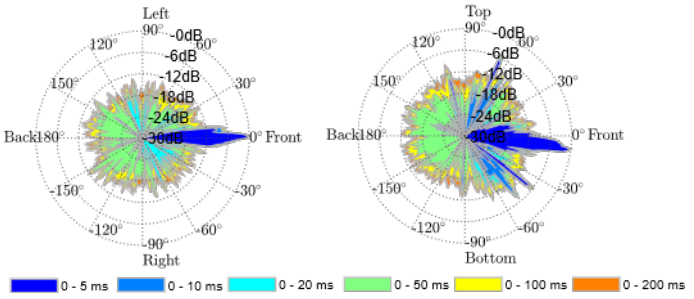
$$\boldsymbol{\tau} = \Big[\tau_{1,2}+c_{1,2}, \tau_{1,3}+c_{1,3}, ..., \tau_{M-1,M}+c_{M-1,M}\Big]^{\mathrm{T}} \tag{4}$$

$$\boldsymbol{V} = \Big[\boldsymbol{r}_1-\boldsymbol{r}_2, \boldsymbol{r}_1-\boldsymbol{r}_3, ..., \boldsymbol{r}_{M-1,M}\Big]^{\mathrm{T}} \tag{5}$$

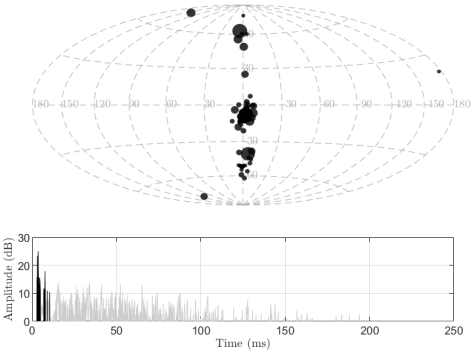
$$\boldsymbol{k} = \boldsymbol{V}^{\dagger}\boldsymbol{\tau} \tag{6}$$

$$\hat{\boldsymbol{\theta}} = \frac{\boldsymbol{k}}{||\boldsymbol{k}||}, \tag{7}$$

where $\tau_{i,j}$ are the time differences of arrival, $c_{i,j}$ are the means of Gaussian functions fitted to the cross-correlated functions $R_{i,j}$ and $\boldsymbol{V}^{\dagger}$ is the Moore-Penrose pseudo-inverse of the matrix of position vector differences of the microphone capsules, wherein the position vectors $\boldsymbol{r}_i$ are the locations of the microphone capsules in cartesian coordinates. $\boldsymbol{k}$ is called the slowness vector. Note that one does not need to know the speed of sound, nor the sampling rate for TDoA estimation itself.



(a) Cumulative energy plots of the lateral plane (left) and the median plane (right), where floor and ceiling reflections are visible.



(b) Interactive Map Visualizer

Figure 2. Two ways of representing SDM analysis data. Direct sound as well as floor and ceiling reflection are easily visible on both plots.

Pseudo Intensity Vector PIV as an alternative estimation methods for DRIR [4] was not yet mentioned in [18], but is available in the SDM toolbox[1]. To calculate the PIV, one needs to know the sound pressure and an estimate of sound velocity along the three coordinate axes. Conveniently, the components of the first order Ambisonics response (also called B-Format response) are proportional to pressure and velocities, so that the (instantaneous, broadband) PIV can be computed as

$$\hat{\boldsymbol{\theta}}(t) = h_{\mathrm{w}}(t) \begin{bmatrix} h_{\mathrm{x}}(t) & h_{\mathrm{y}}(t) & h_{\mathrm{z}}(t) \end{bmatrix}^{\top}. \tag{8}$$

3.2 ... can be the basis for reflection visualization

The simple sound field model of SDM makes it possible to visualize reflections and thereby directional analysis of the response (see Fig. 2). This has been used extensively in the analysis of concert halls [14] by overlaying polar histograms of the acoustic energy in several time intervals.

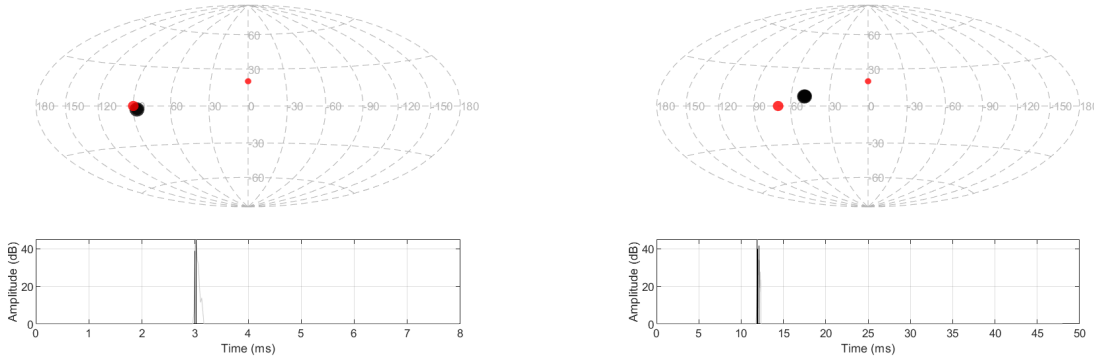
3.3 ... cannot estimate two sound events in short succession

TDoA estimation From the sound field model and the directional estimator, it should be clear that directional analysis fails if two sound events occur simultaneously, wherein it is important to note that simultaneously does not mean that the sound events need too occur within one sampling interval, but within one analysis block. For TDoA estimation, the block size should be selected in accordance with the array size. The minimal block size N_{min} should correspond to at least twice the TDoA for sound incidence from a direction parallel to the largest distance between two microphones,

$$N \overset{!}{\geq} N_{\mathrm{min}} = T_{\mathrm{min}} f_{\mathrm{s}} = 2 \frac{d_{\mathrm{max}}}{c} f_{\mathrm{s}}. \tag{9}$$

This means that the spatio-temporal limit depends on the size of the array. The larger the array is, the larger the window size that should be selected, so the larger the distance required between two reflections in order to separate them. Looking at T_{min} it becomes clear that increasing the sampling rate f_{s} does not improve the separation of the reflections, which explains the observations from [3]. In TDoA-based proccessing, the sampling rate influences the result so that the possible TDoA values are quantized in time. Thereby, also the possible DoA estimates are quantized to a grid, where the shape depends on the geometry of the array and the density depends on the product of array size and sampling rate. However, the applied interpolated cross-correlation solves this problem, so that sub-sample accuracy can be reached, provided sufficient SNR. In practice, a window function is applied to the frame and the frame size is chosen to be slightly larger than N_{min}.

[1] https://se.mathworks.com/matlabcentral/fileexchange/56663-sdm-toolbox



(a) TDoA estimation for a cube array (3D intensity probe) (b) PIV based on a tetrahedral array (FOA microphone)

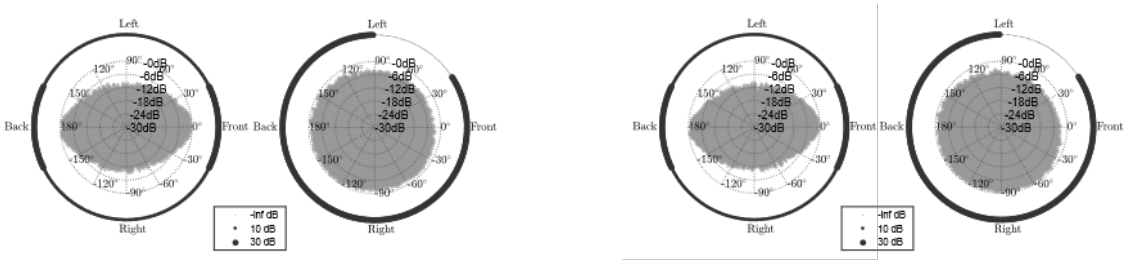
Figure 3. Estimation of two plane waves arriving in very short succession ($\Delta t = 0.05$ ms), much shorter than the window size of 0.5 ms. The plane waves have a level difference of 6dB. TDoA estimation ideally returns the direction of the largest peak, while PIV estimation returns a weighted average direction.

PIV estimation Equation (8) shows an instantaneous PIV that returns one direction per sample. It should be noted that, even though not directly apparent, the temporal resolution is limited by the array size here as well. In PIV estimation, the window used for estimation of the direction at each sample needs to be long enough for a sound wave to pass through the complete array. PIV estimation is not necessarily done in blocks, but in [22] for example, the response is filtered using an 8th order zero-phase bandpass with its upper cutoff at the spatial aliasing frequency of the array, which for a first order array is $f_{\mathrm{sa}} = c/(2\pi r)$. Low-pass filtering the response before estimation leads to temporal integration, which first makes estimation possible.

TDoA and PIV estimations show different behavior for multiple reflections found within one analysis window. While TDoA ideally returns the direction of the largest peak, PIV returns a weighted mean of the directions, see Figure 3. Recently, a variant using a new directional estimator, is capable of estimating two directions at the same time [5]. The method is based on the normalized PIV.

3.4 ... can approximate the directional energy distribution in an anisotropic late field

Given these results, it is striking that estimation under the sound field model of one direction at the time still results in some success also in the late part of the response, where clearly more than one sound event arrives in each analysis window. An example for TDoA and PIV behaviour have been provided in [11]. In case of PIV estimation, it can even be shown analytically how the distribution in an anisotropic diffuse field influences the directional statistics of the estimate, which will be presented in upcoming work. Estimation in a simulated field, modelled with Gaussian noise sources with direction dependent energy distribution is shown in Fig. 4.



(a) TDoA estimation for a cube array (3D intensity probe) (b) PIV based on a tetrahedral array (FOA microphone)

Figure 4. DoA estimation in a simulated anisotropic diffuse field consisting of 180 Gaussian noise sources around the receiver with different energy, that is indicated by the circles outside of the polar histogram plots. The estimate follows the directional distribution only to some extend. Arrays with $r = 1.25$ cm.

4 SDM RENDERING...

4.1 ... can be performed to loudspeakers or headphones, using arbitrary spatialization techniques

SDM rendering is done by distributing the pressure response to a number of reproduction channels. In case rendering is done to loudspeakers, this can be thought of as finding a selection function $w_l(t)$ for each loudspeaker l, which modulates the pressure response to create the loudspeaker response as in

$$g_l(t) = w_l(t)p(t). \tag{10}$$

Nearest Loudspeaker Synthesis Nearest Loudspeaker Synthesis (NLS) was the panning approach used in SDM [18]. It is done by finding the loudspeaker index l_{NL}, which is closest to the directional estimate at each sample

$$w_l(t) = \begin{cases} 1 & l = l_{NL}(t) \\ 0 & l \neq l_{NL}(t) \end{cases}, \qquad \text{where} \quad l_{\mathrm{NL}}(t) = \arg\min_l ||\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}_l(t)||. \tag{11}$$

Ambisonics Encoding More recently, SDM analysis results have been used to render Ambisonics RIR. If the input is a first order Ambisonics DRIR (ARIR), and rendering is also done to Ambisonics, the method is also referred to Ambisonics SDM (ASDM), or "upmixing". This creates the modulation functions for each Ambisonics channel, as opposed to each loudspeaker channel as before. Encoding is a straightforward operation in the SH domain and an arbitrarily high encoding order N can be used

$$\boldsymbol{w}_N(t) = \boldsymbol{y}_N(\hat{\boldsymbol{\theta}}(t)), \tag{12}$$

where $\boldsymbol{y}_N(\hat{\boldsymbol{\theta}}(t))$ are the real spherical harmonics evaluated at the direction $\hat{\boldsymbol{\theta}}(t)$, stacked into a vector using Ambisonics Channel Numbering (ACN). Decoding to loudspeakers can be done with all available Ambisonics decoding approaches [23].

Headphone Rendering Any of the above approaches can be used to create a binaural response $\boldsymbol{b}(t)$, by treating the L measurement points of an HRTF measurement as virtual loudspeakers and convolving the responses for these loudspeakers with the HRIR $\boldsymbol{h}_l(t)$

$$\boldsymbol{b}(t) = \sum_{l=1}^{L} \sum_{n=1}^{N} g_l(n)\, \boldsymbol{h}_l(t-n). \tag{13}$$

The loudspeaker, Ambisonics or binaural responses are then convolved with an arbitrary source signal.

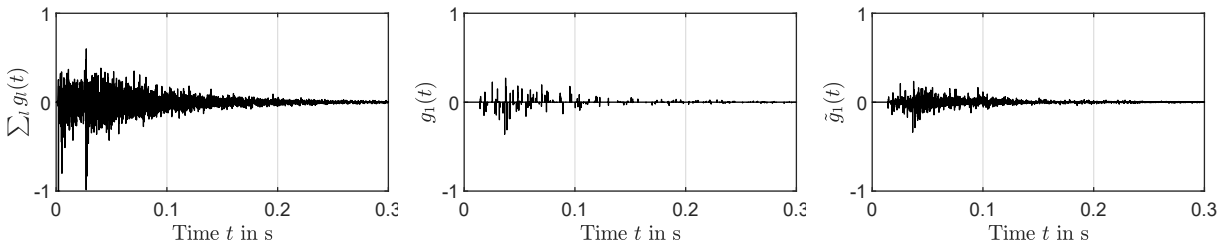
4.2 ... cannot render very transient sounds without compensation for roughness/graininess

The inevitable result of this rendering stage is that the response of each loudspeaker channel is sparse in time. This means that the envelope of each loudspeaker channels varies quickly and when listening to one loudspeaker signal alone, it sounds rough or "grainy". In the SDM literature, it has been described that since the modulation functions sum up to 1,

$$\sum_l w(t) = 1, \tag{14}$$

also all the individual sparse reproduction channels $g_l(t)$ should also add up to the omnidirectional response $p(t)$. While this is true if the responses are summed digitally, before reproduction, it is not true in a realistic listening situation or in binaural rendering. In [12] it has been demonstrated by distributing different kinds of noise to several loudspeakers, that time and level differences introduced by the head of the listener alone are able to destroy perfectly coherent summation of the reproduction channels. This either occurs in binaural rendering, Eq. (13), but also naturally, when listening to the rendering with a loudspeaker array.

The first consequence of this fact is that the response that is summed at the listeners ears still partly maintains the envelope fluctuations of the sparse reproduction channels, so that when rendering very transient sounds,



(a) Sum of all loudspeaker responses after SDM rendering.

(b) One loudspeaker response

(c) One loudspeaker response after roughness compensation.

Figure 5. Every loudspeaker response following NLS rendering to $L = 48$ loudspeakers is sparse in time.

the rendering can sound rough or "grainy". It should be noted that this problem vanishes for continuous source signals. Recently, modifications to SDM have been introduced to improve roughness. In [2], the rendered binaural response is convolved with the output of a cascade of three Schroeder allpass filters with delay length $M = \{37, 113, 215\}$. Figure 5 shows the sum of all channels, and one separate channel $g_1(t)$ before and after ($\tilde{g}_1(t)$) such a roughness compensation filter.

Essentially, the goal of the roughness compensation filter is to densify the response of each reproduction channel without changing the relations between the channels and without increasing the reverberation time. This means that ideally, the filters should be dense, short, and have a flat frequency response. This is approached by the Schroeder cascade, which becomes rather dense after a while and has a frequency response with well distributed zeros. Since computational efficiency is less important when rendering the responses, which are still to be convolved with the source signal, any noise sequence can be used for roughness compensation. One short, dense and spectraly flat alternative could be an exponentially decaying sequence of random binary noise, which has a unit pulse at each sample with randomly chosen sign. Further, new approaches for Ambisonics rendering with SDM are being developed, which offer reduced roughness by rendering additional directions [6].

4.3 ... cannot preserve the spectral content of a SRIR without proper equalization

Another consequence of the sparsity of each channel and the imperfect summation at the listeners ears is a change in spectrum, which was first noticed in [17]. Clearly, each of the sparse loudspeaker responses has a different spectrum compared to the initial, omnidirectional response in that the pseudo-random amplitude modulation has spread the energy in frequency, such that additional high frequency content is present. If now all these "whitened" channels are added incoherently, their spectral content is still retained in the final rendered response.

To compensate the whitening effect, equalization should be applied to each block of the rendered response. Such equalization needs to be done carefully, in order to avoid time aliasing artefacts and audible amplitude modulation. The omnidirectional channel and the reproduction channels are therefore split into blocks of length M. We could denote for example, $p^{(k)}(t) = p(t + kM)$, for $t = \{0, M-1\}$ and zero elsewhere. Then, an equalization filter response $s_l^{(k)}$ is determined for each channel and each block as from the DFTs of pressure $p^{(k)}(f)$ and channel responses

$$s_l^{(k)} = \mathscr{F}^{-1}\left\{ \frac{\|p^{(k)}(f)\|}{\|g_l^{(k)}(f)\|} \exp\left(\imath \phi_l^{(k)}(f)\right) \right\} \quad g_l^{\mathrm{EQ}}(t) = \sum_k g_l^{(k)}(t - kM) * s_l^{(k)}(t). \tag{15}$$

Therein, the phase ϕ can be selected for beneficial time behaviour of the correction filter s, as for example linear or minimum phase. The correction filters are then convolved with the respective blocks and added. When implementing this equalization it is important to assure that at least M zeros are appended to each block in order to avoid time aliasing. What should further be mentioned is that there are also other variants of this equalization. In [22] for example, equalization is instead done with an octave filter-bank on each of the Ambisonics channels, as obtained by eq. (12). In [2], instead of equalizing the response, only the decay of the rendered binaural response is adjusted.

4.4 ... cannot create fully authentic stimuli in general

We could say that a virtual acoustics rendering is “authentic", if it can not be distinguished from a real world stimulus. Concerning SDM, one way to test its authenticity would be to compare a binaural SDM rendering to a binaural dummy head reference. Given the properties of rendering described above, which need to be carefully corrected, it is not surprising that in listening tests using direct comparison to a binaural reference, SDM is usually found to be noticeably different [1, 11, 13].

4.5 ... can create plausible auralizations

Even though, in a direct comparison to a binaural reference, differences are often audible, SDM can still produce high quality renderings. Plausibility can be defined as evoking the auditory illusion of a rendered sound to be real [7]. If this illusion even occurs when both rendered sources and real sources are present, either simultaneously or non-simultaneously, one may speak of transfer-plausible rendering [20]. In [2], binaural SDM renderings were used in a test akin to a transfer-plausibility design. They were believed to be real as often as a real loudspeaker.

5 CONCLUSION AND FUTURE DIRECTIONS

We have demonstrated several aspects of what SDM can do, and some that it can not. It has became clear that the broadband analysis stage of SDM is a robust way to identify reflections in the early part of the response, but that the temporal resolution is limited by the size of the used microphone array, regardless of the estimation method. Also, we have seen that despite the obvious violation of the sound field model, the late energy distributions can be approximated in a particular sense. More potential problems lie on the synthesis side, for example when rendering very transient signals. Here, the simplicity of the sound field model poses a problem, but we have described strategies for roughness and whitening compensation. Still, all previous research has shown that SDM can robustly provide renderings that possess the main characteristics of a measured space, which makes it a valuable tool for further research. In the near future, a new version of the SDM toolbox will be released, which incorporates all the discussed processing steps, including roughness compensation. Future directions for parametric processing of impulse responses in general are incorporating more advanced sound field models with multiple sources and developing strategies for reliable objective and subjective testing of SRIR algorithms.

ACKNOWLEDGEMENTS

This work was part of the VRACE project that has received funding from the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska Curie actions (grant agreement number 812719.)

REFERENCES

[1] J. Ahrens. Perceptual Evaluation of Binaural Auralization of Data Obtained from the Spatial Decomposition Method. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, USA, Oct. 2019.

[2] S. V. Amengual Garí, P. T. Calamia, and P. W. Robinson. Optimizations of the spatial decomposition method for binaural reproduction. *J. Audio Eng. Soc.*, 68(12):959–976, Dec. 2020.

[3] S. V. Amengual Garí, W. Lachenmayr, and E. Mommertz. Spatial analysis and auralization of room acoustics using a tetrahedral microphone. *J. Acoust. Soc. Am.*, 141(4):EL369–EL374, Apr. 2017.

[4] M. Frank and F. Zotter. Spatial impression and directional resolution in the reproduction of reverberation. In *DAGA - Fortschritte der Akustik*, Mar. 2016.

[5] L. Gölles and F. Zotter. Directional enhancement of first-order ambisonic room impulse responses by the 2+2 directional signal estimator. In *15th International Conference on Audio Mostly*, Sept. 2020.

[6] E. Hoffbauer and M. Frank. 4-directional ambisonic spatial decomposition method with reduced temporal artifacts. 2022. accepted.

[7] A. Lindau and S. Weinzierl. Assessing the plausibility of virtual acoustic environments. *Acta Acust*, 98(5):804–810, Sept. 2012.

[8] T. Lokki, J. Pätynen, A. Kuusinen, and S. Tervo. Concert hall acoustics: Repertoire, listening position, and individual taste of the listeners influence the qualitative attributes and preferences. *J. Acoust. Soc. Am.*, 140(1):551–562, July 2016.

[9] L. McCormack, V. Pulkki, A. Politis, O. Scheuregger, and M. Marschall. Higher-order spatial impulse response rendering: investigating the perceived effects of spherical order, dedicated diffuse rendering, and frequency resolution. *J. Audio Eng. Soc*, 68(5):368–354, May 2020.

[10] J. Merimaa. *Analysis, synthesis, and perception of spatial sound - binaural localization modeling and multichannel loudspeaker reproduction*. PhD thesis, Helsinki University of Technology, 2006.

[11] N. Meyer-Kahlen, S. J. Schlecht, and T. Lokki. Parametric late reverberation from broadband directional estimates. *Int. Conf. on Immersive and 3D Audio (I3DA)*, Sept. 2021.

[12] N. Meyer-Kahlen, S. J. Schlecht, and T. Lokki. Perceptual roughness of spatially assigned sparse noise for rendering reverberation. *J. Acoust. Soc. Am.*, 150(5):3521–3531, Nov. 2021.

[13] A. Pawlak, H. Lee, T. Lund, and A. Makivirta. Subjective evaluation of spatial analysis and synthesis methods using different microphone arrays. In *Int. Conf. on Immersive and 3D Audio (I3DA)*, Sept. 2021.

[14] J. Pätynen, S. Tervo, and T. Lokki. Analysis of concert hall acoustics via visualizations of time-frequency and spatiotemporal responses. *J. Acoust. Soc. Am.*, 133(2):842–857, Feb. 2013.

[15] J. Riionheimo. Movie sound, part 2: Preference and attribute ratings of six listening environments. *J. Audio Eng. Soc.*, 69(1), Jan. 2021.

[16] S. Tervo and T. Lokki. Interpolation methods for the SRP-PHAT algorithm. In *International Workshop on Acoustic Signal Enhancement*, Jan. 2008.

[17] S. Tervo, J. Pätynen, N. Kaplanis, and e. al. Spatial analysis and synthesis of car audio system and car cabin acoustics with a compact microphone array. *J. Audio Eng. Soc.*, 63(11):914–925, Feb. 2015.

[18] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki. Spatial decomposition method for room impulse responses. *J. Audio Eng. Soc.*, 61(1):17–28, Mar. 2013.

[19] S. Tervo, J. Saarelma, J. Pätynen, I. Huhtakallio, and P. Laukkanen. Spatial analysis of the acoustics of rock and nightclubs. In *Proceedings of the Institute of Acoustics*, volume 37, pages 551–558, Oct. 2015.

[20] S. Wirler, N. Meyer-Kahlen, and S. J. Schlecht. Towards transfer-plausibility for evaluating mixed reality audio in complex scenes. In *AES International Conference on Audio for Virtual and Augmented Reality*, Aug. 2020.

[21] Y. Yamasaki and T. Itow. Measurement of spatial information in sound fields by closely located four point microphone method. *J. Acoust. Soc. Jpn (E)*, 10(2):101–110, 1989.

[22] M. Zaunschirm, M. Frank, and F. Zotter. Binaural rendering with measured room responses: first-order ambisonic microphone vs. dummy head. *Applied Sciences*, 10(5):1631, Feb. 2020.

[23] F. Zotter and M. Frank. *Ambisonics: A practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*, volume 19 of *Springer Topics in Signal Processing*. Springer International Publishing, 2019.