OPTIMIZING THE SPATIAL DECOMPOSITION METHOD FOR BINAURAL RENDERING

Sebastià V. Amengual Garí[1] **Johannes M. Arend**[1,2]
Paul Calamia[1] **Philip Robinson**[1]
[1] Facebook Reality Labs, Redmond, USA
[2] Technische Hochschule Köln, Cologne, Germany
samengual@fb.com

ABSTRACT

The spatial decomposition method (SDM) aims at parameterizing a sound field as a succession of plane waves, allowing the analysis and rendering of multichannel room impulse responses (RIRs). The method was originally developed for the use with open microphone arrays, utilizing time differences of arrival to compute directional estimates. A later version introduced the use broadband pseudo-intensity vectors from B-format RIRs. Through simulations and measurements, we explore optimal values for the various processing parameters such as array size and temporal processing window size and compare the results of TDOA and PIV DOA estimation. We introduce spatial clustering of reflections as a post-processing step, which reduces the un-natural direction-of-arrival spread of late reflections at the expense of spatial distortion for consecutive reflections. To address whitening of late reverberation, we introduce RTMod+AP, an equalization approach specifically designed for the correction of binaural RIRs (BRIRs), allowing the use of dense HRTF datasets for the synthesis of SDM data. In a perceptual experiment we investigate the links between spatial resolution and plausibility of binaural SDM auralizations by directly comparing head-tracked renderings against real loudspeakers. [1]

1. INTRODUCTION

The spatial decomposition method (SDM) aims at parametrizing a sound field as a succession of plane waves, by assigning a direction of arrival (DOA) to each of the samples of a room impulse response (RIR). As such, it can be used for the analysis and reproduction of a sound field based on measured multichannel room impulse responses (RIRs). Originally, the method was developed for the use with open microphone arrays, exploiting the time differences of arrival (TDOA) [2] between microphones to generate DOA estimates. The same authors released an open implementation of the algorithm [3], which included an alternative analysis approach based on broadband pseudo-intensity vectors (PIVs) of B-format RIRs. In this paper we compare simulation and measurement results results from PIV arrays to those of TDOA and investigate optimal parameters for the analysis using TDOA and open arrays.

Rendering SDM data using dense Head Related Transfer Function (HRTF) datasets allows achieving a high degree of spatial resolution. However, we show that this can result in severe timbral degradations. As the RIR progresses into the late reverberation and multiple reflections overlap in time, the sound field assumption of consecutive plane waves is violated and DOA estimates become unstable and less reliable. In turn, consecutive samples of the RIR are mapped to disparate locations, destroying narrow band information and resulting in an increase of broadband energy. This effect is accentuated when using spatially dense HRTFs, as small fluctuations in the DOA estimates result in incorrect mapping of samples onto several adjacent HRTF directions. A potential mitigation of this artifact is the spatial clustering of reflections. Here we investigate the application of regular quantization grids on the DOA estimates of early reflections and late reverberation.

Another consequence of rapidly varying DOA estimates when directly rendering of RIRs analyzed with SDM is the whitening of late reverb, which was further addressed by Tervo et al. by proposing a time-frequency equalization [4]. This equalization is best suited for loudspeaker reproductions, as the resulting time-varying filters are generated by comparing the rendered loudspeaker streams with the original omnidirectional RIR. When applied to binaural rendering with a spatially dense HRTF dataset this approach becomes impractical due to computing limitations, given that a large number of intermediate loudspeaker streams need to be rendered prior to binauralization using a virtual loudspeaker approach. In this paper we propose an alternative equalization comprising a reverberation correction process (RTMod) and the processing of the resulting Binaural Room Impulse Responses (BRIR) with a cascade of all-pass filters (RTMod+AP).

It has been suggested previously that the use of synthetic spatial data and the reduction of spatial resolution results in only minor impairments when auralizing SDM data [5]. As such, in perceptual experiments we investigate the minimum required spatial resolution of binaural SDM renderings by directly comparing binaural signals against real loudspeakers. Preliminary results suggest that auralizations with binaural room impulse responses using high

[1] Due to changes in the conference schedule caused by COVID-19, an extended version of this manuscript is in press at the time of publication [1]. Parts of the present text and figures are extracted from the cited manuscript.

spatial resolution for the direct sound but only 14 DOAs for reflections are rated as being as plausible as real loudspeakers.

2. DOA ESTIMATION: SIMULATIONS

The DOA analysis method presented in the original SDM paper proposes an estimation approach relying on time differences of arrival between windowed signals originating from closely spaced microphones in an open array configuration [2]. A public code release by the same authors [3] also includes an estimation method based on the instantaneous or windowed pseudo-intensity of a B-format array. In this section we evaluate the performance of both approaches using simulations.

2.1 Simulated RIRs

We simulated multichannel RIRs of 500 shoebox rooms with side lengths uniformly distributed between 2 and 25 m, in order to cover a meaningful range of room sizes. The source and receiver locations are randomly chosen in each simulation. Image Source Method (ISM) simulations were conducted with AKTools [6] and include frequency dependent material absorption and air absorption. The materials for each wall are different and are kept constant for all the room configurations. We expanded the functionality of the simulator to include ideal first-order microphones to generate B-format signals for their subsequent analysis using broadband PIV. The simulated RIRs contain 64 sound events, corresponding to the direct sound and specular reflections up to 3^{rd} order. Note that the simulations in this case contain purely specular information and the generalization of the analysis to measured RIRs might be limited.

2.2 Evaluation Metric

We evaluate the performance of the DOA estimation using an objective metric ϵ_{DOA} that evaluates the DOA error of each sample in the RIR. The angular distance between the ground truth DOA $\mathbf{D}_{ISM}(n)$ and the estimated direction $\mathbf{D}(n)$ is computed and weighted by the energy of the corresponding sample. The resulting total error is normalized by the total energy of the RIR.

$$\epsilon_{DOA} = \frac{\sum_{n=1}^{N} \arccos\{\mathbf{D}(n)^{\mathrm{T}}\mathbf{D}_{ISM}(n)\}p(n)^2}{\sum_{n=1}^{N} p(n)^2} \quad (1)$$

where p_n is the instantaneous pressure of sample n and N is the number of samples in the RIR. The DOA vectors are unit vectors in cartesian coordinates. Note that since this metric evaluates the estimation error of each sample it is only suitable for the analysis of RIRs that do not present overlapping reflections.

2.3 Time differences of arrival

DOA estimation using TDOA requires a compact microphone array with at least four microphones arranged in a 3D space. If the data is intended for auralization, one of the microphones must be omnidirectional - for analysis only, other directivities are in principle acceptable. Multiple microphone array configurations have been reported in the literature, such as arrays of microphone pairs arranged in orthogonal directions - with or without a center microphone [4, 7–10], tetrahedral arrays with a center (physical or virtual) omnidirectional microphone [9, 11], or a 12 element star shaped array [9].

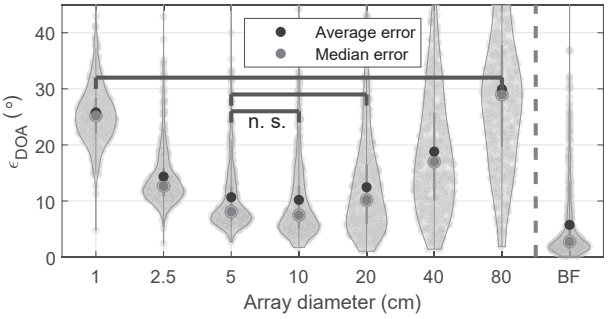

Figure 1. DOA error for various array configurations (500 ISM simulations at $f_s = 48$ kHz). Different diameters refer to TDOA analysis using an open microphone array (7 microphones with a center capsule and 6 capsules arranged in pairs on orthogonal axes) and BF refer to PIV analysis using an ideal B-format array. Brackets denote statistically non-significant differences.

2.3.1 Array size

One of the most commonly used topologies is that of an array composed of 6 or 7 omnidirectional microphones (3 orthogonal pairs, 1 center microphone) [4, 8–10]. We investigated the optimal size of this geometry by performing DOA analysis on the 500 simulated ISM RIRs. We performed the analysis using the function `SDMpar` of the SDM Toolbox [3] and the smallest allowed window size for each array configuration. The results in Fig. 1 show that, for a sampling rate of 48 kHz, a microphone diameter of 10 cm presents the smallest DOA estimation error among the evaluated dimensions. Perceptual validations found that, at this sampling rate, auralizations using arrays of 6 sensors with diameters of 10 and 20 cm result in smaller perceptual differences when compared to arrays of 4 cm diameter or other configurations such as a tetrahedral array or a 12-element array [9]. When using smaller arrays it might be beneficial to utilize higher sampling rates, as in [4] in order to increase the time resolution and avoid spatial quantization of the DOA estimates. A formal comparison of array sizes for other sampling rates is left for future work.

2.3.2 Window size

When performing DOA estimation using TDOA analysis, the size of the sliding window determines the compromise between temporal and spatial resolution. In [2], the authors recommend the use of a time window slightly longer than the time needed for an acoustic event to travel along the longest array dimension. In Fig. 2, the DOA estimation er-

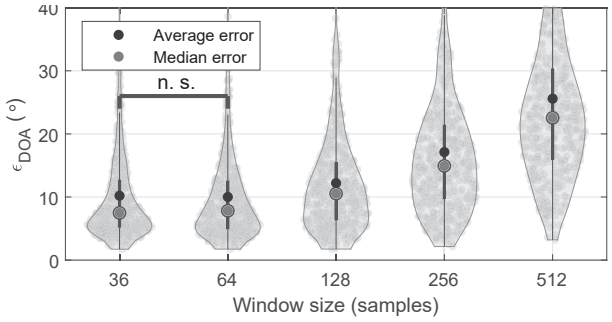

Figure 2. DOA error for various analysis window sizes using signals from an open microphone array (500 ISM simulations at $f_s = 48\,\text{kHz}$). The array has a diameter of 10 cm (7 microphones with a center capsule and 6 capsules arranged in pairs on orthogonal axes). Brackets denote statistically non-significant differences.

rors for various window sizes are reported. The estimation error grows proportionally with the window length. For the evaluated case of a 10 cm diameter array at 48 kHz, a window size of 36 samples seems appropriate, although the DOA estimation error using a window of 64 samples is not statistically significantly larger.

2.4 Pseudo-intensity vectors

The directional analysis can also be conducted using signals from first order coincident array configurations (B-format arrays). This analysis was first applied in the Spatial Impulse Response Rendering (SIRR) method [12], using narrow-band analysis. In SDM, typically broadband PIVs are used. However, in some cases SDM auralizations using broadband PIVs for the DOA estimation have been found to be perceptually unsatisfactory [13, 14], presenting lower ratings when directly compared against open microphone array auralizations and a reference [9]. These studies have not explicitly investigated the root cause of the perceptual impairments of PIV SDM auralizations, which could be caused partially by a subpar DOA estimation but also due to shortcomings in the rendering process. Nonetheless, in [9] the analysis results using instantaneous PIVs are significantly worse than those of TDOA analysis. Other investigations using band-pass filtered and smoothed the DOA estimates reported that binaural auralizations are perceptually very similar to a dummy head reference [15].

As with the open microphone arrays, we conducted a DOA analysis on 500 simulated RIRs featuring ideal B-format receivers. We obtained the DOA estimates through the `SDMbf` function from the SDM Toolbox [3] using instantaneous (non-windowed) PIVs. As reported in Fig. 1, in an ideal simulation the results from a B-Format array are significantly better than those from the best tested open array configuration. However, note that when using real measurements other factors such as imperfect directional properties, spatial aliasing exhibited by B-format microphones and gain mismatch between sensors will contribute to degrading the performance of the PIV DOA analysis.

3. DOA ESTIMATION: MEASUREMENTS

ISM simulations represent the best case scenario for DOA estimation using SDM, as the method assumes a sound-field model composed of successive specular events. In addition, the sensors exhibit ideal characteristics and the RIRs are free of noise. In order to evaluate the performance of the TDOA and PIV approach in a practical scenario, we conducted RIR measurements using a tetrahedral microphone (CoreSound Tetramic) and then used either the A-format signals (4 cardioid microphones at the edges of the tetrahedron) or the B-format signals to estimate the DOA. The A-to-B conversion is performed using individually calibrated encoding matrices, as provided by the manufacturer. The room used for this measurements is an apartment-like scene with a high absorptive ceiling.

The results for various window sizes are presented in Fig. 3. Both methods present considerable agreement for the DOA of more energetic samples, although they seem to differ by a slight angular offset. Longer windows provide more stable estimates. As ground truth data for the DOAs is not available, it is not straightforward to assess which of the estimations is closer to the actual DOAs. However, it seems reasonable to conclude that PIV analysis using instantaneous signals ($w = 1$ sample) results in a significantly poorer performance than with longer windows. It is observed that estimated DOAs corresponding to the direct sound are scattered around the entire sphere. In the PIV analysis the window is convolved with the PIVs, effectively acting as a low pass filter. Longer windows contribute to generating more stable DOA estimates without signifficantly affecting the DOA of the strongest events. However, longer windows contribute to creating trailing patterns betwen DOAs of strong events.

For the TDOA approach, trailing patterns are less pronounced. With the use of longer windows, the estimated DOAs tend to concentrate around strong events. However, the estimated DOA of one specific reflection (light green) changes significantly depending on the window size.

4. RENDERING

Binaural room impulse responses (BRIRs) can be synthesized as a weighted sum of HRIRs corresponding to each DOA, appropriately delayed and weighted by the amplitude of the instantaneous pressure of the omnidirectional RIR. However, assigning and spatializing a DOA for each of the samples in a RIR is only conceptually correct for a sound field composed of consecutive broadband pulses, in which each of the samples of the RIR corresponds to one separate acoustic event. This does not occur in practice, where reflections are band limited - with each pulse spanning several samples - and overlap with each other, eventually resulting in a diffuse sound field.

The violation of this sound-field model when synthesizing BRIRs results in two artifacts: spectral distortion of discrete reflections and late reverberation whitening. These two artifacts are more accentuated when the mapping of consecutive samples is assigned and rendered to

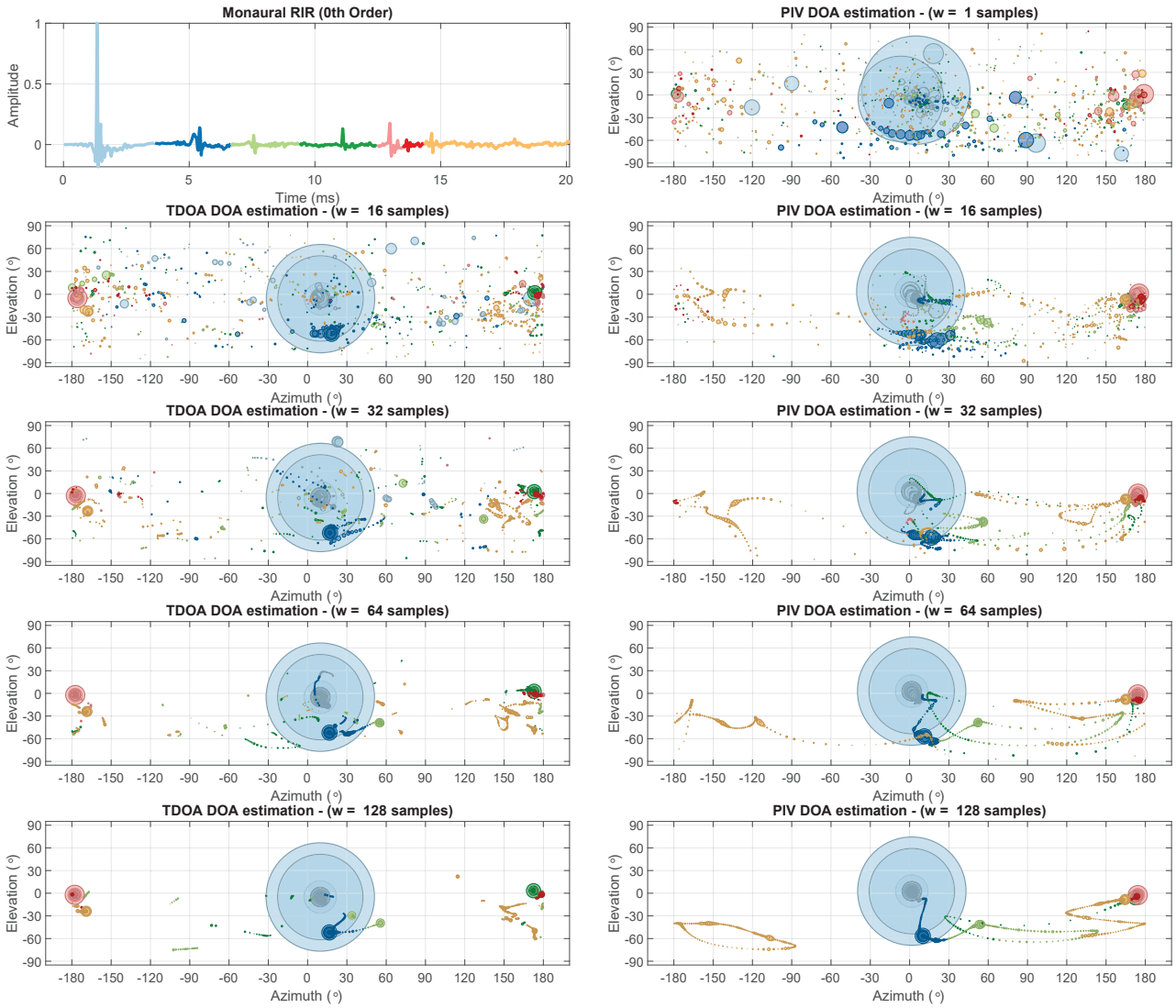

Figure 3. DOA estimation results from TDOA (left) and PIV (right) analysis at various window sizes ($f_s = 48\,\text{kHz}$) using a tetrahedral array (Tetramic). Each circular marker represents a sample of the monaural RIR, with area proportional to the instantaneous energy of each sample. Note that in PIV analysis windows are convolved with the PIVs, effectively low-passing the directional estimates.

several locations, and thus it is expected to be more severe when the rendering is done with dense HRTF datasets, as opposed to loudspeaker layouts. In other words, when DOA estimates of consecutive samples differ ever so slightly, they will be mapped to several adjacent HRTF directions, resulting in a spatially spread succession of broadband pulses.

4.1 DOA Postprocessing

When rendering SDM data using loudspeaker reproduction, a common approach is the use of Nearest Loudspeaker Synthesis [4, 8], which assigns the DOA to the closest loudspeaker. This reduces the spatial spread of single events by collapsing nearby DOA values to a single location, but might also result in noticeable localization shifts, as a loudspeaker might not be present at the position of the direct sound. An optimization method of the loudspeaker layout is available in [16]. With regard to binaural synthesis, previous studies propose the use of a moving average filter to smooth the DOA estimates [5, 10, 15].

4.1.1 Direct sound

As demonstrated in Fig. 3, the DOA of the most energetic samples seem to be reliably estimated, although it is not uncommon to observe trailing patterns between them. To further investigate the timbral effects of these artifacts on the rendering of the direct sound we auralized the first 160 samples of a RIR measured with a 7-microphone array (10 cm diameter, one central microphone).

The data are reported in Fig. 4, showing the DOA in four different cases: raw DOA estimates, smoothed DOA estimates with a moving average filter of 16 samples, an arbitrary deviation for illustrative purposes and a reference case with perfectly stable DOA. It is observed that even

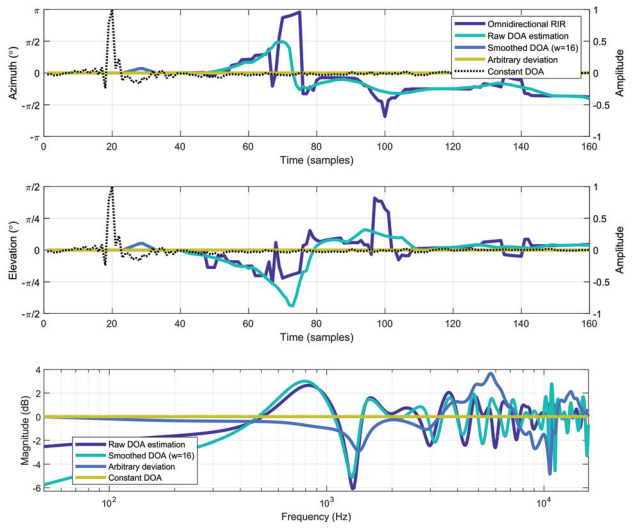

Figure 4. DOA of the direct sound (top two panels) and spectral deviations resulting from mapping samples to multiple locations (bottom panel).

when the DOAs of the most energetic samples are appropriately estimated, spectral deviations of up to 6 dB are present.

Given that the direct sound is arguably one of the main acoustic events that define the timbral properties of the RIR, we suggest that the DOA for the direct sound should be fixed for a number of samples sufficiently high to ensure that its spectral properties are preserved. However, care needs to be taken to ensure that the DOA estimation of the first event after the direct sound is preserved.

4.1.2 DOA Quantization

The same spatial-timbral trade-off discussed for direct sound is present for reflections and late reverberation. This becomes especially severe when low-passed and overlapped reflections appear in the RIR, rendering the high temporal resolution of the DOAs unusable - using all the available information results in spreading specular reflections onto multiple directions, resulting in spatial and timbral degradations.

As a straightforward approach to reduce the timbral artifacts we investigate the use of an arbitrary grid to reduce the spatial spread of early reflections - while maintaining full spatial resolution for the direct sound. In Section 5 we investigate the use of Lebedev grids and the minimum required resolution through perceptual tests.

Other clustering approaches are suitable in this case, such as the optimization of the virtual loudspeaker layout using weighted spatial energy maps [16]. Another promising approach is the use of Density Based Spatial Clustering (DBSCAN) [17], which could be used to identify portions of the RIR with meaningful DOA data and cluster them. Additionally, DBSCAN could also be used to identify unreliable DOA data which could then be rendered as a diffuse component, in a similar fashion to SIRR [12]. However, this is out of the scope of this manuscript.

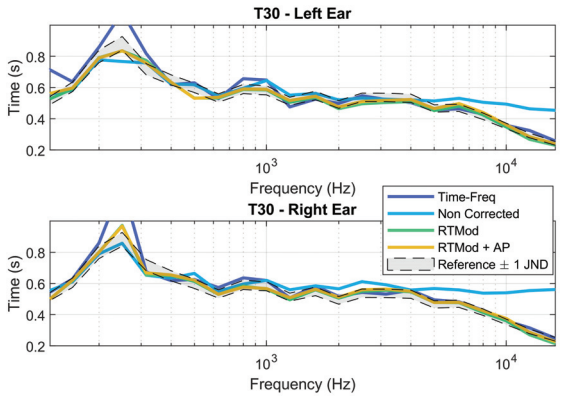

Figure 5. Reverberation time (T30) of various reverberation equalization techniques. The shaded grey area refers to deviations of ± 1 JND from the reference ($\pm 5\%$ of the reference T30, as defined in ISO 3382-1:2009).

4.2 Reverb equalization

Direct auralization of an RIR using DOA data to either map the energy to discrete loudspeakers or to generate BRIRs causes a perceivable whitening of the late reverb. In this section we propose a new method to equalize the late reverberation of rendered BRIRs and compare it to the original time-frequency equalization [4].

The reverb equalization algorithm proposed in [4] equalizes each of the rendered streams by comparing their time-frequency properties to those of the original RIR and applying a time-varying equalization to the rendered RIRs. However, this method is not well suited for BRIRs rendered with dense HRTF datasets, as it performs an independent equalization for each of the directions used in the rendering, resulting in impractical time and memory constraints. When downsampling the DOA data, the method runs satisfactorily. In this case, the SDM-rendered BRIRs are first rendered as virtual loudspeaker streams, equalized and convolved with HRTFs corresponding to the loudspeaker directions.

Instead of performing a time-frequency equalization, we propose a two-step approach applied directly to the rendered BRIRs. The first step is a reverberation-time modification (RTMod) of the rendered BRIRs by multiplying them with an exponential function, whose parameters are derived from comparing the reverberation time (T30) of the omnidirectional RIR and the rendered BRIRs. The details of the computation are described in [1, 10]. This step results in BRIRs that have the correct time-frequency properties, although the late reverberation presents more coarse properties than a reference measured BRIR, due to discretely mapping consecutive samples to several HRTFs, instead of rendering a diffuse sound field. This is easily audible when directly listening to the rendered BRIRs, although informal listening revealed a negligible effect for continuous signals and very small artifacts for highly impulsive sounds.

To reduce the signal-dependent quality of the late reverb we compensate for the lack of diffuseness by applying

a cascade of 3 all-pass (AP) filters to the late reverb (RTMod+AP). For this, early reflections and late reverberation are split at the mixing time and the late reverb is processed using the AP filter cascade. This results in increased diffuseness of the late reverberation, but since the same filters are used for both ears the Inter-Aural Cross Correlation is not affected. The early reflections and processed late reverberation are summed back together using cosine ramps.

In the examples investigated in this paper we used a mixing time of 3800 samples (80 ms) and crossfade ramps of 1024 samples. Note that to implement AP filters without creating audible modulations, the delays should be coprime (37, 113 and 215 samples in the investigated examples). We fix a reverberation time of the AP filters at 0.1 s. The entire rendering process, from pressure RIR and DOA data to equalized RTMod+AP BRIRs takes approximately 0.15 s, while synthesis and time-frequency equalization as in [4] takes 8.9 s on the same machine (laptop PC, Intel core i7, running Matlab 2018 and Windows 10).

4.2.1 Validation

To compare the performance of the equalization methods we compared a dummy head reference measurement (KEMAR) with BRIR renderings generated using HRTF measurements of the same mannequin. We used a microphone array of 10 cm diameter with a central microphone and 6 pairs arranged on orthogonal axes, an analysis window of 62 samples and a moving average window of 16 samples to smooth the DOA estimates. For the RTMod and RTMod+AP examples the spatial resolution of the BRIR has been downsampled to 50 directions using a Lebedev grid while keeping the first 160 samples fixed to the original direct sound direction.

Examples of estimated T30 for a BRIR processed with the presented methods are shown in Fig. 5. It can be clearly observed that the non-corrected case (simply convolving delayed and weighted HRTFs with an omnidirectional response) yields an excessive high frequency reverberation. The time-frequency equalization presents T30 results closer to the reference, although slightly overestimated at low frequencies (approx 250 Hz) and around 1 kHz. Finally, both RTMod and RTMod+AP methods present the closest results to the reference, with deviations within $\pm$1JND in most of the frequency bands.

Results of Inter-Aural Cross Correlation (IACC) are depicted in Fig 6. The time-freq method presents higher deviations at low frequencies in all three cases (full, early and late). Non Corrected, RTMod and RTMod+AP present deviations generally within $\pm$1JND, except for larger deviations below 700 Hz in the late IACC.

In Fig. 7 we present the absolute amplitude and envelope of the left ear signal for each of the rendering methods. It is observed that the RTMod+AP yields a better approximation of the envelope than Time-Frequency and RTMod equalizations. Informal listening also confirmed that the roughness of the late reverberation is largely corrected.

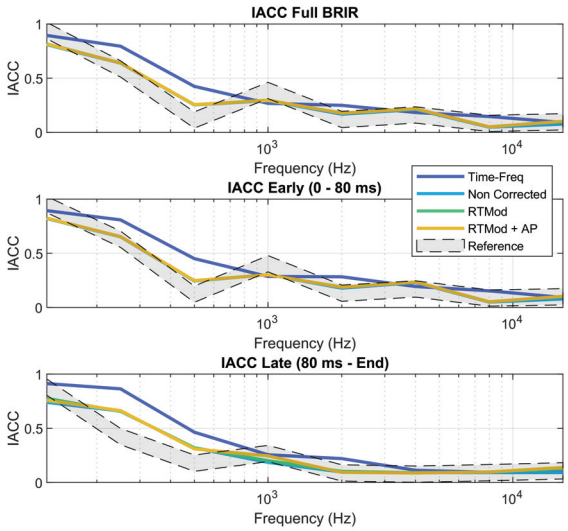

Figure 6. Inter-Aural Cross Correlation of the reference and rendered BRIRs with various equalization methods. The shaded grey area refers to deviations of $\pm$1 JND from the reference ($\pm$0.075 as defined in ISO 3382-1:2009).

5. LISTENING TESTS

To investigate the minimum detectable DOA grid quantization we conducted a pilot listening test where listeners wearing non-occluding headphones (AKG K1000) were asked to compare a real loudspeaker and binaural renderings directly. Generic (KEMAR) HRTFs and headphone equalization were used. Dynamic binaural (2DOF - yaw + pitch) presentation was implemented by tracking both the loudspeaker and headphones using OptiTrack (see [10] for more details on the experimental set-up).

The experiment was conducted in the same room used for objective comparisons (see Figs 5, and 6 for acoustical parameters). We generated renderings using the RTMod equalization method (without AP cascade filtering) with various DOA quantization configurations using Lebedev grids. 7 different stimuli were presented to the listeners (real loudspeaker or BRIR renders with 1, 2, 6, 14, 26, or 50 DOA). The direction of the direct sound was kept intact and fixed for the first 128 samples. The reverberation was rendered separately and quantized as well, although it was presented statically (head locked).

In the experiment, 8 listeners were presented with a single source hidden behind a visually opaque but acoustically transparent curtain. In a 2-Alternative Forced Choice (2AFC) test they were asked to answer which of the two presented stimuli was a real loudspeaker behind the curtain. The audio content used in the test was a sequence of castanets. They then input their answers on a touchscreen. The total number of trials per subject was 21 (pairwise comparison, no repetitions) and the order or presentation was randomized. Users were encouraged to make use of natural head rotations and had unlimited listening time.

Results are reported in Fig. 8 as the number of selec-

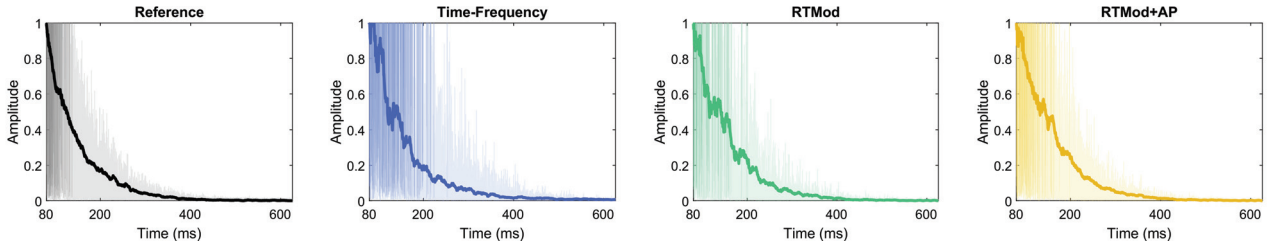

Figure 7. Absolute pressure (thin lines) and envelopes (thick lines) of a reference BRIR (dummy head measurement) and various binaural SDM rendering methods. Only left ear signal is shown.

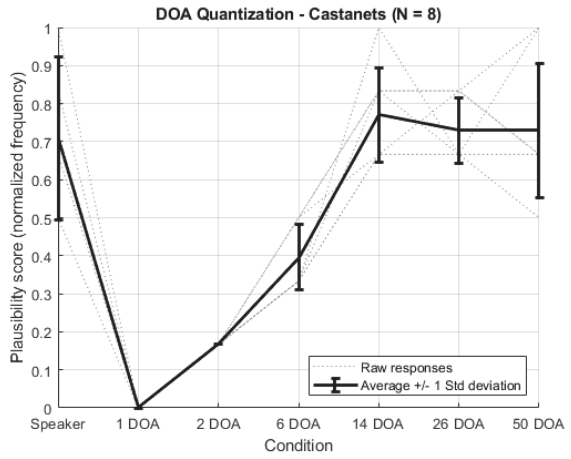

Figure 8. Perceived plausibility of binaural SDM (RTMod) renderings with various degrees of spatial resolution as compared to a real loudspeaker.

tions of one stimulus normalized by the total number of appearances. We label the metric as 'Plausibility Score', and a value of 1 would indicate that a certain stimulus is always more plausible than the rest. The results suggest that listeners perceive renderings with 14 or more DOAs as being as plausible as the real loudspeaker. The small spread at conditions 1, 2 and 6 DOAs suggests that all listeners reliably discriminated between these cases and the real loudspeaker or higher resolution renderings. In addition to showing that increased DOA resolution is not necessarily audible, the results suggest that the rendering improvements presented in the manuscript allow for the rendering of plausible virtual sources, even in the explicit presence of and comparison to real sources.

6. CONCLUSIONS

In this paper we presented several improvements for the rendering of SDM to binaural, including the exploration of optimal analysis parameters, the quantization of spatial information and the development of a binaural equalization approach to address the reverberation whitening.

ISM simulations confirmed that, at a sampling rate of 48 kHz, the optimal parameters for an open array configuration are a diameter of 10 cm and an analysis window between 36 and 64 samples. While in simulations B-format array analysis performs better than open arrays, it is likely that microphone array imperfections significantly affect the performance of the PIV analysis.

Using the same array (Tetramic) for the PIV and TDOA analysis of measured RIRs confirmed that both methods are suitable to obtain reliable estimates of the most prominent events of the RIR. When using PIV analysis it is desirable to use longer convolution windows in the analysis.

We introduced the RTMod+AP equalization approach, which acts directly on re-synthesized BRIRs, by modifying their reverberation time and processing the BRIRs using a cascade of all-pass filters. This results in better objective results than other state of the art equalization methods while requiring fewer computational resources and scales independently from the spatial resolution of the used HRTF dataset.

Perceptual results suggest that equalization with RTMod provides perceptually plausible results when comparing dynamic binaural auralizations to real loudspeakers. Additionally, the spatial resolution of the reflections and reverberation can be downsampled to 14 directions on a Lebedev grid without impairing the plausibility of the rendered BRIRs. Complete perceptual evaluation of RTMod+AP is left for future work.

7. ACKNOWLEDGEMENTS

We want to thank Henrik Hassager, Nils Meyer-Kahlen, and Prof. Tapio Lokki for fruitful discussions and valuable feedback on the manuscript.

8. REFERENCES

[1] S. V. Amengual Gari, J. M. Arend, P. T. Calamia, and P. W. Robinson, "Optimizations of the spatial decomposition method for binaural reproduction," *J. Audio Eng. Soc*, 2020 (In press).

[2] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, "Spatial decomposition method for room impulse responses," *J. Audio Eng. Soc*, vol. 61, no. 1/2, pp. 17–28, 2013.

[3] S. Tervo, "SDM Toolbox." `https://www.mathworks.com/matlabcentral/fileexchange/56663-sdm-toolbox`.

[4] S. Tervo, J. Pätynen, N. Kaplanis, M. Lydolf, S. Bech, and T. Lokki, "Spatial analysis and synthesis of car au-

dio system and car cabin acoustics with a compact microphone array," *J. Audio Eng. Soc*, vol. 63, no. 11, pp. 914–925, 2015.

[5] J. Ahrens, "Auralization of omnidirectional room impulse responses based on the spatial decomposition method and synthetic spatial data," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 146–150, May 2019.

[6] F. Brinkmann and S. Weinzierl, "Aktools—an open software toolbox for signal acquisition, processing, and inspection in acoustics," in *Audio Engineering Society Convention 142*, May 2017.

[7] S. V. Amengual Garí, D. Eddy, M. Kob, and T. Lokki, "Real-time auralization of room acoustics for the study of live music performance," in *Fortschritte der Akustik - DAGA 2016*, (Aachen, Germany), March 2016.

[8] N. Kaplanis, S. Bech, T. Lokki, T. van Waterschoot, and S. Holdt Jensen, "Perception and preference of reverberation in small listening rooms for multi-loudspeaker reproduction," *The Journal of the Acoustical Society of America*, vol. 146, no. 5, pp. 3562–3576, 2019.

[9] J. Ahrens, "Perceptual evaluation of binaural auralization of data obtained from the spatial decomposition method," in *2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 65–69, Oct 2019.

[10] S. V. Amengual Garí, W. O. Brimijoin, H. G. Hassager, and P. W. Robinson, "Flexible binaural resynthesis of room impulse responses for augmented reality research," in *EAA Spatial Audio Signal Processing Symposium*, (Paris, France), pp. 161–166, Sept. 2019.

[11] S. V. Amengual Garí, W. Lachenmayr, and E. Mommertz, "Spatial analysis and auralization of room acoustics using a tetrahedral microphone," *The Journal of the Acoustical Society of America*, vol. 141, no. 4, pp. EL369–EL374, 2017.

[12] J. Merimaa and V. Pulkki, "Spatial impulse response rendering i: Analysis and synthesis," *J. Audio Eng. Soc*, vol. 53, no. 12, pp. 1115–1127, 2005.

[13] C. F. Hold, "Spatial Decomposition Method on non-uniform reproduction layouts," Master's thesis, TU Berlin, Germany, 2019.

[14] 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*, vol. 68, pp. 338–354, June 2020.

[15] M. Zaunschirm, M. Frank, and F. Zotter, "Brir synthesis using first-order microphone arrays," in *Audio Engineering Society Convention 144*, May 2018.

[16] O. Puomio, J. Pätynen, and T. Lokki, "Optimization of virtual loudspeakers for spatial room acoustics reproduction with headphones," *Applied Sciences*, vol. 7, p. 1282, Dec 2017.

[17] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise.," in *Kdd*, vol. 96, pp. 226–231, 1996.