EVALUATION OF SENSOR SELF-NOISE IN BINAURAL RENDERING OF SPHERICAL MICROPHONE ARRAY SIGNALS

Hannes Helmholz[⋆]*, Jens Ahrens*[⋆]*, David L. Alon*[†]*, Sebastià V. Amengual Garí*[†]*, Ravish Mehra*[†]

[⋆] Chalmers University of Technology, 412 58 Gothenburg, Sweden
{hannes.helmholz,jens.ahrens}@chalmers.se
[†] Facebook Reality Labs, Facebook, 1 Hacker Way, Menlo Park, CA 94025, USA

ABSTRACT

Spherical microphone arrays are used to capture spatial sound fields, which can then be rendered via headphones. We use the Real-Time Spherical Array Renderer (ReTiSAR) to analyze and auralize the propagation of sensor self-noise through the processing pipeline. An instrumental evaluation confirms a strong global influence of different array and rendering parameters on the spectral balance and the overall level of the rendered noise. The character of the noise is direction independent in the case of spatially uniformly distributed noise. However, timbre of the rendered self-noise changes with head orientation in the case of spatially non-uniform noise. We determine audibility thresholds of the coloration artifact during head rotations for different array configurations in a perceptual user study.

Index Terms— Sensor self-noise, Spherical microphone arrays, Binaural rendering, Real-time signal processing

1. INTRODUCTION

The accurate capture and reproduction of spatial sound fields is of great interest to many fields of research and the industry, e.g. virtual acoustics, psychoacoustics and entertainment technology. One way of achieving this is the auralization of signals from a spherical microphone array (SMA) over headphones for a single listener. The perceptual properties of binaural rendering based on measured array impulse responses have been studied extensively [1, 2, 3].

Recently, implementations based on block-wise processing have been presented [4, 5, 6], which are able to render dynamic sound scenes containing sensor self-noise within captured or live-rendered SMA signals. Propagation properties of such noise through the rendering pipeline could not be systematically investigated so far. This is because previous (impulse-response-based) implementations exclusively assumed conditions that are violated by the time-variant nature of additive noise. Along these lines, theoretical considerations of the white-noise-gain (WNG) of SMAs have been presented before [7, 2], but drawing conclusions on perceptual properties of *additive noise* was difficult. Insight on the latter is particularly useful since real-world applications, *e.g.* recordings for immersive experiences, will inevitably contain sensor noise. This knowledge can be utilized in the design, specification and calibration process of SMAs.

ReTiSAR[1] [6] is a real-time implementation of the entire processing pipeline from the microphones of the array to the ears of the listener including tracking of head rotations around the vertical axis. The rendering pipeline can also emulate additive microphone self-noise that is incoherent between the sensors. Due to malfunction or electrical interference, microphones might exhibit increased noise levels in a non-uniform manner, whereof perceptual properties have not been investigated so far. We use ReTiSAR to systematically analyze the influence on real-time binaural rendering by performing an instrumental and perceptual evaluation with respect to propagated noise level, timbre, and spatial attributes.

[1] https://github.com/AppliedAcousticsChalmers/ReTiSAR

2. RENDERING METHOD

An incoming sound field that is captured by a SMA can be transformed into the spherical harmonics (SH) domain by means of plane wave decomposition [8], which results in what is also referred to as Ambisonics signals. A set of compensating radial filters is applied to account for the spatial extent and the scattering properties of the array, revealing varying properties for different radii and SH processing orders. Theoretically, these filters exhibit very large gains at low and at high frequencies, which we restrict through soft-clipping [2].

Ideally, the binaural rendering of SMA signals reproduces the corresponding ear signals as if the listener was placed and exposed to the sound field at the position of the array. In the following, we will refer to these signals as *wanted* or *target* signals, as opposed to the additive sensor noise. A generic or individual head model is characterized by a set of head-related transfer functions (HRTFs). ReTiSAR realizes the binaural rendering by combining all components in the SH domain taking into account the listener's head rotation, while pre-computing static components for best performance [6].

ReTiSAR can auralize anechoic sounds in the space of measured Array Room Impulse Response (ARIR) sets as described in [9], providing a very high signal-to-noise ratio (SNR). The resulting emulated microphone signals are free of additive noise so that sensor self-noise can be emulated by adding arbitrary synthetic noise to the *pre-rendered* microphone array signals. We use normally distributed white noise, which is turned pink by means of an IIR filter to resemble the typical spectral density of microphone self-noise. The noise amplitude can be adjusted in real-time via an OSC (Open Sound Control) interface.

3. DATA

We investigate a range of different rendering and microphone array parameters such as the number M of sensors in the array, maximum rendered SH order N, array radius r, gain limitation of the generated radial filters $\hat{a}$, and potentially employed HRTF equalization. The investigated configurations are based on two typical spherical microphone arrays employed in other publications *i.e.*, `mh acoustics Eigenmike` and `CR1_VSA_110RS_L` [9] with 32

and 110 channels respectively. Refer to Table 1 for the full list of parameter sets that were investigated.

We used the HRTFs of a `Neumann KU100` dummy head [10] in the auralization. The measured high-density grid allows for rendering of up to $N=35$. We investigate the basic processing pipeline without timbre or spatial enhancements such as [11, 12, 13, 14]. A manual direction-independent HRTF equalization similar to [3] is applied in some configurations (cf. column *EQ* in Table 1).

4. INSTRUMENTAL EVALUATION

This section provides insight into the propagation properties of wanted binaural signals as well as sensor self-noise independently through the rendering pipeline. The properties of coherence and uniformity of the noise have fundamental impact on the output.

All following instrumental measurements are performed at a block size of $l=4096$ samples of the real-time processing pipeline. Informal listening and instrumental evaluation showed that there was no difference to rendering with $l=1024$. We therefore assume that the processing block size does not have an influence on the self-noise propagation despite marginal fluctuations in the statistical properties of the generated noise. Wanted signals were rendered based on simulated impulse responses of an ideal plane wave ($N=40$) impinging from frontal direction on a rigid SMA under anechoic conditions.

All presented data represents the transmission path from the output of one SMA channel (pointing exactly into the direction of the rendered sound source) to the output of the binaural renderer *i.e.*, the ear signals. The additive noise makes the processing pipeline under consideration violate the assumption of linearity and time invariance, hence we cannot use transfer functions to represent the characteristics of the pipeline. We use an equivalent representation by the ratio of output over input spectra, independently for target and noise signals.

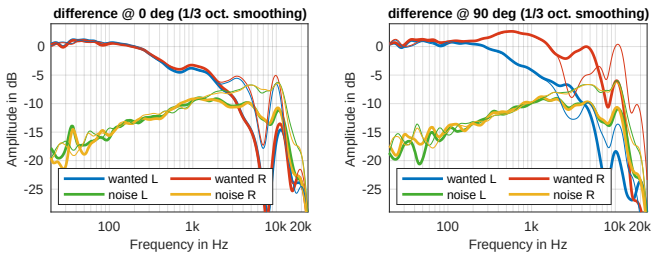


Fig. 1: Signal alteration from one sensor to referenced ear for stimuli $N=4$ *(IV)* in **thick *vs.*** $N=8$ *(III)* in **thin lines**

These were captured in real-time from the rendering pipeline over a duration of 2 s and smoothed in 1/3-octave bands. The source content for the wanted binaural signals in the instrumental evaluation was recorded white noise. AKtools [15] was used to generate all figures.

4.1. Spatially Uniformly Distributed Incoherent Noise

In this case, we create individual *incoherent* noise for each of the SMA sensors of identical root-mean-square (RMS) level. We use the term *spatially uniform* for convenience, but note that the spatial distribution of microphones is not necessarily exactly uniform.

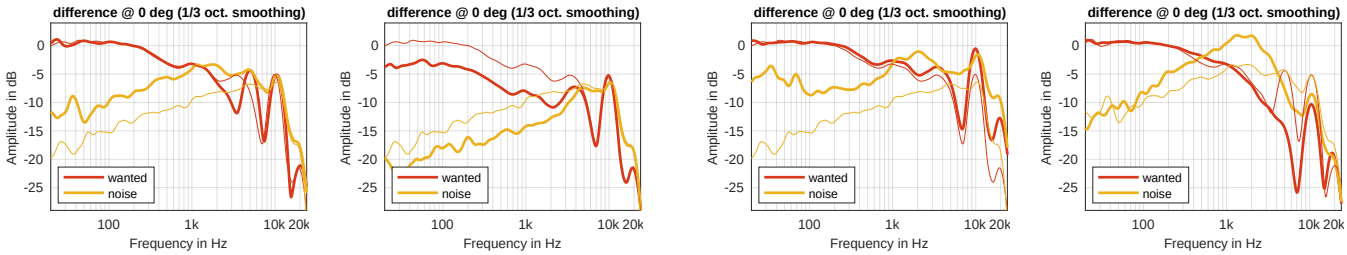
Informal listening shows that the overall perception is similar to dichotic playback of noise [16]. The noise is perceived diffuse and externalized around the listener without an apparent location and, most importantly, exhibits no dependency on head orientation [17, 18]. This effect is depicted in Fig. 1 where the WNG (green/yellow) is identical for both depicted head orientations (left *vs.* right sub-figures). The wanted binaural signals (blue/red) are congruent for frontal sound incidence, but change as expected according to head orientation. For the purpose of further WNG analysis, plots in Fig. 2 depict only results for frontal direction to the right ear.

It is also apparent from Fig. 1 that the SH order truncation attenuates the output signals at high frequencies (in the presented case above approx. 2 kHz). This effect is well known [1, 13]. The comparison of different orders (thick *vs.* thin) in Fig. 1 also shows an influence of the order on the self-noise propagation.

A comparable effect regarding high frequency components occurs when decreasing the number of sensors while preserving the overall array dimensions *i.e.*, its radius. The combination of spatial aliasing and SH order truncation almost covers the influence on the wanted signals in Fig. 2a, but is more pronounced at other head orientations. Remarkably, the overall WNG improves with a higher sensor count (Fig. 2a, thin). This observation is in accordance with the predictions from [7, Eq. 29] and [19], in the order of $\approx$6 dB below 1 kHz for the chosen parameter sets. The resulting SNR gain at the output occurs due to wanted signals being added coherently whereas noise signals being added incoherently in the pipeline.

A common practice when facing order truncation in the SH domain is the application of global equalization to counterbalance the unwanted attenuation of the signals at higher frequencies. This does not have an effect on the direction dependency of the error but on the overall timbre of the reproduction. As confirmed in Fig. 2b, the equalization affects target signals and self-noise uniformly.

The application of a gain limitation to the radial filters is performed for numerical stability and effectively reduces the SH order at the corresponding frequencies where the limitation is effective [2]. The effect of a higher limit *i.e.*, allowing higher filter gains, can be clearly seen in Fig. 2c with an otherwise identical array configuration.



(a) $M,N=32,4$ *(V) vs.* $110,8$ *(III)* (b) EQ *(I) vs.* No EQ *(III)* (c) $\hat{a}=18\,\mathrm{dB}$ *(II) vs.* $0\,\mathrm{dB}$ *(III)* (d) $r=4.2\,\mathrm{cm}$ *(VIII) vs.* $8.75\,\mathrm{cm}$ *(V)*

Fig. 2: Signal alteration from one sensor to either ear for various value *(stimulus)* in **thick *vs.*** value *(stimulus)* in **thin lines**

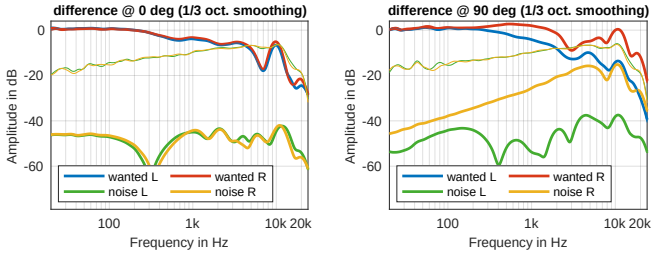


Fig. 3: Signal alteration from one sensor to referenced ear for *spatially non-uniform* noise in **thick *vs.*** *uniform* noise *(III)* in **thin lines**

stimulus	M	N	r in cm	$\hat{a}$ in dB	EQ
(I)	110	8	8.75	18	yes
(II)	110	8	8.75	18	no
(III)	110	8	8.75	0	no
(IV)	110	4	8.75	0	no
(V)	32	4	8.75	0	no
(VI)	32	4	4.2	18	no
(VII)	32	4	4.2	0	yes
(VIII)	32	4	4.2	0	no

Table 1: Investigated SMA configurations; $M=110$ in *Lebedev* and $M=32$ in *over-sampled equal area* grid; $\hat{a}$: radial filter gain limit

The wanted signal (red) is hardly affected whereas the self-noise (yellow) experiences a significant increase, varying over the entire frequency range.

Differences occur at high frequencies when altering the array radius while keeping all other parameters identical. The wanted signals (red) experience changes that are more pronounced for certain head orientations as seen for $M=32$ in Fig. 2d. The radial filters have a broadband effect on the self-noise (yellow). The filters for smaller SMAs are less forgiving in that they increase self-noise particularly around the aliasing frequency.

4.2. Spatially Non-Uniformly Distributed Incoherent Noise

Real-world SMAs will not comprise perfectly spatially uniform noise output at all channels. For simplicity, we exclusively consider the case of one sensor exhibiting a higher level of self-noise. In that case the perceived spatial properties and the timbre of the noise change with the head orientation. The otherwise diffuse noise is rendered with some superposed components of different timbre with their perceived direction changing strongly even during slight head movements. This effect on the position is particularly apparent for strong disparities of the sensor levels.

If inter-channel differences in noise level are small, the perceived additive noise components become less pronounced. Informal listening showed that they can only be heard around certain head orientations. This is particularly the case for head orientations where the microphone that exhibits the higher noise level is located along the listener's axis through both ears. Fig. 3 visualizes an extreme case of only one single microphone channel exhibiting additive noise. The strong influence of the listeners head rotation on the rendered signal is clearly visible.

5. PERCEPTUAL EVALUATION

We conducted an experiment to determine the thresholds at which the noise coloration due to head rotation becomes audible in the case of one microphone exhibiting a higher noise level than the others. Specifically, the microphone whose normal points towards frontal direction of the array was chosen. Its relative level increase ΔL was varied systematically to determine perceptual thresholds where artifacts become (in)audible. The investigated combinations of the SMA parameters are listed in Table 1.

5.1. Method

A 3AFC (three alternative forced choice) paradigm user study was performed to determine the location of the audibility threshold on a psychometric function [20]. An adaptive listening test with the staircase method of 2-down-1-up was used to converge at a beforehand unknown individual threshold with reasonable effort and experiment duration [21]. During the study head tracking was applied and only sensor-noise *i.e.*, no binaural target signal, was being rendered.

Sec. 4.1 shows that the noise levels in the ear signals differ significantly for the rendering configurations. The rendered noise outputs were aligned to a comfortable listening level by adjusting the RMS amplitudes as shown in Fig. 4.

A pilot test suggested that the thresholds will vary significantly for different array configurations. We therefore used distinct staircase parameters for the SMAs with different sensor counts. Arrays with $M=110$ start at $\Delta L=+30\,\mathrm{dB}$ and a step size of $\pm 6\,\mathrm{dB}$ while configurations of $M=32$ with $\Delta L=+18\,\mathrm{dB}\pm 5\,\mathrm{dB}$. According to the adaptive procedure, the individual step size is decreased by $1\,\mathrm{dB}$ after each successful step down. Both the step size and ΔL are limited to a minimum of $1\,\mathrm{dB}$. The adaptive staircase procedure for each condition is terminated after the subject reached a number of 7 *direction reversals*, as exemplarily depicted in Fig. 5. The resulting perceptual threshold is defined by the mean of ΔL values of the last 4 reversals.

The graphical user interface and adaptive staircase procedure are realized in MATLAB®. The order of test conditions is randomized for each subject. ReTiSAR is executed as a Python package by an operating system command call with the respective rendering configuration. After start-up, the individual output level ΔL of the microphone with the higher noise level is alternated via the OSC interface according to the subjects' responses. All conditions are rendered in real-time with a block length of $l=1024$ samples. Every 3AFC-trial had a random head azimuth offset of $\pm 45°$ relative to the array assigned. This ensures variation of the orientation in which the strongest artifacts occurred.

The ear signals are delivered via an `Antelope Orion 32` audio interface, a `Lake People G109` headphone amplifier and `AKG K702` headphones without further equalization. The instantaneous head orientation of the listener is provided via a `Polhemus Patriot` tracker. The playback chain has been set to a comfortable listening volume, which was measured to be $58.2\,\mathrm{dBSPL}$ with a `GRAS 45BB KEMAR` head and torso simulator.

The participants performed a short training to get familiar with the user interface and overall impression of the rendered signals. The training examples were based on stimulus *(III)* and demonstrated the artifact for $\Delta L=+30\,\mathrm{dB}$, $+14\,\mathrm{dB}$ and $+9\,\mathrm{dB}$.

A total of 11 expert listeners participated in the experiment. The subjects were instructed to use a conscious *listening strategy* with slow, sweeping head movements. This made sure that the subjects would adopt a head orientation where the artifact is most prominent. The duration of the experiment was approximately two hours including extended breaks.

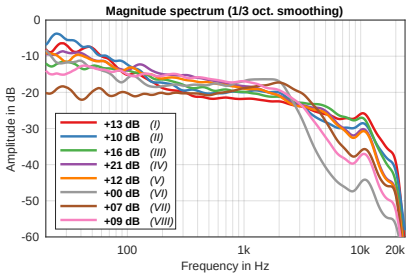


Fig. 4: Ear signals according to Table 1 including specified level adjustments for uniform noise before applying ΔL

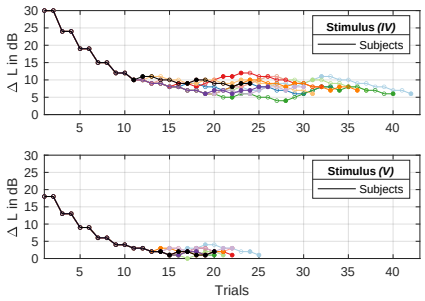


Fig. 5: Resulting 3AFC adaptive staircase for example stimuli according to Table 1

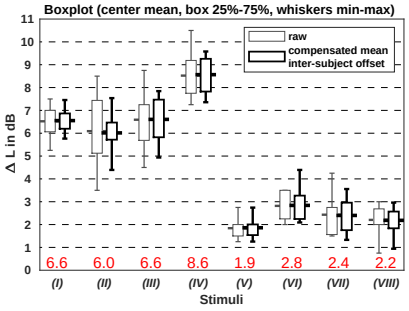


Fig. 6: Perceptual thresholds according to Table 1; compensated mean values as text

5.2. Results and Discussion

The results are shown in the box plot in Fig. 6 with and without compensation of individual thresholds according to inter-subject offsets. For the latter global differences in the average over all 8 conditions were counteracted from spanning 3.7 to 5.5 dB before and 4.6 dB after the compensation. This leads to an overall decrease in variation with negligible effect on the resulting mean thresholds.

A strong influence of the number of microphones on the audibility threshold can be seen when comparing stimuli *(I)* to *(IV)* and *(V)* to *(VIII)*. The SMA configurations with $M=110$ exhibit higher thresholds than configurations with $M=32$. This is intuitive as the contribution of a single microphone is less significant in case of a higher sensor count.

The observed difference is not moderated by the SH rendering order. On the contrary, reducing from $N=8$ in stimulus *(III)* to $N=4$ in *(IV)* yielded the overall highest threshold. *(IV)* and *(V)* differ only in the number of microphones. Using the lowest possible number of 32 microphones for the given 4th order resulted in the lowest threshold. The *relative* difference of $\approx$7 dB between *(IV)* and *(V)* seems particularly large but should be interpreted under consideration of the applied equalization of *noise playback volumes*. Directly comparable *absolute* audibility levels in real-life conditions, *i.e.*, levels being aligned for identical *target signal volumes* excluding the gains specified in Fig. 4, cannot be predicted from that.

The average difference between the configurations with different sensor counts, stimuli *(I)-(IV) vs. (V)-(VIII)*, is 4.6 dB, closely matching the power level of the sensor ratio $10\log_{10}{}^{110}/_{32}$. If this scaling, comparable to the increase in WNG due to a higher channel count [7, 19], holds true, will have to be investigated with a wider range of array configurations.

An influence of the array radius on the audibility threshold cannot be seen (cf. *(V)* with $r=8.75$cm and *(VIII)* with $r=4.2$cm). The same applies to different radial filter gain limits as can be concluded hen comparing *(II)* and *(III)* as well as *(VI)* and *(VIII)*. No influence of the applied gain limitation ($\hat{a}=0$dB *vs.* 18dB) is apparent. Recall however, that the radial filter gain limit has a considerable effect on the absolute noise level (cf. Fig. 2c and Fig. 4).

Global HRTF equalization does show an influence on the audibility thresholds, which become evident when comparing the thresholds of stimuli *(I)(II)* and *(VII)(VIII)* respectively. Even though equalization can significantly improve overall timbre of the binaural decoding, the fact that they are global does not mitigate the dependency of the noise on the head orientation. We do therefore not expect an influence from other direction-independent equalization approaches such as [13]. Other mitigation approaches for spatial aliasing [11, 12] and SH order truncation [14, 22] aim at better matching reference ITDs and ILDs of the rendered HRTF set. They yield an effect on the spatial impression of the rendered scene eventually so that conclusion regarding the audibility thresholds of spatially non-uniformly distributed noise cannot be drawn.

It should be highlighted that the gathered thresholds apply only to the considered, very critical rendering conditions. Lower absolute noise levels, the presence of a wanted binaural signal and room reverberation are likely to hinder detectability of the artifact. We also expect higher thresholds for non-expert listeners.

6. CONCLUSIONS

The presented instrumental evaluation of the propagation of microphone self-noise through a real-time binaural rendering pipeline illustrated important properties: 1) The noise level and spectral balance are independent of the head orientation if all microphones of the array exhibit incoherent additive noise with identical statistical properties (cf. Fig. 1) [17, 18]. 2) The SNR at the output increases with the number of microphones (cf. Fig. 2a) [7, 19].

The perceptual evaluation manifested audibility thresholds for head-orientation-dependent properties of the noise in the binaural signals for the case that one microphone in the array exhibits an increased noise level. As expected, a major influence was seen by the number of microphones in the SMA. The audibility thresholds under critical conditions were in the order of a few dB, which is higher than the tolerances of professional microphone capsules. This suggests that real-world microphone arrays are not likely to cause audible head-orientation-dependent artifacts. To consolidate this, artifacts due to small global variations in the sensor self-noise *i.e.*, more than one channel, as well as the influence of the spherical sampling grid [23] should be investigated.

If low-cost capsules like MEMS microphones are employed that exceed the required tolerances [24], or if a capsule in an array is faulty, a potential mitigation strategy might be reducing the quadrature weight of the according microphone. Channels in the vicinity might be able to compensate for a loss in wanted binaural signal while resulting in a lower (noise) contribution the impaired channel. It is subject to future work to analyse the benefits of an increased WNG *vs.* the impairment of the target signal.

7. ACKNOWLEDGMENT

We thank Facebook Reality Labs for financing this project. We thank Tim Lübeck for help with preparation of the user study and all subjects for the voluntary participation.

8. REFERENCES

[1] Amir Avni, Jens Ahrens, Matthias Geier, Sascha Spors, Hagen Wierstorf, and Boaz Rafaely, "Spatial perception of sound fields recorded by spherical microphone arrays with varying spatial resolution," *The Journal of the Acoustical Society of America*, vol. 133, no. 5, pp. 2711–2721, 2013.

[2] Benjamin Bernschütz, *Microphone Arrays and Sound Field Decomposition for Dynamic Binaural Recording*, Phd thesis, Technische Universität Berlin, 2016.

[3] Jens Ahrens and Carl Andersson, "Perceptual Evaluation of Headphone Auralization of Rooms Captured with Spherical Microphone Arrays with Respect to Spaciousness and Timbre," *The Journal of the Acoustical Society of America*, vol. 145, no. 4, pp. 2783–2794, 2019.

[4] Leo McCormack and Archontis Politis, "SPARTA & COMPASS: Real-time implementations of linear and parametric spatial audio reproduction and processing methods," in *Conference on Immersive and Interactive Audio*, York, 2019, pp. 1–12, Audio Engineering Society.

[5] Franz Zotter and Matthias Frank, *Ambisonics: A Practical 3D Audio Theory for Recording*, Springer Open, Cham, Switzerland, 2019.

[6] Hannes Helmholz, Carl Andersson, and Jens Ahrens, "Real-Time Implementation of Binaural Rendering of High-Order Spherical Microphone Array Signals," in *Fortschritte der Akustik – DAGA 2019*, Rostock, Germany, 2019, pp. 1–4, Deutsche Gesellschaft für Akustik.

[7] Boaz Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 1, pp. 135–143, 2005.

[8] Munhum Park and Boaz Rafaely, "Sound-field analysis by plane-wave decomposition using spherical microphone array," *The Journal of the Acoustical Society of America*, vol. 118, no. 5, pp. 3094–3103, 2005.

[9] Philipp Stade, Benjamin Bernschütz, and Maximilian Rühl, "A Spatial Audio Impulse Response Compilation Captured at the WDR Broadcast Studios," in *27th Tonmeistertagung – VDT International Convention*, Cologne, Germany, 2012, pp. 551–567, Verband Deutscher Tonmeister e.V.

[10] Benjamin Bernschütz, "A spherical far field HRIR/HRTF compilation of the Neumann KU 100," in *Fortschritte der Akustik – AIA/DAGA 2013*, Meran, Italy, 2013, pp. 592–595, Deutsche Gesellschaft für Akustik.

[11] Benjamin Bernschütz, "Bandwidth Extension for Microphone Arrays," in *AES Convention 133*, San Francisco, USA, 2012, pp. 1–10, Audio Engineering Society.

[12] David Alon and Boaz Rafaely, *Spatial Decomposition by Spherical Array Processing*, chapter 2, pp. 25–47, Wiley, 2017.

[13] Zamir Ben-Hur, Fabian Brinkmann, Jonathan Sheaffer, Stefan Weinzierl, and Boaz Rafaely, "Spectral equalization in binaural signals represented by order-truncated spherical harmonics," *The Journal of the Acoustical Society of America*, vol. 141, no. 6, pp. 4087–4096, 2017.

[14] Markus Zaunschirm, Christian Schörkhuber, and Robert Höldrich, "Binaural rendering of Ambisonic signals by head-related impulse response time alignment and a diffuseness constraint," *The Journal of the Acoustical Society of America*, vol. 143, no. 6, pp. 3616–3627, 2018.

[15] Fabian Brinkmann and Stefan Weinzierl, "AKtools - an open software toolbox for signal acquisition, processing, and inspection in acoustics," in *Audio Engineering Society Convention 142*, Berlin, Germany, 2017, pp. 1–6, Audio Engineering Society.

[16] Jens Blauert and Werner Lindemann, "Spatial mapping of intracranial auditory events for various degrees of interaural coherence," *The Journal of the Acoustical Society of America*, vol. 79, no. 3, pp. 806–813, 1986.

[17] Earl G Williams, *Fourier Acoustics*, Academic Press, London, 1 edition, 1999.

[18] Daniel P. Jarrett, Emanuël A. P. Habets, and Patrick A. Naylor, *Theory and Applications of Spherical Microphone Array Processing*, vol. 9, Springer International Publishing, 1 edition, 2017.

[19] Sébastien Moreau, Jérôme Daniel, and Stéphanie Bertet, "3D Sound Field Recording with Higher Order Ambisonics Objective Measurements and Validation of Spherical Microphone," in *AES Convention 120*, Paris, France, 2006, pp. 1–24, Audio Engineering Society.

[20] Harry Levitt, "Transformed UpDown Methods in Psychoacoustics," *The Journal of the Acoustical Society of America*, vol. 49, no. 2, pp. 467–477, 1971.

[21] Bernhard Treutwein, "Adaptive Psychophysical Procedures," *Vision research*, vol. 35, no. 17, pp. 2503–2522, 1995.

[22] Christoph Hold, Hannes Gamper, Ville Pulkki, Nikunj Raghuvanshi, and Ivan J Tashev, "Improving Binaural Ambisonics Decoding by Spherical Harmonics Domain Tapering and Coloration Compensation," in *International Conference on Acoustics, Speech and Signal Processing*, Brighton, UK, 2019, pp. 261–265, IEEE.

[23] Franz Zotter, "Sampling Strategies for Acoustic Holography/Holophony on the Sphere," in *International Conference on Acoustics (NAG/DAGA)*, Rotterdam, Netherlands, 2009, pp. 1–4, Deutsche Gesellschaft für Akustik.

[24] Raimundo González, Joshua Pearce, and Tapio Lokki, "Modular Design for Spherical Microphone Arrays," in *International Conference on Audio for Virtual and Augmented Reality*, Redmond, US, 2018, pp. 1–7, Audio Engineering Society.