Instrumental Evaluation of Sensor Self-Noise in Binaural Rendering of Spherical Microphone Array Signals

Hannes Helmholz, David Alon, Sebastia Gari, Jens Ahrens

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

Hannes Helmholz[1] **David Lou Alon**[2] **Sebastià V. Amengual Garí**[2] **Jens Ahrens**[1]

[1] Division of Applied Acoustics, Chalmers University of Technology, Sweden
[2] Facebook Reality Labs, 1 Hacker Way, Menlo Park, 94025, USA

hannes.helmholz@chalmers.se jens.ahrens@chalmers.se

ABSTRACT

Sensor self-noise is an important aspect in the design of spherical microphone arrays. We consider the application of binaural rendering of spherical microphone array signals in this paper. We use the *Real-Time Spherical Array Renderer* (ReTiSAR) to analyze the frequency-dependent white-noise-gain i.e., the improvement of the *signal-to-noise ratio* (SNR) between a selected microphone of the array and the binaural output signals. The configuration of the array as well as of the processing pipeline have been shown to strongly influence the spectral properties and the overall level of the noise in the binaural signals. We found that arrays with small radii reduce the SNR by approximately $8\,\mathrm{dB}$ unless a strong limitation of the radial filter gain is applied. Some array configurations with larger radii and number of microphones (above spherical harmonics order 4) are capable of increasing the SNR in the rendered signals by up to $9\,\mathrm{dB}$. We also found variations in the spectral properties of the rendered self-noise as a function of the head orientation for higher-order *Lebedev* grids.

1. INTRODUCTION

Spherical microphone arrays (SMAs) can employ large amounts of microphones. For example, the recently built open-source *HØSMA-7N* array comprises 64 sensors [1]. The microphone signals can be combined for various applications, such as a beamformer or the reproduction aiming at preserving the original spatial properties. However, each microphone produces and contributes a certain amount of sensor self-noise, raising the question as to what is the *signal-to-noise ratio* (SNR) in the processed output signals of the array. In the present case, we investigate the effect on binaural rendering i.e., the ear signals of a virtual listener in the sound field captured by the SMA. For the reproduced target sound field a spherical harmonics rendering order of around 8 or higher was shown to produce binaural signals that were hardly distinguishable from the actual ground truth ear signals [2, 3].

The *wanted target signals* in the array sensors i.e., any signal other than *unwanted additive noise*, are the result of waves impinging on the array surface. Thus, they are related through the wave equation adding up coherently between microphones, while the self-noise adds up incoherently. This leads to an improvement in SNR with an increasing number of microphones, as predicted in theory [4]. Assuming all microphones contributing an identical amount of noise, further theoretical assessments investigated the dependency of *white-noise-gain* (WNG) on both the array and the processing pipeline configuration [2, 4]. The rendering process of sound fields captured with a SMA involves so-called radial filters. These have been shown to exhibit large gains of several $10\,\mathrm{dB}$ in certain frequency ranges, boosting the noise relative to the wanted signal substantially. Consequently, different approaches have been suggested in order to restrict the gain of the employed radial filters [5–7]. We showed in [8] based on a real-time implementation that the number of microphones as well as the radial filters yield the strongest influence on the resulting level and spectral properties of the noise in the binaural signals.

The rendering of SMA data can be performed based on live-captured array signals or measured impulse response data sets. The latter are also termed *array room impulse responses* (ARIRs) and encode a static representation of a single sound source and the surrounding environment. Such data sets can be acquired from actual multi-channel SMAs or from sequential measurements with a single microphone (potentially mounted on a suitable scattering object), e.g. [9]. By means of sweep-based acquisition methods, the effect of noise in the measurement signals on the resulting IRs can be well controlled by adjusting the sweep duration and by performing averages over multiple acquisitions [10].

Unfortunately, such noise mitigation techniques are not available for the rendering of streamed data from live-captured or recorded microphone array signals. A continuous stream of SMA sensor data allows for the encoding of dynamic scenes, including an arbitrary number of moving sources, changes in the environmental conditions and position of the SMA itself. A high overall amplification of the signals from the SMA is required if the target source has a low level, inherently raising the absolute noise level resulting from the sensors. The presence of environmental and sensor self-noise is therefore a particular challenge for the rendering of streamed SMA signals.

In the present paper we employ the *Real-Time Spherical Array Renderer (ReTiSAR)* [1] [11, 12] to determine the SNR in the ear signals for practically relevant configurations.

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

In particular, we evaluate the SNR at a selected reference sensor of the SMA against the binaural output signals. Also, we will outline some spectral attributes of the rendered self-noise in dependency on the parameters of the rendering pipeline. The gained insight is helpful and important to customize array designs for specific applications and for identifying a favorable dynamic range.

2. RENDERING METHOD

An incoming sound field can be sampled by a spherical microphone array by capturing the signals arising at discrete sensor positions on the surface of the sphere. Such signals can be transformed into the *spherical harmonics* (SH) domain by means of plane wave decomposition [13]. This results in what is also referred to as *Ambisonics B-Format* signals or *Higher Order Ambisonics* (HOA) signals [14, 15]. In our case, the decomposition into SHs is performed via a discretization of the transformation integral [16] under consideration of the according *quadrature weights* for each SMA configuration.

To account for the spatial extent and the scattering of the physical array body, a set of compensating radial filters is introduced. These exhibit properties that strongly depend on the SMA radius and employed SH processing order. In theory, these filters are required to apply very large amplification gains at low and at high frequencies. Practically, this would result in numeric instabilities as well as an excessive amplification of microphone sensor self-noise. We employ a soft-clipping approach to restrict excessive gains with variable limitation levels [2, 5].

The goal of binaural rendering of SMA signals is to reproduce ear signals as if the listener was placed at the position of the array. To achieve this, a virtual head model is exposed to the sound field extracted from the SMA. Such head models are described by either generic or individual sets of *head-related transfer functions* (HRTFs). For this investigation, we employ a high resolution data set of a *Neumann KU100* dummy head [17]. The binaural rendering approach, as described in the following paragraph, takes the instantaneous head orientation of the listener into account. Thereby, the possible movement of the virtual head is restricted to the rotation around the vertical axis (one degree of freedom). Static elements of the computation i.e., those that do not depend on the input to the rendering pipeline, are pre-computed and optimized during startup to yield the best real-time performance [11].

A common method for the binaural rendering of Ambisonics is the encoding of the signals to a set of virtual loudspeakers. These are then convolved with the HRTF set and summed to produce the ear signals [18]. Instead of employing discrete virtual loudspeakers, *ReTiSAR* renders the binaural signals by convolving sound field with the HRTF components directly in the SH domain [16], without the intermediate explicit decoding to virtual loudspeakers.

The (binaural) rendering of SMA signals in spherical harmonics faces some inherent limitations. On the one hand, the number of sensors in the employed SMA restricts the maximum SH order that can be extracted. Since the captured sound field will usually not be order constrained, *spatial aliasing* arises in the SH decomposition. To that effect, spatial ambiguities lead to an unnatural increase of the signal energy above the temporal spatial aliasing frequency [4]. On the other hand, the extracted order-limited sound field coefficients will limit the spatial resolution of the employed virtual head model as higher-order components of the HRTFs are not triggered. This SH *order truncation* leads to an overall loss in spatial resolution and causes an attenuation of high frequency content, which shows some amount of dependency on the head orientation [19]. Several approaches have been proposed to mitigate such *spatial undersampling errors*, which were evaluated instrumentally and perceptually [20].

3. RENDERING CONFIGURATIONS

We analyse a representative set of configurations in terms of array geometry and the parameters of the rendering pipeline. This allows for finding any potential trends in the results when scaling those features. The individual parameters are investigated in an isolated manner and were selected in accordance to a former investigation [8]. We chose a maximum permitted amplification of the radial filters of $\hat{a} = 0\,\mathrm{dB}$ and $18\,\mathrm{dB}$, respectively. The chosen configurations are presented in Tab. 1 and will be referred to by their respective abbreviations in the following.

The rendering scenario that we consider is an ideal plane wave ($N = 40$) impinging from the frontal direction $(0°, 0°)$ on a SMA under anechoic conditions. The microphones are assumed to be located on the surface of a rigid spherical body in all configurations. The array response is represented by simulated microphone impulse responses and are therefore virtually free of noise or measurement errors.

Various strategies exist for generating a grid of sensor positions on the surface of a SMA [21]. We cover three fundamentally different grid types as indicated in Tab. 1:

Abbrv.	**Grid type**	M	N	r in cm	$\hat{a}$ in dB
EM32	*Eigenmike*	32	4	4.2	18
					0
LE38	*Lebedev*	38	4	4.2	18
					0
GL50	*Gauss-Legendre*	50	4	4.2	18
					0
GL50L	*Gauss-Legendre*	50	4	8.75	18
					0
LE110	*Lebedev*	110	8	8.75	18
					0
GL162	*Gauss-Legendre*	162	8	8.75	18
					0
LE230	*Lebedev*	230	12	8.75	18
					0
GL338	*Gauss-Legendre*	338	12	8.75	18
					0

Table 1: Investigated rendering configurations.

The same sensor locations of the *Eigenmike* [22], which are based on an Archimedean solid i.e., the truncated icosahedron, as well as *Lebedev* [23] and *Gauss-Legendre* [24] sampling schemes for different orders. As indicated in Tab. 1, the number of required sensors M increases with the desired rendering order N.

The SMA configurations exhibit different radii to yield a representative density according to the number of sensors on the surface. All arrays that support encoding at $N = 4$ exhibit a radius of $r = 4.2\,\mathrm{cm}$ in order to be directly comparable to the *Eigenmike* SMA. The arrays evaluated at higher SH orders, complemented by the `GL50L` configuration, exhibit a radius of $r = 8.75\,\mathrm{cm}$, which has been shown to be a convenient size for the rendering target binaural signals [2]. This choice was also inspired by the configurations in the *WDR Cologne* SMA room measurement data set [9].

All contained instrumental evaluations are based on a processing block size of 4096 samples. Despite marginal fluctuations in the statistical properties of the generated noise, previous investigations did not show any relevant influence of the block length [8, 11]. We evaluated all data that are presented in this paper also for 1024 samples and found entirely congruent results.

4. ACQUISITION METHOD

We are primarily interested in the practical implications of the microphone self-noise in terms of the *signal-to-noise ratio* (SNR) at the output of the binaural rendering pipeline. A *root-mean-square* (RMS) level can be calculated from discrete time domain input and output signals of length L:

$$\mathrm{RMS}_{\mathrm{dB}}\{x\} = 20\,\log_{10}\sqrt{\frac{1}{L}\sum_{i=1}^{L}|x(i)|^2} \qquad (1)$$

We evaluate the RMS level independently for the target (*wanted*) and the noise (*unwanted*) signals components to allow for the computation of the respective SNR in dB:

$$\mathrm{SNR}_{\mathrm{in,dB}} = \mathrm{RMS}_{\mathrm{in,wanted,dB}} - \mathrm{RMS}_{\mathrm{in,noise,dB}}$$
$$\mathrm{SNR}_{\mathrm{out,dB}} = \mathrm{RMS}_{\mathrm{out,wanted,dB}} - \mathrm{RMS}_{\mathrm{out,noise,dB}}$$

Comparing the measure of the input and the output expresses the overall SNR influence of the system under consideration:

$$\Delta L_{\mathrm{SNR}} = \mathrm{SNR}_{\mathrm{out,dB}} - \mathrm{SNR}_{\mathrm{in,dB}}$$

This evaluation is performed for all configurations listed in Tab. 1. Each channel of the SMA contributes to the binaural output signals. We pick the microphone pointing exactly into the direction of the simulated sound source as the reference for the SNR at the input. Accordingly, this channel yields the best possible input SNR of the array. The output signals of the system are the binaural signals, which vary strongly with the instantaneous head orientation. For that reason, we capture the output at representative head orientations in the horizontal plane i.e., $0°$, $+45°$ and $+90°$ azimuth. The positions are expressed in relation to a right handed spherical coordinate system of the array, with the source impinging from $0°$ azimuth in the horizontal plane.

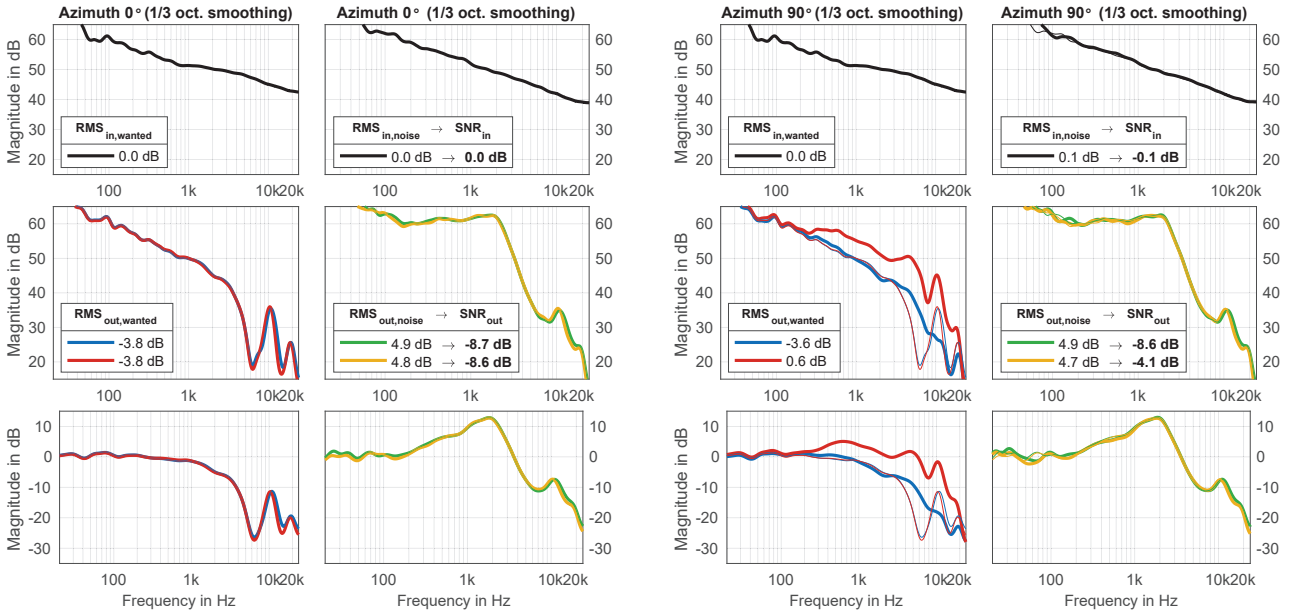
The input and output signals are captured simultaneously in real-time. ReTiSAR allows to render the wanted sound field and self-noise components in succession so that they can be measured independently. We use pink noise for both the emulated self-noise (cf. Appendix) and the auralized target signal (cf. typical long-term average magnitude spectra [25, 26]). The noise coloration is achieved by means of an IIR filter applied to continuous Gaussian noise that is generated live [12]. In order to reduce measurement uncertainties due to potential variations in the statistical noise properties, the signals are recorded and analysed for a duration of $2\,\mathrm{s}$. For ease of illustration, all resulting depictions of frequency domain data are smoothed in 1/3-octave bands and generated in *Matlab* with the *AKtools* toolbox [27].

5. RESULTS

Exemplary for all configurations contained in Tab. 1, we discuss the *Eigenmike* configuration with $\hat{a} = 18\,\mathrm{dB}$ in detail. Fig. 1a depicts the input and output signals in frequency domain for the target sound field (left column) and the emulated self-noise (right column). The top row comprises the respective signals arising at the chosen reference sensor of the SMA i.e., at the input stage of the processing pipeline. Contrary to the noise component, the signal originating from the captured target sound field is not entirely pink anymore, which is expected due to the scattering off the physical SMA body. For convenience, both components have been normalized to exhibit an RMS level and thus also an input SNR of $0\,\mathrm{dB}$.

The middle row of Fig. 1a comprises the binaural signals produced by the `EM32 (18 dB)` processing pipeline. The wanted signals for left (blue) and right (red) ear are identical for a head orientation of $0°$ and typical characteristics of the employed HRTFs are apparent. We also observe that the noise magnitude (green and yellow) is identical for both ear signals (but with different phase). The bottom row in Fig. 1 depicts the smoothed output over input spectra i.e., the difference between top and middle plots for each column. This methodology to generate the difference plot is identical to [8] and is also used to illustrate the configurations in Fig. 2. For the `EM32 (18 dB)` configurations at $0°$ head orientation, the resulting difference in SNR due to the binaural rendering is $-8.6\,\mathrm{dB}$ i.e., the SNR in the binaural signals is $8.6\,\mathrm{dB}$ lower than at the reference microphone.

Fig. 1b contains the results for the same `EM32 (18 dB)` configuration with a $90°$ head orientation. The signals on the input stage are identical (besides some negligible fluctuations due to the real-time noise emulation process). This is intended, as we capture the identical reference SMA sensor while the instantaneous head orientation is introduced after the input signal decomposition. The rendered target signals clearly exhibit the expected characteristic of a lateral sound source with the signal at the contralateral i.e., left, ear being strongly attenuated at mid to high frequencies. However, the rendered self-noise on the output is similar to the frontal head orientation. This confirms the observation that



(a) Head orientation 0° (frontal source)

(b) Head orientation 90° (lateral source)

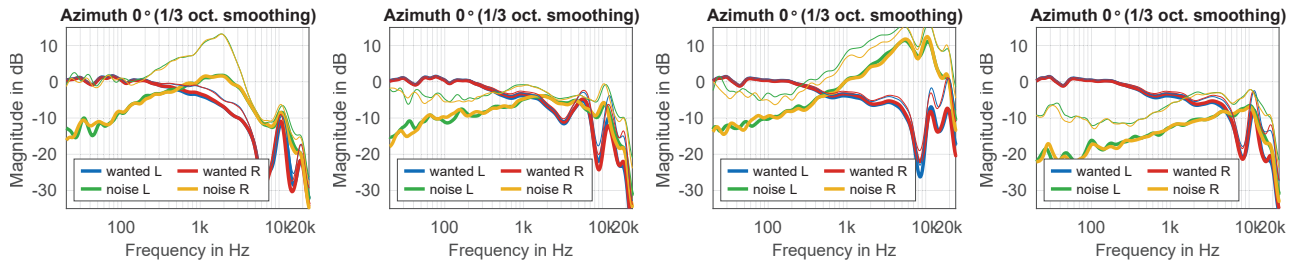
Figure 1: Spectra of *wanted* (left) and *noise* (right) signal components of `EM32 (18 dB)` configuration. Sensor input (top), ear output (middle) and resulting system influence (bottom). Blue and green curves refer to the left ear, red and yellow curves refer to the right ear.

auralization of the uniformly contributing noise is largely independent of the head orientation [8]. Although the noise RMS level is unchanged, the overall system SNR varies with the rendered target signals. This configuration yields $\Delta L_{\text{SNR}} = \{-8.6\,\text{dB}, -4.1\,\text{dB}\}$ for the left and right ear respectively. Compared to the frontal source position, this is an identical result for the contralateral and a considerable improvement for the ipsilateral ear.

The `EM32 (0 dB)` configuration incorporates a stronger constraint on the radial filters, which results in an average $\Delta L_{\text{SNR}} = -0.1\,\text{dB}$ for the frontal source position. Thereby, the measured SNR in the binaural signals is improved by $+8.6\,\text{dB}$ compared to `EM32 (18 dB)`, but is substantially smaller than the employed limitation delta of $18\,\text{dB}$. This is due to the frequency dependency of the rendered self-noise as shown in Fig. 2a. The restriction of the radial filters yields approximately $10\,\text{dB}$ of noise attenuation below but no improvement above the configuration-specific spatial aliasing frequency (around $5\,\text{kHz}$ according to [4]).

Fig. 3 summarizes the results for all rendering conditions from Tab. 1 and three head orientations. We observe an improvement in SNR for all configurations when comparing $\hat{a} = 0\,\text{dB}$ vs. $18\,\text{dB}$. As expected, a lower maximum gain of the radial filters leads to a higher SNR. In addition, all configurations at $r = 4.2\,\text{cm}$ exhibit an increase in SNR of $+8.6\,\text{dB}$ on average when reducing the maximum gain of the radial filters. This improvement is consistent for all investigated small SMAs and head orientations. The configurations with $r = 8.75\,\text{cm}$ show improvements of $+3.7\,\text{dB}$ to $+5.6\,\text{dB}$ due to a limitation from $\hat{a} = 18\,\text{dB}$ to $0\,\text{dB}$. The gain according to the restriction is therefore smaller for the larger SMAs, as indicated also by Fig. 2b and 2d.

Furthermore, Fig. 3 confirms a general improvement in SNR for higher sensor count SMA configurations, as theoretically predicted [4, 28] and practically shown [8]. This is apparent from the comparison of `GL50L`, `GL162` and `GL338` with improvements of $\Delta L_{\text{SNR}} = \{+0.4\,\text{dB}, +1.1\,\text{dB}, +3.6\,\text{dB}\}$ (averaged for



(a) `EM32 (0 dB, 18 dB)` (b) `GL50L (0 dB, 18 dB)` (c) `LE230 (0 dB, 18 dB)` (d) `GL338 (0 dB, 18 dB)`

Figure 2: Resulting influence on *wanted* and *noise* components for configurations according to Tab. 1 at 0° relative head orientation. Comparison of radial filter limitation of $0\,\text{dB}$ (thick) and $18\,\text{dB}$ (thin).

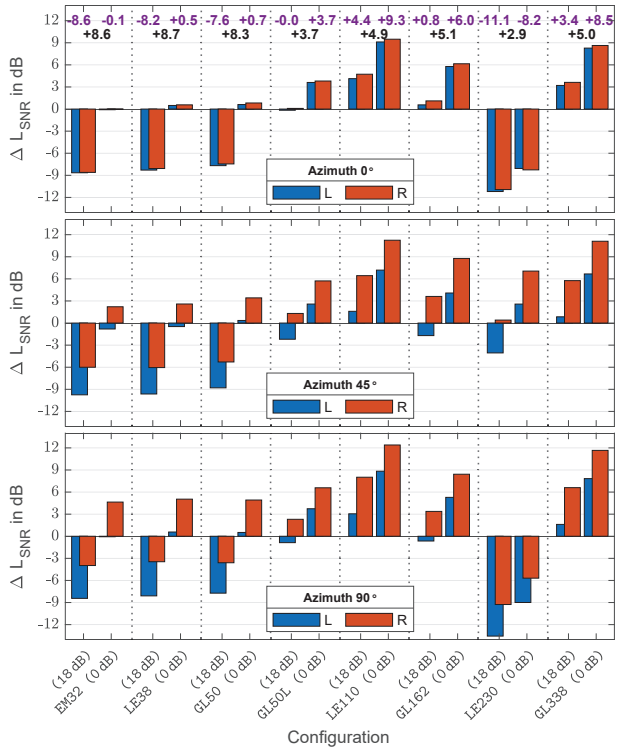


Figure 3: Resulting SNR changes for configurations according to Tab. 1 at different relative head orientations. Mean values (purple) and every other difference (black) as text for the 0° head orientation (top).

$\hat{a} = 18\,\mathrm{dB}$ over all three head orientations).

Remarkably, `LE110` yields the overall best SNR of all configurations with $\Delta L_{\mathrm{SNR}} = +9.3\,\mathrm{dB}$ for $\hat{a} = 0\,\mathrm{dB}$ (cf. Fig. 3). Moreover, the associated higher-order `LE230` configuration exhibits the overall worst noise influence with $\Delta L_{\mathrm{SNR}} = -11.1\,\mathrm{dB}$. As indicated by Fig. 3, the three investigated head orientations of that specific *Lebedev* configuration yield very different SNRs. The 45° head orientation seems to behave comparably well with a ΔL_{SNR} close to the higher-order *Gauss-Legendre* grids. The frontal and lateral source positions however exhibit vastly increased self-noise levels (cf. Fig. 2c), resulting in an impaired SNR in the ear signals.

6. DISCUSSION

Although a multitude of sensors simultaneously contributes self-noise in case of the array processing, we observe that the SNR in the ear signals for larger SMA configurations is actually improved compared to the reference sensor (cf. Fig. 3). Note that the SNR at the reference sensor is very similar to the SNR that the signal would exhibit when the given scenario is captured with a single microphone. However, a sophisticated SMA reproduction (either loudspeaker or headphone based) will render both self-noise and target signals spatially. In contrast to a single-channel superposition of both components, the binaural SMA reproduction yields an entirely diffuse and externalized (out of head) rendition of the additive noise [8]. Accordingly,

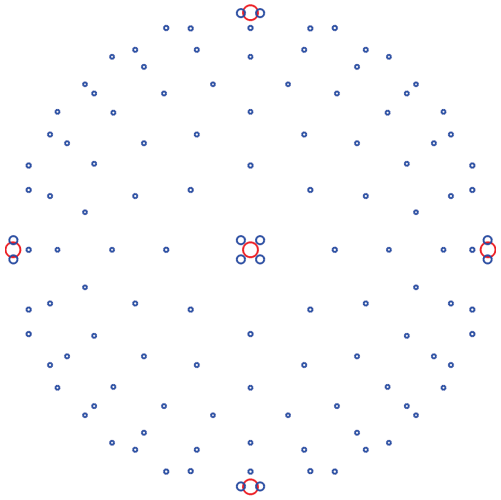


Figure 4: Sensor locations of the `LE230` configuration with positive (blue) and negative (red) *quadrature weights* encoded as circle radius. Grid symmetry and point clustering around poles are well visible.

psychoacoustic abilities of the auditory system like spatial unmasking will be triggered in the listener [29]. Therefore, the *perceived impairment* due to the presence of additive noise at identical RMS levels is suspected to be different.

The smaller SMA configurations ($N = 4$, $r = 4.2\,\mathrm{cm}$) yield considerably less SNR in the rendered binaural signals (cf. Fig. 3) compared to the larger configurations. However, a strong restriction of the radial filters amplification gains ($\hat{a} = 0\,\mathrm{dB}$) achieves comparable SNRs to the less restricted larger configurations ($\hat{a} = 18\,\mathrm{dB}$, $N \geq 8$, $r = 8.75\,\mathrm{cm}$). Comparing the results for `GL50`, `GL50L` and `GL162` reveals that the observed improvement in SNR arises mostly due to the increase in SMA radius and not the number of sensors with the according SH rendering order. From the perspective of rendered self-noise, the larger `GL50L` configuration is clearly preferred over the `GL50` configuration. The smaller configuration however will evoke less spatial aliasing in the rendered target sound field, due to the denser positioning of sensors on the surface of the SMA. This may be favorable for the target signal.

We performed the analysis from Fig. 3 also with an A-weighting of all involved signals in order to avoid a potentially misleading influence of the signal energy at very low and very high frequencies. The observed tendencies remained the same while all resulting SNR changes due to the binaural rendering turned out to be a few dB lower overall compared to the weighted case.

Although all SMA sensors are contributing noise in a uniform manner, the ear signals exhibit strong variations over different head orientations for the `LE230` configuration (cf. Fig. 3). We can confirm that the rendered target signals do not exhibit any observable artifacts. Fig. 2c shows strongly increased noise levels compared to other configurations and also e.g. the 45° head orientation of the same condition. Informal listening to this `LE230` configuration with head-tracking immediately reveals clearly noticeable fluctuations of the rendered self-noise when rotating the

head. We found such artifacts i.e., very pronounced changes in noise level and coloration, also occur for other *Lebedev* configurations at higher SH orders.

Remarkably, the *Lebedev* grid at higher orders can yield distributions with negative *quadrature weights*. This also happens for other spherical sampling methods not included here, e.g. *Fliege-Maier* [30]. The sensor placement of the `LE230` configuration is depicted in Fig. 4, with a distinct clustering of points at orthogonal pole positions. The negative quadrature weights occur at the center of the clusters, and the magnitude of the weights in the clusters can differ significantly from other positions on the grid. Note that this specific grid would not be very suitable for a real-world spherical array anyhow, due to the immediate proximity of sensors in the clusters. The combination of sensor distribution as well as sign and magnitude of the quadrature weights seems to yield a different interaction for the target sound field on the one hand and incoherent noise components on the other hand. To our knowledge, this behaviour has not been documented so far.

7. CONCLUSIONS

The spectral properties of the processed binaural target sound field is influenced by the SMA and rendering configuration [20]. As reported previously, also the overall level and the coloration of binaurally rendered sensor self-noise is strongly dependent on the configuration [8]. In this contribution, we established how the combination of these factors translates into a signal-to-noise ratio in the output signals of a binaural rendering pipeline (cf. Fig. 3).

The input SNRs of this investigation have been adjusted to $0\,\mathrm{dB}$. Accordingly, the observed SNR changes (cf. Fig. 3) can be used as an addend to determine the resulting output SNR in dependence of any reference input SNR. This can be helpful to assess requirements for signal headroom in practical deployments according to an actual input SNR at the reference microphone i.e., the SMA sensor facing into the main direction of sound incidence.

A limitation of the radial filter gains can exert a major improvement of the resulting SNR in the ear signals. The employed soft-limiting is only one of the approaches exactly designed for that purpose. Besides, increasing the number of sensors (and the SH rendering order accordingly) yields gradual improvements in SNR. Also an increase of the SMA radius alone showed to be very favourable for the rendered noise levels at medium frequencies in particular (cf. Fig. 2a and 2b).

High-order *Lebedev* grids turned out to exhibit a very unfavourable noise transmission behaviour, which includes overall high magnitudes and a strong dependency on head orientation. To get a better understanding of the behaviour, a larger variety of rendering orders as well as other spherical sampling grids should be considered in the future. We will provide a more detailed evaluation of coloration changes over the entire range of horizontal head rotation based on the *Composite Loudness Level* [31] in future work.

Our investigation assumed identical noise contributions by all channels of different spherical microphone arrays. We analysed the self-noise of a real-world example, the *Eigenmike* SMA, under anechoic conditions and found that both the noise coloration as well as amplitudes of the individual sensors vastly depend on the employed pre-amplification gain (cf. Appendix). Therefore, further analysis is required to relate the conclusions from the present paper to real-world scenarios considering practical target, environmental and equivalent input noise levels.

The software tools and data evaluation scripts that we employed in this study as well as detailed results for all investigated configurations (also Appendix) are available [2] .

8. ACKNOWLEDGMENT

We thank Facebook Reality Labs for financing this project. We thank Gary W. Elko of *mh acoustics LLC* for assisting the interpretation of the measurement data in the Appendix.

A. APPENDIX

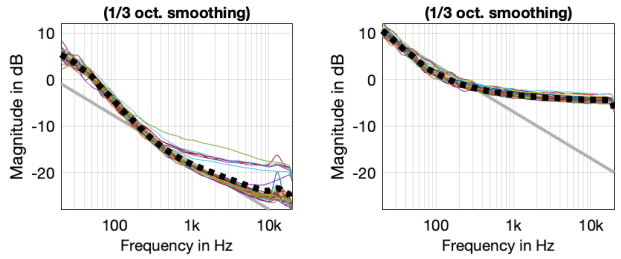
We measured the self-noise level of the microphones in our *mh acoustics Eigenmike 32* (unit serial number 28) SMA to be able to predict the SNR of binaurally rendered recordings obtained from the SMA. It is important to note that the manufacturer provides a set of frequency-independent calibration weights to compensate for mismatch in the sensitivity of individual channels. The calibration weights are unique for each array unit. They have to be applied to the signals before any subsequent processing. We exclusively consider the microphone signals with the weights applied in the following unless stated as different.

We performed $10\,\mathrm{s}$-long calibrated measurements of all SMA channels in the anechoic chamber at Chalmers University of Technology. The captured signals were filtered with an 8th-order Butterworth bandpass from $15\,\mathrm{Hz}$ to $23\,\mathrm{kHz}$. The RMS *sound pressure level* (SPL, re $20\,\mu\mathrm{Pa}$) and *power spectral density* (PSD) were calculated according to Eqn. (1) and [32], respectively.

Equivalent noise levels: The *equivalent input noise* (EIN) at the highest pre-amplification gain was approximately $23\,\mathrm{dB_{SPL}}$ (cf. Tab. 2). This is a superposition of noise components from the measurement chain as well as the background noise in the room at $17\,\mathrm{dB_A}$, as measured with a *Brühl & Kjær 2260 Investigator*. Note that this is also at the lower limit of what can be measured with this device. The EIN increases considerably for lower input gains up to approximately $39\,\mathrm{dB_{SPL}}$ or $36\,\mathrm{dB_{SPL}}$ (cf. Tab. 2). This is due to the relative increase in noise contributions from the *analog-to-digital converter* (ADC), in this case gain-independent at around $-95\,\mathrm{dB_{FS}}$ (RMS and when A-weighted). The array manufacturer confirmed the figures.

Channel Mismatch: For the lowest pre-amplification gain, the differences in RMS noise level between channels are minuscule before application of the calibration weights, as evident from Fig. 6 (left, yellow). The weights may therefore be considered to directly represent the inter-channel level mismatch at the input of a binaural renderer. The

[2] `http://doi.org/10.5281/zenodo.3711626`



(a) $+30\,\mathrm{dB}$: pronounced inter-channel differences dominated by internal electric and microphone self-noise.

(b) $-10\,\mathrm{dB}$: minor inter-channel differences dominated by equalization weights over uniform ADC noise.

Figure 5: PSD of individual (colored) and averaged (black) equivalent input noise of *Eigenmike (weighted)* at different pre-amplification gains. Pink noise (grey) for comparison.

statistical distribution of the weighted input noise for our array unit is depicted in Fig. 6 (right).

Noise Color: Fig. 5 depicts the PSD of the EIN for all individual SMA channels. For the highest available input gain, the PSD decreases significantly for increasing frequencies (cf. Fig. 5a). Our assumption is that the noise color is determined by the apparent acoustic background noise in the anechoic chamber due to the roll off towards high frequencies. For the lowest possible input gain, the PSD flattens out towards higher frequencies (cf. Fig. 5b). This supports our interpretation that the EIN towards lower gains is dominated by noise from the ADC input stage, which is usually white [33].

Conclusion: The observed self-noise levels, color and inter-channel differences of the investigated SMA measurement chain vastly depend on the employed pre-amplification gain. All observed EIN contributions (cf. Tab. 2) will be superimposed with acoustical noise that can potentially exhibit a higher SPL ($27\,\mathrm{dB_A}$ averaged from [34]). Whether acoustical noise or self-noise dominates in a practical application depends therefore on the scenario. In case that self-noise dominates, then the calibration weights determine the inter-channel differences. Recall from Fig. 3 how the input SNR is altered by the application of binaural rendering.

Gain	**Signal Level**		**EIN Level**	
in dB	$\mathrm{dB_{FS}}^{*}$	$\mathrm{dB_{SPL}}^{**}$	$\mathrm{dB_{SPL}}$	$\mathrm{dB_A}$
+30	−8.2	102.2	**22.9**	**17.3**
+20	−15.1	109.1	23.7	18.1
+10	−23.6	117.6	25.7	20.5
0	−33.2	127.2	30.6	25.7
−10	−43.0	137.0	**38.9**	**35.6**

*@ $94\,\mathrm{dB_{SPL}}$ **@ $0\,\mathrm{dB_{FS}}$

Table 2: Signal level from/at the sensor facing the source (2nd and 3rd column) and equivalent input noise level at the input of the renderer (average of all sensors, 4th and 5th column) of *Eigenmike (weighted)* at different pre-amplification gains (1st column). The signal level increases with the gain in steps of $\{9.8, 9.6, 8.5, 6.9\}\,\mathrm{dB}$ while the SPL of the EIN decreases in steps of $\{8.3, 4.9, 2.0, 0.8\}\,\mathrm{dB}$.

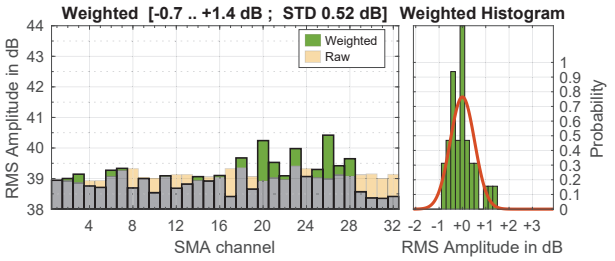


Figure 6: $-10\,\mathrm{dB}$: inter-channel differences (left) and histogram (right, green) with probability density function of approximated normal distribution (right, red) of *Eigenmike (weighted)* equivalent input noise.

9. REFERENCES

[1] O. Moschner, D. Dziwis, T. Lübeck, and C. Pörschmann, "Development of an Open Source Customizable High Order Rigid Sphere Microphone Array," in *AES Convention 148*, (Vienna, Austria), pp. 1–5, AES, 2020.

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

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

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

[5] B. Bernschütz, C. Pörschmann, S. Spors, and S. Weinzierl, "Soft-Limiting der modalen Amplitudenverstärkung bei sphärischen Mikrofonarrays im Plane Wave Decomposition Verfahren," in *Fortschritte der Akustik – DAGA 2011*, (Düsseldorf, Germany), pp. 661–662, DEGA, 2011.

[6] F. Zotter, "A Linear-Phase Filter-Bank Approach to Process Rigid Spherical Microphone Array Recordings," in *IcETRAN*, (Palić, Serbia), pp. 1–8, IEEE, 2018.

[7] C. Langrenne, E. Bavu, and A. Garcia, "A linear phase IIR filterbank for the radial filters of ambisonic recordings," in *Spatial Audio Signal Processing Symposium*, (Paris, France), pp. 127–132, EAA, 2019.

[8] H. Helmholz, J. Ahrens, D. L. Alon, S. V. Amengual Garí, and R. Mehra, "Evaluation of Sensor Self-Noise in Binaural Rendering of Spherical Microphone Array Signals," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (Barcelona, Spain), pp. 161–165, IEEE, 2020.

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

[10] S. Müller and P. Massarani, "Transfer-Function Measurement with Sweeps," *Journal of the Audio Engineering Society*, vol. 49, no. 6, pp. 443–471, 2001.

[11] H. Helmholz, C. Andersson, and J. Ahrens, "Real-Time Implementation of Binaural Rendering of High-Order Spherical Microphone Array Signals," in *Fortschritte der Akustik – DAGA 2019*, (Rostock, Germany), pp. 1462–1465, DEGA, 2019.

[12] H. Helmholz, T. Lübeck, J. Ahrens, S. V. Amengual Garí, D. L. Alon, and R. Mehra, "Updates on the Real-Time Spherical Array Renderer (ReTiSAR)," in *Fortschritte der Akustik – DAGA 2020*, (Hannover, Germany), pp. 1–4, DEGA, 2020.

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

[14] M. A. Gerzon, "Periphony: With-Height Sound Reproduction," *Journal of the Audio Engineering Society*, vol. 21, no. 1, pp. 2–10, 1973.

[15] D. Malham, "Higher order ambisonic systems for the spatialisation of sound," in *International Computer Music Conference Proceedings*, (Beijing, China), pp. 484–487, Michigan Publishing, 1999.

[16] B. Rafaely and A. Avni, "Interaural Cross Correlation in a Sound Field Represented by Spherical Harmonics," *Journal of the Acoustical Society of America*, vol. 127, no. 2, pp. 823–828, 2010.

[17] B. Bernschütz, "A spherical far field HRIR/HRTF compilation of the Neumann KU 100," in *Fortschritte der Akustik – AIA/DAGA 2013*, (Meran, Italy), pp. 592–595, DEGA, 2013.

[18] J.-M. Jot, S. Wardle, and V. Larcher, "Approaches to Binaural Synthesis," in *AES Convention 105*, vol. 861, (San Francisco, USA), pp. 4861–4875, 1998.

[19] Z. Ben-Hur, D. L. Alon, B. Rafaely, and R. Mehra, "Loudness Stability of Binaural Sound with Spherical Harmonic Representation of Sparse Head-Related Transfer Functions," *Eurasip Journal on Audio, Speech, and Music Processing*, vol. 2019, no. 5, pp. 1–14, 2019.

[20] T. Lübeck, H. Helmholz, J. M. Arend, C. Pörschmann, and J. Ahrens, "Perceptual Evaluation of Mitigation Approaches of Impairments due to Spatial Undersampling in Binaural Rendering of Spherical Microphone Array Data," *Journal of the Audio Engineering Society*, vol. 68, no. 6, pp. 428–440, 2020.

[21] F. Zotter, "Sampling Strategies for Acoustic Holography/Holophony on the Sphere," in *International Conference on Acoustics (NAG/DAGA)*, (Rotterdam, Netherlands), pp. 1107–1110, DEGA, 2009.

[22] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 2, (Orlando, USA), pp. 1781–1784, IEEE, 2002.

[23] V. I. Lebedev, "Spherical quadrature formulas exact to orders 25-29," *Siberian Mathematical Journal*, vol. 18, no. 1, pp. 99–107, 1977.

[24] I. Bogaert, "Iteration-free computation of gauss-legendre quadrature nodes and weights," *SIAM Journal on Scient. Comp.*, vol. 36, no. 3, pp. 1008–1026, 2014.

[25] L. E. Cornelisse, J.-P. Gagné, and R. C. Seewald, "Long-term Average Speech Spectrum at the Chest-level Microphone location," *Journal of Speech-Language Pathology and Audiology*, vol. 15, no. 3, pp. 7–12, 1991.

[26] A. Elowsson and A. Friberg, "Long-Term Average Spectrum in Popular Music and its Relation to the Level of the Percussion," in *AES Convention 142*, (Berlin, Germany), pp. 1–12, AES, 2017.

[27] F. Brinkmann and S. Weinzierl, "AKtools - an open software toolbox for signal acquisition, processing, and inspection in acoustics," in *AES Convention 142*, (Berlin, Germany), pp. 1–6, AES, 2017.

[28] S. Moreau, J. Daniel, and S. Bertet, "3D Sound Field Recording with Higher Order Ambisonics – Objective Measurements and Validation of Spherical Microphone," in *AES Convention 120*, (Paris, France), pp. 1–24, AES, 2006.

[29] A. Kohlrausch, J. Braasch, D. Kolossa, and J. Blauert, "An Introduction to Binaural Processing," in *The Technology of Binaural Listening* (J. Blauert, ed.), ch. 1, pp. 1–32, Berlin, Germany: Springer Berlin, 2013.

[30] J. Fliege and U. Maier, "The distribution of points on the sphere and corresponding cubature formulae," *IMA Journal of Numerical Analysis*, vol. 19, no. 2, pp. 317–334, 1999.

[31] K. Ono, V. Pulkki, and M. Karjalainen, "Binaural Modeling of Multiple Sound Source Perception: Coloration of Wideband Sound," in *AES Convention 112*, (Munich, Germany), pp. 1–8, AES, 2002.

[32] J. Ahrens, C. Andersson, P. Höstmad, and W. Kropp, "Tutorial on Scaling of the Discrete Fourier Transform and the Implied Physical Units of the Spectra of Time-Discrete Signals," in *AES Convention 148*, (Vienna, Austria), pp. 1–5, AES, 2020.

[33] B. Widrow and I. Kollár, "Spectrum of Quantization Noise and Conditions of Whiteness," in *Quantization Noise: Roundoff Error in Digital Computation, Signal Processing, Control, and Communications*, ch. 20, pp. 529–562, Cambridge, UK: Cambridge University Press, 2008.

[34] P. Marie, C. H. Jeong, J. Brunskog, and C. M. Petersen, "Audience Noise in Concert Halls During Musical Performances," in *41st International Congress and Exposition on Noise Control Engineering (INTER-NOISE)*, (New York, US), pp. 2231–2240, ICA, 2012.