# On the Impact of Noise Introduced by Spherical Beamforming Techniques on Data-Based Binaural Synthesis

Till Rettberg and Sascha Spors

*Institute of Communications Engineering, University of Rostock, 18191 Rostock, Germany*

*Email: {till.rettberg, sascha.spors}@uni-rostock.de*

## Introduction

The combination of sound field analysis techniques and binaural synthesis has been a subject of recent research. Typically, the acoustic scene is captured by a spherical microphone array and decomposed into plane waves employing modal beamforming. The sound field is subsequently auralized by the means of head-related transfer functions (HRTFs). For head-tracked, real-time binaural synthesis, translatory and rotational movements can be conveniently taken into account within the plane wave representation. Ideally, this allows to combine room transfer functions of arbitrary listening environments with (possibly individualized) free-field HRTFs for binaural room auralization.

Spherical microphone arrays suffer from practical limitations [1]: Spherical harmonic expansion is only possible up to a finite order $N$. Typically, the sound field on the sphere can only be captured at discrete microphone positions, introducing spatial aliasing for higher frequencies. Equipment noise is present in the recorded microphone signals. For modal beamforming, its attenuation or amplification is frequency-dependent, introducing colored noise in the resulting HRTFs. Furthermore, imperfect sensor positioning and calibration impair the orthogonal decomposition.

Some of these practical limitations and their implications for data-based binaural synthesis have been discussed in prior publications: The effects of a finite spherical expansion for a continuous sensor has been investigated in [2]: Directional perception seems not to be compromised for orders $N \geq 5$, while coloration and distance perception are still influenced at orders $N \geq 10$. Sampling at discrete microphone positions was considered in [3], showing coloration artifacts and deviation in perceived direction even at a maximum order of $N = 23$. The effects of translatory head-movement are subject of current research [4]. This paper aims to extend the results by taking equipment noise into account.

## Spherical Beamforming

From the complex sound pressure $P(\phi_s, \theta_s, \omega)$ at each of the $S$ microphone positions $(\phi_s, \theta_s)$, the modal spectrum $\mathring{P}_n^m(\omega)$ is obtained by the discrete spherical harmonic transform [1]:

$$\mathring{P}_n^m(\omega) = \sum_{s=1}^{S} \alpha_s P(\phi_s, \theta_s, \omega) Y_n^{m*}(\phi_s, \theta_s)\,, \qquad (1)$$

where $Y_n^m(\phi, \theta)$ denote the spherical harmonics of degree $m$ and order $n$ up to $N$ and $\alpha_s$ denote quadrature weights, depending on the employed sampling grid. Then, decomposition plane waves $\bar{P}(\phi_0, \theta_0, \omega)$ can be understood as beamforming in the modal domain:

$$\bar{P}(\phi_0, \theta_0, \omega) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \frac{d_n(\omega)}{b_n(\omega)} \mathring{P}_n^m(\omega) Y_n^m(\phi_0, \theta_0)\,, \qquad (2)$$

where the fraction $\frac{d_n(\omega)}{b_n(\omega)}$ represent the modal gain, where the radial function $b_n(\omega)$ depends on the array setup, i.e. for cardioid microphones on an open sphere of radius $r$:

$$b_n(\omega) = 4\pi i^n j_n(\frac{\omega}{c} r) - i j_n'(\frac{\omega}{c} r)\,, \qquad (3)$$

where $j_n$ is the spherical Bessel function of order $n$.

## Modal Limiting

The numerator $d_n(\omega)$ in (2) accounts for a limiting factor:[1] In theory, an ideal plane wave decomposition is achieved with $d_n(\omega) = 1$ [1]. However, since the spherical Bessel functions decay rapidly for $\omega \to 0$, the amplification of higher modes at low frequencies grows reciprocally. To ensure robustness, the maximum order $n$ has to be reduced at lower frequencies, or equivalently, the modal amplitude $\frac{d_n(\omega)}{b_n(\omega)}$ has to be limited to a maximum value $a_{max}$. We followed the latter approach proposed in [5], limiting the modal gain with soft-knee shelving curve to ensure a smooth spatio-temporal response.

The robustness against diffuse white noise is typically expressed in terms of the white noise gain (WNG)

$$\mathrm{WNG}(\omega) = 10 \log_{10} \left( \frac{S}{(4\pi)^2} \frac{\left| \sum_{n=0}^{N} d_n(\omega)(2n+1) \right|^2}{\sum_{n=0}^{N} \frac{|d_n(\omega)|^2}{|b_n(\omega)|^2}(2n+1)} \right) \qquad (4)$$

which expresses the SNR improvement from the array input to its output: a negative WNG values indicate amplification, positive ones suppression of noise.

## Results

Modal filters were computed for different maximum order $N$ and maximum modal gain $a_{max}$, using the Sound

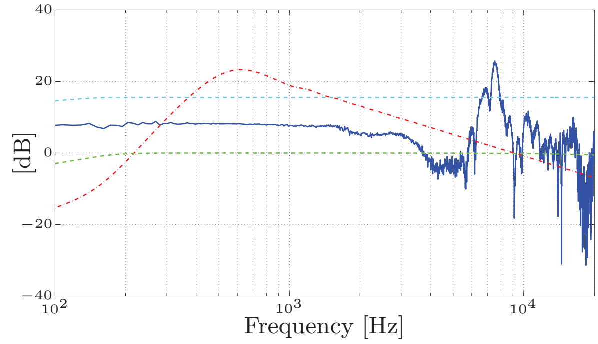
[1] This notation follows [1]. Care must be taken, since [5] and [6] use $d_n(\omega)$ not for the numerator, but for the entire fraction $d_n(\omega) := \frac{x}{b_n(\omega)}$, where $x$ depends on limiting and the desired beam.

Field Analysis Toolbox (SOfiA) [6]. For comparison with prior results, the same array setup as in [3] was used: a Lebedev grid of 770 cardioid microphones on an open sphere surface of $r = 0.5$m. The beam-filters were combined with measured HRTFs (3m distance, $1^\circ$ azimuth resolution [7]) to synthesize binaural room transfer functions (BRTFs) for varying signal-to-noise ratios.

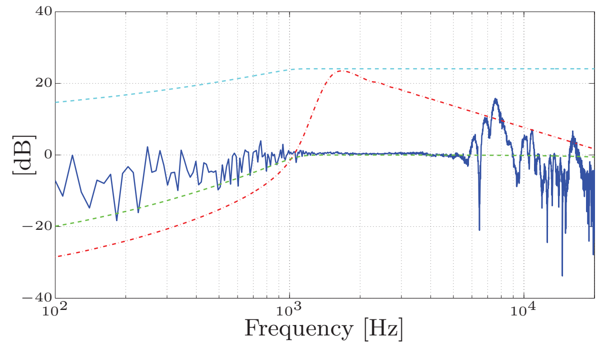
The deviation in magnitude between the computed BRTFs and the original HRTFs (which would idealistically be zero over the whole audible range) is shown in fig. 1 for the left ear, along with the common array performance measures. Above 5kHz for $N = 20$ spatial aliasing is prominent. The modal soft-limiting attenuates contributions of higher orders $n$ for low frequencies, imposing a high-pass characteristic and reducing directivity. Note that high orders are also attenuated for very high frequencies, but this is of less concern since the upper range is dominated by spatial aliasing. Fig. 1b shows severe noise amplification up to 1kHz, corresponding to the region of negative WNG. This is reduced by limiting $N$ (either explicitly as in 1a, or implicitly in terms of a reduced $a_{max}$ in 1c), but comes at the cost of reduced directivity: The omnidirectional contributions account for the difference between the beamformer's normalized magnitude response and the $\frac{|BRTF|}{|HRTF|}$ magnitude.

# Conclusion

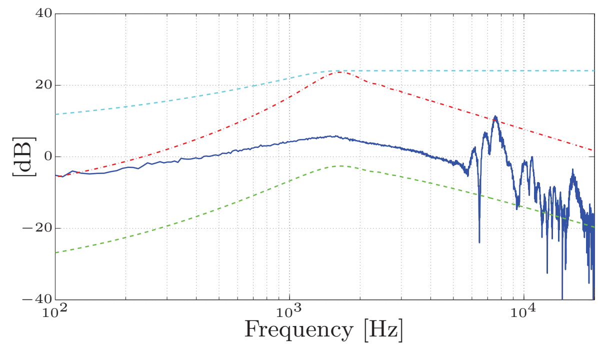
For data-based binaural synthesis, noise amplification appears to be the crucial limiting factor. For practical SNRs, the required orders $N \geq 10$ [3], cannot be straightforwardly achieved. The perceptual implications of modal windowing and of colored noise in HRTFs in general are not fully investigated yet. Listening experiments will be necessary to find a reasonable trade-off between noise and omnidirectional contributions in the resulting BRIRs. Furthermore, the WNG measures the robustness usually against spatially uncorrelated white noise. Modeling the diffuse noise field in terms of impinging plane waves might be advantageous to treat equipment noise and spatially correlated noise in the same analytical framework.




**(a)** max. order $N = 5$, modal gain limit $a_{max} = 40$ dB




**(b)** $N = 15$, $a_{max} = 40$ dB




**(c)** $N = 15$, $a_{max} = 10$ dB

**Figure 1:** Magnitude deviation between computed BRTFs and original HRTFs (solid blue) with 60dB SNR at array input. Dashed lines: WNG (red), directivity index (cyan) and beam-filter magnitude in looking direction (green).

# References


[1] B. Rafaely. Analysis and design of spherical microphone arrays. *Speech and Audio Processing, IEEE Transactions on*, 13(1):135–143, 2005.

[2] S. Spors and H. Wierstorf. Evaluation of Perceptual Properties of Phase-Mode Beamforming in the Context of Data-Based Binaural Synthesis. In *International Symposium on Control, Communications, and Signal Processing*, Rome, Italy, May 2012.

[3] S. Spors, H. Wierstorf, and M. Geier. Comparison of modal versus delay-and-sum beamforming in the context of data-based binaural synthesis. In *132nd Convention of the Audio Engineering Society*, Budapest, Hungary, April 2012.

[4] F. Schultz and S. Spors. Data-based Binaural Synthesis Including Rotational and Translatory Head-Movements. In *52nd International Conference of the Audio Engineering Society*, Guildford, UK, September 2013.

[5] B. Bernschütz, C. Pörschmann, S. Spors, and S. Weinzierl. Soft-Limiting der modalen Amplitudenverstärkung bei sphärischen Mikrofonarrays im Plane Wave Decomposition Verfahren. In *German Annual Conference on Acoustics (DAGA)*, Düsseldorf, Germany, March 2011.

[6] B. Bernschütz, C. Pörschmann, S. Spors, and S. Weinzierl. SOFiA - Sound Field Analysis Toolbox. In *International Conference on Spatial Audio*, Detmold, Germany, November 2011.

[7] H. Wierstorf, M. Geier, A. Raake, and S. Spors. A Free Database of Head Related Impulse Response Measurements in the Horizontal Plane with Multiple Distances. In *130th Convention of the Audio Engineering Society*, London, UK, May 2011.