
# A Software Tool For Auralization of Simulated Sound Fields

# A SOFTWARE TOOL FOR AURALIZATION OF SIMULATED SOUND FIELDS

Jens Ahrens Chalmers University of Technology, Gothenburg, Sweden

## 1 INTRODUCTION

Wave-based methods like Finite Element Method (FEM), Boundary Element Method (BEM), and Finite-Difference Time-Domain (FDTD) are becoming increasingly popular for room acoustic simulation applications. Contrary to geometric acoustics-based simulations where the spatial information is available in a tangible form, it is not straightforward to auralize wave-based simulations, and a variety of methods have been proposed. We present in this paper an auralization software toolbox that implements a selection of these methods in MATLAB. Research on the perceptual properties of the different approaches is limited at this stage so that we cannot yet provide guidelines on what approach may be most preferable in what situation. Our primary intent is providing the implementations to support such research.

The implemented algorithms work based on the following sampling strategies of the simulated sound field:

- *Volumetric sampling*: The simulated sound pressure is sampled with equal spacing inside a cubical volume.
- *Surface sampling*: The sound pressure and the normal sound pressure gradient (or, equivalently, the particle velocity) are sampled along a closed surface. If the sound pressure gradient is not available straightforwardly from the simulation framework, it is computed by sampling the sound pressure along two layers.

Both volumetric and surface sampling can be used to produce the following output formats:

- A *spherical harmonic (SH) representation* of the simulated sound field, or, in other words, an ambisonic representation. Our implementation is compatible with the modern ambisonics standards to that the established software tools like SPARTA[1] and the IEM Plugin Suite[2] can be used for rendering on both headphones and loudspeakers.
- *Direct binaural rendering*: Binaural output signals are computed directly from the sampled data without an intermediate ambisonic representation. This gives away some freedom regarding the playback formats and makes headtracking less straightforward, but it may have advantages.

Fig. 1 depicts a graphical representation of the signal processing pipeline.

Our implementation targets primarily simulations of the acoustic impulse response of a space. Neither the concepts themselves nor the implementation are limited to that. Running signals can be processed, too, but we will assume impulse responses in this paper to keep the language simple. We also focus on binaural rendering for convenience.

The present paper provides an overview of the software toolbox. Audio examples and further technical details are available from the online repository. Cf. Sec. 6.

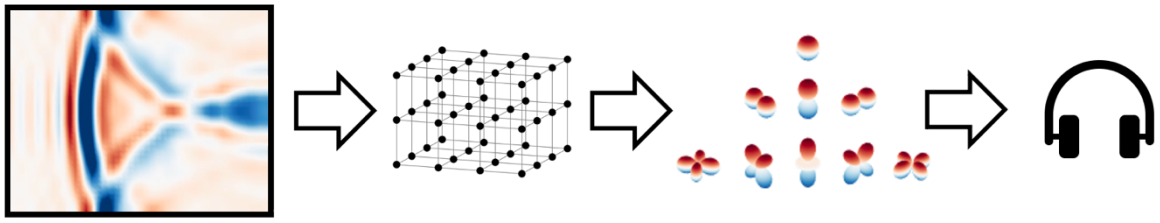

Fig. 1: Graphical representation of the pipeline of volumetric sampling for obtaining a spherical harmonic representation of the simulated sound pressure field that is rendered binaurally. The pipeline for direct binaural rendering is similar but does not comprise the third step. The data for the sound field simulation on the very left were obtained from[3].

# 2 LITERATURE

As a first method for binaural auralization of simulated sound fields, it was proposed in[4] to monitor the simulated sound field at two locations in space that coincide with the ear positions of a notional listener. The computation of an orthogonal representation from volumetrically sampled simulation data by means of an intermediate discrete plane wave decomposition (PWD) was originally proposed in[5]. An extension of[5] that uses a continuous PWD was presented in[6], which is the basis of our implementation of volumetric sampling with SH output. Direct binaural rendering of volumetrically sampled data was proposed in[7], which is also the basis of our implementation of this concept.

Surface sampling of pressure and gradient for producing an SH representation of a sound field was presented in[8] for spherical surfaces in the context of microphone array-based sound field analysis. The method can be applied in the present context without modification. As to our awareness, the solution for sampling pressure and gradient along a cubical surface (as in Fig. 2(c)) for obtaining an SH representation as well as direct binaural rendering of surface-sampled pressure and gradient along any kind of geometry have not been presented in the literature and are our original contribution.

A variety of other methods for auralization of simulated sound fields have been proposed that are not yet included in our toolbox. This includes[9,10], where the sound field is observed of over an infinitesimal volume for obtaining an SH representation, and[11], which obtains an SH representation by numerically evaluating a boundary integral of pressure and velocity. [12] employs single-layer surface sampling of the pressure for obtaining an SH representation via virtual microphones.

There is a branch of signal processing that aims at finding the optimal sampling strategy for sound fields to minimize the number of sampling points (or *nodes*) and the error of the sound field reconstruction. Examples are[13,14]. So far, these methods have only been demonstrated at low frequencies, and it seems that further work is required to extend them to the entire audible frequency range so that we do not consider such methods here.

# 3 TECHNICAL BACKGROUND

## 3.1 General

Independent of whether one seeks to compute an SH representation or the binaural signal directly, sampling the pressure along a single-layer surface produces ambiguities, which manifests as a bad conditioning of the problem, and the solution is generally not useful. Two options exist for avoiding this ambiguity: 1) Volumetric sampling and 2) sampling of the pressure and the pressure gradient along a surface whereby the gradient needs to be taken in direction normal to the surface. It is unclear at this point, which of the two produces the most favorable result.

The circumstance that the sound pressure field inside a simply connected surface is uniquely defined by the pressure and the normal gradient was demonstrated in[15]. In principle, pressure and pressure gradient can be combined (i.e. added) in a variety of ways such that the uniqueness is maintained. It seems most favorable to add them as

$$S_{\mathrm{card}}(\vec{\boldsymbol{x}}_0, \omega) = S(\vec{\boldsymbol{x}}_0, \omega) + \frac{1}{\mathrm{i}\frac{\omega}{c}} \frac{\partial}{\partial \vec{\boldsymbol{n}}} S(\vec{\boldsymbol{x}}, \omega)|_{\vec{\boldsymbol{x}}=\vec{\boldsymbol{x}}_0} \quad , \tag{1}$$

whereby $S(\vec{\boldsymbol{x}}_0, \omega)$ is the sound pressure at point $\vec{\boldsymbol{x}}_0$ on the surface at radian frequency $\omega$, $c$ is the speed of sound, i is the imaginary unit, and $\frac{\partial}{\partial \vec{\boldsymbol{n}}} S(\vec{\boldsymbol{x}}, \omega)|_{\vec{\boldsymbol{x}}=\vec{\boldsymbol{x}}_0}$ is the derivative of $S(\vec{\boldsymbol{x}}, \omega)$ taken in direction of the outward pointing normal $\vec{\boldsymbol{n}}$ on the surface and evaluated at $\vec{\boldsymbol{x}}_0$. This results in a virtual cardioid sensor that is located at $\vec{\boldsymbol{x}}_0$ and that points outward in direction $\vec{\boldsymbol{n}}$. This approach has been employed with microphone arrays, too[8].

Some simulation frameworks provide direct access to the particle velocity $V_{\vec{\boldsymbol{n}}}(\omega)$, which relates to the pressure gradient $\frac{\partial S(\omega)}{\partial \vec{\boldsymbol{n}}}$ via[16]

$$V_{\vec{\boldsymbol{n}}}(\omega) = -\frac{1}{\mathrm{i}\rho\omega} \frac{\partial S(\omega)}{\partial \vec{\boldsymbol{n}}} \quad , \tag{2}$$

whereby $\rho$ is the density of air. This can be directly plugged into (1).

Other simulation frameworks provide only direct access to the sound pressure. The gradient $\frac{\partial}{\partial \vec{\boldsymbol{n}}} S(\vec{\boldsymbol{x}}, \omega)|_{\vec{\boldsymbol{x}}=\vec{\boldsymbol{x}}_0}$ in (1) has to be approximated in these cases by means of the finite difference between two pressure sampling points: One essentially has to have access to the sound pressure along two surfaces that need to be close to each other compared to the wavelength. One speaks also of a double-layer surface.

We will consider the sampling grids depicted in Fig. 2 in the remainder. Surface sampling is usually formulated for spherical surfaces like in Fig. 2 (middle). This is certainly the most favorable geometry for this method. Some acoustic simulation frameworks such as FDTD usually use Cartesian sampling grids so that spherical sampling is not possible without interpolation. For these cases, we reformulated the surface sampling for arbitrary simply-connected surfaces including cubical surfaces like the one depicted in Fig. 2 (right).

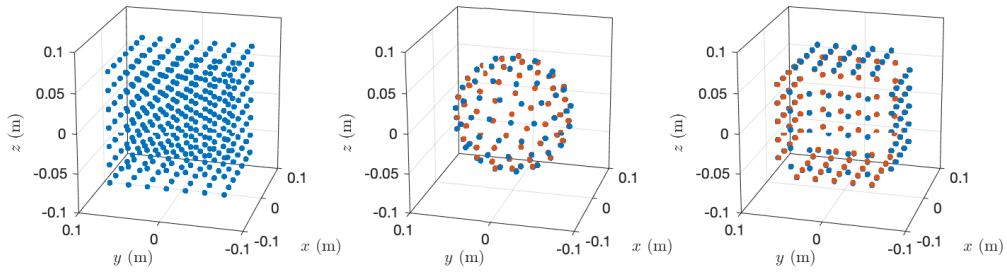


Fig. 2: The sampling grids considered in this paper. Left: Volumetric with 343 nodes. Middle: Spherical surface (Fliege grid with 100 double nodes). Right: Cubical surface with 150 double nodes. The surface grids are depicted as double layers, which allows for computing the normal pressure gradient from the pressure without knowledge on the particle velocity. Red color indicates the inner layer. All grids have a diameter of approx. 140 mm.

All methods that we implemented provide the least-squares solution to the problem in the frequency domain. I.e., all data are transformed to the frequency domain, and the problem is solved separately for each frequency bin. The result of the least-squares solution is a set of finite impulse response (FIR) filters that the acoustic simulation data are filtered with to obtain either the SH representation or the binaural signal. Fig. 3 depicts the signal flow charts for both cases.

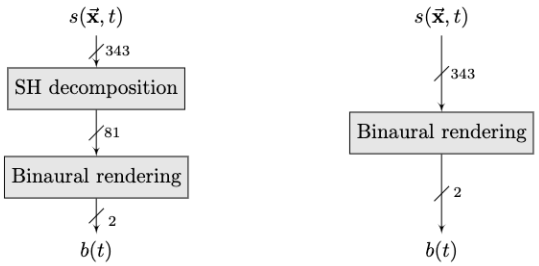


Fig. 3: Signal flow charts. Left: Decomposition of a simulated sound field $s(\vec{\boldsymbol{x}}, t)$ that was sampled at 343 points in space into an 8th-order (81-channel) SH representation that is rendered binaurally. Right: Direct binaural rendering of a simulated sound field $s(\vec{\boldsymbol{x}}, t)$ that was sampled at 343 points in space. $b(t)$ is the two-channel binaural output signal. The blocks *SH decomposition* and *Binaural rendering* represent multichannel FIR filtering operations that can be carried out in time domain or block-wise in frequency domain.

# 4 EVALUATION

All methods that our implementation comprises are linear with respect to the input data. We can therefore assess the performance based on elementary sound fields and predict the performance for complex sound fields based on that. The accuracy of the binaural output signals is an indicator for the effectiveness of the method. Accurate binaural output signals (i.e. output signals that are identical to the ground truth) are the ultimate proof that the method is perceptually transparent. It is to be expected that perfect accuracy is not feasible (or it may require impractical efforts) so that deviations of the actual binaural output signals from the ground truth binaural signals may arise. Such deviations can only ultimately be assessed by means of perceptual experiments and are subject to future work.

Head-related transfer functions (HRTFs) are, by definition, the acoustic response of the human ears to plane wave sound incidence. We can therefore simulate individual plane waves propagating over the sampling grid and assess the accuracy of the auralization pipeline by comparing the binaural output of the pipeline to the HRTFs corresponding to the incidence direction of the plane wave under consideration[7]. This assumes, of course, that we use the same set of HRTFs for the binaural rendering against which we compare the output of the pipeline. Since any sound field can be represented by a superposition of plane waves, it is sufficient to only analyze the plane wave case to evaluate the general numerical accuracy.

It is unclear at this point what set of parameters – i.e. what combination of size of the sampling grid and number of sampling points – is required for perceptual transparency (i.e. for the circumstance that ground truth and output of the processing pipeline are perceptually indistinguishable). We evaluate exemplarily the grids from Fig. 2 that are all of a size that is similar to that of a human head and all comprise 343 pressure sampling points or 100 or 150 pairs of pressure and gradient. The intermediate ambisonic representations that this yields are of 8th order.

Fig. 4-7 depict the ground truth HRTFs for a variety of horizontal incidence directions, the binaural output of the processing pipeline for a plane wave of identical incidence direction, and the difference between the two. Data are presented for volumetric and spherical surface sampling and intermediate ambisonic representation as well as for volumetric and spherical surface sampling and direct binaural rendering. 0° incidence angle in the figures corresponds to incidence from straight ahead; 90° corresponds to lateral incidence from the left. All data are for the left ear. The results for the cubical

surface sampling are very similar to the results for the spherical surface sampling and are omitted here. The results for non-horizontal sound incidence are very similar to the results for horizontal sound incidence and are omitted, too.

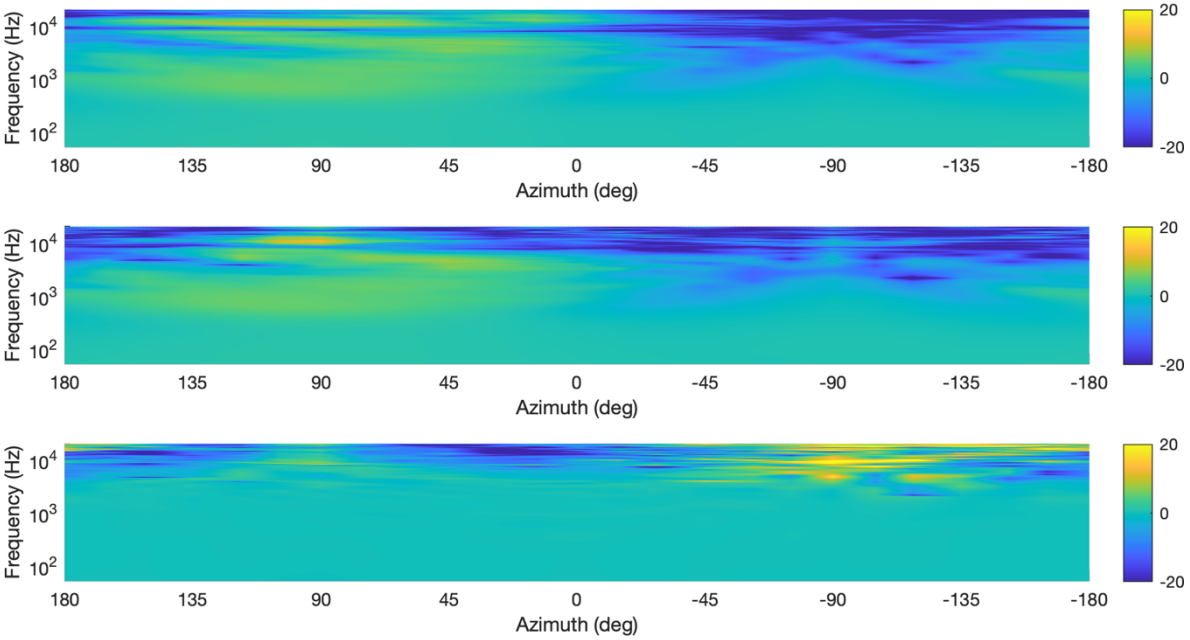


Fig. 4: Magnitudes in dB of HRTFs (top), binaural output of the processing pipeline using volumetric sampling (Fig. 2(a)) with intermediate ambisonics (middle), and the difference between the two (bottom) for horizontal sound incidence.

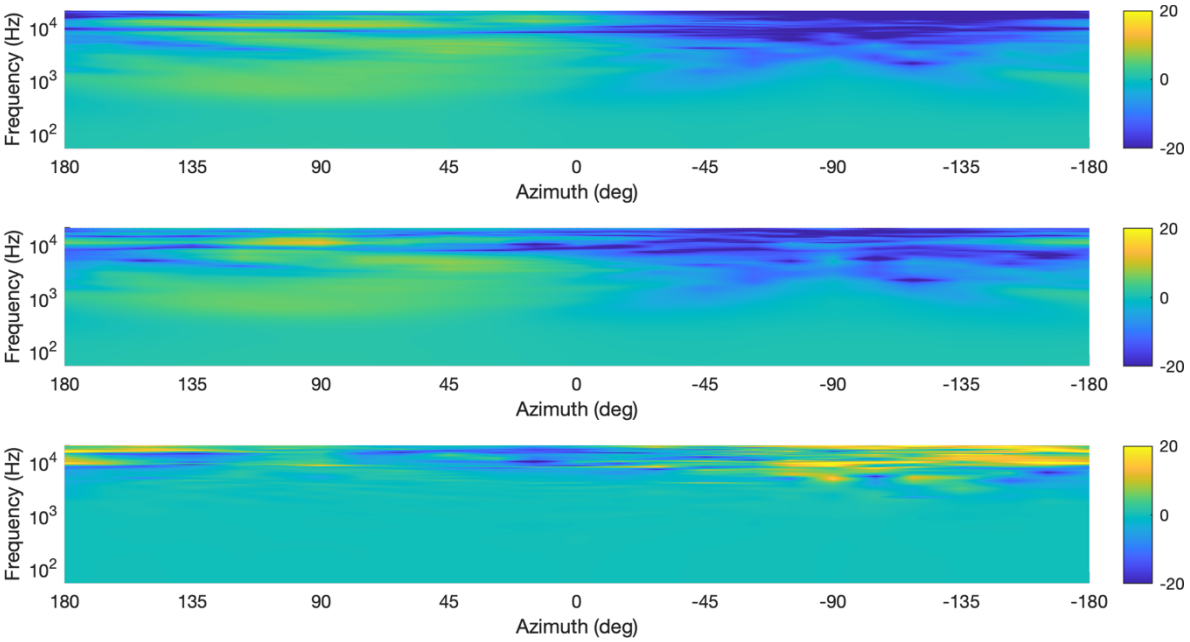


Fig. 5: Magnitudes in dB of HRTFs (top), binaural output of the processing pipeline using spherical surface sampling (Fig. 2(b)) with intermediate ambisonics (middle), and the difference between the two (bottom) for horizontal sound incidence. The results for cubical surface sampling (Fig. 2(c)) are nearly identical.

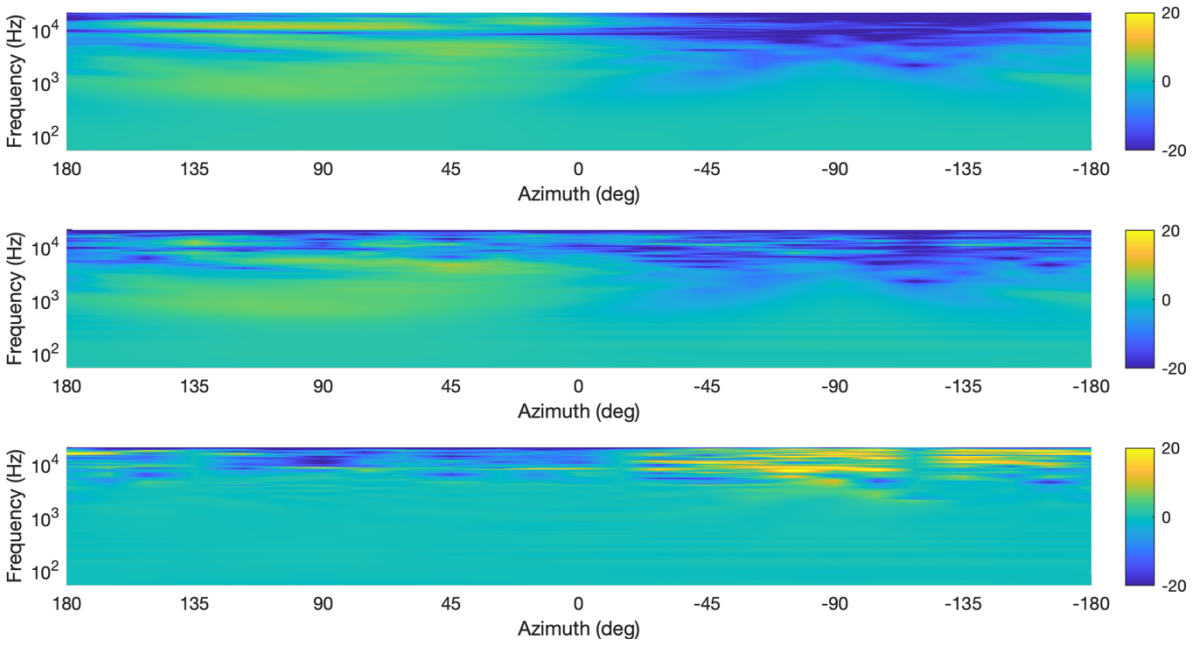


Fig. 6: Magnitudes in dB of HRTFs (top), binaural output of the processing pipeline using direct rendering of volumetrically sampled data (middle, cf. Fig. 2(a)), and the difference between the two (bottom) for horizontal sound incidence.

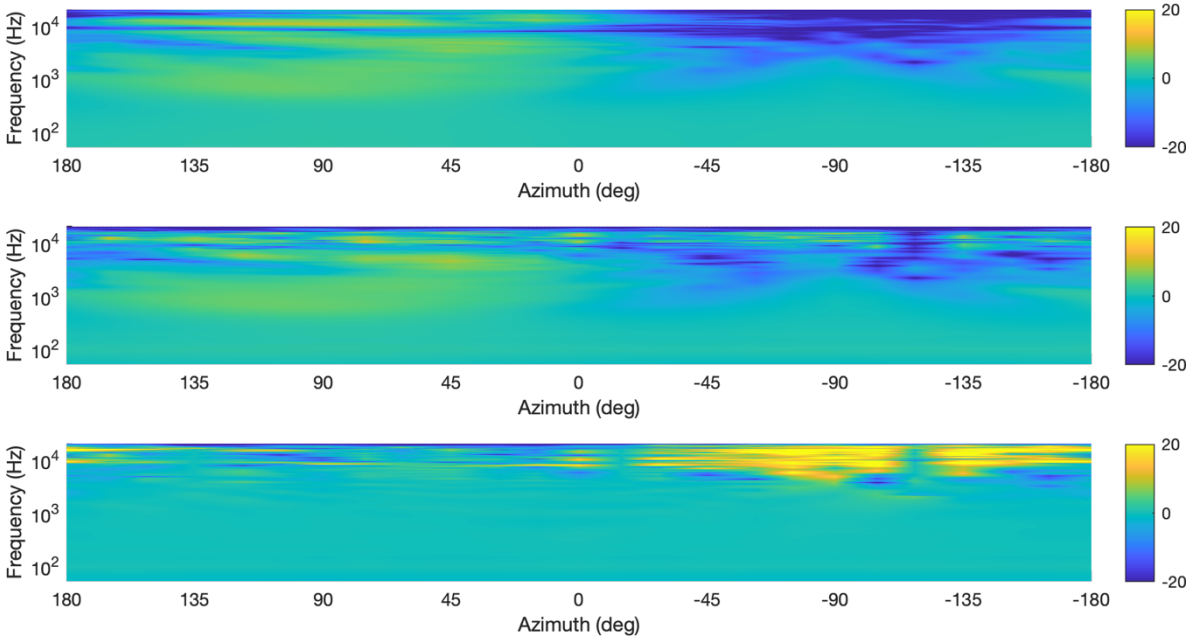


Fig. 7: Magnitudes in dB of HRTFs (top), binaural output of the processing pipeline using direct rendering of spherical surface-sampled pressure and pressure gradient (middle, cf. Fig. 2(b)), and the difference between the two (bottom) for horizontal sound incidence. The results for cubical surface sampling (Fig. 2(c)) are nearly identical.

It is evident from Fig. 4-7 that the deviation of the binaural output of the signal processing pipeline is very low for frequencies below 6 kHz but can deviate considerably at frequencies above that. This critical frequency is primarily dependent on the combination of number of sampling points and the grid dimensions. It is noteworthy that the largest deviations tend to occur at the contralateral ear (i.e. around -90° incidence direction). Deviations there are typically perceptually less critical. We also highlight that whether the deviations at high frequencies tend to manifest as too high magnitude or too low magnitude depends on the combination of algorithm, grid size, number of sampling points, and the regularization of the algorithm. We cannot provide explicit guidelines at this point.

The fact that the deviations occur exclusively at higher frequencies is also promising as typical signals like speech and music exhibit only a small part of their energy there. Refer to the audio examples that we provide online to experience this (see Sec. 6). The audio examples comprise also reverberant data besides the free-field data that are depicted in Fig. 4-7. Informal listening suggests that the perceptual difference between the output of the processing pipeline and the ground truth is small.

# 5 SUMMARY AND CONCLUSIONS

We demonstrated an implementation of a set of methods for auralization of simulated sound fields. For the chosen grids of head size and of 343 pressure points or 100 or 150 sampling points for both pressure and pressure gradient, respectively, the binaural output signals of the methods deviate from the ground truth only at high frequencies and mostly on the contralateral side, which makes the output perceptually very similar to the ground truth.

What combination of grid type, grid dimensions, and number of sampling points produces perceptually transparent auralization (i.e. an auralized signal that is perceptually indistinguishable from the ground truth) is unclear at this stage and is subject to further research. It is conceivable that such a combination can be found. The presented methods will then constitute a unified approach to auralization as they can be applied to any acoustical simulation method, be it wave-based, geometric, or energy-based, so long as the simulation method allows for computing volumetric sound pressure data. Simulation methods can thereby be compared perceptually without uncertainty regarding the influence of the auralization method on the result.

# 6 RESOURCES

Find our software toolbox as well as a variety of anechoic and reverberant binaural audio examples here: https://github.com/AppliedAcousticsChalmers/auralization-toolbox

# 7 REFERENCES

1. https://leomccormack.github.io/sparta-site/
2. https://plugins.iem.at/
3. L. Müller, W. Kropp, and J. Forssén, ‘Measurement, Simulation and Auralization of Indoor Road Traffic Noise,’ in Proc. of DAGA, Stuttgart, Germany (March 2022)
4. M. Kleiner, E. Granier, B.-I. Dahlenbäck, and P. Svensson, ‘Coupling of Low and High Frequency Models in Auralization,’ in Proc. 15th Int. Congr. On Acoustics, Trondheim, Norway (June 1995)
5. B. Støfringsdal and U.P. Svensson, ‘Conversion of Discretely Sampled Sound Field Data to Aualization Formats,’ J. Audio. Eng. Soc. 54(5), pp. 380-400 (May 2006)
6. J. Sheaffer, M. van Walstijn, B. Rafaely, and K. Kowalczyk, ‘Binaural Reproduction of Finite Difference Simulations Using Spherical Array Processing’, IEEE/ACM TASLP 23(12), pp. 2125-2135 (Dec. 2015)

7. M. A. Poletti and U. P. Svensson, 'Beamforming Synthesis of Binaural Responses From Computer Simulations of Acoustic Spaces,' J. Acoust. Soc. Am. 124, pp. 301–315 (2008)
8. I. Balmages and B. Rafaely, "Open-sphere designs for spherical microphone arrays," IEEE TASLP, vol. 15, no. 2, pp. 727– 732 (2007)
9. R. Mehra, L. Antani, S. Kim, and D. Manocha, "Source and listener directivity for interactive wave-based sound propagation," IEEE Trans. on Visualization and Computer Graphics, vol. 20, no. 4, (2014)
10. S. Bilbao, A. Politis, and B. Hamilton, "Local time-domain spherical harmonic spatial encoding for wave-based acoustic simulation," IEEE Signal Processing Letters, vol. 26, no. 4, pp. 617–621 (2019)
11. J. A. Hargreaves, L. R. Rendell, and Y. W. Lam, "A framework for auralization of boundary element method simulations including source and receiver directivity," JASA, vol. 145 (2019)
12. A. Southern, D. T. Murphy, and L. Savioja, "Spatial encoding of finite difference time domain acoustic models for auralization," IEEE TASLP, vol. 20, no. 9, pp. 2420–2432 (2012)
13. G. Chardon, W. Kreuzer, and M. Noisternig, "Design of Spatial Microphone Arrays for Sound Field Interpolation," IEEE Journal of Selected Topics in Signal Processing 9(5), pp. 780 - 790 (August 2015)
14. S. A. Verburg Riezu, F. Elvander, T. van Waterschoot, and E. Fernandez Grande, "Optimal sensor placement for sound field reconstruction," in Proc. of 24th Int. Congress on Acoustics (2022)
15. A. J. Burton and G. F. Miller, "The application of integral equation methods to the numerical solution of some exterior boundary-value problems," Proc. of the Royal Soc. of London. A., vol. 323, no. 1553, pp. 201–210 (1971)
16. E. G. Williams, "Fourier Acoustics," Academic Press, San Diego (1999)