The Chalmers Auralization Toolbox

Jens Ahrens

Technical note v. 2024/12/06

Chalmers University of Technology

jens.ahrens@chalmers.se

Abstract

This document presents a summary of the methods that are implemented in the Chalmers Auralization Toolbox. There are two types of solutions that are available (direct auralization and auralization with an intermediate ambisonic representation) that can be applied with three different types of grids (cubical volumetric, spherical surface, and cubical surface) on which the sound field to be auralized is observed.

1 Introduction

Please refer to (Ahrens, 2024) for a comprehensive introduction to the Chalmers Auralization Toolbox[1] and its purposes. The present document focuses on the mathematical and signal processing-related details of the methods that are implemented.

2 Ambisonic Auralization

Methods that we term here *ambisonic methods* produce a representation of the sampled sound field in terms of spherical harmonic basis functions (SHs). This representation is also referred to as ambisonic representation and can be rendered binaurally or using a loudspeaker array (Zotter and Frank, 2019). The advantage of the ambisonic methods is that the problem of auralization is decomposed into two independent stages that can be optimized separately. From a purely practical perspective, head-tracked rendering of the ambisonic representation can be carried out straightforwardly by applying rotation operations on either the ambisonic sound field or the ambisonic HRTF representation.

2.1 Volumetric Sampling

Our implementation follows (Sheaffer et al., 2015) where volumetric sampling of the simulated sound pressure is performed for obtaining the spherical harmonic (SH) coefficients. We summarize this method in the following. (Sheaffer et al., 2015) may be considered an extension of (Støfringsdal and Svensson, 2006).

A sound pressure $S(\vec{\mathbf{x}}, \omega)$ at a point $\vec{\mathbf{x}}$ and at radian frequency ω can be represented by a sum of SH coefficients $\breve{S}_{n,m}(\omega)$ as (Gumerov and Duraiswami, 2005)

$$S(\vec{\mathbf{x}}, \omega) = \sum_{n=0}^{N} \sum_{m=-n}^{n} 4\pi \mathrm{i}^n \breve{S}_{n,m}(\omega)\, j_n\left(\omega \frac{r}{c}\right) Y_{n,m}(\beta, \alpha)\,. \tag{1}$$

$j_n(\cdot)$ denotes the spherical Bessel function of n-th order, and c denotes the speed of sound. The coordinate system is depicted in Fig. 1. The maximum order N of the decomposition in (1) is infinity for arbitrary sound fields but must be limited in the signal processing.

We use the following definition of the surface SHs $Y_{n,m}(\beta, \alpha)$, which is popular in the ambisonics community:

$$Y_{n,m}(\beta, \alpha) = (-1)^m \sqrt{\frac{2n+1}{4\pi} \frac{(n-|m|)!}{(n+|m|)!}}\, P_{n,|m|}(\cos\beta) \times \begin{cases} \sqrt{2}\sin|m|\alpha, & \forall m < 0 \\ 1 & , \forall m = 0 \\ \sqrt{2}\cos|m|\alpha, & \forall m > 0 \end{cases}. \tag{2}$$

[1] The Chalmers Auralization Toolbox is available here: https://github.com/AppliedAcousticsChalmers/auralization-toolbox

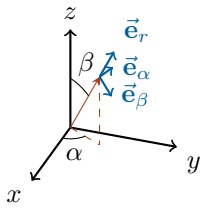

Figure 1: Spherical coordinate system with radius r, azimuth α, and colatitude β and their orientations $\vec{\mathbf{e}}_r$, $\vec{\mathbf{e}}_\alpha$, and $\vec{\mathbf{e}}_\beta$

$P_{n,m}(\cdot)$ are the associated Legendre polynomials (Gumerov and Duraiswami, 2005).

When sampling $S(\vec{\mathbf{x}}, \omega)$ volumetrically at Q points, (1) can be formulated separately for each sampling point $\vec{\mathbf{x}}_q$ ($q = 1, 2, \dots Q$). The resulting set of equations can then be formulated for each angular frequency ω using matrix notation as

$$\mathbf{s} = \mathbf{B}\mathbf{s}_{\mathbf{n},\mathbf{m}} \,, \tag{3}$$

whereby $\mathbf{s}$ comprises the sampled sound pressure, $\mathbf{s} = [S(\vec{\mathbf{x}}_1, \omega)\ S(\vec{\mathbf{x}}_2, \omega)\ \dots\ S(\vec{\mathbf{x}}_Q, \omega)]^T$, $\mathbf{B}$ is a $Q \times (N+1)^2$ matrix that contains the terms $4\pi \mathrm{i}^n j_n\left(\omega \frac{r_q}{c}\right) Y_{n,m}(\beta_q, \alpha_q)$ with q increasing downwards and (n, m) increasing towards the right according to the summation in (1), and $\mathbf{s}_{\mathbf{n},\mathbf{m}}$ contains the SH coefficients $\mathbf{s}_{\mathbf{n},\mathbf{m}} = [\breve{S}_{0,0}(\omega)\ \breve{S}_{1,-1}(\omega)\ \dots\ \breve{S}_{N,N}(\omega)]^T$.

An estimate $\hat{\mathbf{s}}_{n,m}$ of $\mathbf{s}_{n,m}$ can be obtained from the sampled sound pressure $\mathbf{s}$ as

$$\hat{\mathbf{s}}_{n,m} = \mathbf{B}^\dagger \mathbf{s} \,, \tag{4}$$

whereby $\mathbf{B}^\dagger$ is the Moore-Penrose pseudoinverse of $\mathbf{B}$. We refer to $\mathbf{B}^\dagger$ as the auralization matrix and use the symbol $\mathbf{C}_{n,m}$ for it in the remainder.

The Bessel functions $j_n\left(\omega \frac{r}{c}\right)$ in $\mathbf{B}$ exhibit zeros and take on very small values for small ω when $n > 0$. (Sheaffer et al., 2015) proposes soft-clipping for regularizing the problem, which limits how small the magnitude of the elements of $\mathbf{B}$ is permitted to be. It was originally proposed in (Bernschütz, 2016) for limiting the magnitude of a quantity and was reformulated in (Sheaffer et al., 2015) for the present purpose as $B_{\text{clipped}} = \left(\frac{2g}{\pi} \frac{|B|}{B} \arctan \frac{\pi}{2g|B|}\right)^{-1}$, whereby B is the quantity to be soft-clipped, and g is the minimum permitted magnitude on a linear scale after soft-clipping. This greatly improves the conditioning.

While this can indeed produce perceptually very useful results, we found the results to be even more favorable when inverting $\mathbf{B}$ in (3) via a singular value decomposition (SVD) (Cohen, 2021) that is regularized by clipping the dynamic range of the singular values to $40\,\mathrm{dB}$. This is similar to what was proposed in (Politis and Gamper, 2017).

An SH representation of order N comprises $(N+1)^2$ SH coefficients $\breve{S}_{n,m}(\omega)$ that need to be deduced from the sampled sound pressure. Avoiding that the equation system in (3) is underdetermined requires $Q \geq (N+1)^2$ sampling points ($\mathbf{B}$ in (3) needs to be a square or a tall matrix).

2.2 Surface Sampling

The Kirchhoff-Helmholtz integral (Gumerov and Duraiswami, 2005) demonstrates that the sound pressure field inside a source-free domain is uniquely described by the sound pressure $S(\vec{\mathbf{x}}, \omega)$ and normal sound pressure gradient $\frac{\partial}{\partial \vec{\mathbf{n}}} S(\vec{\mathbf{x}}, \omega)$ distributions along the simply connected surface that encloses the domain. We therefore seek to establish a method that requires only sampling of that surface rather than inside the entire volume.

It is proven in (Burton and Miller, 1971, p. 207) that a weighted sum $S(\vec{\mathbf{x}}, \omega) + \gamma \frac{\partial}{\partial \vec{\mathbf{n}}} S(\vec{\mathbf{x}}, \omega)$ uniquely defines the sound pressure inside the volume that is enclosed by the surface so long as $\mathrm{Im}\{\gamma\} \neq 0$. We choose $\gamma = 1/(\mathrm{i}\frac{\omega}{c})$ here as this fulfills the uniqueness criterion, it assures that pressure and gradient are added with similar magnitudes, and it allows for the weighted sum to be interpreted as a virtual sensor with (far-field) cardioid directivity, which connects our formulation well to the literature on microphone arrays (Balmages and Rafaely, 2007; Thomas, 2019).

For ease of notation, we first formulate our proposed method for sampling points on a spherical surface that is centered around the coordinate origin and generalize the formulation to arbitrary surfaces in Sec. 2.2.2.

Table 1: The components of (10). $P'_{n,|m|}(\cdot)$ is the derivative of $P_{n,|m|}(\cdot)$ with respect to the argument, which can be computed via a recurrence relation (Gumerov and Duraiswami, 2005, Eq. (2.1.53)).

$\frac{\partial}{\partial r} j_n\left(\omega\frac{r}{c}\right) = \frac{\omega}{c} j'_n\left(\omega\frac{r}{c}\right)$
$\frac{\partial}{\partial\beta} Y_{n,m}(\beta,\alpha) = (-1)^{m+1}\sqrt{\frac{2n+1}{4\pi}\frac{(n-\lvert m\rvert)!}{(n+\lvert m\rvert)!}}\, P'_{n,\lvert m\rvert}(\cos\beta)\sin\beta \begin{cases}\sqrt{2}\sin\lvert m\rvert\alpha, & \forall m<0\\ 1, & \forall m=0\\ \sqrt{2}\cos\lvert m\rvert\alpha, & \forall m>0\end{cases}$
$\frac{\partial}{\partial\alpha} Y_{n,m}(\beta,\alpha) = (-1)^{m}\sqrt{\frac{2n+1}{4\pi}\frac{(n-\lvert m\rvert)!}{(n+\lvert m\rvert)!}}\, P_{n,\lvert m\rvert}(\cos\beta)\lvert m\rvert \begin{cases}\sqrt{2}\cos\lvert m\rvert\alpha, & \forall m<0\\ 0, & \forall m=0\\ (-1)\sqrt{2}\sin\lvert m\rvert\alpha, & \forall m>0\end{cases}$

2.2.1 Spherical Surfaces

The signal $S_{\text{card.}}(\vec{x}_0,\omega)$ from a virtual sensor at position $\vec{\mathbf{x}}_0$ with frequency-independent far-field cardioid directivity facing outward in radial direction is obtained via (Balmages and Rafaely, 2007)

$$S_{\text{card.}}(\vec{\mathbf{x}}_0,\omega) = S(\vec{\mathbf{x}}_0,\omega) + \frac{1}{\mathrm{i}\frac{\omega}{c}}\frac{\partial}{\partial r} S(\vec{\mathbf{x}},\omega)\Big|_{\vec{\mathbf{x}}=\vec{\mathbf{x}}_0}. \tag{5}$$

Sampling $S_{\text{card.}}(\vec{\mathbf{x}},\omega)$ along a spherical surface that is centered around the coordinate origin assures that the gradient in (5) is computed in outward-facing normal direction so that the uniqueness requirements are fulfilled. A similar case was treated in the literature on SMAs (Balmages and Rafaely, 2007; Thomas, 2019) and their 2-dimensional equivalent (Hulsebos et al., 2002) where the good conditioning was demonstrated. Using the equation of motion (Gumerov and Duraiswami, 2005, Eq. (1.1.3)), (5) can be expressed based on the normal particle velocity $V(\vec{\mathbf{x}},\omega)$ as

$$S_{\text{card.}}(\vec{\mathbf{x}}_0,\omega) = S(\vec{\mathbf{x}}_0,\omega) - \rho c\, V(\vec{\mathbf{x}_0},\omega)\,, \tag{6}$$

whereby ρ is the mass density of air. In the following, we demonstrate how the equivalent to (3) can be formulated for this setting.

Expressing (5) in terms of SHs reads (Balmages and Rafaely, 2007)

$$S_{\text{card.}}(\vec{\mathbf{x}}_0,\omega) = \sum_n \sum_m 4\pi\mathrm{i}^n \breve{S}_{n,m}(\omega)\ \left[j_n\left(\omega\frac{r_0}{c}\right) - \mathrm{i} j'_n\left(\omega\frac{r_0}{c}\right)\right] Y_{n,m}(\beta_0,\alpha_0)\,. \tag{7}$$

$j'_n(\cdot)$ is the derivative of $j_n(\cdot)$ with respect to the argument, which can be computed via a recurrence relation (Gumerov and Duraiswami, 2005, Eq. (2.1.87)). The equivalent of $\mathbf{B}$ in (3) now comprises all terms in the summation on the right hand side of (7) other than $\breve{S}_{n,m}(\omega)$. The equivalent of $\mathbf{s}$ comprises $S_{\text{card.}}(\vec{\mathbf{x}}_0,\omega)$, which we obtain from the simulated sound pressure according to (5) while approximating the gradient via the corresponding difference quotient as

$$\frac{\partial}{\partial r} S(\vec{\mathbf{x}},\omega)\Big|_{\vec{\mathbf{x}}=\vec{\mathbf{x}}_0} = \frac{S(\vec{\mathbf{x}}_2,\omega) - S(\vec{\mathbf{x}}_1,\omega)}{\|\vec{\mathbf{x}}_2 - \vec{\mathbf{x}}_1\|}\,, \tag{8}$$

which we can compute directly from the simulation data. $\vec{\mathbf{x}}_0 = \frac{\vec{\mathbf{x}}_2 - \vec{\mathbf{x}}_1}{2}$ is the midpoint between $\vec{\mathbf{x}}_1$ and $\vec{\mathbf{x}}_2$, which have to lie accordingly on a radial line.

2.2.2 Arbitrary Surfaces

As the uniqueness criterion $\mathrm{Im}\{\gamma\} \neq 0$ from (Burton and Miller, 1971) holds for arbitrary simply-connected surfaces, we generalize the formulation from Sec. 2.2.1 here and establish the equivalents of (5) and (7) – and consequently of (3) and (4) – for such arbitrary surface shapes.

We need to compute the gradient in direction $\vec{\mathbf{n}} = \vec{\mathbf{n}}(\vec{\mathbf{x}}_0)$ that is normal to the arbitrary surface at the point $\vec{\mathbf{x}}_0$. The partial derivative with respect to r in (5) needs to be replaced with the derivative in direction of the unit-length vector $\vec{\mathbf{n}}$ (pointing in direction $(\beta_{\vec{\mathbf{n}}}, \alpha_{\vec{\mathbf{n}}})$), which is given by

$$\frac{\partial}{\partial\vec{\mathbf{n}}} S(\vec{\mathbf{x}},\omega)\Big|_{\vec{\mathbf{x}}=\vec{\mathbf{x}}_0} = \begin{bmatrix}\cos\alpha_{\vec{\mathbf{n}}}\sin\beta_{\vec{\mathbf{n}}}\\ \sin\alpha_{\vec{\mathbf{n}}}\sin\beta_{\vec{\mathbf{n}}}\\ \cos\beta_{\vec{\mathbf{n}}}\end{bmatrix}^T \nabla\, S(\vec{\mathbf{x}},\omega)\Big|_{\vec{\mathbf{x}}=\vec{\mathbf{x}}_0} \tag{9}$$

whereby ^{T} denotes the matrix transpose operator and (Arfken and Weber, 2005, Sec. 2.5)

$$\nabla = \vec{\mathbf{e}}_r \frac{\partial}{\partial r} + \vec{\mathbf{e}}_\beta \frac{1}{r} \frac{\partial}{\partial \beta} + \vec{\mathbf{e}}_\alpha \frac{1}{r \sin \beta} \frac{\partial}{\partial \alpha} \tag{10}$$

with

$$\vec{\mathbf{e}}_r = \begin{bmatrix} \cos\alpha \sin\beta \\ \sin\alpha \sin\beta \\ \cos\beta \end{bmatrix}, \; \vec{\mathbf{e}}_\beta = \begin{bmatrix} \cos\alpha \cos\beta \\ \sin\alpha \cos\beta \\ -\sin\beta \end{bmatrix}, \; \vec{\mathbf{e}}_\alpha = \begin{bmatrix} -\sin\alpha \\ \cos\alpha \\ 0 \end{bmatrix}.$$

Fig. 1 visualizes the unit vectors $\vec{\mathbf{e}}_r, \vec{\mathbf{e}}_\beta, \vec{\mathbf{e}}_\alpha$ of the coordinate dimensions. Recall that (r, β, α) denote the position $\vec{\mathbf{x}}$. The first vector on the right-hand side of (9) is the Cartesian form of $\vec{\mathbf{n}}$.

All of the above equations in this section need to be inserted into one another to establish the equivalents of (5) and (7) and consequently the equivalents of (3) and (4). Space restrictions prevent us from providing explicit expressions, and we can only state their components. Refer to Tab. 1 for the explicit expressions for the gradients in (9)-(10) expressed in SHs.

$P'_{n,|m|}(\cdot)$ in Tab. 1 and the factor $1/\sin\beta$ in (10) are not defined if they are evaluated at locations on the z-axis ($\beta = 0, \pi$). This is not a limitation so long as we choose the surface such that its normal at the locations where the surface intersects with the z-axis is parallel to the z-axis. Because then, the trigonometric functions in (9) vanish for those undefined cases.

As to our awareness, the surface sampling along arbitrary simply connected surfaces presented above has not been presented elsewhere in the literature.

2.2.3 A Note on the Regularization

As mentioned in Sec. 2.1, we regularize the inversion of $\mathbf{B}$ in (3) by restricting the dynamic range of the singular values. For both cubical volumetric and cubical surface grids, a dynamic range of $40\,\mathrm{dB}$ is useful. Spherical surface grids provide better conditioning, and the result is perceptually ever so slightly more favorable for high-resolution spherical grids when using a dynamic range of $70\,\mathrm{dB}$ for the regularization.

3 Direct Auralization

The direct auralization methods are based on (Poletti and Svensson, 2008). The auralization is modelled as a MIMO system the input to which is the sampled sound field data.

Consider the sampled sound field $\mathbf{s} = [S(\vec{\mathbf{x}}_1, \omega)\; S(\vec{\mathbf{x}}_2, \omega)\; \ldots\; S(\vec{\mathbf{x}}_Q, \omega)]^T$, whereby $S(\vec{\mathbf{x}})$ can be either the observed sound pressure or the signal from a virtual cardioid sensor. Similarly to (4), we establish a relation between the observed sound field $\mathbf{s}$ and the estimation of the binaural output signal $\hat{b}(\omega)$

$$\hat{b}(\omega) = \mathbf{C}_{\text{direct}} \mathbf{s}\,. \tag{11}$$

$\mathbf{C}_{\text{direct}}$ in (11) is actually a vector, but we still term it auralization matrix and notate it as a matrix to highlight the correspondence between (4) and (11). $\mathbf{C}_{\text{direct}}$ represents the transfer function between the observed sound field and the according binaural output signal. Eq. (11) therefore has to be formulated separately for the left and right binaural output signal. $\mathbf{C}_{\text{direct}}$ can be found through a least squares (LS) fit onto example data. Recalling that HRTFs are defined as the acoustic ear signals due to a plane wave in free-field conditions makes sampled plane waves and HRTFs a convenient set of input/output data for computing the LS solution. Formulating eq. (11) for more planes waves with different incidence directions than sampling points Q makes the resulting equation system overdetermined. Like any other method, direct auralization suffers from spatial aliasing. The LS solution of that equation system can exhibit large errors in the frequency range where spatial aliasing occurs. We found that the solution is perceptually more favorable if the LS solution is replaced with a MagLS solution in the frequency range where the spatial aliasing occurs. Our implementation corresponds to what is termed end-to-end MagLS, variant 2, proposed in (Deppisch et al., 2021). As with ambisonic auralization, the LS and end-to-end MagLS problems are regularized via clipping the dynamic range of the singular values of the underlying SVD to $40\,\mathrm{dB}$.

The original method was demonstrated for volumetrically sampled sound pressure. Applying it to spherical and cubical surface sampled data does not require any modifications. Note that the listener orientation is encoded in $\mathbf{C}_{\text{direct}}$, which means that a new $\mathbf{C}_{\text{direct}}$ has to be computed from scratch for each head orientation of interest.

References

Jens Ahrens. Perceptually transparent binaural auralization of simulated sound fields. *JAES*, 2024. (submitted).

George B. Arfken and Hans J. Weber. *Mathematical Methods for Physicists*. Elsevier, Amsterdam, 6 edition, 2005.

Ilya Balmages and Boaz Rafaely. Open-sphere designs for spherical microphone arrays. *IEEE TASLP*, 15(2): 727–732, 2007.

Benjamin Bernschütz. Microphone arrays and sound field decomposition for dynamic binaural recording. PhD thesis, Technische Universität Berlin, 2016.

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.*, 323(1553):201–210, 1971.

Mike X. Cohen. *Linear Algebra: Theory, Intuition, Code*. sincXpress, 2021.

Thomas Deppisch, Hannes Helmholz, and Jens Ahrens. End-to-end magnitude least squares binaural rendering of spherical microphone array signals. In *Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, pages 1–7, 2021. doi: 10.1109/I3DA48870.2021.9610864.

Nail Gumerov and Ramani Duraiswami. *Fast Multipole Methods for the Helmholtz Equation in Three Dimensions*. Elsevier, Amsterdam, 2005.

Edo Hulsebos, Diemer de Vries, and Emmanuelle Bourdillat. Improved microphone array configurations for auralization of sound fields by wave-field synthesis. *JAES*, 50(10):779–790, October 2002.

Mark A. Poletti and U. Peter Svensson. Beamforming synthesis of binaural responses from computer simulations of acoustic spaces. *The Journal of the Acoustical Society of America*, 124(1):301–315, 2008. doi: 10.1121/1.2924206.

Archontis Politis and Hannes Gamper. Comparing modeled and measurement-based spherical harmonic encoding filters for spherical microphone arrays. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 224–228, 2017. doi: 10.1109/WASPAA.2017.8170028.

Jonathan Sheaffer, Maarten van Walstijn, Boaz Rafaely, and Konrad Kowalczyk. Binaural reproduction of finite difference simulations using spherical array processing. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23(12):2125–2135, 2015.

Bård Støfringsdal and Peter Svensson. Conversion of discretely sampled sound field data to auralization formats. *JAES*, 54(5):380–400, may 2006.

Mark R. P. Thomas. Practical concentric open sphere cardioid microphone array design for higher order sound field capture. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 666–670, 2019.

Franz Zotter and Matthias Frank. *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*. Springer, Heidelberg, 2019.