Directional Enhancement of First-Order Ambisonic Room Impulse Responses By The 2+2 Directional Signal Estimator

Lukas Gölles
Graz University of Technology &
University of Music and Performing Arts
Graz, Austria
lukas.goelles@student.tugraz.at

Franz Zotter
Institute of Electronic Music and Acoustics,
University of Music and Performing Arts
Graz, Austria
zotter@iem.at

ABSTRACT

Tetrahedral microphone arrays offer a relatively efficient means for virtualization of a loudspeaker in a room. However their first-order Ambisonic room impulse responses (ARIRs) are only poorly resolved, directionally. One would typically desire the directional resolution of third-order ARIRs for binaural or fifth-order ARIRs for loudspeaker-based reproduction. The enhancement or sharpening of the directional resolution from first order to higher orders has recently been done by the Ambisonic Spatial Decomposition Method (ASDM). This contribution proposes a related method that outperforms ASDM. Instead of the pseudo-intensity vector, it uses a dual-direction detector and a 2+2 directional signal estimator that are inspired by the High Angular Plane Wave Expansion (HARPEX). After revising the geometry of the dual-direction signal model, the paper introduces a short-term averaged detector, an optimal complementary direction pair, and a regularized signal estimator. Both estimator and detector are tested in a technical analysis and the paper concludes with a comparative listening experiment indicating the advantage of the new method.

CCS CONCEPTS

• **Applied computing → Engineering**.

KEYWORDS

virtual acoustics, room impulse response, first-order Ambisonics, resolution enhancement

ACM Reference Format:
Lukas Gölles and Franz Zotter. 2020. Directional Enhancement of First-Order Ambisonic Room Impulse Responses By The 2+2 Directional Signal Estimator. In *Proceedings of the 15th International Audio Mostly Conference (AM'20), September 15–17, 2020, Graz, Austria.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3411109.3411131

1 INTRODUCTION

With the increased use of virtual reality, which it makes possible to represent rooms with VR glasses, the interest in acoustically virtualizing loudspeakers in rooms increases accordingly. Tetrahedral microphone arrays, as an example of acoustic counterparts to 3D cameras, allow to measure first-order Ambisonic room impulse responses (ARIR). However, the directional resolution of these ARIR is limited. In practice, different algorithms are therefore used to improve the directional resolution.

A common and simple algorithm, *Ambisonic Spatial Decomposition Method* (ASDM), tries to estimate the direction by the pseudo-intensity vector [1, 13, 14, 17]. Another method is *Spatial Impulse Response Rendering* (SIRR). It is a bit more complex by operating in the short-term Fourier domain, in contrast to ASDM, and it considers a decorrelated diffuse stream [10, 11].

By contrast, the simplicity in ASDM is the assumption that only one direction is present in the impulse response at a time. While this may be precise for direct sound and early reflections, the assumption is most likely violated by later parts of the reverberation that gradually become dense and finally diffuse. Interestingly, the late diffuse part is not problematic, or at least the spectral deficiency caused by the wrong assumption is easily corrected and produces consistent results [14]. Our goal here is to better represent the transition between the direct and early sounds and the diffuse late parts of the impulse response. Practice showed that when feeding the enhanced ARIR with highly impulsive sounds, and only in such cases, a temporally and directionally structured, granulated reverberation becomes audible, which is worth avoiding. For this purpose, we propose a time-domain algorithm that is able to enhance 2 detected and 2 complementary directional signals at a time, in order to exploit all of the four first-order responses.

The concept appears in the STFT-domain *High Angular Plane Wave Expansion* algorithm (HARPEX [2, 3]) that detects two directions and combines them with two complementary directions to extract 4 directional signals in the constellation of an irregular tetrahedron. We re-fine and elaborate on the concept with new ideas in this paper, to come up with a stable time-domain algorithm suitable for first-order ARIRs, as opposed to STFT-domain processing of recorded signals, for which HARPEX is designed. We propose a new geometric and short-term averaging concept of a dual-direction estimator (DDE) in this work, an optimal completion by two complementary directions ensuring a well-conditioned irregular tetrahedron beamformer/decoder to extract the four directional signals, which are subsequently upmixed into higher-order Ambisonic by re-encoding and spectral correction as in ASDM. We refer to the entire procedure as *Directional Enhancement By The 2 + 2 Directional Signal Estimator* (2DSE2).

AM'20, September 15–17, 2020, Graz, Austria

ACM ISBN 978-1-4503-7563-4/20/09…$15.00
https://doi.org/10.1145/3411109.3411131

2 METHOD

2.1 Dual Direction Estimator (DDE)

We assume a perfectly encoded first-order ARIR as B-format impulse response $\boldsymbol{b}$, assuming that their pickup patterns peak at unity (omni) or $\sqrt{3}$ (figures of eight), i.e. the signal is N3D normalized. It will turn out beneficial further below to normalize this impulse response by the omnidirectional channel w, so that the entire directional information lies compactly in the channles x, y and z,

$$\frac{\boldsymbol{b}}{w} = \frac{1}{w}\begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ \sqrt{3}\,\boldsymbol{\beta} \end{bmatrix}. \tag{1}$$

The benefit of this normalization was also discovered in the work of Daniel *et al.* [4], which came out while authoring this paper. For stable direction detection, a bandpass between 200 Hz and 4 kHz avoids noise or unequal omnidirectional vs. directional levels at low frequencies, or spatial aliasing above $kr \le 1$, with the wave number k and a tetrahedral microphone array radius of $r \approx 1.5$ cm. Every ARIR sample is modelled to consist of two signals s_1 and s_2 from the directions $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, with $||\boldsymbol{\theta}_1|| = ||\boldsymbol{\theta}_2|| = 1$, yielding the normalized ARIR

$$\frac{\boldsymbol{b}}{w} = \frac{s_1}{s_1+s_2}\begin{bmatrix} 1 \\ \sqrt{3}\,\boldsymbol{\theta}_1 \end{bmatrix} + \frac{s_2}{s_1+s_2}\begin{bmatrix} 1 \\ \sqrt{3}\,\boldsymbol{\theta}_2 \end{bmatrix}. \tag{2}$$

The two fractional scalars $\frac{s_1}{s_1+s_2}$ and $\frac{s_2}{s_1+s_2}$ sum up to unity in the first row $1 = \frac{s_1}{s_1+s_2} + \frac{s_2}{s_1+s_2}$, so that the first row becomes trivial. As we are interested in directions rather than the signals, here, we may simplify the equations of the remaining rows by replacing $\frac{s_1}{s_1+s_2} =: \gamma$, $\frac{s_2}{s_1+s_2} =: 1-\gamma$, and $\frac{1}{w}\begin{bmatrix} x & y & z \end{bmatrix}^{\mathrm{T}} =: \sqrt{3}\boldsymbol{\beta}$, after omitting the common factor $\sqrt{3}$,

$$\boldsymbol{\beta} = \gamma\,\boldsymbol{\theta}_1 + (1-\gamma)\,\boldsymbol{\theta}_2. \tag{3}$$

Obviously $\boldsymbol{\beta}$ of the model may only be a point on the line $\boldsymbol{\beta}_g$ passing through the unit vector positions $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. The line $\boldsymbol{\beta}_g$ can also be defined from two independent samples $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ in the ARIR. The model directions $\boldsymbol{\theta}_{1,2}$ are the intersections of the observed line $\boldsymbol{\beta}_g$ with the unit sphere, $\boldsymbol{\theta}_{1,2} = \left\{\boldsymbol{\beta}_g \;\middle|\; ||\boldsymbol{\beta}_g||^2 = 1\right\}$. The normalized difference of the two independent observations $\boldsymbol{v} = \frac{\Delta\boldsymbol{\beta}}{||\Delta\boldsymbol{\beta}||} = \frac{\boldsymbol{\beta}_2-\boldsymbol{\beta}_1}{||\boldsymbol{\beta}_2-\boldsymbol{\beta}_1||}$ yields a unit vector to establish the equation of the observed line $\boldsymbol{\beta}_g$,

$$\boldsymbol{\beta}_g = \boldsymbol{\beta}_1 + g\cdot\boldsymbol{v}, \qquad g \in \mathbb{R}. \tag{4}$$

To obtain a position on the line that is less signal-dependent than $\boldsymbol{\beta}_1$, we can alternatively start the line at the tangent point $\boldsymbol{\beta}_{min}$ of the smallest sphere touching the line. As the position vector $\boldsymbol{\beta}_{min}$ is a radial and therefore orthogonal to the tangent $\boldsymbol{v}$, cf. Fig. 1, we find it by Gram-Schmidt orthogonalization of $\boldsymbol{\beta}_1$ with regard to $\boldsymbol{v}$:

$$\boldsymbol{\beta}_{min} = \boldsymbol{\beta}_1 - \left(\boldsymbol{v}^T\boldsymbol{\beta}_1\right)\boldsymbol{v}. \tag{5}$$

With this point $\boldsymbol{\beta}_{min}$ the line $\boldsymbol{\beta}_g$ is redefined to

$$\boldsymbol{\beta}_g = \boldsymbol{\beta}_{min} + h\cdot\boldsymbol{v}, \qquad h \in \mathbb{R}. \tag{6}$$

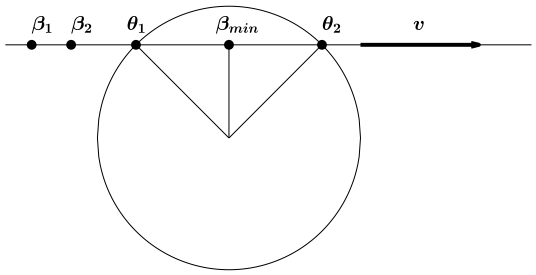

Figure 1: Constellation of $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, $\boldsymbol{v}$, $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, and $\boldsymbol{\beta}_{min}$

Finally the two directions $\boldsymbol{\theta}_{1,2}$ can be calculated as intersection of the line $\boldsymbol{\beta}_g$ with the unit sphere. We get a simple expression for the Euclidean distance between $\boldsymbol{\theta}_{1,2}$ and $\boldsymbol{\beta}_{min}$ that describes the parameter h of the model directions, i.e. $h = \pm\|\boldsymbol{\theta}_{1,2} - \boldsymbol{\beta}_{min}\|$, because the origin $\boldsymbol{0}$, the position $\boldsymbol{\beta}_{min}$, and the unit-vector directions $\boldsymbol{\theta}_{1,2}$ form a rectangular triangle,

$$h = \pm\sqrt{1 - ||\boldsymbol{\beta}_{min}||^2}, \tag{7}$$

and it yields the direction pair:

$$\boldsymbol{\theta}_{1,2} = \boldsymbol{\beta}_{min} \pm \sqrt{1 - ||\boldsymbol{\beta}_{min}||^2}\cdot\boldsymbol{v} \tag{8}$$

Short-time averaged DDE:. To get a stationary and stabilized solution for $\boldsymbol{\theta}_{1,2}$ under time-varying signals depending on the discrete-time index n, it is convenient to use averages weighted by the omnidirectional signal strength $w[n]^2$ determining both a stable position $\boldsymbol{\beta}$ on the line estimated from the instantaneous $\boldsymbol{\beta}[n]$ and a stable line vector $\Delta\boldsymbol{\beta}$ estimated from the first backwards difference $\Delta\boldsymbol{\beta}[n] = \boldsymbol{\beta}[n] - \boldsymbol{\beta}[n-1]$. This could be done within a time window from $-\frac{N}{2} \le n \le \frac{N}{2}$ involving a window function $a[n]$,

$$\overline{\boldsymbol{\beta}} = \frac{\sum_{n=-\frac{N}{2}}^{\frac{N}{2}} a[n]\,w^2[n]\,\boldsymbol{\beta}[n]}{\sum_{n=-\frac{N}{2}}^{\frac{N}{2}} a[n]\,w^2[n]}, \tag{9}$$

$$\overline{\Delta\boldsymbol{\beta}} = \frac{\sum_{n=-\frac{N}{2}}^{\frac{N}{2}} a[n]\,w^2[n]\,\boldsymbol{\beta}[n]\,\mathrm{sign}\left(\Delta\boldsymbol{\beta}_{ref}^T\Delta\boldsymbol{\beta}[n]\right)}{\sum_{n=-\frac{N}{2}}^{\frac{N}{2}} a[n]\,w^2[n]}, \tag{10}$$

where $\Delta\boldsymbol{\beta}_{ref}$ is a reference direction that can be chosen to maximize $a[n]\,w^2[n]\,\boldsymbol{\beta}[n]$ within the time window. The sign operator $\mathrm{sign}\left(\Delta\boldsymbol{\beta}_{ref}^T\Delta\boldsymbol{\beta}[n]\right)$ avoids destructive averaging over opposing directions. By replacing $\boldsymbol{\beta}$ and $\Delta\boldsymbol{\beta}$ of equation (5) by the averaged solutions,

$$\overline{\boldsymbol{\beta}_{min}} = \overline{\boldsymbol{\beta}} - \left(\boldsymbol{v}^T\overline{\boldsymbol{\beta}}\right)\boldsymbol{v} \tag{11}$$

the two directions can be calculated similarly as in equation (8),

$$\boldsymbol{\theta}_{1,2} = \overline{\boldsymbol{\beta}_{min}} \pm \sqrt{1 - ||\overline{\boldsymbol{\beta}_{min}}||^2}\cdot\boldsymbol{v}. \tag{12}$$

If the line $\boldsymbol{\beta}_g$ misses the unit sphere, the two directions are replaced by the single direction of $\overline{\boldsymbol{\beta}}$. Therefore a crossover with

$d = \frac{0.01}{\sqrt{1-||\overline{\boldsymbol{\beta}_{min}}||^2}}$ is used,

$$\boldsymbol{\theta}_{1,2} = \frac{d}{1+d} \cdot \left(\overline{\boldsymbol{\beta}_{min}} \pm \sqrt{1 - ||\overline{\boldsymbol{\beta}_{min}}||^2} \cdot \boldsymbol{v} \right) + \frac{1}{1+d} \cdot \overline{\boldsymbol{\beta}}. \quad (13)$$

2.2 Complementing to a tetrahedron

As in [2], the stable directional-signal pair $\boldsymbol{\theta}_{1,2}$ (if there are linearly independent two directions) can be complemented with a complementary direction pair $\boldsymbol{\theta}_{3,4}$ carrying diffuse signals. This is done based on the idea to form a tetrahedron, cf. figure 2. A regular tetrahedron with the directions front-left, front-right, back-up, back-down (FL, FR, BU, BD) is formalized by the 4 unit direction vectors,

$$\begin{bmatrix} \boldsymbol{FL} \\ \boldsymbol{FR} \\ \boldsymbol{BU} \\ \boldsymbol{BD} \end{bmatrix} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & +\sqrt{2} & 0 \\ 1 & -\sqrt{2} & 0 \\ -1 & 0 & +\sqrt{2} \\ -1 & 0 & -\sqrt{2} \end{bmatrix}. \quad (14)$$

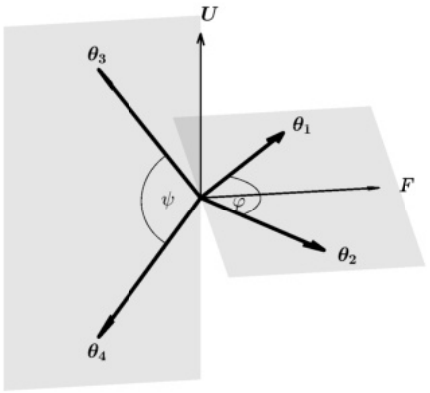

Figure 2: Regular tetrahedron prototype $\varphi = \psi = 109.5°$

Even after replacing $\boldsymbol{FL}$ and $\boldsymbol{FR}$ by two irregular positions $\boldsymbol{\theta}_{1,2}$, the frontal direction is calculated by

$$\boldsymbol{F} = \frac{\boldsymbol{\theta}_1 + \boldsymbol{\theta}_2}{||\boldsymbol{\theta}_1 + \boldsymbol{\theta}_2||}, \quad (15)$$

the vertical direction by

$$\boldsymbol{U} = \frac{\boldsymbol{\theta}_1 \times \boldsymbol{\theta}_2}{||\boldsymbol{\theta}_1 \times \boldsymbol{\theta}_2||}, \quad (16)$$

and the left direction by

$$\boldsymbol{L} = \boldsymbol{U} \times \boldsymbol{F}. \quad (17)$$

In the $\boldsymbol{FL}$, $\boldsymbol{FR}$, $\boldsymbol{BU}$, $\boldsymbol{BD}$-orientation, the front corresponds to $\boldsymbol{F}^T = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$, up to $\boldsymbol{U}^T = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$, and left to $\boldsymbol{L}^T = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$. From the matrix in equation (14), we find what also holds for arbitrary orientation: how to construct $\boldsymbol{\theta}_{3,4}$ from $\boldsymbol{F}$ and $\boldsymbol{U}$

$$\boldsymbol{\theta}_{3,4} = -\frac{1}{\sqrt{3}} \boldsymbol{F} \pm \frac{\sqrt{2}}{\sqrt{3}} \boldsymbol{U} = -\frac{1}{\sqrt{3}} \frac{\boldsymbol{\theta}_1 + \boldsymbol{\theta}_2}{||\boldsymbol{\theta}_1 + \boldsymbol{\theta}_2||} \pm \frac{\sqrt{2}}{\sqrt{3}} \frac{\boldsymbol{\theta}_1 \times \boldsymbol{\theta}_2}{||\boldsymbol{\theta}_1 \times \boldsymbol{\theta}_2||}. \quad (18)$$

For later beamforming applications it is desirable that the condition number of the matrix of spherical harmonics, evaluated at $\boldsymbol{\theta}_{1,2,3,4}$, is as low as possible and therefore it is necessary to consider the direction pair $\boldsymbol{\theta}_{3,4}$. We denote this matrix in N3D to get a unity condition number for the regular tetrahedron. For a given horizontal opening angle $\varphi = \angle(\boldsymbol{\theta}_1; \boldsymbol{\theta}_2)$, the vertical opening angle $\psi = \angle(\boldsymbol{\theta}_3; \boldsymbol{\theta}_4)$ should be adjusted accordingly. The N3D tetrahedral encoding matrix, which depends on φ and ψ, cf. Appendix A, is defined as

$$\tilde{\boldsymbol{Y}}_1 = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \cos\frac{\varphi}{2} & \cos\frac{\varphi}{2} & -\cos\frac{\psi}{2} & -\cos\frac{\psi}{2} \\ \sin\frac{\varphi}{2} & -\sin\frac{\varphi}{2} & 0 & 0 \\ 0 & 0 & \sin\frac{\psi}{2} & -\sin\frac{\psi}{2} \end{bmatrix}. \quad (19)$$

The calculation of the condition number of the matrix $\tilde{\boldsymbol{Y}}_1$ for a given horizontal opening angle $\varphi \in [0; \pi]$ for all horizontal opening angles $\psi \in [0; \pi]$ numerically obtains an ideal horizontal angle ψ with a minimum condition number. For $\varphi \leq 110°$, Appendix A derives the optimal horizontal opening angle analytically as

$$\psi_{opt} = 2 \arccos\left(\frac{1}{3\cos\frac{\varphi}{2}} \right). \quad (20)$$

For greater vertical opening angles the following approximation is sufficient and delivers a very similar condition number as the numerical optimum,

$$\psi_{approx} = \begin{cases} 2 \arccos\left(\frac{1}{3\cos\frac{\varphi}{2}} \right) & \text{if } \varphi \leq 110° \\ 26 \arctan\left(\frac{109° - \varphi}{26°} \right) + 109° & \text{else.} \end{cases} \quad (21)$$

Both cases (numerical optimum and approximation) are shown in figure 3.

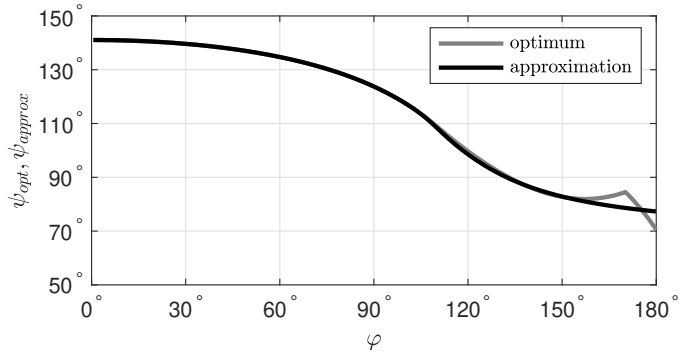

Figure 3: Optimal and approximated vertical opening angle

We denote the directions of the optimal irregular tetrahedron $\hat{\boldsymbol{T}}$ in the prototype orientation of figure 2, with the detected horizontal

angle φ and the optimal vertical angle $\psi = \psi_{approx}$, as

$$\hat{\boldsymbol{T}} = \begin{bmatrix}\hat{\boldsymbol{\theta}}_1 & \hat{\boldsymbol{\theta}}_2 & \hat{\boldsymbol{\theta}}_3 & \hat{\boldsymbol{\theta}}_4\end{bmatrix} = \begin{bmatrix}\cos\frac{\varphi}{2} & \cos\frac{\varphi}{2} & -\cos\frac{\psi}{2} & -\cos\frac{\psi}{2} \\ \sin\frac{\varphi}{2} & -\sin\frac{\varphi}{2} & 0 & 0 \\ 0 & 0 & \sin\frac{\psi}{2} & -\sin\frac{\psi}{2}\end{bmatrix}. \tag{22}$$

In the orientation with a reference frame $\boldsymbol{F}, \boldsymbol{L}, \boldsymbol{U}$ consistent with the observed directions $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ by eqs. (15), (16), (17) we obtain the four signal directions $\boldsymbol{T}$ by rotation

$$\boldsymbol{T} = \begin{bmatrix}\boldsymbol{\theta}_1 & \boldsymbol{\theta}_2 & \boldsymbol{\theta}_3 & \boldsymbol{\theta}_4\end{bmatrix} = \boldsymbol{R} \cdot \hat{\boldsymbol{T}} \tag{23}$$

$$\text{with } \boldsymbol{R} = \begin{bmatrix}\boldsymbol{F} & \boldsymbol{L} & \boldsymbol{U}\end{bmatrix} = \begin{bmatrix}F_x & L_x & U_x \\ F_y & L_y & U_y \\ F_z & L_z & U_z\end{bmatrix}. \tag{24}$$

2.3 Beamforming and Re-encoding

For re-encoding, it is necessary to extract the directional signals at the tetrahedral directions $\boldsymbol{T}$. To get the components a *Least Square Beamformer* is used, where the error signal $||\boldsymbol{e}||^2 = \boldsymbol{e}^T\boldsymbol{e} = (\boldsymbol{b} - \boldsymbol{Y}_1\boldsymbol{s})^T(\boldsymbol{b} - \boldsymbol{Y}_1\boldsymbol{s})$, with the first-order spherical harmonic matrix $\boldsymbol{Y}_1 = \begin{bmatrix}\boldsymbol{y}_1(\boldsymbol{\theta}_1) & \boldsymbol{y}_1(\boldsymbol{\theta}_2) & \boldsymbol{y}_1(\boldsymbol{\theta}_3) & \boldsymbol{y}_1(\boldsymbol{\theta}_4)\end{bmatrix}$, is minimized [7]. The solution of optimal weight $\boldsymbol{s}$ is given by a hypercardioid beamformer $\boldsymbol{Y}_1^T\boldsymbol{b}$ with subsequent directional-signal crosstalk cancellation matrix $\left[\boldsymbol{Y}_1^T\boldsymbol{Y}_1\right]^{-1}$,

$$\boldsymbol{s} = \left[\boldsymbol{Y}_1^T\boldsymbol{Y}_1\right]^{-1}\boldsymbol{Y}_1^T\boldsymbol{b}. \tag{25}$$

If the horizontal opening angle φ is small, inverting the matrix $\boldsymbol{Y}_1^T\boldsymbol{Y}_1$ gets problematic; one singular value of this matrix is $2\sin^2\frac{\varphi}{2}$ and will therefore vanish. To avoid inverting a singular matrix, a singular value decomposition of the matrix $\boldsymbol{Y}_1^T\boldsymbol{Y}_1$ is performed, $\boldsymbol{Y}_1^T\boldsymbol{Y}_1 = \boldsymbol{V}\boldsymbol{S}\boldsymbol{V}^T$. This allows to modify the diagonal singular value matrix $\boldsymbol{S}$ directly by adding a small identity matrix,

$$\tilde{\boldsymbol{S}} = (1-c)\,\boldsymbol{S} + \frac{c}{\sqrt{\pi}}\,\boldsymbol{I}, \tag{26}$$

whereby $c \in [10^{-12}; 10^{-4}]$ denotes the linear crosstalk parameter. Therefore the solution of the beamforming problem is given by

$$\boldsymbol{s} = \left[\boldsymbol{V}\,\tilde{\boldsymbol{S}}\,\boldsymbol{V}^T\right]\boldsymbol{Y}_1^T\boldsymbol{b} = \boldsymbol{V}\,\tilde{\boldsymbol{S}}^{-1}\boldsymbol{V}^T\boldsymbol{Y}_1^T\boldsymbol{b}, \tag{27}$$

and finally an upscaled ARIR of arbitrary ambisonic order N is computed by re-encoding

$$\boldsymbol{b}_N = \boldsymbol{Y}_N\,\boldsymbol{V}\,\tilde{\boldsymbol{S}}^{-1}\boldsymbol{V}^T\boldsymbol{Y}_1^T\boldsymbol{b}, \tag{28}$$

with the N^{th}-order spherical harmonic matrix $\boldsymbol{Y}_N = \begin{bmatrix}\boldsymbol{y}_N(\boldsymbol{\theta}_1) & \boldsymbol{y}_N(\boldsymbol{\theta}_2) & \boldsymbol{y}_N(\boldsymbol{\theta}_3) & \boldsymbol{y}_N(\boldsymbol{\theta}_4)\end{bmatrix}$.

To compensate for the effect that the encoding of equation (28) typically causes an unnatural high frequency boost in the later parts of reverberation, spectral correction is necessary, like it is done in [14] by ASDM.

For some applications it can be useful to apply a source widening algorithm to the diffuse parts of the impulse response. Therefore the frequency dependent z-rotation matrix $\boldsymbol{R}$ from [17] is used,

$$\boldsymbol{R}(m\hat{\phi}\cos(\omega\tau)) = \begin{bmatrix}\cos(m\hat{\phi}\cos(\omega\tau)) & \sin(m\hat{\phi}\cos(\omega\tau)) \\ -\sin(m\hat{\phi}\cos(\omega\tau)) & \cos(m\hat{\phi}\cos(\omega\tau))\end{bmatrix}, \tag{29}$$

whereby $\hat{\phi}$ is the parameter to adapt the magnitude of the frequency-varying rotation and τ is its change rate over frequency. In [15] a filter, which is used for phantom source widening, is separated in two parts,

$$H_1(\omega) = \cos(m\hat{\phi}\cos(\omega\tau)) \qquad H_2(\omega) = \sin(m\hat{\phi}\cos(\omega\tau)). \tag{30}$$

For the two parts in the time domain,

$$\begin{aligned} h_1(t) &= \sum_{q=-\infty}^{\infty} J_{|q|}(m\hat{\phi})\cos\left(\frac{\pi}{2}|q|\right)\delta(t-q\tau) \\ h_2(t) &= \sum_{q=-\infty}^{\infty} J_{|q|}(m\hat{\phi})\sin\left(\frac{\pi}{2}|q|\right)\delta(t-q\tau), \end{aligned} \tag{31}$$

with Bessel functions J_ν typically the values for q can be limited $q \in [-5;5]$. As the algorithm places successive frequencies at slightly different directions, the source width increases. In [16] it is noted that a small spectral coloration due to the restriction to the causal part $q \geq 0$ is to a great extent outweighed by its benefit of a significantly more natural sound.

3 RESULTS

3.1 Technical Analysis

For test purposes a virtual environment is introduced, where the direction of time-shifted Hann impulses is estimated by the pseudo intensity vector (PIV) and the dual-direction estimator (DDE). By means of worldmap plots, the results can be represented and matched. To compare the impulse responses of ASDM and 2DSE2, two error measures, reconstruction and energy error, are used,

$$\begin{aligned} \boldsymbol{E}_R &= 20\cdot\log_{10}\left(\|\boldsymbol{b}_{rec} - \boldsymbol{b}\|\right) \\ \boldsymbol{E}_E &= 20\cdot\log_{10}\left(\frac{\|\boldsymbol{b}_{rec}\|_F}{\|\boldsymbol{b}\|_F}\right), \end{aligned} \tag{32}$$

whereby $\boldsymbol{b}_{rec}$ denotes the reconstructed impulse response, $\boldsymbol{b}$ the original impulse response, and $\|.\|_F$ the Frobenius norm. For calculating these measured for $N > 1$, the signal components up to first-order are used for the detection and after upmixing compared to each other.

In the following test scenarios a noise of Amplitude 10^{-5} is added to the Hann impulse(s). The first impulse of Amplitude 1 starts at sample $n = 1$. The average constant of equation (10) was set to 4 and a crosstalk parameter $c = 10^{-12}$ of equation (26) was used.

In the first scenario a Hann impulse at $(-90°, 90°)$ (azimuth, zenith) should be detected by the two algorithms. The results of the detection is shown in figure 4, where the desired direction is marked as circle, the detected directions are marked as filled circles, where the amplitude defines colour and size, i.e. the larger and darker the mark, the louder the signal. For both algorithms (PIV and DDE) a stable detection of the desired directions can be noticed. For both algorithms results of the error measures can be declared as irrelevant because they are relatively low in the interesting time range (high amplitude), cf. figure 6.

If only one direction is present in the signal, the results for both direction detection algorithms in Ambisonics 5^{th} order are similar. To show the limits of detection by the PIV, two Hann impulses with amplitudes 1 and 0.75 and directions $\boldsymbol{\theta}_1 = (-90°, 90°)$ and $\boldsymbol{\theta}_2 = (45°, 45°)$ are used, and the second impulse starts 4 samples

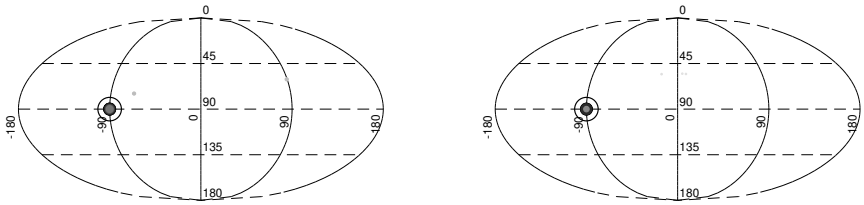

Figure 4: Detected directions for one encoded direction (marked as circle) with PIV (left) and DDE (right)

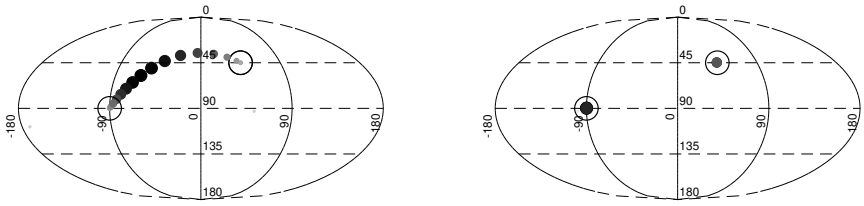

Figure 5: Detected directions for two encoded directions (marked as circles) with PIV (left) and DDE (right)

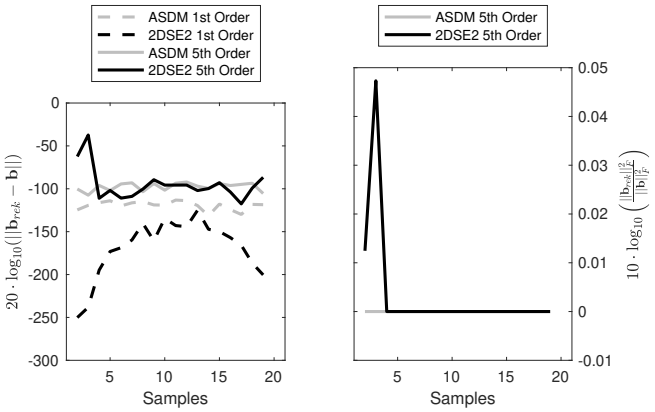

Figure 6: Error measures of ASDM and 2DSE2 for Ambisonics Order 1 and 5 and one direction

later. In figure 5 it turns out that the PIV is influenced by the second direction, while the DDE identifies two directions correctly. In terms of the reconstruction error 2DSE2 gives better results in parts of high amplitudes, cf. figure 8. In terms of the energy error 2DSE2 causes the smaller error here, which is negligibly small, in contrast to PIV.

When adding a third direction and impulse at (100°, 135°) with the amplitude 0.25 and an additional delay of 4 samples to the second impulse, also the results for DDE begin to drift. However, in contrast to the floating direction detected by PIV, DDE keeps identifying the stronger signal direction correctly, it appears. By contrast, with PIV also the detection of the strongest direction is affected by the two other, weaker directional signals. The DDE apparently indicates a compromise between two weaker (second and third) Hann impulses, cf. figure 7. The reconstruction error, cf. figure 9, is significantly lower in the case of re-encoding into 1st order for the 2DSE2, with 5th order mainly the 2DSE2 also delivers better results. It can be determined that the reconstruction error can be kept reasonably low by the 2DSE2 as long as 2 directions are present in the signal. If a third is added, the limits are reached for the 2DSE2. Looking at the energetic error, it can also be seen that the 2DSE2 causes the smaller error. It should be noted that the errors for three concurrent sounds are significantly higher than those for two.

3.2 Listening experiment

Preparation

A listening experiment was conducted to underline the theoretical results. A design like a Multi-Stimulus Test with Hidden Reference and Anchor (MUSHRA) [8] for the trials appears useful.

In order to obtain comparable results, like in [13], two different signals were used, music and speech. An excerpt from *What's Trumps* (0:38-0:42, composer: Lukas Lohner, played by the *RhythmusSport-Gruppe*) [5] was selected as music signal; an excerpt from EBU's female English speech reference recording [6] as speech signal. To use two different rooms with different room acoustical properties, receiver and source directional impulse response measurements from the György-Ligeti Hall (reverberation time: 1.1 s) of the University of Music and Performing Arts Graz and the Dom im Berg Graz (reverberation time: 0.65 s) were used. These room impulse responses were calculated from sweep responses recorded with the

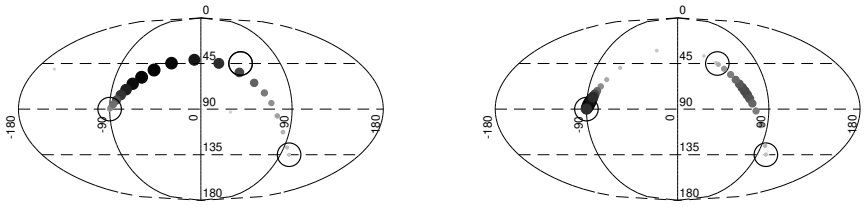

Figure 7: Detected directions for three encoded directions (marked as circles) with PIV (left) and DDE (right)

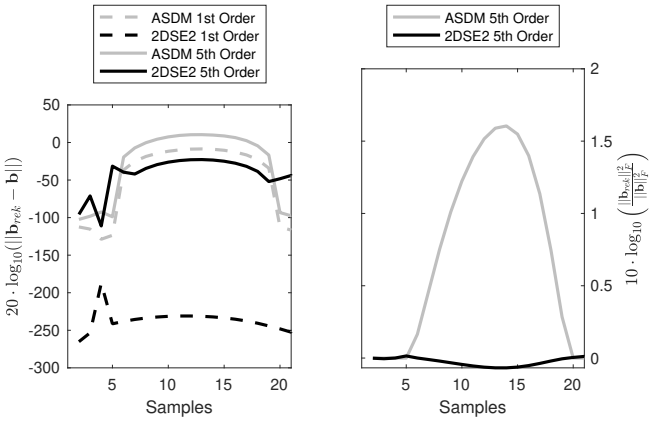

Figure 8: Error measures of ASDM and 2DSE2 for Ambisonics Order 1 and 5 for two shifted directions

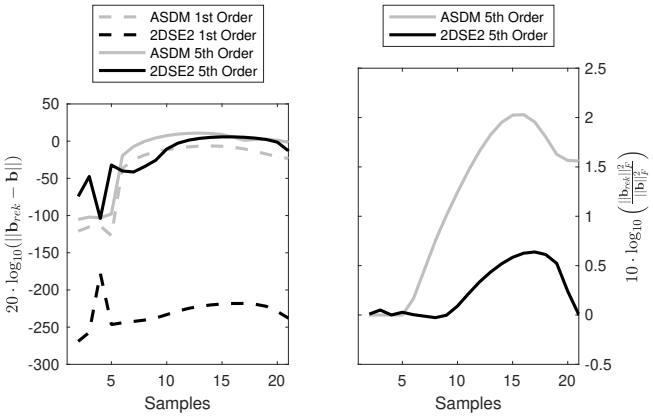

Figure 9: Error measures of ASDM and 2DSE2 for Ambisonics Order 1 and 5 for three directions

Soundfield ST450 microphone.
In both rooms Ambisonic loudspeaker systems are installed, from which the listening experiment only uses the frontal loudspeakers. This choice was made as their binaural reproduction appears to differ most clearly from a dummy-head reference. In György-Ligeti Hall 3 loudspeakers, and in Dom im Berg 5 loudspeakers are driven by a signal for binaural virtualization. Geometry and arrangement of the afore-mentioned loudspeakers are shown in figure 10.

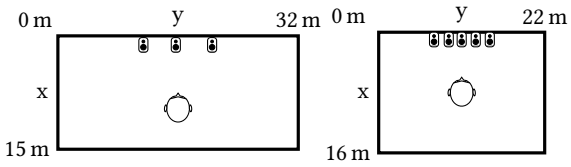

Figure 10: Geometry, listener and loudspeaker position(s) in virtual György-Ligeti Hall (left) and Dom im Berg (right).

The first-order Ambisonic room impulse responses (ARIR1) were processed with two different algorithms (ASDM and 2DSE2), each with and without phantom source widening in later parts of reverberation, so that 5$^{\text{th}}$ order Ambisonic room impulse responses could be used for the listening experiment. Impulse Response measurements with the dummy head, *Neumann KU100* served as a reference, and were only available for trials in the virtual György-Ligeti Hall. All stimuli were statically rendered binaural audio signals.
`StereoEncoder`, `AllRADDecoder` and `BinauralDecoder` from the IEM Plugin-Suite [12] were used for encoding and decoding, while Matthias Kronlachners `mcfx_convolver_36` from his mcfx-Plugin-Suite [9] was used for convolving the signal with the ARIRs.

György-Ligeti Hall
In the first room, György-Ligeti Hall, listeners were asked to rate the naturalness of the music binaural signals on a scale from 0 (unnatural) to 10 (very natural). The next step was to compare the algorithms with the dummy head recording on a scale from 0 (different) to 10 (identical). The dummy head recording was also hidden in the examples to be assessed, in order to enable a subsequent assessment of the reliability of the listeners. The trials with speech stimuli were done in the same way. In addition, a fingersnaping signal was used as test signal for the György-Ligeti Hall to evaluate the performance with transient sounds, comparing the diffuse reverberation of the different algorithms with the dummy head.

Dom im Berg
Because the measurement with the dummy head was not available for the Dom im Berg, the trials in this virtual scene were limited to assessing the naturalness of the stimuli rendered via ASDM and

2DSE2. Again, listeners were asked to rate music and speech on a scale of 0 (unnatural) to 10 (natural). As the differences between the algorithms are small for transient signals in an acoustically dry environment, the Dom im Berg trials excluded the finger-snapping signal and only consisted of the speech and music trials about naturalness.

Results

12 male listeners took part in the experiment, at the average age of 29 years, and the average time required was 12 min 29 s. Figure 12 shows the result of the listening experiment, rating the naturalness, for the György-Ligeti Hall. The dummy head stimuli were rated most naturally for all three signals (music, speech and snapping). As expected, the 2DSE2 delivers better results than the ASDM in all three cases. For music and speech, the widening for late and diffuse ARIR parts did not make any obvious difference. Not only considering the median confirms this, but also results of a sign-rank test with Bonferroni correction indicates: There are no significant differences ($p > 2.5$) for the ASDM and ASDM widened or 2DSE2 and 2DSE2 widened for music and speech, while regardless of widening, ASDM and 2DSE2 exhibit significant differences ($p < 0.02$). For the finger-snapping signal, widening of the later ARIR parts significantly improves the results ($p < 0.01$).

Figure 11 shows the results for the assessment of the similarity to the dummy head. The difference between the algorithms ASDM and 2DSE2 is not as distinct as before. Nevertheless, 2DSE2 is classified as more similar to the dummy head than ASDM. One reason for this result is that the timbre between ASDM/2DSE2 and the dummy head is not identical. It was also noted that the spatial auditory image of the dummy head always contained the least disturbing amount of lateral signal components. Some listeners remarked that they had heard the most of such lateral disturbance in what they rated worst. On the other hand, clear results can be shown for Dom im Berg, cf. figure 13. While the 2DSE2 was rated quite natural, ASDM was often rated on the lower end of the scale. Again, the results show that source widening in the later parts of the ARIR yields no advantage with music and speech signals. Between the ASDM and 2DSE2 stimuli with widening or without, the sign-rank test accordingly indicates that there are no significant differences ($p > 2.3$), while the the two methods ASDM and 2DSE2 are significantly different from each another ($p < 0.01$), regardless of employing widening or not.

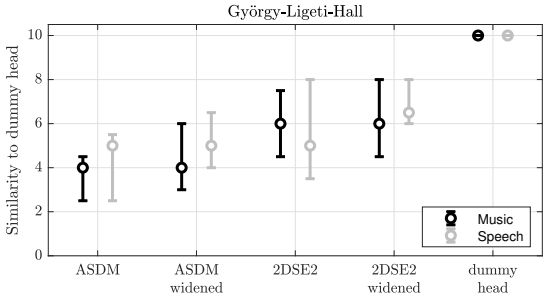

Figure 11: Result: Rate of similarity to the dummy head for the György-Ligeti Hall (authenticity)

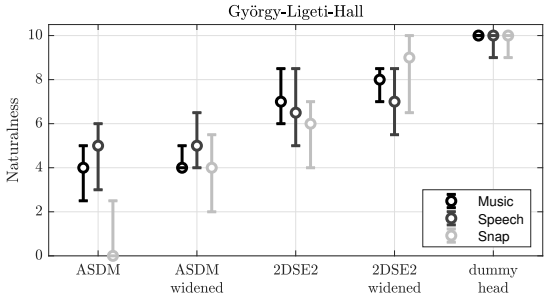

Figure 12: Result: Rate of naturalness for the György-Ligeti Hall (plausibility)

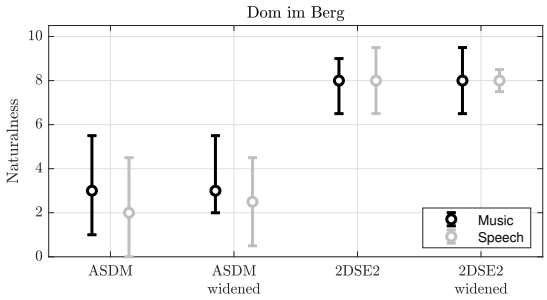

Figure 13: Result: Rate of naturalness for the Dom im Berg (plausibility)

4 CONCLUSION

In this paper we could present and test a new method for directional enhancement of first-order Ambisonic room impulse responses. A technical comparison to the *Ambisonic Spatial Decomposition Method* and the pseudo-intensity vector as its direction estimator shows the advantages of the dual-direction estimator, especially when detecting two directions or more. By two error measures, reconstruction and energy error, the basic technical properties of both directional enhancement algorithms could be compared to each other.

For comparison of their perceptual implications, we undertook comparative listening experiment according to which the directional ARIR enhancement of 2DSE2 yields a significantly more natural/plausible result than ASDM for both György-Ligeti Hall and Dom im Berg ARIRs. For transient sounds, widening of the later ARIR parts significantly improves the results, but could be shown to be ineffective for music or speech.

With the results of the technical study and the comparative listening experiment, we conclude that 2DSE2 is preferable over ASDM as directional enhancement method for first-order Ambisonic room impulse responses.

To support reproducible research, the underlying ARIRs, the 2DSE2/ASDM-upmixed ARIRs, experimental stimuli and responses are made available under https://phaidra.kug.ac.at/o:107155.

We gratefully thank the listeners of our experimental study for their voluntary participation.

REFERENCES

[1] J. Ahrens. 2019. Auralization of Omnidirectional Room Impulse Responses Based on the Spatial Decomposition Method and Synthetic Spatial Data. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 146–150.

[2] Natasha Barrett and Svein Berge. 2010. A New Method for B-Format to Binaural Transcoding. In *Audio Engineering Society Conference: 40th International Conference: Spatial Audio: Sense the Sound of Space*. http://www.aes.org/e-lib/browse.cfm?elib=15527

[3] Svein Berge and Natasha Barrett. 2010. High Angular Plane Wave Expansion. In *Proc. of the 2nd International Symposium on Ambisonics and Spherical Acoustics*.

[4] Jérôme Daniel and Srđan Kitić. 2020. Time Domain Velocity Vector for Retracing the Multipath Propagation. In *Proc. ICASSP*. Barcelona.

[5] Deutsche Gesellschaft für Akustik. 2020. Stimulus-Datenbank: Jazz/Funk Rhythmussportgruppe: What's Trumps. https://www.dega-akustik.de/fachausschuesse/va/stimulus-datenbank/, https://doi.org/10.5281/zenodo.3601032, Accessed 2020-04-16.

[6] EBU. 2008. EBU SQAM CD: Sound Quality Assessment Material recordings for subjective tests. (2008). https://tech.ebu.ch/news/ebu-cds-now-online-31oct08.

[7] Philippe-Aubert Gauthier, Eric Chambatte, Cedric Camier, Yann Pasco, and Alain Berry. 2011. Derivation of Ambisonics Signals and Plane Wave Description of Measured Sound Field Using Irregular Microphone Arrays and Inverse Problem Theory.

[8] International Telecomunication Union. 2015. Recommendation ITU-R BS.1534-3, Method for the subjective assessment of intermediate quality level of audio systems. (10 2015).

[9] Matthias Kronlachner. 2020. mcfx Plugin Suite. http://www.matthiaskronlachner.com/?p=1910 Accessed 2020-04-16.

[10] Juha Merimaa and Ville Pulkki. 2005. Spatial Impulse Response Rendering I: Analysis and Synthesis. *J. Audio Eng. Soc* 53, 12 (2005), 1115–1127. http://www.aes.org/e-lib/browse.cfm?elib=13401

[11] Ville Pulkki and Juha Merimaa. 2006. Spatial Impulse Response Rendering II: Reproduction of Diffuse Sound and Listening Tests. *J. Audio Eng. Soc* 54, 1/2 (2006), 3–20. http://www.aes.org/e-lib/browse.cfm?elib=13664

[12] Daniel Rudrich. 2020. IEM Plugin Suite. https://plugins.iem.at Accessed 2020-04-16.

[13] Sakari Tervo, Jukka Pätynen, Antti Kuusinen, and Tapio Lokki. 2013. Spatial Decomposition Method for Room Impulse Responses. *J. Audio Eng. Soc* 61, 1/2 (2013), 17–28. http://www.aes.org/e-lib/browse.cfm?elib=16664

[14] Markus Zaunschirm, Matthias Frank, and Franz Zotter. 2020. Binaural Rendering with Measured Room Responses: First-Order Ambisonic Microphone vs. Dummy Head. *Applied Sciences* 10, 5 (Feb 2020), 1631. https://doi.org/10.3390/app10051631

[15] Franz Zotter and Matthias Frank. 2013. Efficient Phantom Source Widening. *Archives of Acoustics* 38 (03 2013), 27–37. https://doi.org/10.2478/aoa-2013-0004

[16] Franz Zotter and Matthias Frank. 2017. Phantom Source Widening by Filtered Sound Objects. In *Audio Engineering Society Convention 142*. http://www.aes.org/e-lib/browse.cfm?elib=18605

[17] Franz Zotter and Matthias Frank. 2019. *Ambisonics - A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality* (first edition ed.). Springer, Heidelberg (Deutschland).

A OPTIMAL CONDITION NUMBER

The condition of the first-order tetrahedron encoding matrix $\boldsymbol{Y}_1$ is given by,

$$\operatorname{cond}(\boldsymbol{Y}_1) = \frac{\max \boldsymbol{\sigma}_{\boldsymbol{Y}_1}}{\min \boldsymbol{\sigma}_{\boldsymbol{Y}_1}}, \tag{33}$$

where $\boldsymbol{\sigma}_{\boldsymbol{Y}_1}$ denotes the singular value vector of $\boldsymbol{Y}_1$. By taking the square of the transposed matrix, two squared singular values are given,

$$\tilde{\boldsymbol{Y}}_1\tilde{\boldsymbol{Y}}_1^T = \begin{bmatrix} 4a^2 & 2ac_1 - 2ac_2 & 0 & 0 \\ 2ac_1 - 2ac_2 & 2c_1^2 + 2c_2^2 & 0 & 0 \\ 0 & 0 & 2s_1^2 & 0 \\ 0 & 0 & 0 & 2s_2^2 \end{bmatrix}, \tag{34}$$

whereby $a = \frac{1}{\sqrt{3}}, c_1 = \cos\frac{\varphi}{2}, s_1 = \sin\frac{\varphi}{2}, c_2 = \cos\frac{\psi}{2}$ and $s_2 = \sin\frac{\psi}{2}$, and a eigenvalue problem of the left upper 2×2 Matrix remains,

$$\begin{aligned} (4a^2 - x)(2c_1^2 + 2c_2^2 - x) - (2ac_1 - 2ac_2)^2 &= 0 \\ \sqrt{4a^4 - 8a^2c_1c_2 + c_1^4 + 2c_1^2c_2^2 + c_2^4} &= x_{1,2}\,. \end{aligned} \tag{35}$$

Therefore the singular values are given by

$$\boldsymbol{\sigma} = \begin{bmatrix} 2a^2 + c_1^2 + c_2^2 - \sqrt{4a^4 - 8a^2c_1c_2 + c_1^4 + 2c_1^2c_2^2 + c_2^4} \\ 2a^2 + c_1^2 + c_2^2 + \sqrt{4a^4 - 8a^2c_1c_2 + c_1^4 + 2c_1^2c_2^2 + c_2^4} \\ 2s_1^2 \\ 2s_2^2 \end{bmatrix}. \tag{36}$$

To calculate the optimal vertical opening angle in dependence to the horizontal opening angle, the condition number shall be minimized. For opening angles $\psi \leq 110°$ the condition number is given by $\frac{\sigma_2}{\sigma_3}$,

$$\operatorname{cond}(\tilde{Y}_1\tilde{Y}_1^T) = \frac{2a^2 + c_1^2 + c_2^2 + \sqrt{4a^4 - 8a^2c_1c_2 + c_1^4 + 2c_1^2c_2^2 + c_2^4}}{2s_1^2}. \tag{37}$$

To get the minimal condition number, the expression in the numerator has to be minimized,

$$\begin{aligned} x &= 2a^2 + c_1^2 + c_2^2 + \sqrt{4a^4 - 8a^2c_1c_2 + 2c_1^2c_2^2 + c_1^4 + c_2^4} \\ \frac{dx}{d\psi} &= -\sin\psi\left[2c_2 + \frac{-8a^2c_1 + 4c_1^2c_2 + 4c_2^3}{2\sqrt{4a^4 - 8a^2c_1c_2 + 2c_1^2c_2^2 + c_1^4 + c_2^4}}\right] = 0\,. \end{aligned} \tag{38}$$

The non-trivial solution is considered,

$$\begin{aligned} 2c_2 + \frac{-8a^2c_1 + 4c_1^2c_2 + 4c_2^3}{2\sqrt{4a^4 - 8a^2c_1c_2 + 2c_1^2c_2^2 + c_1^4 + c_2^4}} &= 0 \\ c_2\sqrt{4a^4 - 8a^2c_1c_2 + 2c_1^2c_2^2 + c_1^4 + c_2^4} &= 2a^2c_1 - c_1^2c_2 - c_2^3 \\ a^2c_2^2 - c_1c_2^3 + c_1^3c_2 - a^2c_1^2 &= 0\,. \end{aligned} \tag{39}$$

It is obvious that a solution to the equation is $c_2 = c_1$, i.e. $\varphi = \psi$ and a polynomial division is done,

$$\begin{array}{rl} (-c_1c_2^3 + a^2c_2^2 + c_1^3c_2 + a^2c_1^2) : (c_2 - c_1) = & -c_1c_2^2 \quad + (a^2 - c_1^2)c_2 \quad + a^2c_1\,. \\ \underline{\qquad -c_1c_2^3 + c_1^2c_2^2} & \\ (a^2 - c_1^2)c_2^2 + c_1^3c_2 + a^2c_1^2 & \\ \underline{(a^2 - c_1^2)c_2^2 - (a^2 - c_1^2)c_1c_2} & \\ a^2c_1c_2 - a^2c_1^2 & \end{array} \tag{40}$$

This quadratic equation $c_2^2 + \frac{c_1^2 - a^2}{c_1}c_2 - a^2 = 0$ is solved by the quadratic formula,

$$\begin{aligned} c_2 &= \frac{a^2 - c_1^2}{2c_1} \pm \sqrt{\frac{a^4 - 2a^2c_1^2 + c_1^4 + 4a^2c_1^2}{4c_1^2}} \\ &= \frac{a^2 - c_1^2}{2c_1} \pm \sqrt{\frac{(a^2 + c_1)^2}{(2c_1)^2}} = \left\{\frac{a^2}{c_1}, -c_1\right\}. \end{aligned} \tag{41}$$

The positive solution $c_2 = \frac{a^2}{c_1}$ is considered,

$$\psi_{opt} = 2\arccos\left(\frac{a^2}{\cos\frac{\varphi}{2}}\right) \qquad \text{for } \varphi \leq 110°. \tag{42}$$