EFFICIENT PHANTOM SOURCE WIDENING AND DIFFUSENESS IN AMBISONICS

Franz Zotter, Matthias Frank

Institute of Electronic Music and Acoustics
Univ. of Music and Performing Arts Graz
Austria, {zotter, frank}@iem.at

Matthias Kronlachner,

Univ. of Music and Performing Arts Graz
Austria, m.kronlachner@gmail.com
http://matthiaskronlachner.com

Jung-Woo Choi,

Dept. of Mechanical Engineering, KAIST
Korea Advanced Institute of Science and Technology
khepera@kaist.ac.kr

ABSTRACT

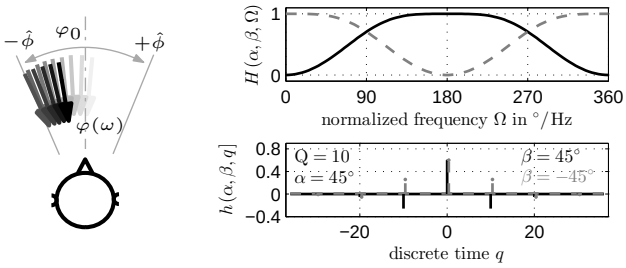
Object-based spatial audio considers virtual sound sources having a width/diffuseness parameter. This parameter aims at controlling the perceived width or diffuseness of the auditory object, or phantom source, created by the renderer. Width/diffuseness provides an important salience parameter that is independent of perceived direction and timbre. A highly efficient sparse filter structure for two-channel stereophony was described and tested recently, but it becomes ineffective for most parts of a large audience. This paper presents phantom source width/diffuseness control for Ambisonics. The new approach is a remarkably elegant application of the previously described stereo phantom source widening on Ambisonics. Compared with former experimental data, our experiments show a greater freedom of increasing the width and widening that works for a larger listening area.

1. INTRODUCTION

Special filter structures for increasing the width of two-channel stereophonic phantom sources have been discussed since the nineteen fifties, e.g. [1, 2, 3]. More general frequency domain fast block filter all-pass designs with random phase were most successful for phantom source widening, [4, 5, 6], until recently a special sparse FIR phase-based and amplitude-based filter structure could be proposed [7, 8]. However, experiment and practice show that two-channel stereo widening or diffuseness degrades for more than 30cm laterally shifted listening positions. Moreover, widening/diffuseness does not exceed the loudspeaker pair.

For phantom source widening on many loudspeakers, vector-base amplitude panning (VBAP) using a frequency-dependent panning angle was presented in [9]. The effectiveness of the approach could be heard at audio engineering workshops[1]. The practical implementation employs fast block filters that represent the frequency-dependent panning. However, the steep stop band transitions of these filters aren't continuously differentiable, hence careful FIR approximation becomes necessary.

Recent work [10] lets us expect alternative, simple filters in Ambisonics, but we currently lack suitable algorithms. We therefore consider Ambisonic encoding of a plane wave, cf. Fig. 1(a),



(a) Basic idea. (b) $H(\alpha, \beta, \Omega)$ has a sparse time response.

Figure 1: Sound with frequency-dispersed arrival direction (a) is perceived as widened. Can it be created by the sparse filter in (b)?

$$p(\omega, \boldsymbol{r}) = e^{\mathrm{i}k\,\boldsymbol{s}^{\mathrm{T}}\boldsymbol{r}} \tag{1}$$

whose direction of arrival (DOA) shall be frequency-dependent $\boldsymbol{s} = \boldsymbol{s}(\omega)$. k is the wave number ω/c, ω is 2π times frequency f, c is the speed of sound, $\boldsymbol{s}$ is a direction vector $\|\boldsymbol{s}\| = 1$, and $\boldsymbol{r}$ is a point of observation. Despite frequency dependency disperses the DOA, the sound pressure stays perfectly unitary $|p(\omega, \boldsymbol{r})| = 1$.

The Fourier transform pair of a phase modulated cosine [11] yields a sparse time response when used to define a time-invariant frequency response, cf. appendix, with the normalized frequency $\Omega = \omega\mathrm{T}$,

$$H(\alpha, \beta, \Omega) = \cos[\alpha\,\cos(\Omega) + \beta], \tag{2}$$
$$\overset{\mathcal{IFT}}{\longrightarrow} h(\alpha, \beta, t) = \sum_{\lambda=-\infty}^{\infty} \cos(\tfrac{\pi}{2}\,|\lambda| + \beta)\, J_{|\lambda|}(\alpha)\, \delta(t + \lambda).$$

Its simplest discrete-time implementation $h(\alpha, \beta, q]$, cf. Fig. 1(b), uses the integer sample index q, and is nonzero only at $q = \lambda\mathrm{Q}$,

$$h(\alpha, \beta, \lambda\mathrm{Q}] = \cos(\tfrac{\pi}{2}|\lambda| + \beta)\, J_{|\lambda|}(\alpha), \tag{3}$$

with $\lambda, q \in \mathbb{Z}$, $\mathrm{Q} \in \mathbb{N}$. For small α the filter is truncated in $|\lambda|$ and implemented efficiently as sparse FIR filter. Such filters were employed in [8] for two-channel stereo phantom source widening.

[1] workshop by Ville Pulkki, "Time-frequency processing of spatial audio", presented at AES 128th Conv., London; AES 40th Int. Conf. on Spatial Audio, Tokyo; AES 129th Conv., San Francisco; EAA Winter School Cutting Edge in Spatial Audio, Merano 2013

Assuming a dispersion constant $\hat{\phi}$ and an average DOA φ_0, this filter might intuitively provide a frequency-dispersed DOA

$$\boldsymbol{s}(\Omega) = \begin{pmatrix} \cos[\hat{\phi}\cos(\Omega) + \varphi_0] \\ \sin[\hat{\phi}\cos(\Omega) + \varphi_0] \end{pmatrix}. \tag{4}$$

Based on this, an Ambisonic representation of a dispersive DOA encoding system is developed in section two. The third section describes a listening experiment to clarify whether the algorithm widens the phantom source at three different listening positions more effectively than a comparable two-channel stereo implementation. After discussing how good results of the perceptual study correlate with the $\mathrm{IACC_{E3}}$ measure, section four highlights generalization of the Ambisonic widening algorithm. Non-sinusoidal DOA dispersion curves might be favorable for Ambisonic orders higher than 3 to avoid DOA accumulation points. Moreover, to widen/diffuse entire Ambisonic productions, rotation matrices rather than encoders are defined. For 3D Ambisonics, spherical-cap-filling and isotropic widening is briefly sketched.

2. FREQUENCY-VARYING AMBISONIC ENCODER

The incoming field p is classically represented in 2D Ambisonics by the encoding coefficients b_m that drive the equation

$$p(\omega, r, \varphi) = \underbrace{\sum_{m=-\infty}^{\infty} \mathrm{i}^m \, J_m(kr)\, \Phi_m(\varphi)}_{\text{represented by playback system}} \; b_m(\varphi_\mathrm{s}), \tag{5}$$

with the orthonormal circular harmonics

$$\Phi_m(\varphi) = \sqrt{\tfrac{2-\delta_m}{2\pi}} \begin{cases} \cos(m\varphi), & \text{for } m \geq 0, \\ -\sin(m\varphi), & \text{for } m < 0. \end{cases} \tag{6}$$

A plane-wave field arriving from the direction of the polar angle φ_0 would just be encoded by a set of frequency-independent scalars $b_m(\varphi_0) = \Phi_m(\varphi_0)$. A DOA-dispersed plane wave around φ_0 requires encoding by $\Phi_m[\hat{\phi}\,\cos(\Omega) + \varphi_0]$ and hence responses are related to Eq. (2) by $\alpha = m\hat{\phi}$ and $\beta = m\varphi_0$ or $\beta = m\varphi_0 + \pi/2$,

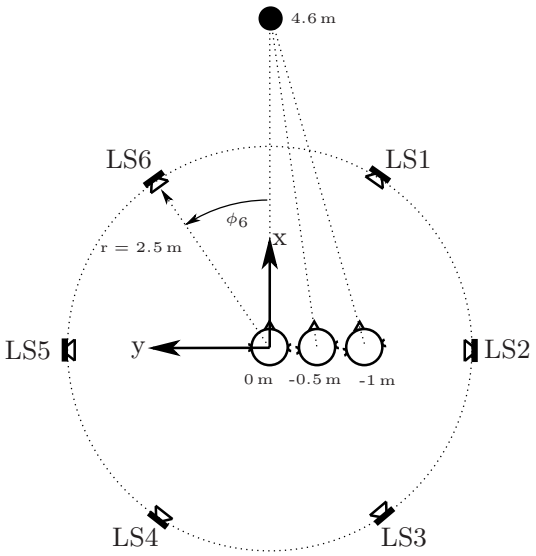
$$b_m(\Omega) = \sqrt{\tfrac{2-\delta_m}{2\pi}} \begin{cases} \cos[m\hat{\phi}\cos(\Omega) + m\varphi_0], & m \geq 0, \\ -\sin[m\hat{\phi}\cos(\Omega) + m\varphi_0], & m < 0. \end{cases}$$

Correspondingly, an encoding system is defined by the transform pair in Eq. (2) and permits efficient sparse FIR implementation using Eq. (3). Herewith, impulse responses $b_m[q]$ of Ambisonic encoding with dispersed DOA in the inverval $[\varphi_0 - \hat{\phi},\, \varphi_0 + \hat{\phi}]$ become nonzero only at $\lambda\mathrm{Q}$. For causal and finite responses, λ is shifted by a causality-bringing truncation offset $\Lambda \in \mathbb{N}$, and it is truncated to $0 \leq \lambda \leq 2\Lambda$. The $2\Lambda + 1$ nonzero entries of $b_m[q]$ are defined by Eq. (3) and lie in the discrete-time inverval $q \in [0,\, 2\Lambda\mathrm{Q}]$,

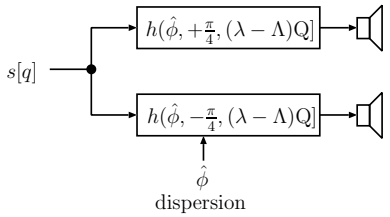
$$b_m[\lambda\mathrm{Q}] = \sqrt{\tfrac{2-\delta_m}{2\pi}} \begin{cases} h(m\hat{\phi}, m\varphi_0, (\lambda-\Lambda)\mathrm{Q}], & m \geq 0, \\ h(m\hat{\phi}, m\varphi_0 + \pi/2, (\lambda-\Lambda)\mathrm{Q}]. & m < 0. \end{cases} \tag{7}$$

Discrete-time Ambisonic signals $\chi_m[q]$ encoding a sound signal $s[q]$ are obtained by convolution with this single-input-multiple-output (SIMO) system of impulse responses

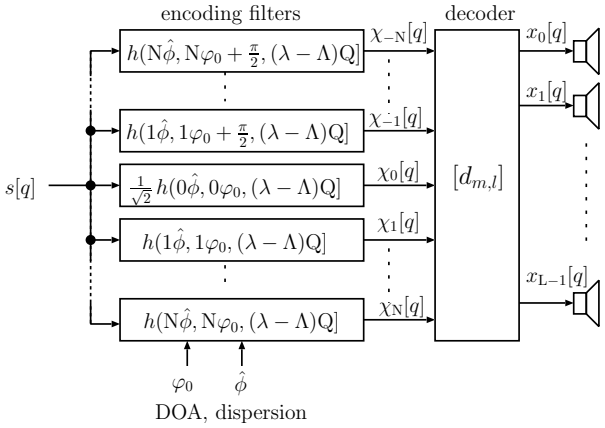
$$\chi_m[q] = s[q] \star b_m[q]. \tag{8}$$



(a) Setup of the listening experiment.



(b) Stereo widening around $\varphi_0 = 0$.



(c) Proposed Ambisonic widening around φ_0.

Figure 2: Experimental setup and signal processing schemes for testing widened perceived (phantom) sources with dispersion $\hat{\phi}$. Stereo widening is used to supply LS6 and LS1, and Ambisonic widening supplies LS1…LS6 and uses N=2.

The experimental evaluation that follows uses second-order Ambisonics, i.e., it restricts m to $|m| \leq \mathrm{N}$ with $\mathrm{N} = 2$. We chose the truncation offset $\Lambda = 9$ and a time grid of $\mathrm{Q} = 110$ samples between nonzero entries, i.e. $\mathrm{T} \approx 2.5\,\mathrm{ms}$ at the sample rate $f_\mathrm{s} = 44.1\,\mathrm{kHz}$.

3. EXPERIMENT

The listening experiment evaluates the perceived source extent of the Ambisonic widening and compares it to the stereo widening. The experiment hereby investigates the controllability of the widening effect at different listening positions.

3.1. Setup and Conditions

Fig. 2(a) shows the experimental setup consisting of a hexagonal ring of 6 Genelec 8020 loudspeakers at a radius of $\mathrm{r} = 2.5\,\mathrm{m}$. The hexagonal layout is suitable for second order Ambisonics. As it contains a loudspeaker pair at $\pm 30^\circ$, comparison to a two-channel stereo widening algorithm is possible. The height of all loudspeakers was set to $1.2\,\mathrm{m}$ which is also the ear height of the subjects. The experiment was performed in the IEM CUBE, a $10.3\,\mathrm{m} \times 12\,\mathrm{m} \times 4.8\,\mathrm{m}$ large room with a mean reverberation time of $470\,\mathrm{ms}$ that fulfills the recommendation for surround reproduction in ITU-R BS.1116-1 [12]. The central listening position lies within the effective critical distance. The subjects were asked to fix a reference point at $4.6\,\mathrm{m}$ in the frontal direction at all listening positions.

For the regular hexagonal setup, an optimal $\max\text{-}\boldsymbol{r}_\mathrm{E}$ Ambisonic decoder is obtained by $d_{lm} = \cos[\frac{m\pi}{2(\mathrm{N}+1)}]\,\Phi_m[-\frac{\pi\,(1+2l)}{6}]$, cf. [13], to get six loudspeaker signals from five Ambisonic signals $\chi_m(t)$

$$x_l[q] = \sum_{m=-2}^{2} d_{lm}\;\chi_m[q], \quad \forall l = 0\ldots 5. \tag{9}$$

Dispersive encoding Eq. (7) is used with $\mathrm{Q} = 110$ and $\Lambda = 9$ ($0 \le \lambda \le 2\Lambda$). The DOA dispersion $\hat{\phi}$ is investigated for $\varphi_0 = 0$.

As a reference, an amplitude-based two-channel stereo widener was tested using filters derived from Eq. (3) to generate signals for loudspeaker one and six, aiming at widened frontal sound, cf. [8],

$$x_{6,1}[q] = s[q] \star h(\hat{\phi}, \pm\tfrac{\pi}{4}, (\lambda - \Lambda)\mathrm{Q}], \tag{10}$$

also using $\mathrm{Q} = 110$, $\Lambda = 9$ ($0 \le \lambda \le 2\Lambda$), and a varying $\hat{\phi}$.

The algorithms were fed with $22\,\mathrm{s}$ of English speech from EBU SQAM CD [14] and adjusted to 65dB(A). Speech was less sensitive to subtle perceivable differences that do not affect the perceived spatial extent but could indirectly distort the results.

Table 1: *Stereophonic and Ambisonic conditions of the listening experiment share cond. 1 and were adjusted to cover similar widths.*

cond.	stereo	Ambisonics
1	$\hat{\phi} = 0^\circ$	$\hat{\phi} = 0^\circ$
2	$\hat{\phi} = 30^\circ$	-
3	$\hat{\phi} = 50^\circ$	-
4	$\hat{\phi} = 70^\circ$	-
5	$\hat{\phi} = 90^\circ$	-
6	-	$\hat{\phi} = 20^\circ$
7	-	$\hat{\phi} = 35^\circ$
8	-	$\hat{\phi} = 47^\circ$
9	-	$\hat{\phi} = 65^\circ$

For both stereo and Ambisonic algorithms, 5 different conditions with varying dispersion $\hat{\phi}$ were tested, cf. Tab. 1. Dispersion values of stereo conditions were chosen to be comparable to known data on phase-based stereo widening [7, 8] that used the same input speech signal. Dispersion values of the Ambisonic conditions were adjusted in a preliminary test at the central listening position as to obtain a perceived extent similar to the extent of the stereophonic conditions. The zero-dispersion condition $\hat{\phi} = 0^\circ$ produces exactly the same loudspeaker signals for stereo and Ambisonics on the given setup. Hence this condition was only tested once, cf. condition 1 in Tab. 1. Each subject completed 6 MUSHRA-like comparisons, each of which displaying the 9 conditions in random order. In this way, each subject gave a full comparison twice at each of the three listening positions.

In the MUSHRA-like comparison, subjects were allowed to seamlessly switch between the playback of the 9 conditions. The user interface was placed on their lap. In each comparison, subjects rated the perceived spatial extent of the 9 selectable stimuli on 9 corresponding quasi-continuous sliders from "narrow" to "wide" (the test used the German terms „schmal" and „breit"). To support subjects in organizing their ratings, they were provided a button to re-arrange the 9 selectable conditions by ascending slider values.

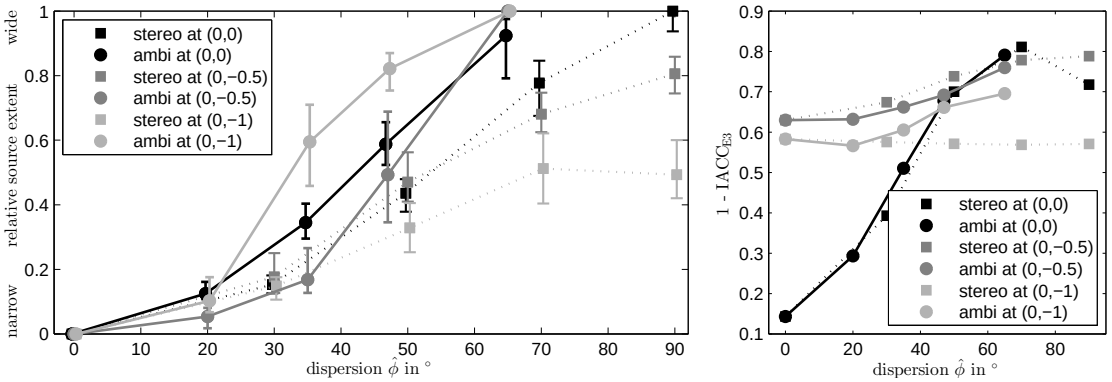
All of the 12 subjects (all male, age range: 27-39 years, median: 31 years) were members of the Institute of Electronic Music and Acoustics and familiar with spatial audio.

3.2. Results

Within each subject, the repetitions of the same comparison task correlated at least with 73%. The mean correlation was 92%. Thus, the repeated comparisons of all subjects were summarized in the analysis below, yielding statistics using 24 answers for each condition and listening position.

Stereo widening: Fig. 3(a) shows the medians and corresponding confidence intervals of perceived source extent in dotted lines for conditions using stereo. For all listening positions an analysis of variance (ANOVA) reveals the dispersion as a significant factor (probability $> 99.9\%$). However, not all neighboring conditions yield significantly different extent for the off-center positions. At the position $(0, -1)\,\mathrm{m}$, the conditions with 70° and 90° yield no significantly different mean values (13%). Comparing the median values, the conditions with 70° and 90° at $(0, -0.5)\,\mathrm{m}$, as well as 50° and 70° are not significantly different. Thus, the lateral distance to central listening position decreases the controllability of stereo widening. While the perceived differences remain between conditions with small dispersion, differences for the greatest dispersion get lost. Note that each perceptual scale in Fig. 3(a) is relative and covers all the 9 conditions (Stereo and Ambisonics) but is specific to one listening position. The diagram does not permit direct comparison of scale values for different listening positions.

Ambisonics: For the Ambisonic widening, dispersion is still a significant factor ($> 99.9\%$). Furthermore the neighboring conditions yield significantly different mean ($> 95\%$) and median values for all listening positions. Other than before, the nearly linear slope between 20° and 65° gets steeper and not shallower for off-center positions. This is plausible as both dispersion and lateral proximity of the listening position to LS2 increase lateral sound.



(a) Relative perceptual scales for efficient widening at different listening poitions. (b) $\mathrm{IACC}_{\mathrm{E3}}$ for efficient widening.

Figure 3: Perceptual, relative scales at each listening position for efficient stereo and Ambisonic widening and dependency of measured $1 - \mathrm{IACC}_{\mathrm{E3}}$ on dispersion $\hat{\phi}$. Perceptual scales show medians and corresponding 95% confidence intervals and their values are normalized to the maximum extent at each listening position.

While the perceived extent of stereo and Ambisonic widening is similar at the central listening position, the less controllable extent of stereo widening is strongly reduced at the off-center listening positions when compared to Ambisonic widening.

Inter-aural cross correlation coefficients (IACC) were calculated from binaural impulse responses of a B&K HATS 4128C in the experimental setup. The IACC is defined as the maximum of the inter-aural cross correlation function (IACF), cf. [15]. This article uses the $\mathrm{IACC}_{\mathrm{E3}}$, which is the average of the early IACCs for three octave bands around 500 Hz, 1 kHz, and 2 kHz, cf. [16]. For all the 9 conditions, the $\mathrm{IACC}_{\mathrm{E3}}$ values are correlated to the perceived source extent by 91% at $(0, 0)$ m, and by 92% for the off-center listening position at $(0, -0.5)$ m and 82% at $(0, -1)$ m, cf. Tab. 1 and Fig. 3(b). Neither relative scales nor correlations inter-relate the perceived extent at different listening positions.

4. WIDENING AND DIFFUSION FOR MASTERING

An obvious extension of the algorithm is an enlarged time grid $\mathrm{Q}/f_s \geq 10\,\mathrm{ms}$. For dispersion $\hat{\phi} \approx 80°$, this causes a diffuse spatial reverb characterized by an equally long attack and decay. Hereby, relatively short diffuse reverberation can be created.

Below, efficient Ambisonic widening/diffusion is extended to yield high-quality 2D and 3D mastering effects, see Figs. 4 and 6.

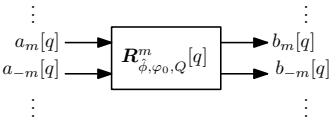
4.1. Frequency-dependent rotation.

The widening algorithm is applicable in mastering of Ambisonic productions or recordings after re-expressing it as a rotation. Using $\cos[\hat{\phi}\cos\Omega + \varphi_0] = \sum_{l=-\infty}^{\infty} \cos(\frac{\pi}{2}\,|l| + \varphi_0)\; J_{|l|}(\hat{\phi})\, e^{\mathrm{i}l\Omega}$ to define frequency-dependent Ambisonic z-rotation matrices achieves the effect. All Ambisonic signals $a_{\pm m}[q]$ belonging to cos and sin harmonics of same m are rotated in pairs by convolution with 2×2 matrices, see Fig. 4(a),

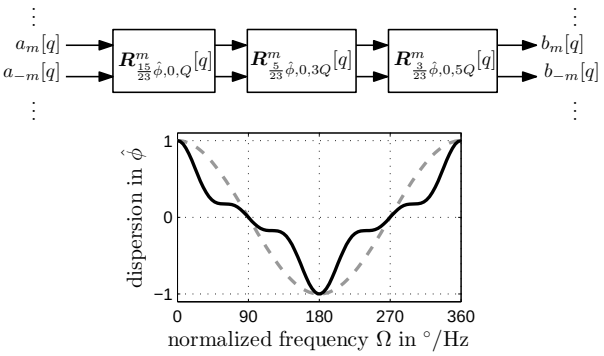
$$\begin{bmatrix} b_m[q] \\ b_{-m}[q] \end{bmatrix} = \underbrace{\begin{bmatrix} \cos(m\,\zeta) & -\sin(m\,\zeta) \\ \sin(m\,\zeta) & \cos(m\,\zeta) \end{bmatrix}}_{:=\boldsymbol{R}^m_{\hat{\phi},\varphi_0,\mathrm{Q}}[q]} \star \begin{bmatrix} a_m[q] \\ a_{-m}[q] \end{bmatrix}. \qquad (11)$$

Setting the rotation angle to $\zeta = \hat{\phi}\,\cos(\Omega) + \varphi_0$ creates a time-domain 2×2 filter matrix containing the sparse responses of Eq. (3)

$$\boldsymbol{R}^m_{\hat{\phi},\varphi_0,\mathrm{Q}}[\lambda\mathrm{Q}] = \begin{bmatrix} h(m\hat{\phi}, m\varphi_0, (\lambda-\Lambda)\mathrm{Q}] & h(m\hat{\phi}, m\varphi_0 + \frac{\pi}{2}, (\lambda-\Lambda)\mathrm{Q}] \\ h(m\hat{\phi}, m\varphi_0 - \frac{\pi}{2}, (\lambda-\Lambda)\mathrm{Q}] & h(m\hat{\phi}, m\varphi_0, (\lambda-\Lambda)\mathrm{Q}] \end{bmatrix}.$$

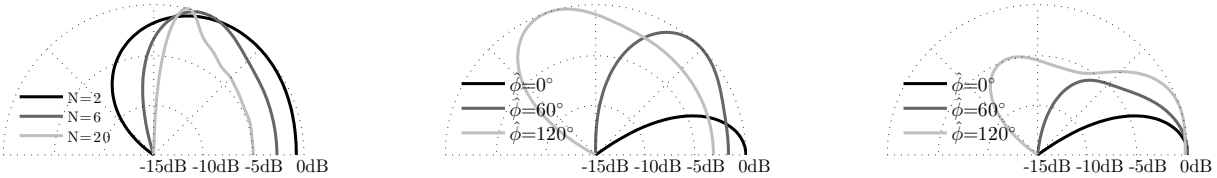


(a) scheme for rotational widening/diffusion.



(b) scheme for depicted, alternative dispersion curve.

Figure 4: (a) processing for dispersive rotation, (b) chained up to produces alternative dispersion curves.



(a) cos-shaped widening at dispersion $\hat{\phi} = 80°$ and varying Ambisonic orders.

(b) cos-shaped widening at varying dispersion and Ambisonic order $\mathrm{N} = 5$.

(c) harmonic-series-shaped widening at varying dispersion and Ambisonic order $\mathrm{N} = 5$.

Figure 5: Broad-band energy distribution $E(\varphi)$ for widening drawn over angle: (a) Ambisonic widening with cosine curve and varying orders, (b) for fifth order and varying dispersion, (c) for fifth order and modified dispersion curve.

4.2. Chaining-up rotations for other dispersion curves

The cos-shaped dispersion around $\varphi_0 = 0$ yields accumulation points in its angularly symmetric broad-band energy distribution

$$E(\varphi) = \int_0^{2\pi} \left\{ \sum_{m=0}^{\mathrm{N}} \cos[\tfrac{m\pi}{2(\mathrm{N}+1)}] \, \cos[m(\varphi - \hat{\phi}\cos\Omega)] \right\}^2 \mathrm{d}\Omega,$$

especially in higher-order Ambisonics, cf. Fig. 5(a). Fig. 4(b) shows a harmonically chained-up rotational dispersion to obtain the curve $\zeta = \sum_{j=0}^{2} \frac{15\,\hat{\phi}}{23(2j+1)} \cos[(2j+1)\Omega)]$,

$$\hat{\boldsymbol{R}}^m_{\hat{\phi},0,\mathrm{Q}}[q] = \boldsymbol{R}^m_{\frac{15}{23}\hat{\phi},0,\mathrm{Q}}[q] \star \boldsymbol{R}^m_{\frac{5}{23}\hat{\phi},0,3\mathrm{Q}}[q] \star \boldsymbol{R}^m_{\frac{3}{23}\hat{\phi},0,5\mathrm{Q}}[q]. \quad (12)$$

The resulting energy distribution in Fig. 5(c) shows less prominent extrema and a stable center compared to Fig. 5(b).

4.3. Application to 3D Ambisonics

The same 2×2 filter matrices are involved when dispersively rotating 3D Ambisonic signals $a_{n,\pm m}$ around the z axis, because the signals are related to the same azimuth harmonics as in 2D. We set $\varphi_0 = 0$ and use

$$\begin{bmatrix} b_{n,m}[q] \\ b_{n,-m}[q] \end{bmatrix} = \boldsymbol{R}^m_{\hat{\phi},0,\mathrm{Q}}[q] \star \begin{bmatrix} a_{n,m}[q] \\ a_{n,-m}[q] \end{bmatrix}. \quad (13)$$

However, sounds close to the z axis remain entirely unaffected.

As a remedy, rotational dispersion can be sequentially applied around z, y, and x. The same filter matrix is involved after using static rotations intermediately aligning y or x with the vertical direction. Given such rotation matrices $\boldsymbol{R}^{(\cdot z)}$ and $\boldsymbol{R}^{(z\cdot)}$, we define:

$$\begin{aligned} \boldsymbol{R}^{(z)}_{\hat{\phi},0,\mathrm{Q}}[q] &= [\boldsymbol{R}^m_{\hat{\phi},0,\mathrm{Q}}[q]]_{n,m}, \quad (14)\\ \boldsymbol{R}^{(y)}_{\hat{\phi},0,\mathrm{Q}}[q] &= \boldsymbol{R}^{(yz)}\, \boldsymbol{R}^{(z)}_{\hat{\phi},0,\mathrm{Q}}[q]\, \boldsymbol{R}^{(zy)}, \\ \boldsymbol{R}^{(x)}_{\hat{\phi},0,\mathrm{Q}}[q] &= \boldsymbol{R}^{(xz)}\, \boldsymbol{R}^{(z)}_{\hat{\phi},0,\mathrm{Q}}[q]\, \boldsymbol{R}^{(zx)}, \end{aligned}$$

and execute to manipulate the set of Ambisonic signals $\boldsymbol{a}[q]$

$$\boldsymbol{b}[q] = \boldsymbol{R}^{(x)}_{\hat{\phi},0,1.9\mathrm{Q}}[q] \star \boldsymbol{R}^{(y)}_{\hat{\phi},0,1.3\mathrm{Q}}[q] \star \boldsymbol{R}^{(z)}_{\hat{\phi},0,\mathrm{Q}}[q] \star \boldsymbol{a}[q]. \quad (15)$$

Different time grids $\{1.9\mathrm{Q}, 1.3\mathrm{Q}, \mathrm{Q}\}$ for x, y, and z achieve a trajectory over frequency that evenly fills a spherical cap of the angular size $\approx 2\hat{\phi}$, as shown in Fig. 6.

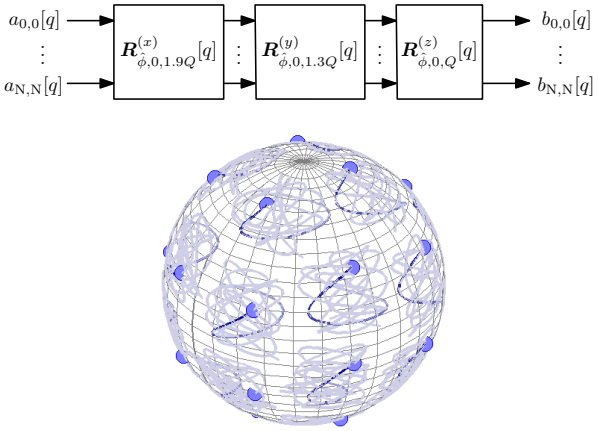


Figure 6: Scheme and trajectories for 3D widened/diffused providing sphere-cap-shaped dispersion.

5. CONCLUSION AND OUTLOOK

A highly effective Ambisonic encoding system for phantom source widening has been presented that disperses the direction of arrival of the sound over frequency.

Listening experiments for horizontal second-order Ambisonics have successfully proven that the new algorithm not only efficiently but also effectively controls the perceived source extent, which can be described by the $\mathrm{IACC}_{\mathrm{E3}}$.

Remarkably good control was also gained outside the central listening spot, where the effect of the new algorithm outperforms its stereophonic counterpart, even for a 1 m displacement.

In higher orders, the simple version of the algorithm might yield perceivable accumulation points of the distributed sound at the turning points of the angular dispersion curve. An alternative was given using a series of rotational dispersions of different frequency normalization / time grid Q.

Moreover, the algorithm can be generalized to work as 3D Ambisonic rotation matrix dispersing around all the three Cartesian axes.

It is subject to future research to investigate simple mathematical models of the decorrelation and widening for multi-channel audio, see also [17]. In particular, investigation of the matrix of inter-channel correlation coefficients appears promising. In addition, it is a perceptually interesting question how the rotational widening algorithm performs (a) for different dispersion contours

and (b) for rotation along other than the vertical axis. To assess how far coloration is a problem, a promising strategy of evaluation is to regard the binaural composite loudness levels, cf. [18].

6. REFERENCES

[1] Manfred R. Schröder, "An artificial stereophonic effect obtained from a single audio signal," *J. Audio Eng. Soc.*, vol. 6, no. 2, 1958.

[2] Benjamin B. Bauer, "Phasor analysis of some stereophonic phenomena," *The Journal of the Acoustical Society of America*, vol. 33, no. 11, pp. 1536–1539, 1961.

[3] Michael A. Gerzon, "Stereophonic signal processor generating pseudo stereo signals," *Patent, WO 93/25055*, 1993.

[4] Gary S. Kendall, "The decorrelation of audio signals and its impact on spatial imagery," *Computer Music J.*, vol. 19, no. 4, 1995.

[5] Maurice Bouéri and Chris Kyriakakis, "Audio signal decorrelation based on a critical band approach," in *Preprint 6291, 117th Conv. Audio Eng. Soc.*, San Francisco, 2004.

[6] Guillaume Potard, *3D-audio object oriented coding*, Ph.D. thesis, University of Wollongong, 2006.

[7] Franz Zotter, Matthias Frank, Georgios Marentakis, and Alois Sontacchi, "Phantom source widening with deterministic frequency dependent time delays," in *DAFx-11*, 2011.

[8] Franz Zotter and Matthias Frank, "Efficient phantom source widening," *Archives of Acoustics*, vol. 38, no. 1, pp. 27–37, 2013.

[9] Mikko-Ville Laitinen, Tapani Philajamäki, Cumhur Erkut, and Ville Pulkki, "Parametric time-frequency representation of spatial sound in virtual worlds," *ACM Trans. Appl. Percept.*, vol. 9, no. 2, 2012.

[10] Jung-Woo Choi, "Source-width extension technique for sound field reproduction systems," in *Proc. 52nd Int. AES Conf.: Sound Field Control - Engineering and Perception, Guildford*, Sept. 2013.

[11] John R. Carson, "Notes on the theory of modulation," *Proceedings of the Institute of Radio Engineers*, vol. 10, no. 1, pp. 57–64, 1922.

[12] ITU, "ITU-R BS.1116-1: Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems," 1997.

[13] Jérôme Daniel, *Représentation de champs acoustiques, application à la transmission et à la reproduction de scénes sonores complexes dans un contexte multimédia*, Phd thesis, Université Paris 6, 2001.

[14] EBU, "Tech 3253 - Sound Quality Assessment Material (SQAM)," http://tech.ebu.ch/publications/sqamcd.

[15] ISO, "ISO 3382-1:2009: Acoustics - measurement of room acoustic parameters - part 1: Performance spaces," 2009.

[16] Takayuki Hidaka, Leo L. Beranek, and Toshiyuki Okano, "Interaural cross-correlation, lateral fraction, and low- and high-frequency sound levels as measures of acoustical quality in concert halls," *The Journal of the Acoustical Society of America*, vol. 98, no. 2, pp. 988–1007, 1995.

[17] Matthias Frank and Franz Zotter, "Simple technical prediction of phantom source widening," in *AIA/DAGA, Fortschritte der Akustik*, Meran, 2013.

[18] Kazuho Ono, Ville Pulkki, and Matti Karjalainen, "Binaural modeling of multiple sound source perception: Methodology and coloration experiments," in *Audio Engineering Society Convention 111*, 11 2001.

[19] "NIST Digital Library of Mathematical Functions," http://dlmf.nist.gov/, Release 1.0.6 of 2013-05-06.

[20] F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, Eds., *NIST Handbook of Mathematical Functions*, Cambridge University Press, New York, NY, 2010.

7. APPENDIX

In order to observe the Fourier components of the $\cos[\alpha\cos(\Omega)]$ and $\sin[\alpha\cos(\Omega)]$ functions above, we can regard

$$\begin{aligned}\cos[\alpha\cos(\Omega)+\beta] &= \frac{1}{2}\left\{e^{\mathrm{i}[\alpha\sin(\Omega)+\beta]} + e^{-\mathrm{i}[\alpha\cos(\Omega)+\beta]}\right\}\\ &= \underbrace{\frac{e^{\mathrm{i}\beta}e^{\mathrm{i}\frac{\alpha}{2}[e^{\mathrm{i}\Omega}+e^{-\mathrm{i}\Omega}]}}{2}}_{:=f(\beta,\alpha,\Omega)} + \underbrace{\frac{e^{-\mathrm{i}\beta}e^{-\mathrm{i}\frac{\alpha}{2}[e^{\mathrm{i}\Omega}+e^{-\mathrm{i}\Omega}]}}{2}}_{f^\star(\beta,\alpha,\Omega)=f(-\beta,-\alpha,\Omega)},\end{aligned}$$

which is further, using $J_\lambda(\alpha) = (-1)^\lambda J_{-\lambda}(\alpha)$, [19, 20, Eq. 10.4.1],

$$\begin{aligned}f(\beta,\alpha,\Omega) = \frac{e^{\mathrm{i}\beta}}{2}\,e^{\mathrm{i}\frac{\alpha}{2}[e^{\mathrm{i}\Omega}+e^{-\mathrm{i}\Omega}]} &= \frac{e^{\mathrm{i}\beta}}{2}\sum_{\lambda=-\infty}^{\infty}\mathrm{i}^\lambda\,J_\lambda(\alpha)\,e^{\mathrm{i}\lambda\Omega}\\ &= \frac{e^{\mathrm{i}\beta}}{2}\sum_{\lambda=-\infty}^{\infty}\mathrm{i}^{|\lambda|}J_{|\lambda|}(\alpha)\,e^{\mathrm{i}\lambda\Omega}.\end{aligned}$$

Using $J_\lambda(-\alpha) = (-1)^\lambda J_\lambda(\alpha)$, [19, 20, Eq. 10.11.1], we obtain

$$\begin{aligned}f(-\beta,-\alpha,\Omega) = &= \frac{e^{-\mathrm{i}\beta}}{2}\sum_{\lambda=-\infty}^{\infty}\mathrm{i}^{|\lambda|}J_{|\lambda|}(-\alpha)\,e^{\mathrm{i}\lambda\Omega}\\ &= \frac{e^{-\mathrm{i}\beta}}{2}\sum_{\lambda=-\infty}^{\infty}\mathrm{i}^{-|\lambda|}J_{|\lambda|}(\alpha)\,e^{\mathrm{i}\lambda\Omega},\end{aligned}$$

and are finally able to re-expand $\cos[\alpha\cos\Omega+\beta]$ in terms of

$$\begin{aligned}\cos[\alpha\cos\Omega+\beta] &= \frac{1}{2}\sum_{\lambda=-\infty}^{\infty}\left[e^{\mathrm{i}\beta}\,\mathrm{i}^{|\lambda|} + e^{-\mathrm{i}\beta}\,\mathrm{i}^{-|\lambda|}\right]J_{|\lambda|}(\alpha)\,e^{\mathrm{i}\lambda\Omega}\\ &= \sum_{\lambda=-\infty}^{\infty}\cos(\beta+\tfrac{\pi}{2}\,|\lambda|)\,J_{|\lambda|}(\alpha)\,e^{\mathrm{i}\lambda\Omega}. \qquad (16)\end{aligned}$$

Transformation into the time domain just uses the transform pair $e^{\mathrm{i}\lambda\Omega} \leftrightarrow \delta(t-\lambda)$ after frequency scaling $e^{\mathrm{i}\lambda\mathrm{T}\Omega} \leftrightarrow \delta(t-\lambda\mathrm{T})$.