Bibliographical reference

Energy-Preserving Ambisonic Decoding, F. Zotter, H. Pomberger and M. Noisternig, Acta Acustica **vol. 98** (Number 1), 2012, pp. 37-47

DOI

https://doi.org/10.3813/AAA.918490

ACTA ACUSTICA UNITED WITH ACUSTICA
Vol. 98 (2012) 37 – 47

DOI 10.3813/AAA.918490

Energy-Preserving Ambisonic Decoding

F. Zotter[1)], H. Pomberger[1)], M. Noisternig[2)]

[1)] Institute of Electronic Music and Acoustics, University of Music and Performing Arts, Graz, Austria. zotter@iem.at

[1)] Acoustic and Cognitive Spaces Research Group, IRCAM – CNRS UMR STMS, Paris, France

Summary

Ambisonics with height is a three-dimensional sound field reproduction technique for spherical loudspeaker arrangements surrounding the reproduction area. It employs spherical harmonics up to a given order to expand incident sound fields with a limited angular resolution. The expansion coefficients describe the spatial sound scene. For reproduction, these coefficients are decoded to a set of surrounding loudspeakers. Common decoding approaches either sample the spherical harmonic excitation at the loudspeaker positions or match the excitation modes of a continuous sound field to those of the loudspeakers. For well-designed spherical loudspeaker arrays, both decoding approaches achieve good perceptual localization of virtual sound sources. However, both approaches perform unsatisfactorily with non-uniformly arranged arrays. Sounds from directions with only sparse loudspeaker coverage appear with altered loudness levels. This distracting effect results from variations in the decoded energy. The present article demonstrates an improved decoding technique, which preserves the decoded energy. Using available objective estimators, the localization qualities of these energy-preserving decoders are shown to lie between both common decoding approaches.

PACS no. 43.60.Gk, 43.60.Md, 43.60.Sx, 43.38.Tj

1. Introduction

Several techniques exist to reproduce three-dimensional sound fields with surrounding loudspeaker arrays. These techniques can be categorized into physical reproduction approaches and hearing-related model approaches. Physical reproduction approaches, as higher-order Ambisonics (HOA) and wave field synthesis (WFS) [1], aim at holophonic reconstruction of wave fields. In contrast, hearing-related models, as vector base amplitude panning (VBAP) [2], rely on a psychoacoustical phenomenon known as *phantom source*. It denominates the impression of one single sound event localized between loudspeakers. To evoke this impression, suitable weights are employed to distribute an audio signal to a multichannel loudspeaker arrangement; rules for these weights are called *panning functions*.

The present article focuses on Ambisonics, which is based on an expansion of incident sound fields to spherical harmonics up to a given order, whereby the angular resolution is limited. As Ambisonics was primarily limited to first order, cf. [3, 4], it was extended to higher-order systems (HOA) with increased spatial resolution, cf. [5, 6, 7, 8, 9, 10]. The excitation of a spatial sound scene is described by coefficients of the spherical harmonics expansion. These coefficients are the Ambisonic signals, and an idealized excitation with a spherical continuum of surrounding sound sources is approximated by decoding the Ambisonic signals to a surrounding loudspeaker array. There are two common approaches for designing Ambisonic decoders. One approach is to sample the spherical harmonic excitation at the loudspeaker positions. The other most frequently discussed decoder design approach is known as *mode-matching* [6], which aims at matching the spherical harmonic modes excited by the loudspeaker signals with the modes of the Ambisonic sound field decomposition.

Ambisonics is usually regarded as a physical reproduction approach that achieves three-dimensional holophonic sound reproduction. However, the size of the area free of reproduction errors is limited. In practice, for high frequencies this error-free *sweet area* can be much smaller than a single listener's head [7, 10, 11]. Therefore it seems that three-dimensional holophony for larger auditoria is an overambitious aim. Interestingly, the perceived localization and sound coloration outside the sweet area is reported as reasonable [12, 13, 14], despite holophonic reproduction errors. Taking these psychoacoustical considerations into account, the interpretation of Ambisonics in terms of a panning function proves useful for larger auditoria.

In general, a panning function should provide good auditory properties of phantom sources, independent of the panning direction. These properties are a stable and adequate localization, constant perceived source width,

Received 20 February 2011,
accepted 3 July 2011.

Figure 1. The IEM-CUBE is a concert room providing a 24-channel hemispherical Ambisonic system for 50 listeners.

and constant perceived loudness. After the introduction of Ambisonics by Gerzon *et al.* in the 1970s [15, 3, 4], these properties were usually improved by manual fine-tuning of the decoder. To avoid the *"black art"* [16] of manual optimization, several publications proposed objective measures to estimate the perceived localization, cf. [16, 17, 18], and objectives for its optimization, cf. also [19]. With the advent of HOA, optimized mode-matching decoders were proposed, e.g. by Daniel *et al.* [5], to avoid large mislocalization towards proximal loudspeakers at off-center listening positions.

Phantom sources with constant perceived loudness independent of the panning direction are achieved for uniformly arranged loudspeaker arrays with both decoding approaches, sampling and mode-matching. However, when decoding to non-uniform arrangements, phantom sources exhibit altered loudness levels for panning directions pointing to regions of sparse loudspeaker coverage.

Because loudness alteration is very distracting, the present article addresses this problem. As in other panning approaches, the objective energy measure is proposed to gain control of the perceived loudness. Meeting the expectation of a constant loudness, an improved decoding technique is presented that preserves the decoded energy. The resulting objective localization estimators are shown to lie between the results of the mode-matching and the sampling decoder, using this new technique.

Notice that this article focuses on a purely technical concept of the objective localization measures, which were originally intended to provide an estimation of the perceived localization. Accordingly, the objective localization measures are employed for the comparison of the mapping properties regarding common decoding approaches and the newly developed energy-preserving decoder.

Non-uniform loudspeaker arrangements covering only a part of a sphere, e.g. hemispherical arrays, need to be treated differently to achieve similar results. Nevertheless, even for such incomplete arrangements, energy preservation is shown to be feasible using a set of alternative basis functions.

This article is organized as follows. Section two reviews the Ambisonic related theory and the advantages and limitations of holophonic HOA rendering, especially regarding the accessible region of accurate synthesis. Section three presents a pragmatic view to Ambisonics with height, reviewing it as an amplitude-panning technique for small to large-scale auditoria, e.g. the IEM-CUBE shown in Figure 1. Section four reviews the vector measures known from Ambisonics literature and introduces estimators for the directional error, the decoded energy, and the energy spread. Section five takes up the idea of employing energy-preserving matrices to retain loudness, as introduced by Gerzon [17]. Based on such matrices, a new decoder design technique is proposed. Section six presents a new set of basis functions that allows for energy-preserving decoding on incomplete arrangements, e.g. a hemisphere.

2. Holophonic decoding

Ambisonics can be regarded as the discretization of a spherical distribution of sources at a given radius r_L driven by the angular amplitude distribution $f(\boldsymbol{\theta})$, cf. [20]. Within this article, we define the Cartesian direction vector as

$$\boldsymbol{\theta} = \left[\cos(\varphi)\sin(\vartheta), \sin(\varphi)\sin(\vartheta), \cos(\vartheta)\right]^\mathrm{T}$$

with φ and ϑ being the azimuth and zenith angle, respectively.

2.1. Continuous surrounding sources

Ambisonics is based on the nonhomogeneous Helmholtz equation with the continuous excitation $f(\boldsymbol{\theta})$ on a sphere with radius r_L, cf. [11, 21],

$$\left(\Delta + k^2\right) p = -f\left(\boldsymbol{\theta}\right) \frac{\delta(r - r_\mathrm{L})}{r_\mathrm{L}^2}, \tag{1}$$

where Δ is the Laplacian, p is the sound pressure, $\delta(\cdot)$ is the Dirac delta distribution, and k is the wave number.

The homogeneous solution to equation (1), i.e. $\left(\Delta + k^2\right) p = 0$, within a source-free spherical volume $\Omega = \{\boldsymbol{x} \in \mathbf{R}^3 : \|\boldsymbol{x}\| < r_\mathrm{L}\}$ is

$$p = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} b_{nm}\, j_n(kr)\, Y_n^m(\boldsymbol{\theta}), \tag{2}$$

where $j_n(kr)$ are the spherical Bessel functions, $Y_n^m(\boldsymbol{\theta})$ are the real-valued spherical harmonics (SHs), illustrated in Figure 2, and b_{nm} are the SH spectral coefficients. The real-valued SHs are defined as

$$Y_n^m(\boldsymbol{\theta}) = N_n^{|m|} P_n^{|m|}(\cos\theta) \begin{cases} \cos(m\phi) & \text{for } m \geq 0, \\ \sin(m\phi) & \text{for } m < 0, \end{cases}$$

where $P_n^{|m|}$ denotes the associated Legendre function, and $N_n^{|m|}$ is the scalar energy-normalization of the SHs. Any sound field within Ω that is generated by sources located outside Ω, i.e. any *incident sound field*, can be expressed in the form of equation (2).

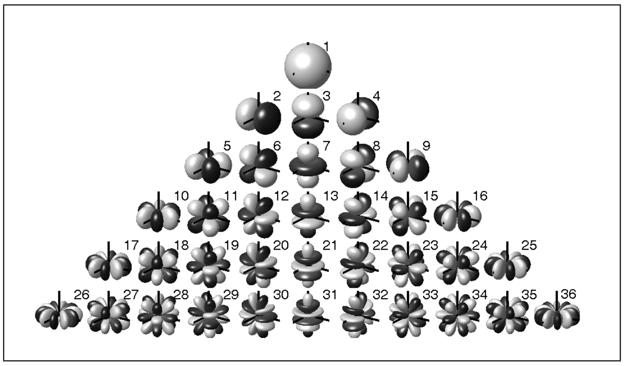

Figure 2. Balloon plots of real-valued spherical harmonics of order $n \leq 5$ plotting magnitude and sign as radius and gray values, respectively.

The particular solution of equation (1) is obtained by *variation of parameters*, cf. [11], and it becomes

$$p = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} -\mathrm{i}k\, \phi_{nm}\, h_n(kr_\mathrm{L})\, j_n(kr)\, Y_n^m(\boldsymbol{\theta}), \qquad (3)$$

where $h_n(kr_\mathrm{L})$ are the spherical Hankel functions and ϕ_{nm} are the SH spectral coefficients of $f(\boldsymbol{\theta})$ obtained by integral transformation on the unit sphere $\mathbb{S}^2$, $\phi_{nm} = \int_{\mathbb{S}^2} f(\boldsymbol{\theta}) Y_n^m(\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\theta}$. Comparing equation (2) and (3) reveals that any incident sound field within Ω can be generated by choosing $f(\boldsymbol{\theta})$ such that

$$\phi_{nm} = \frac{\mathrm{i}}{k} \frac{b_{nm}}{h_n(kr_\mathrm{L})}.$$

For notational simplicity, the double sum in equation (3) can be expressed as a weighted scalar product of two column vectors[1]

$$p = -\mathrm{i}k\, \boldsymbol{y}(\boldsymbol{\theta})^\mathrm{T} \operatorname{diag}\{\boldsymbol{j}(kr)\} \operatorname{diag}\{\boldsymbol{h}(kr_\mathrm{L})\}\, \boldsymbol{\phi}, \qquad (4)$$

with

$$\begin{aligned} \boldsymbol{\phi} &:= [\phi_{00}, \cdots, \phi_{nm}, \cdots]^\mathrm{T}, \\ \boldsymbol{y}(\boldsymbol{\theta}) &:= \left[Y_0^0(\boldsymbol{\theta}), \cdots, Y_n^m(\boldsymbol{\theta}), \cdots\right]^\mathrm{T}, \\ \boldsymbol{j}(kr) &:= [j_0(kr), \cdots, \underbrace{j_n(kr), \cdots, j_n(kr)}_{2n+1}, \cdots]^\mathrm{T}, \\ \boldsymbol{h}(kr) &:= [h_0(kr), \cdots, \underbrace{h_n(kr), \cdots, h_n(kr)}_{2n+1}, \cdots]^\mathrm{T}. \end{aligned}$$

The SH spectrum in vector-matrix notation becomes

$$\boldsymbol{\phi} = \int_{\mathbb{S}^2} f(\boldsymbol{\theta})\, \boldsymbol{y}(\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\theta}. \qquad (5)$$

This spectrum is considered as the infinite set of *Ambisonic signals* defined in the frequency or time domain.

Synthetic sound scenes are traditionally composed of virtual point sources. A point source driven by the signal s and located at $\boldsymbol{\theta}_s$ on a sphere with radius r_L is described as an angular Dirac delta distribution

$$f_s(\boldsymbol{\theta}) = s\, \delta(\boldsymbol{\theta} - \boldsymbol{\theta}_s). \qquad (6)$$

Inserting equation (6) into equation (5), the Ambisonic representation $\boldsymbol{\phi}_s$ of this point source yields

$$\boldsymbol{\phi}_s = s \int_{\mathbb{S}^2} \delta(\boldsymbol{\theta} - \boldsymbol{\theta}_s)\, \boldsymbol{y}(\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\theta} = s\, \boldsymbol{y}(\boldsymbol{\theta}_s). \qquad (7)$$

This step is usually referred to as *encoding*. For sources located at a different radius $r_s > r_\mathrm{L}$, Ambisonic signals are converted by filtering, cf [22],

$$\boldsymbol{\phi}_s = s \operatorname{diag}\{\boldsymbol{h}(kr_\mathrm{L})\}^{-1} \operatorname{diag}\{\boldsymbol{h}_n(kr_s)\}\, \boldsymbol{y}(\boldsymbol{\theta}_s).$$

This indirectly proves that the spectrum $\boldsymbol{\phi}$ can represent any incident sound field within the spherical volume Ω.

2.2. Discrete surrounding sources

Real-world Ambisonic playback facilities use loudspeakers at discrete locations $\{\boldsymbol{\theta}_l\}_{l=1\ldots\mathrm{L}}$, which are controlled by the corresponding driving signals g_l. In this article, loudspeakers are modeled as point sources, i.e., their directivity is neglected. According to equation (6), the angular excitation of the facility corresponds to the superposition of all loudspeakers

$$\hat{f}(\boldsymbol{\theta}) = \sum_{l=1}^{\mathrm{L}} \delta\left(\boldsymbol{\theta} - \boldsymbol{\theta}_l\right)\, g_l, \qquad (8)$$

or equivalently in the spherical harmonics domain, cf. equation (7),

$$\hat{\boldsymbol{\phi}} = \sum_{l=1}^{\mathrm{L}} \boldsymbol{y}(\boldsymbol{\theta}_l)\, g_l. \qquad (9)$$

For notational simplicity, the above equation can be expressed in matrix-vector form

$$\hat{\boldsymbol{\phi}} = \boldsymbol{Y}\, \boldsymbol{g}, \qquad (10)$$

with

$$\begin{aligned} \boldsymbol{g} &:= \left[g_1, \ldots g_\mathrm{L}\right]^\mathrm{T}, \\ \boldsymbol{Y} &:= \left[\boldsymbol{y}(\boldsymbol{\theta}_1), \ldots, \boldsymbol{y}(\boldsymbol{\theta}_\mathrm{L})\right]. \end{aligned}$$

The signals $\boldsymbol{g}$ driving the loudspeakers are determined from the given Ambisonic signals $\boldsymbol{\phi}$ by a decoder matrix $\boldsymbol{D}$,

$$\boldsymbol{g} = \boldsymbol{D}\, \boldsymbol{\phi}. \qquad (11)$$

The decoding matrix $\boldsymbol{D}$ forces the discrete angular excitation of the playback facility $\hat{f}(\boldsymbol{\theta})$ to approximate the sound field of any continuous excitation $f(\boldsymbol{\theta})$. The design of this matrix constitutes the *decoding problem*.

A perfect reconstruction of $f(\boldsymbol{\theta})$ by $\hat{f}(\boldsymbol{\theta})$ can be expressed as matching the modal source strengths or equivalently the Ambisonic signals $\boldsymbol{\phi}_s = \hat{\boldsymbol{\phi}}$. This is referred to as

[1] A single sum $\sum_{l=0}^{\mathrm{L}} a_l\, b_l\, c_l$ equates to $\boldsymbol{a}^\mathrm{T} \operatorname{diag}\{\boldsymbol{b}\}\, \boldsymbol{c}$.

mode-matching in literature, e.g. [6]. The mode-matching problem can only be solved for a finite number of Ambisonic signals. Thus a maximum order $n \leq \mathrm{N}$ is assumed, which implies angular band-limitedness of the target sound field, i.e. an ideally limited angular resolution. The finite order versions of the vectors $\boldsymbol{y}(\boldsymbol{\theta})$, $\boldsymbol{\phi}$, and the matrix $\boldsymbol{Y}$ are denoted as $\boldsymbol{\phi}_\mathrm{N}$, $\boldsymbol{y}_\mathrm{N}$, and $\boldsymbol{Y}_\mathrm{N} \in \mathbf{R}^{K\times L}$, obtaining the finite size $\mathrm{K} = (\mathrm{N}+1)^2$. Theoretically, finite order mode-matching also demands an angularly band-limited playback facility $\hat{\boldsymbol{\phi}} = \hat{\boldsymbol{\phi}}_\mathrm{N}$. Insertion of equation (10) into equation (11) shows that mode-matching is equivalent to

$$\boldsymbol{Y}_\mathrm{N}\, \boldsymbol{D}_\mathrm{ma} = \boldsymbol{I}. \tag{12}$$

This equation is most conveniently solved starting with the compact singular value decomposition (SVD)

$$\boldsymbol{Y}_\mathrm{N} = \hat{\boldsymbol{U}}\, \hat{\boldsymbol{S}}\, \hat{\boldsymbol{V}}^\mathrm{T}, \tag{13}$$

where $\hat{\boldsymbol{S}} = \mathrm{diag}(\hat{s}_1, \ldots, \hat{s}_\mathrm{M}) \in \mathbf{R}^{\mathrm{M}\times\mathrm{M}}$ contains singular values in decreasing order $\hat{s}_1 \geq \hat{s}_2 \geq \cdots \geq \hat{s}_\mathrm{M}$, and $\hat{\boldsymbol{U}} \in \mathbf{R}^{\mathrm{K}\times\mathrm{M}}$ and $\hat{\boldsymbol{V}} \in \mathbf{R}^{\mathrm{L}\times\mathrm{M}}$ are matrices with orthogonal columns. For the compact SVD, M is defined such that $\hat{s}_\mathrm{M}$ is greater than zero, and $\mathrm{M} \leq \min\{\mathrm{L}, \mathrm{K}\}$. With this decomposition, the mode-matching decoder becomes

$$\boldsymbol{D}_\mathrm{ma} = \hat{\boldsymbol{V}}\, \hat{\boldsymbol{S}}^{-1}\, \hat{\boldsymbol{U}}^\mathrm{T}. \tag{14}$$

The above solution is exact for $\mathrm{M} = \mathrm{K}$, thus at least K loudspeakers need to be arranged such that $\boldsymbol{Y}_\mathrm{N}$ is of full rank, i.e., it has K non-zero singular values.

In practice, non-uniform or incomplete loudspeaker arrangements can cause very small but non-zero singular values. In this case, the calculation of $\boldsymbol{D}_\mathrm{ma}$ is numerically problematic and yields loud playback signals. Reasonable mode-matching with such a *numerically rank deficient* [23] matrix $\boldsymbol{Y}_\mathrm{N}$ requires suitable regularization, resulting in an approximate match $\boldsymbol{Y}_\mathrm{N}\boldsymbol{D}_\mathrm{ma} \approx \boldsymbol{I}$.

Tikhonov regularization, e.g. $\boldsymbol{D}_\mathrm{ma} = \boldsymbol{Y}_\mathrm{N}^\mathrm{T}(\boldsymbol{Y}_\mathrm{N}\boldsymbol{Y}_\mathrm{N}^\mathrm{T} + \lambda\boldsymbol{I})^{-1}$, has been applied, cf [10]. A more versatile means of regularization is the truncated singular value decomposition (TSVD). In contrast to the compact SVD, M is defined such that $\hat{s}_\mathrm{M}/\hat{s}_1 \geq \alpha$, with the truncation parameter $\alpha \in [0, 1]$. Choosing α or M accordingly, the decoder in equation (14) is calculated using only the numerically relevant singular values of the SVD.

As shown in [23] for applying Tikhonov's method with identity matrix, a regularization parameter λ can always be found such that its solution is close to the TSVD solution for a certain truncation parameter α. The TSVD method is more useful for theoretical considerations as it is much simpler to analyze than Tikhonov's method [23, p.109].

2.3. Finite order and spatial aliasing

Successful decoding reproduces the lower-order components $n \leq \mathrm{N}$ of the Ambisonic signals $\boldsymbol{\phi}_\mathrm{N}$. Holophony can be obtained for, but is not limited to, sound fields of limited order: $j_n(kr)$ vanishes for small arguments kr and for high orders n, and one recognizes the vanishing higher-order components in equation (3) within small radii $r < r_{\mathrm{m}ax,N}$. In the sweet area around the origin, i.e. within $r < r_{\mathrm{m}ax,N}$, any incident sound field can be represented exactly by the lower-order components, cf . [7, 10, 11, 21].

The sound field produced by loudspeaker arrays does, in general, not only consist of lower-order components $\hat{\boldsymbol{\phi}} \neq \hat{\boldsymbol{\phi}}_\mathrm{N}$. In fact, the attempt to control these components produces higher-order components as well. Thus, the reasonable prerequisite of an ideally limited resolution is not accurate. Nevertheless, the known shape of $j_n(kr)$ suppresses uncontrolled components inside the sweet area [11]. The mismatch outside this area is referred to as *spatial aliasing*.

As a rule of thumb, the frequency-dependent size of the holophonic sweet area is estimated by $r_{\mathrm{m}ax,N}/\lambda \approx \mathrm{N}/6$, cf. [7].

In practice, the ideal of an Ambisonic *with height* facility for audio playback within the entire audible frequency range is not achievable with reasonable hardware effort. Currently used loudspeaker facilities, as known to the authors, are capable of decoding signals of orders $1 \leq \mathrm{N} \leq 5$; a loudspeaker facility for decoding Ambisonic up to orders N=9 is currently under construction at IRCAM. Thus, for the upper frequency bands, the sweet area can be estimated to be much smaller than a single listener's head. Nevertheless, Ambisonic playback is frequently applied to mid and large scale concert venues. Most, if not all, listeners are situated outside the sweet area since it can only be increased by increasing the reproduction order. In other words, holophonic errors are present across nearly the whole listening area.

Interestingly enough, however, studies [24] indicate a useful perceived localization, even outside the sweet area, despite the inevitable spatial aliasing. Using more than a critical number of loudspeakers for a given order suppresses spatial aliasing in an area larger than the sweet area. Unexpectedly, the increased playback effort seems to have a negative influence on localization and coloration in this aliasing free area around the sweet area, cf [12, 25, 13]. To underline the fact that the reduction of spatial aliasing without increasing the order is counterproductive, the authors propose the expression *friendly aliasing*.

3. Amplitude panning

Amplitude panning distributes an audio signal $s(t)$ to several loudspeakers using weights $\boldsymbol{g}(t) = \boldsymbol{w}\, s(t)$. Applying suitable the weights $\boldsymbol{w}$ creates the impression of a single sound source that is perceived from a certain direction. In psychoacoustics this effect is called *phantom source* and the direction of this phantom source is controlled by the weights $\boldsymbol{w}$.

For a single sound source, Ambisonics can be interpreted as a continuous amplitude panning function, cf. Figure 3, that is evaluated at each loudspeaker

$$g_l = \frac{4\pi}{\mathrm{L}}\, \boldsymbol{y}_\mathrm{N}(\boldsymbol{\theta}_l)^\mathrm{T}\, \boldsymbol{y}_\mathrm{N}(\boldsymbol{\theta}_s)\, s. \tag{15}$$

As in Malham's work [26], this concept is used to obtain a *decoder by sampling*

$$\boldsymbol{D}_{\mathrm{sp}} = \frac{4\pi}{\mathrm{L}}\, \boldsymbol{Y}_{\mathrm{N}}^{\mathrm{T}}, \tag{16}$$

or in the decomposed form $\boldsymbol{D}_{\mathrm{sp}} = (4\pi/\mathrm{L})\, \hat{\boldsymbol{V}} \hat{\boldsymbol{S}} \hat{\boldsymbol{U}}^{\mathrm{T}}$. Both mode-matching, cf. equation (14), and sampling decoders, cf. equation (16), seem to work fine, at least for uniform loudspeaker arrangements, for which the singular values of $\boldsymbol{Y}_{\mathrm{N}}$ do not largely differ from $\sqrt{\mathrm{L}/4\pi}$. Listeners have reported convincing localization accuracy and sound quality even outside the sweet area. This raises the question: *Does it have to be "clean" holophony, or does a simple interpretation as a panning-law serve our needs?*

Regarding Ambisonics as a panning technique hence *relaxes* the mode-matching conditions towards spherical panning. Several works with a similar concept can be found in literature, cf. [27, 28, 29, 30].

Providing listeners with a convincing impression of phantom sources requires, independent of the panning direction,

- a constant perceived loudness,
- a stable and adequate localization, and
- a constant perceived source width.

The best solution for a large audience might already be simpler, aiming at shaping a reasonable gain pattern for the playback signal on the loudspeaker facility. Malham states in [26] that psychoacoustic optimization of decoders is not productive for large listening areas: *"Designing decoders for large areas, such as concert halls, requires a design strategy aimed at achieving even power distribution."*

4. Objective quality criteria

In classical literature about Ambisonics, practical measures to evaluate decoders can be found. Gerzon, cf. [16], proposes the superposition of the linear and squared loudspeaker signals to measure the playback magnitude,

$$\hat{P} = \sum_{l=1}^{\mathrm{L}} g_l, \qquad \hat{E} = \sum_{l=1}^{\mathrm{L}} |g_l|^2. \tag{17}$$

Moreover, to measure the directional concentration of the playback on the loudspeaker facility, the centroids of the amplitude and its square are proposed

$$\hat{\boldsymbol{r}}_V = \frac{\sum_{l=1}^{\mathrm{L}} g_l\, \boldsymbol{\theta}_l}{\hat{P}}, \tag{18}$$

$$\hat{\boldsymbol{r}}_E = \frac{\sum_{l=1}^{\mathrm{L}} |g_l|^2\, \boldsymbol{\theta}_l}{\hat{E}}. \tag{19}$$

These centroids are weighted sums of each loudspeaker direction vector $\boldsymbol{\theta}_l$ multiplied by its partial contribution, e.g. the partial squared signal $|g_l|^2/\hat{E}$. The vectors $\hat{\boldsymbol{r}}_V$, $\hat{\boldsymbol{r}}_E$, representing both centroids, point into the directions where the playback signals or their squares, respectively, appear strongest on the loudspeaker arrangement. Their lengths measure the directional concentration of panning on the playback facility.

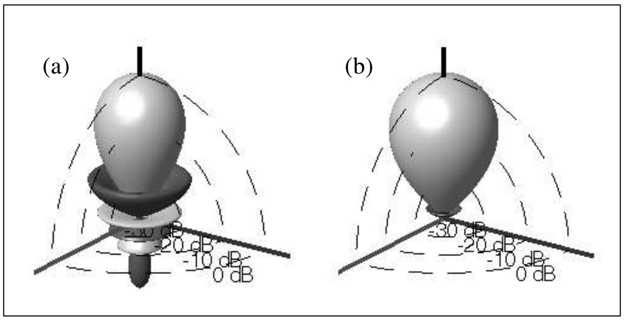

Figure 3. Angular amplitude-distribution for a virtual point-source of the order N = 5 plotted as balloon diagram (a) without and (b) with *max-*r_E smoothing.

For a single active loudspeaker, the centroids coincides with the position on the unit sphere, i.e., its length reaches a maximum $\|\boldsymbol{\theta}_l\| = 1$, whereas for a uniform spherical distribution of loudspeakers driven with equal amplitude signals the centroids point at the origin, $\|\hat{\boldsymbol{r}}_V\| \to 0$ and $\|\hat{\boldsymbol{r}}_E\| \to 0$.

The quantities $\hat{P}$ and $\hat{\boldsymbol{r}}_V$ are based on the superposition with fixed phase relations and thus are only considered relevant within the sweet area for low frequencies.

The quantities $\hat{E}$ and $\hat{\boldsymbol{r}}_E$ are called *energy* and *energy vector* as the squared signals are proportional to the energy radiated by the loudspeakers when the interaction of their signals in the sound field is negligible.

Classical Ambisonic literature uses the measures $\hat{E}$ and $\hat{P}$ to estimate the perceived loudness, and it proposes the length of $\hat{\boldsymbol{r}}_E$ or $\hat{\boldsymbol{r}}_V$ to estimate the perceived source width [16, 31], which is currently under investigation [32]. Nevertheless, the attributes of perceived localization should be accurately estimated using binaural localization models, on which [33] gives an overview. For the estimation of the source width in VBAP [34, 35], simple binaural models have already been applied. Although the detailed information these models aim to provide (direction, width) for different listening positions and head orientations is important, in general, they do not directly yield a simple quality measure for the evaluation of Ambisonic decoders.

As the development of such a measure lies outside the scope of this article, technical measures are used to objectively describe the technical quality of decoding.

4.1. Avoiding mislocalization

A virtual point source located at $(\boldsymbol{\theta}_s, r_{\mathrm{L}})$ is represented by the Ambisonic signals $\boldsymbol{\phi}_{s,\mathrm{N}}(\boldsymbol{\theta}_s) = \boldsymbol{y}_{\mathrm{N}}(\boldsymbol{\theta}_s)\, s$, cf. equation (7). The corresponding angular distribution is a rotationally symmetric sinc-like function that exhibits some distracting side lobes, the largest one opposing the main lobe, cf. Figure 3a. For listeners outside the sweet area, these side lobes may increase the localization errors and the perceived source width.

Side lobes can be successfully diminished by angular smoothing, which is accomplished with a *spherical convolution* [36] that attenuates the higher-order components.

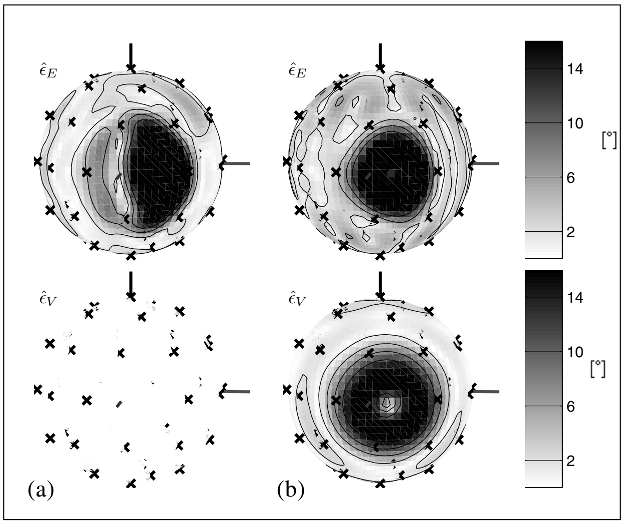

Figure 4. Estimated directional mapping errors in degrees with (a) mode-matching and (b) sampling decoders (speakers removed).

Arranging the order weights a_n in a vector $\boldsymbol{a}_\mathrm{N}$ according to the scheme of equation (4), the smoothed function results in

$$\boldsymbol{\phi}_{s,\mathrm{N}}(\boldsymbol{\theta}_s) = \mathrm{diag}\left\{\boldsymbol{a}_\mathrm{N}\right\} \boldsymbol{y}_\mathrm{N}(\boldsymbol{\theta}_s)\, s \tag{20}$$

and yields the decoded loudspeaker driving signals

$$\boldsymbol{g} = \boldsymbol{D}\, \mathrm{diag}\left\{\boldsymbol{a}_\mathrm{N}\right\} \boldsymbol{y}_\mathrm{N}(\boldsymbol{\theta}_s)\, s. \tag{21}$$

The angularly smoothed version (*max-*r_E) of the amplitude distribution in Figure 3a is depicted in Figure 3b.

Literature on Ambisonics mainly discusses two types of weighting functions [5, 26],

- *In-phase* decoding: full side lobe suppression, and broadest main lobe [26, 8],
- *max-*r_E decoding: best $\|\hat{\boldsymbol{r}}_E\|$-rated angular energy concentration [5],

and frequency-dependent combinations thereof. For first-order (N = 1) decoding, also the unweighted narrowest-amplitude shape *max-*r_V was included. When applying one single weighting function in combination with higher-order Ambisonics, *max-*r_E weighting is preferred in recent literature [13, 24].

4.2. Directional error

Assuming the phase shifts of all arriving signals are equal, the linear vector measure $\hat{\boldsymbol{r}}_V$ is easy to calculate using equation (18). This assumption is meaningful for the central position, where the direction of $\hat{\boldsymbol{r}}_V$ indicates the direction a signal is mapped to, cf. [16]. For higher-order decoders $\hat{\boldsymbol{r}}_V$ might not always be insightful, as it is biased towards the components of the first order.

In contrast, the energy vector measure $\hat{\boldsymbol{r}}_E$ does not contain phase information, cf. [16]. Therefore, it reflects mapping directions for locations around the sweet area, where propagation losses in the arriving loudspeakers signals do not largely differ, but phase relations are regarded as diverse and position-dependent. The difference of mapping directions due to the change of perspective are not reflected in this measure. Still, the evaluation of $\hat{\boldsymbol{r}}_E$ from the central position is considered sufficient, as the change of perspective is natural and is not considered an error. Using both measures, the directional deviation is indicated by

$$\hat{\epsilon}_V = \arccos \frac{\boldsymbol{\theta}_s^\mathrm{T} \hat{\boldsymbol{r}}_V}{\|\hat{\boldsymbol{r}}_V\|}, \tag{22}$$

$$\hat{\epsilon}_E = \arccos \frac{\boldsymbol{\theta}_s^\mathrm{T} \hat{\boldsymbol{r}}_E}{\|\hat{\boldsymbol{r}}_E\|}. \tag{23}$$

To illustrate the objective quality criteria, we use in the following an example of a slightly non-uniform loudspeaker arrangement. This example is based on the uniform layout of 41 loudspeakers, using equal area partitioning [37]. Both directional errors would stay small for this entirely uniform arrangement. Hence, to simulate slight non-uniformity, one loudspeaker was removed for the analysis of different decoder designs. Despite this is still benign compared to non-uniformity of real facilities, it is a starting point to demonstrate the application of the objective quality criteria.

Figures 4a and 4b show the estimated deviation angles $\hat{\epsilon}_E$ and $\hat{\epsilon}_V$ over a variable panning direction $\boldsymbol{\theta}_s$ for the above described non-uniform layout using mode-matching and sampling decoding, respectively. Figure 4a shows that mode-matching precisely matches the $\hat{\boldsymbol{r}}_V$ vector to the panning direction. The sampling decoder produces big errors near the removed speaker for both directional deviation estimators, cf. Figure 4b.

4.3. Energy spread

The directional concentration of the playback signals on the loudspeaker arrangement is measured with the length of the vector $\hat{\boldsymbol{r}}_E$, which is valid outside the sweet area for diverse, position-dependent phase relations of the incident loudspeaker signals $\boldsymbol{g}$. Similar to [31], the measure $\|\hat{\boldsymbol{r}}_E\|$ can be mapped to an angular spread by

$$\hat{\sigma}_E = 2 \arccos \|\hat{\boldsymbol{r}}_E\|\,. \tag{24}$$

In the discrete case, the angular spread measure $\hat{\sigma}_E$ varies except for some particular loudspeaker arrays, whereas its continuous counterpart σ_E is constant. In the ideal continuous case, the invariance follows from the analytic expression [8, p. 312]

$$\|\boldsymbol{r}_E\| = \frac{2\sum_{n=0}^{\mathrm{N}} n\, a_n\, a_{n-1}}{\sum_{n=0}^{\mathrm{N}} (2n+1)\, a_n^2}, \tag{25}$$

e.g. for N = 5 and *max-*r_E the angular spread measure $\sigma_E \approx 42°$.

The bottom row of Figure 5 shows an evaluation of the spread measure $\hat{\sigma}_E$ as a function of the panning direction. The results for the uniform 41 loudspeaker arrangement using a mode-matching and a sampling decoder are depicted in Figure 5a and Figure 5b, respectively. Regarding $\hat{\sigma}_E$, both decoders perform reasonably well.

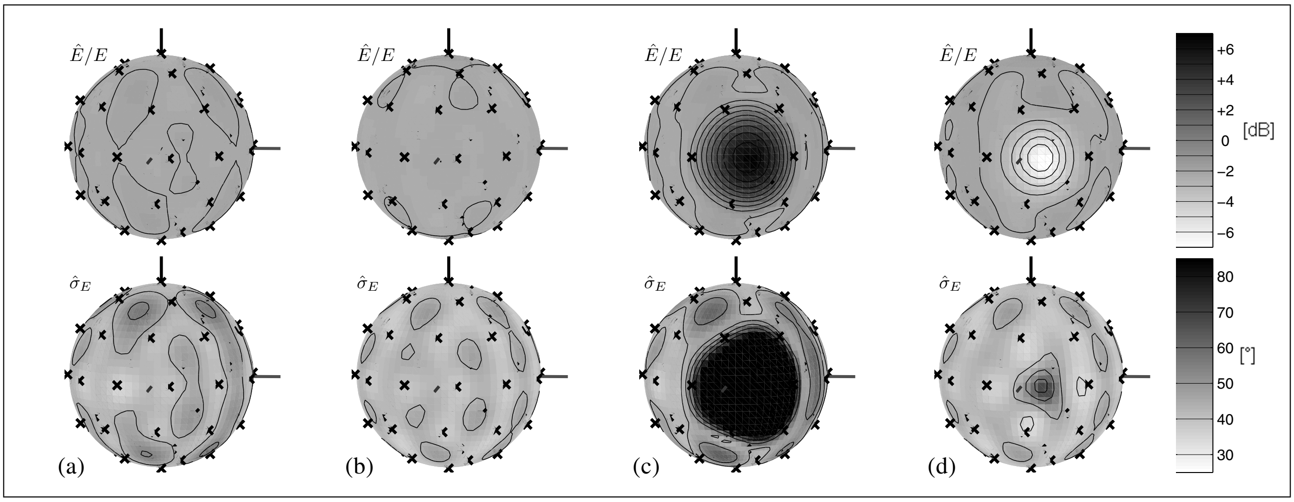

Figure 5. Energy and energy spread measures $\hat{E}/E$ and $\hat{\sigma}_E$ plotted as a function of the panning direction on the unit sphere. Different decoders with N = 5 and *max*-r_E weighting are evaluated for a uniform 41 loudspeaker layout using equal area partitioning [37]: (a) Mode-matching and (b) sampling for all speakers. The loudspeaker in the middle is removed in the diagrams: (c) mode-matching, and (d) sampling.

Figures 5c and 5d show the results for the slightly non-uniform layout with one speaker removed before designing the decoders. In contrast to the uniform layout, both decoders yield an increase of the spread in the vicinity of the gap, as a natural consequence of the missing speaker. The increase is observed to a lesser extent for the sampling decoder.

4.4. Energy

The energy $\hat{E}$ from equation (17) is a common measure in amplitude panning and already provides a fair estimate of the average loudness for interference with diverse, position-dependent phase relations. It is not applicable at low frequencies, e.g. < 200Hz, for which the sweet area encloses several listeners. The given evaluation is considered sufficiently accurate when neglecting the low-frequency range.

The measure of an ideal (continuous) Ambisonic playback equals $\int_{\mathbb{S}^2} |f(\boldsymbol{\theta})|^2 \, d\boldsymbol{\theta} = \|\boldsymbol{\phi}\|^2$, due to Parseval's theorem. It is easy to find from equation (20) that this is independent of the panning direction $\boldsymbol{\theta}_s$ [8, p. 312] and yields

$$E = \left\| \boldsymbol{\phi}_{s,\mathrm{N}} \right\|^2 = \frac{|s|^2}{4\pi} \sum_{n=0}^{\mathrm{N}} (2n+1)\, a_n^2. \tag{26}$$

Normalizing $\hat{E}$ by the measure of the virtual source E yields a relative measure that represents energy alterations of the decoder, exclusively. Equivalently, this normalization is expressed by normalizing the Ambisonic signals $\boldsymbol{\phi}_{s,\mathrm{N}}$

$$\hat{E}/E = \frac{\boldsymbol{\phi}_{s,\mathrm{N}}^{\mathrm{T}}}{\left\|\boldsymbol{\phi}_{s,\mathrm{N}}\right\|} \boldsymbol{D}^{\mathrm{T}} \boldsymbol{D} \frac{\boldsymbol{\phi}_{s,\mathrm{N}}}{\left\|\boldsymbol{\phi}_{s,\mathrm{N}}\right\|}. \tag{27}$$

The range of energy alterations for a decoder is defined as the interval $\hat{E}_{\min} \leq \hat{E} \leq \hat{E}_{\max}$. For mode-matching and sampling decoders, the range of the alterations can be determined from the singular values of $\boldsymbol{Y}_{\mathrm{N}}$. For the mode-matching decoder, cf. equation (14), the matrix product for calculating $\hat{E}/E$ yields

$$\boldsymbol{D}_{\mathrm{ma}}^{\mathrm{T}} \boldsymbol{D}_{\mathrm{ma}} = \hat{\boldsymbol{U}}\, \hat{\boldsymbol{S}}^{-2}\, \hat{\boldsymbol{U}}^{\mathrm{T}}. \tag{28}$$

Similarly for the sampling decoder, cf. equation (16) and equation (13), this matrix is

$$\boldsymbol{D}_{\mathrm{sp}}^{\mathrm{T}} \boldsymbol{D}_{\mathrm{sp}} = \left(\frac{4\pi}{L}\right)^2 \hat{\boldsymbol{U}}\, \hat{\boldsymbol{S}}^{2}\, \hat{\boldsymbol{U}}^{\mathrm{T}}. \tag{29}$$

As $\hat{\boldsymbol{U}}$ is orthogonal, both decoders only achieve a small range of energy alterations if the range of singular values is small. This is the case for uniform loudspeaker layouts, where $\hat{s}_i \to \sqrt{L/4\pi}$. Regarding energy alterations, decoding to non-uniform layouts reveals the different behavior of mode-matching and sampling decoders. For the sampling decoder, small singular values lead to an attenuation of the decoded energy, i.e., the lower bound $\hat{E}_{\min}$ is decreased.

In contrast, the inversion of the singular values in the mode-matching decoder leads to an increase of the decoded energy, i.e. an increase of the upper bound $\hat{E}_{\max}$.

The top row of Figure 5 shows an evaluation of the energy measure $\hat{E}/E$ as a function of the panning direction. Figures 5a and 5b depict the results for the exemplary uniform layout using a mode-matching and sampling decoder, respectively. As expected, both decoders achieve a small range of energy alterations for the uniform layout. The results for the slightly non-uniform layout (missing speaker example) are depicted in Figures 5c and 5d. The mode-matching decoder exhibits an energy increase in the vicinity of the gap, whereas the sampling decoder yields a loss of energy there. As argued above, this behavior is determined by the singular values of $\boldsymbol{Y}_{\mathrm{N}}$.

5. Energy-preserving decoder

For varying panning directions, mode-matching and sampling decoders cannot provide a constant energy under slight non-uniformity. As shown in equations (28) and (29), the range of energy alterations depends on the singular values of $\boldsymbol{Y}_\mathrm{N}$, only. Consequently, in this section a new energy-preserving decoder is presented, which does not contain energy scaling factors, i.e., the singular values have been removed,

$$\boldsymbol{D}_\mathrm{ep} = \hat{\boldsymbol{V}}\hat{\boldsymbol{U}}^\mathrm{T}. \tag{30}$$

As the columns of $\hat{\boldsymbol{U}}$ and $\hat{\boldsymbol{V}}$ are orthogonal, this decoder yields

$$\boldsymbol{D}_\mathrm{ep}^\mathrm{T}\boldsymbol{D}_\mathrm{ep} = \boldsymbol{I}, \tag{31}$$

and thus no energy alteration, cf. equation (27). This always works regardless of the particular loudspeaker placement as long as their number equals at least the number of the spherical harmonics $\mathrm{L} \leq \mathrm{K}$.

Yet for good decoding, energy-preservation alone is not sufficient. The directional mapping of the decoder must also produce reasonable results. For this reason, the measures for energy spread and the directional errors have been introduced above, which provide rational indicators of the directional mapping related quantities.

As predicted, energy-preserving decoding perfectly produces a constant measure $\hat{E}/E$ for the slightly non-uniform example with the gap as shown in Figure 6. The estimated directional mapping error $\hat{\epsilon}_E$ seems to be slightly better than for the other decoders, whereas $\hat{\epsilon}_V$ seems to lie between the other decoding variants, cf. Figure 4. The energy spread $\hat{\sigma}_E$ does not much differ from other decoders. In general, these direction and spread related measures indicate that energy-preserving decoding produces a useful angular mapping of Ambisonic signals.

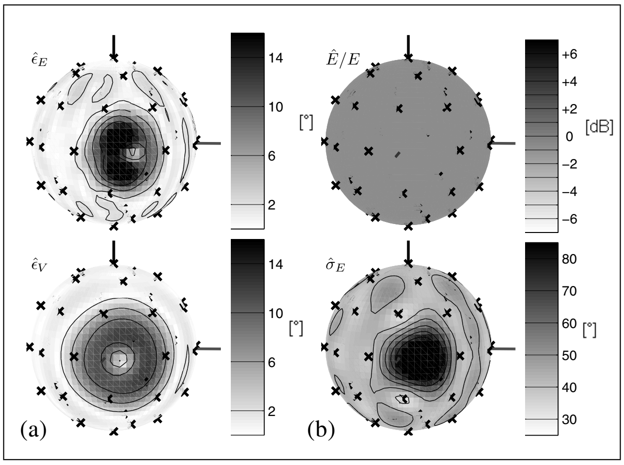

Figure 6. Energy-preserving decoder (removed speaker); (a) estimated directional mapping errors $\hat{\epsilon}_E$ and $\hat{\epsilon}_V$, (b) energy and spread estimators $\hat{E}/E$ and $\hat{\sigma}_E$.

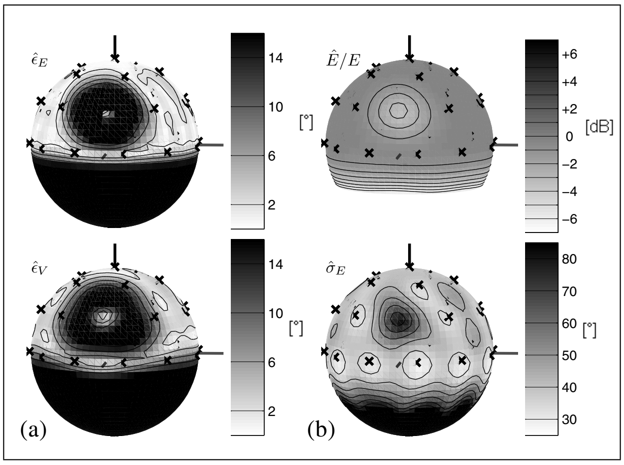

Figure 7. Energy-preserving decoding with $\mathrm{N} = 5$ and a subset of 25 loudspeakers on the hemisphere; (a) estimated directional mapping errors $\hat{\epsilon}_E$ and $\hat{\epsilon}_V$, (b) energy and spread estimators $\hat{E}/E$ and $\hat{\sigma}_E$.

6. Hemispherical decoding

Energy-preserving decoding is possible if there are at least as many loudspeakers as there are Ambisonic signals, i.e. spherical harmonics. Thus an arrangement of $\mathrm{L} = (\mathrm{N}+1)^2$ loudspeakers on the hemisphere allows for energy-preserving playback of the maximum order N. However, it appears costly to use the same number of loudspeakers for a hemispherical setup as required for a full sphere setup with unchanged resolution N. One could naively argue that about the half of the loudspeakers should be sufficient for the hemisphere. This raises the question, what happens if $\mathrm{L} < (\mathrm{N}+1)^2$, i.e., there are fewer loudspeakers than Ambisonic signals?

Figure 7 shows the result for energy measure and directional mapping error for the energy preserving decoder using such a setup. This exemplary setup contains 25 loudspeakers on the upper hemisphere, which are a subset of the previous example setup, from which one loudspeaker is removed to introduce a slight non-uniformity. As a consequence of $\mathrm{L} < (\mathrm{N}+1)^2$, the decoded energy is not preserved for all directions, cf. $\hat{E}/E$ in Figure 7. In particular, the energy is not preserved for regions poorly covered by speakers; in this example, these are the southern hemisphere and the region around the missing speaker. As the southern hemisphere is not considered for playback, the energy loss for virtual sources placed there is usually no problem, whereas around the missing speaker it is disturbing.

In equation (13), the compact singular value decomposition (SVD) was introduced. To investigate the energy loss of the decoder with $\mathrm{L} < (\mathrm{N}+1)^2$, the full SVD of $\boldsymbol{Y}_\mathrm{N}$ is more telling,

$$\boldsymbol{Y}_\mathrm{N} = \left[\hat{\boldsymbol{U}}, \hat{\boldsymbol{U}}_0\right] \begin{bmatrix} \hat{\boldsymbol{S}} \\ \boldsymbol{0} \end{bmatrix} \hat{\boldsymbol{V}}^\mathrm{T},$$

including the full unitary decomposition matrix $\left[\hat{\boldsymbol{U}}, \hat{\boldsymbol{U}}_0\right]$. The matrix $\hat{\boldsymbol{U}}_0 = [\hat{\boldsymbol{u}}_{0,1}, \hat{\boldsymbol{u}}_{0,2}, \ldots, \hat{\boldsymbol{u}}_{0,K-L}]$ contains the left null vectors, satisfying $\hat{\boldsymbol{u}}_{0,i}^{\mathrm{T}} \boldsymbol{Y}_{\mathrm{N}} = \boldsymbol{0}$; therefore $\hat{\boldsymbol{U}}_0$ is called the null space of $\boldsymbol{Y}_{\mathrm{N}}$. The energy preserving decoder, cf. equation (30), can be equivalently expressed in terms of the full SVD,

$$\boldsymbol{D}_{\mathrm{ep}} = \hat{\boldsymbol{V}} \left[\boldsymbol{I}, \boldsymbol{0}\right] \begin{bmatrix} \hat{\boldsymbol{U}}^{\mathrm{T}} \\ \hat{\boldsymbol{U}}_0^{\mathrm{T}} \end{bmatrix}. \tag{32}$$

Inserting this into equation (27) makes obvious that an energy loss occurs if the Ambisonic signal vector lies (at least partly) within the null space, i.e. $\boldsymbol{\phi}_{s,\mathrm{N}}^{\mathrm{T}} \hat{\boldsymbol{U}}_0 \neq \boldsymbol{0}$. In the example above, this is true for panning directions on the southern hemisphere as well as around the missing speaker.

As mentioned earlier, a hemispherical loudspeaker arrangement is usually intended for source directions on one hemisphere only, on which the energy alteration should be minimal. In contrast, energy loss for sources located at the complementary hemisphere is acceptable. As the null space causes the energy loss, it should be designed as to solely affect sources on the complementary (southern) hemisphere. As shown below, this is best achieved by the design of a new set of continuous basis functions that exhibits an energy concentration on the target hemisphere.

6.1. New basis functions

The energy of a continuous panning function is evaluated by

$$E = \boldsymbol{\phi}_{\mathrm{N}}^{\mathrm{T}} \int_{\mathbb{S}^2} \boldsymbol{y}_{\mathrm{N}}(\boldsymbol{\theta}) \boldsymbol{y}_{\mathrm{N}}(\boldsymbol{\theta})^{\mathrm{T}} \, \mathrm{d}\boldsymbol{\theta} \, \boldsymbol{\phi}_{\mathrm{N}}^{\mathrm{T}}.$$

As the spherical harmonics are orthonormal functions on $\mathbb{S}^2$, $\int_{\mathbb{S}^2} \boldsymbol{y}_{\mathrm{N}}(\boldsymbol{\theta}) \boldsymbol{y}_{\mathrm{N}}(\boldsymbol{\theta})^{\mathrm{T}} \, \mathrm{d}\boldsymbol{\theta} = \boldsymbol{I}$, the energy $E = \|\boldsymbol{\phi}_{\mathrm{N}}\|^2$ is unity for any normalized $\boldsymbol{\phi}_{\mathrm{N}}$.

In contrast, the energy of a normalized $\boldsymbol{\phi}_{\mathrm{N}}$ on the hemisphere, or any other continuously connected part of the sphere $\tilde{S}^2 \subset \mathbb{S}^2$, is not constant as the following integral is not the identity matrix

$$\int_{\tilde{S}^2} \boldsymbol{y}_{\mathrm{N}}(\boldsymbol{\theta}) \, \boldsymbol{y}_{\mathrm{N}}(\boldsymbol{\theta})^{\mathrm{T}} \, \mathrm{d}\boldsymbol{\theta} = \boldsymbol{G}. \tag{33}$$

In general, this matrix has to be determined by numerical integration on the sphere, cf. [38]. The properties of $\boldsymbol{G}$ can be investigated using its eigendecomposition

$$\boldsymbol{G} = \boldsymbol{U} \, \mathrm{diag}\left\{\boldsymbol{s}\right\}^2 \boldsymbol{U}^{\mathrm{T}}, \tag{34}$$

where $\boldsymbol{U} \in \mathbb{R}^{\mathrm{K}\times\mathrm{K}}$ is the orthogonal matrix of eigenvectors and $\boldsymbol{s}$ is a vector containing the square roots of the according eigenvalues. Multiplication of equation (33) from both sides with $\boldsymbol{U}$ and its transpose, respectively, yields

$$\int_{\tilde{S}^2} \boldsymbol{U}^{\mathrm{T}} \boldsymbol{y}_{\mathrm{N}}(\boldsymbol{\theta}) \, \boldsymbol{y}_{\mathrm{N}}(\boldsymbol{\theta})^{\mathrm{T}} \boldsymbol{U} \, \mathrm{d}\boldsymbol{\theta} = \mathrm{diag}\left\{\boldsymbol{s}\right\}^2. \tag{35}$$

This reveals that a new orthonormal set of basis functions can be found on $\mathbb{S}^2$

$$\boldsymbol{v}(\boldsymbol{\theta}) = \boldsymbol{U}^{\mathrm{T}} \boldsymbol{y}_{\mathrm{N}}(\boldsymbol{\theta}), \tag{36}$$

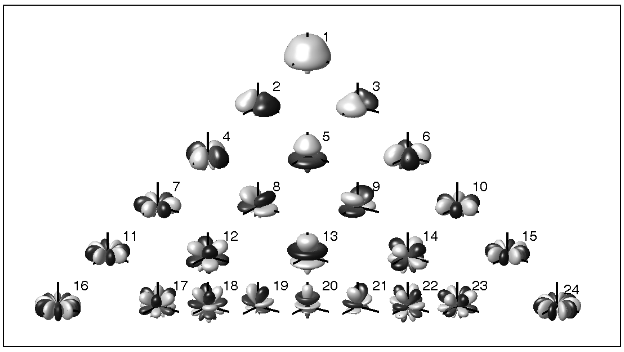

Figure 8. Balloon plots of the new basis functions for N = 5. Magnitude and sign are mapped to radius and gray value, respectively.

whose energies s_i^2 on $\tilde{S}^2$ are sorted in descending order. The new basis functions are equivalent to the spherical harmonics by their inverse relation $\boldsymbol{y}_{\mathrm{N}}(\boldsymbol{\theta}) = \boldsymbol{U} \, \boldsymbol{v}(\boldsymbol{\theta})$.

In contrast to the spherical harmonics, the number of new basis functions can be reduced by choosing only those $v_i(\boldsymbol{\theta})$ with sufficiently large energy $s_i^2/s_1^2 > \beta$, $0 < \beta < 1$. This is most conveniently done by defining an accordingly reduced matrix $\boldsymbol{U} \in \mathbb{R}^{K\times K'}$ with $\mathrm{K}' \leq \mathrm{K}$. The reduced set of new basis functions is not fully equivalent to the spherical harmonics as $\boldsymbol{U}$ does not provide an exact inverse relation $\boldsymbol{U}\,\boldsymbol{U}^{\mathrm{T}} \neq \boldsymbol{I}$ anymore. Nevertheless, only those functions having much energy outside $\tilde{S}^2$ have been discarded. Hence, applying the incomplete reconstruction to a signal $\boldsymbol{U}\,\boldsymbol{U}^{\mathrm{T}}\,\boldsymbol{\phi}_{\mathrm{N}}$ mainly suppresses spatial components outside $\tilde{S}^2$, however not sharply. Practically, attenuation near the border of the target hemisphere can be avoided by enlarging the integration domain $\tilde{S}^2$ in equation (33).

Figure 8 shows a new set of basis functions for the hemisphere for N = 5. This set of $\mathrm{K}' = 24$ functions is obtained by choosing $S^2 = \{(\varphi, \vartheta) \in \mathbb{S}^2 | \vartheta \leq 110°\}$ and $\beta = 0.75$ and ensures a vanishing energy alteration, cf. Figure 9a, for an encoded point source, cf. equation (6). The new functions might to some extent look arbitrary. Therefore, it is convenient to re-group the functions for a more obvious relation to the spherical harmonics, cf. [39, 40]. Figure 8 shows the new set of basis functions after re-grouping.

6.2. Energy-preserving decoding on the hemisphere

In order to apply the new set of basis functions to Ambisonic decoding, the Ambisonic signals $\boldsymbol{\phi}_{\mathrm{N}}$ and the modal representation $\boldsymbol{Y}_{\mathrm{N}}$ of the playback facility are re-encoded into the new orthonormal basis $\boldsymbol{v}(\boldsymbol{\theta})$. According to equation (36), this yields $\boldsymbol{U}^{\mathrm{T}} \boldsymbol{\phi}_{\mathrm{N}}$ and $\boldsymbol{U}^{\mathrm{T}} \boldsymbol{Y}_{\mathrm{N}}$, respectively. With one more singular value decomposition of the new mode matrix

$$\boldsymbol{U}^{\mathrm{T}} \boldsymbol{Y}_{\mathrm{N}} = \check{\boldsymbol{U}} \, \check{\boldsymbol{S}} \, \check{\boldsymbol{V}}^{\mathrm{T}}$$

and the re-encoded Ambisonic signals, the energy-preserving decoder, cf. equation (30), on $\tilde{S}^2$ yields

$$\boldsymbol{D}_{\mathrm{epb}} = \check{\boldsymbol{V}} \check{\boldsymbol{U}}^{\mathrm{T}} \boldsymbol{U}^{\mathrm{T}}. \tag{37}$$

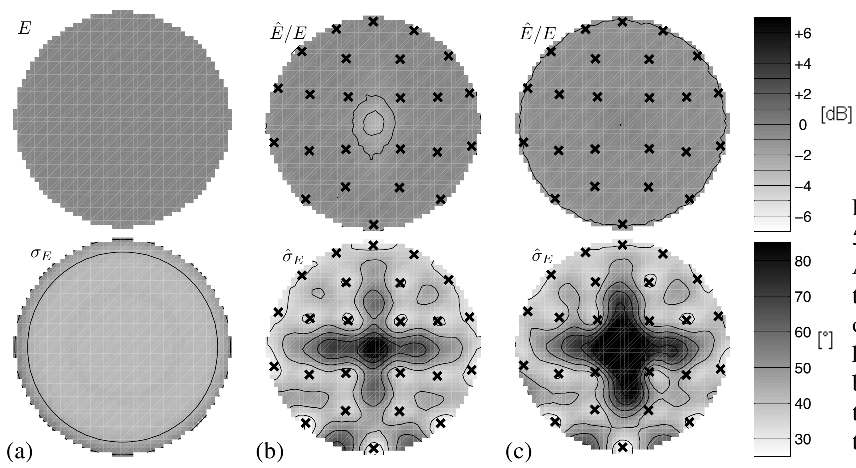

Figure 9. Evaluation of panning with N = 5 basis and decoding for hemispherical Ambisonics. Energy and spread estimators $\hat{E}/E$ and $\hat{\sigma}_E$ are plotted depending on the panning directions of the upper hemisphere (flattened image): (a) new basis with $\beta = 0.75$; (b) decoder equation (30), (c) decoder equation (37) for the IEM-CUBE.

Table I. Loudspeaker coordinates of the IEM-CUBE; its center lies at $x, y = 0$ and $z = 1.34$ m.

loudspeaker	1	2	3	4	5	6	7	8	9	10	11	12
x [m]	4.64	4.60	4.11	1.57	-1.29	-4.38	-4.64	-4.33	-1.07	1.82	4.48	4.71
y [m]	0.00	2.02	4.60	5.03	5.55	3.87	0.02	-3.86	-5.53	-4.94	-4.46	-1.85
z [m]	1.34	1.38	1.40	1.41	1.41	1.36	1.37	1.35	1.40	1.38	1.39	1.39

loudspeaker	13	14	15	16	17	18	19	20	21	22	23	24
x [m]	4.23	1.81	-2.19	-3.62	-3.6	-2.06	1.93	4.22	1.37	-1.25	-1.27	1.40
y [m]	1.77	4.44	4.87	1.48	-1.58	-4.78	-4.21	-1.77	1.46	1.32	-1.34	-1.33
z [m]	3.83	3.94	4.17	3.48	3.49	4.16	3.85	3.77	4.42	4.15	4.14	4.39

Figure 9b shows the energy and energy spread measure for decoding as defined in equation (30) applied to the loudspeaker arrangement of the IEM-CUBE, Table I. Clearly, a noticeable drop of energy appears for virtual sources positioned at the zenith of the playback facility, the center in the diagram of Figure 9b. The achievable improvement involving the new basis functions, cf. equation (37), is demonstrated in Figure 9c. Therein, the energy alteration stays below 1dB at the expense of a larger energy spread for virtual sources near the zenith.

7. Conclusions

We have presented a new decoding method that preserves the energy of the Ambisonic signals over the entire angular domain. By omitting the singular values of the decoding problem, loudness variations of sources with varying directions can be avoided in Ambisonic decoding. This type of decoding always works as soon as there are enough loudspeakers; it is always feasible and by nature numerically stable.

In addition, basis functions adapted to Ambisonics on the hemisphere or other fractions of the sphere have been introduced. These functions are designed in the continuous domain such that they allow for energy-preservation on the desired fraction of the sphere. As the number of functions is smaller than a complete SH set, the angular resolution for energy preserving decoders can be retained for a given loudspeaker spacing.

As other panning approaches, the new decoder finally creates a signal of well-controlled loudness by enforcing a normalized energy output. This consideration is based on the assumption of slightly stochastic interference of the loudspeaker signals in space. On top of that, a better control of the source width is expected, as the underlying spatial resolution is ideally limited using Ambisonics.

In terms of the objective estimators of energy and energy spread, the result of the new approach performs comparably well. Based on an informal listening session in the IEM-CUBE, these estimators seem to be meaningful. However, the proof of their perceptual significance by listening experiments is pending and topic of a current PhD project.

Acknowledgement

The authors wish to acknowledge comments made by the reviewers that greatly helped to improve this article. We acknowledge gratefully Alois Sontacchi and Olivier Warusfel for their support, and Thomas Musil for numerous informative discussions and his assistance on all aspects of implementation. This research was supported in part (second author) by the COMET program of the Austrian Research Promotion Agency (FFG), the Styrian Government, and the Styrian Business Promotion Agency (SFG).

References

[1] A. J. Berkhout: A Holographic Approach to Acoustic Sound Control. J. Audio Eng. Soc. **36** (1988) 977–995.

[2] V. Pulkki: Virtual sound source positioning using vector base amplitude panning. J. Audio Eng. Soc. **45** (1997) 456–466.

[3] M. A. Gerzon: Ambisonics. Part two: Studio techniques. Studio Sound **17** (1975) 24–26.

[4] M. A. Gerzon: The Design of Precisely Coincident Microphone Arrays for Stereo and Surround Sound. prepr. L-20, 50th AES Conv., London, 1975.

[5] J. Daniel, J.-B. Rault, J.-D. Polack: Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions. prepr. 4795, 105th AES Conv., San Francisco, 1998.

[6] M. A. Poletti: A Unified Theory of Horizontal Holographic Sound Systems. J. Audio Eng. Soc. **48** (2000) 1155–1182.

[7] D. B. Ward, T. D. Abhayapala: Reproduction of a plane-wave sound field using an array of loudspeakers. IEEE Trans. on Speech and Audio Proc. **9** (2001) 697–707.

[8] J. Daniel: Représentation de champs acoustiques, application à la transmission et à la reproduction de scénes sonores complexes dans un contexte multimédia. Dissertation. Université Paris 6, 2001.

[9] A. Sontacchi: Dreidimensionale Schallfeldreproduktion für Lautsprecher und Kopfhöreranwendungen. Dissertation. Technische Universität Graz, 2003.

[10] M. A. Poletti: Three-Dimensional Surround Sound Systems Based on Spherical Harmonics. J. Audio Eng. Soc. **53** (2005) 1004–1025.

[11] F. Zotter, H. Pomberger, M. Frank: An Alternative Ambisonics Formulation: Modal Source-Strength Matching and the Effect of Spatial Aliasing. prepr. 7740, 126th AES Conv., Munich, 2009.

[12] A. Solvang: Spectral Impairment of Two-Dimensional Higher Order Ambisonics. J. Audio Eng. Soc. **56** (2008) 267–279.

[13] M. Frank, F. Zotter: Localization Experiments Using Different 2D Ambisonics Decoders. Proc. 25. Tonmeistertagung, Leipzig, 2008.

[14] S. Bertet: Formats audio 3D hiérarchiques: caractérisation objective et perceptive des systèmes Ambionics d'ordres supérieurs. Dissertation. Institut National de Sciences Appliquées de Lyon, 2009.

[15] P. Fellget: Ambisonics. Part one: General system description. Studio Sound **17** (1975) 20–22.

[16] M. A. Gerzon: General Metatheory of Auditory Localization. prepr. 3306, 92nd AES Conv., Vienna, 1992.

[17] M. A. Gerzon: Hierarchical Transmission System for Multispeaker Stereo. J. Audio Eng. Soc. **40** (1992) 692–705.

[18] M. A. Gerzon: Psychoacoustic Decoders for Multispeaker Stereo and Surround Sound. prepr. 3406, 93rd AES Conv., San Francisco, 1992.

[19] A. van Leest: On Amplitude Panning and Asymmetric Loudspeaker Set-Ups. prepr. 6613, 119th AES Conv., New York, 2005.

[20] Y. J. Wu, T. D. Abhayapala: Soundfield reproduction using theoretical continuous loudspeaker. Proc. IEEE Int. Conf. Acoustics, Speech and Signal Proc., ICASSP, Las Vegas, 2008, 377–380.

[21] F. Zotter, H. Pomberger, M. Noisternig: Ambisonics Decoding with and without Mode-Matching: A case study using the hemisphere. Proc. 2nd Int. Symp. Ambisonics Spher. Acoustics, Paris, 2010.

[22] J. Daniel: Spatial Sound Encoding Including Near Field Effect: Introducing Distance Coding Filters and a Viable, New Ambisonic Format. prepr. 16, Proc. 23rd Int. AES Conf., Copenhagen, 2003.

[23] P. C. Hansen: Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion. Society for Industrial and Applied Mathematics, Philadelphia, 1998, (Mathematical Modeling and Computation).

[24] S. Bertet, J. Daniel, E. Parizet, O. Warusfel: Investigation on the restitution system influence over perceived Higher Order Ambisonics sound field: A subjective evaluation involving from first to fourth order systems. Proc. Acoustics-08, Joint ASA/EAA meeting, Paris, 2008.

[25] O. Santala, H. Vertanen, J. Pekonen, J. Oksanen, V. Pulkki: Effect of Listening Room on Audio Quality in Ambisonics Reproduction. prepr. 7664, 126th AES Conv., Munich, 2009.

[26] D. G. Malham, A. Myatt: 3D Sound Spatialization using Ambisonic Techniques. Computer Music Journal **19** (1995) 58–70.

[27] M. Poletti: Robust two-dimensional surround sound reproduction for nonuniform loudspeaker layouts. J. Audio Eng. Soc. **55** (2007) 568–610.

[28] M. Neukom, J. Schacher: Ambisonics equivalent panning. Proc. Int. Computer Music Conf., Belfast, UK, 2008.

[29] J.-M. Batke, F. Keiler: Investigation of Robust Panning Functions for 3D Loudspeaker Setups. prepr. 7979, 128th AES Conv., London, 2010.

[30] F. Zotter, M. Frank, A. Sontacchi: The Virtual t-design Ambisonics-Rig using VBAP. Proc. EAA EUROREGIO, Ljubljana, 2010.

[31] S. Moreau, J. Daniel, S. Bertet: 3D Sound Field Recording with Higher Order Ambisonics–Objective Measurements and Validation of a 4th Order Spherical Microphone. prepr. 6857, 120th AES Conv., Paris, 2006.

[32] M. Frank, G. Marentakis, A. Sontacchi: A simple technical measure for the perceived source width. Fortschritte der Akustik, DAGA, 2011 appear.

[33] R. M. Stern, G. J. Brown, D. Wang: Computational Auditory Scene Analysis. John Wiley & Sons, 2006, Ch. Binaural Sound Localization, 147–186.

[34] V. Pulkki: Spatial Sound Generation and Perception by Amplitude Panning Techniques. Dissertation. Helsinki University of Technology, Espoo, 2001.

[35] V. Pulkki: Uniform Spreading of Amplitude Panned Virtual Sources. Proc. IEEE Workshop on Appl. of Signal Proc. to Audio and Acoustics, 1999, 187–190.

[36] J. R. Driscoll, D. M. J. Healy: Computing Fourier Transforms and Convolutions on the 2-Sphere. Advances in Applied Mathematics **15** (1994) 202–250.

[37] P. Leopardi: Distributing Points on the Sphere. Dissertation. University of New South Wales, Sydney, 2006.

[38] N. Sneeuw: Global spherical harmonic analysis by least-squares and numerical quadrature methods in historical perspective. Geophysical Journal International **118** (1994) 707–716.

[39] R. Pail, G. Plank, W.-D. Schuh: Spatially Restricted Data Distributions on the Sphere: the Method of Ortnonormalized Functions and Applications. Journal of Geodesy **75** (2001) 44–56.

[40] H. Pomberger, F. Zotter: An Ambisoncis Format for Flexible Playback Layouts. Proc. 1st Ambisonics Symposium, Graz, 2009.