All-Round Ambisonic Panning and Decoding

FRANZ ZOTTER AND MATTHIAS FRANK, *AES Student Member*
(zotter@iem.at) (frank@iem.at)

Institute of Electronic Music and Acoustics, University of Music and Performing Arts, Graz, Austria

All-Round Ambisonic Panning (AllRAP) is an algorithm for arbitrary loudspeaker arrangements, aiming at the creation of phantom sources of stable loudness and adjustable width. The equivalent All-Round Ambisonic Decoding (AllRAD) fits into the Ambisonic format concept. Conventional Ambisonic decoding is only simple with optimal loudspeaker arrangements for which it achieves direction-independent energy and energy spread, the estimated phantom source loudness and width. AllRAP/AllRAD is still simple but more versatile and utilizes the combination of a virtual optimal loudspeaker arrangement with Vector-Base Amplitude Panning.

0 INTRODUCTION

An auditory event that is perceived from a certain direction is easily created by playback through a loudspeaker placed there, a *real source*. Less trivially, spatial audio playback techniques use the concept of a source whose location is independent of the playback facility, which is called a *virtual source* or *source object*. The spatial audio playback technique defines a method to evoke an auditory event with the spatial properties of the virtual source, using the playback facility.

Stereophony considers virtual sources having a location within the aperture of a loudspeaker pair. It employs level and/or time-delay differences between the loudspeaker signals to consistently create an auditory event, which is frequently called a *phantom source*. Rules, experimental evidence, and spatial restrictions of stereophonic phantom sources have been clearly shown in the literature on psychoacoustics [1, 2, 3, 4].

Adjusting the location of the phantom source is called *panning*, and it is achieved by level differences in most mixing consoles, i.e., amplitude panning. The work of Pulkki [5, 6, 4] on vector-base amplitude panning (VBAP) has generalized stereophonic amplitude panning to apertures of loudspeaker triplets.

Alternatively, Ambisonics has evolved as amplitude panning relying on analytic mathematical formulations, [7, 8, 9, 10]. Its formulation with height is based on decomposing the sound field excitation into discrete spherical harmonics, originally only up to the first and second order.

In recent research [11, 12, 13, 14], Ambisonics also includes higher-order spherical harmonics. Hereby not only the angular resolution increases, but also reproduction of extended fields becomes meaningful, even from directional virtual sources [15, 16]. On the other hand, today's computational power enables reproduction numerically matching the given listening environment [17, 18, 19].

This paper does not regard highly accurate sound field reproduction. Instead, the pragmatic use of Ambisonics as amplitude panning approach with height is considered. Despite this seeming simple, it has been either requiring mathematical skills or was restricted to dedicated arrangements of loudspeakers when compared to VBAP, so far. A solution circumventing both complications is desirable as it promises seamless operation with emerging Ambisonic microphone techniques [20, 21] without prior directional coding [22]. So far, the complexity of Ambisonics made it less common than VBAP, and its psychoacoustic evaluation was somewhat delayed [23, 24, 25, 26, 27, 28, 29].

In general, both VBAP and Ambisonics are capable of evoking phantom sources using multiple loudspeakers. However, each one has individual strengths and weaknesses. The following sections discuss these aspects and sketch a hybrid approach to rule out remaining shortcomings. The presented approach yields a highly generic algorithm easily applicable to any arrangement of loudspeakers and platforms with only basic computational capacities.

The first section of this paper explains vector-base amplitude panning using a triangulation of the given loudspeaker positions. Moreover, an extension preserving sounds outside the valid panning range is introduced. The second section presents Ambisonic panning with height using polynomials and dedicated loudspeaker arrangements. In the third section, energy and energy spread measures are introduced to compare mapping properties of the discussed amplitude panning methods. The fourth section introduces All-Round Ambisonic Panning (AllRAP) that combines the smooth Ambisonic energy spread with the flexibility of VBAP.

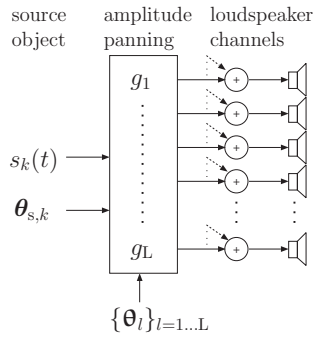


Fig. 1. Block diagram of amplitude panning, in general: a signal is added to the loudspeaker channels by a set of weights.

Energy and energy spread measures are used to compare AllRAP with VBAP in section five. After describing the representation concept of Ambisonics in section six, AllRAP is extended to the equivalent All-Round Ambisonic Decoding (AllRAD). Finally, section seven proposes a method to find an optimal Ambisonic truncation order for a given loudspeaker arrangement, which is adjusted to yield the narrowest and smoothest possible energy spread.

Notation and Naming

In the following sections, positions of loudspeakers $\{\boldsymbol{\theta}_l\}_{l=1,...,L}$ and the panning direction $\boldsymbol{\theta}_s$ are vectors of unit length $\boldsymbol{\theta} = [\cos(\varphi)\sin(\vartheta),\ \sin(\varphi)\sin(\vartheta),\ \cos(\vartheta)]^T$ that depend on the azimuth and zenith angle φ and ϑ, respectively. The panning direction represents the direction of the virtual source. The scalar product between two vectors is denoted as $\langle \boldsymbol{a}, \boldsymbol{b} \rangle = \boldsymbol{a}^T \boldsymbol{b} = \sum_k a_k b_k$. Non-italic symbols express constant values or labels.

In general, amplitude panning distributes each virtual source signal $s(t)$ scaled by the set of gains $\{g_l\}_{l=1,...,L}$ to L loudspeakers. As shown in Fig. 1, the loudspeaker signals $x_l(t)$ are obtained from an input signal $s(t)$ multiplied by the weights g_l

$$x_l(t) = g_l\, s(t), \tag{1}$$

alternatively g_l and $x_l(t)$ can be formulated as vectors

$$\boldsymbol{x}(t) = \boldsymbol{g}\, s(t), \qquad \text{with } \boldsymbol{g} = [g_1, \ldots, g_L]^T. \tag{2}$$

An entire audio scene is described by the parameters and signals of several simultaneous source objects. Such a scene description is called object-based [30]. In the loudspeaker channels several virtual sources signals $s_k(t)$ can be mixed, each one weighted by its own set of gains. This signal representation is referred to as channel-based.

1 VECTOR-BASE AMPLITUDE PANNING (VBAP)

VBAP is well described in the paper of Ville Pulkki [5]. However, for this paper a slight extension of its basic version is required, mainly concerning additional geometrical considerations. Therefore, VBAP is introduced with a brief explanation.

VBAP calculates three weights $\boldsymbol{g}_{ijk} = [g_i,\ g_j,\ g_k]^T$ applicable to creating the impression of a phantom source between three loudspeakers located at $\mathbf{L}_{ijk}=[\boldsymbol{\theta}_i, \boldsymbol{\theta}_j, \boldsymbol{\theta}_k]$. The unnormalized weights $\tilde{\boldsymbol{g}}_{ijk}$ are calculated from the panning direction $\boldsymbol{\theta}_s$ by the following equation

$$\mathbf{L}_{ijk}\, \tilde{\boldsymbol{g}}_{ijk} = \boldsymbol{\theta}_s \qquad \Rightarrow \tilde{\boldsymbol{g}}_{ijk} = \mathbf{L}_{ijk}^{-1}\, \boldsymbol{\theta}_s, \tag{3}$$

and subsequent normalization yields the weights $\boldsymbol{g}_{ijk} = \tilde{\boldsymbol{g}}_{ijk} / \|\tilde{\boldsymbol{g}}_{ijk}\|$ for panning. In order to give good results, the weights must be positive, which is only obtained for virtual sources lying inside a loudspeaker triplet expressed by $\mathbf{L}_{ijk}$. To extend the panning range around this triplet, more loudspeaker triplets need to be attached, cf. [5].

Considering a surrounding arrangement of loudspeakers at the directions $\{\boldsymbol{\theta}_l\}$, $l = 1, \ldots, L$, one would first like to know a definition of suitable loudspeaker triplets. In general, best results can be expected if all triplets are as small as possible. It is desirable that loudspeaker triplets should be available for all surrounding directions. The smallest triplets of neighboring loudspeakers are best defined by a geometric triangulation algorithm [31] that constructs the *convex hull* $\mathcal{H}$ of the set of direction vectors. To make phantom sources work, the loudspeaker triplets of the convex hull $(i, j, k) \in \mathcal{H}$ must not have an aperture larger than *roughly 90°*, cf. [32, p. 29]. Moreover, each oriented triangle of the convex hull has an interior and exterior side. Triangles for which the listener lies at the exterior side must be removed from $\mathcal{H}$. Regarding the oriented aperture angle, this corresponds to angles $\geq 180°$. Therefore only the triplet $(i, j, k)_{(\boldsymbol{\theta}_s)}$ whose aperture is $\leq 90°$, i.e., $(i, j, k)_{(\boldsymbol{\theta}_s)} \in \mathcal{H}_{\leq 90°}$, and whose weights are all non-negative $g_i, g_k, g_k \geq 0$ after evaluating Eq. (3) is selected for VBAP.

The gains g_l for all loudspeakers $l = 1, \ldots, L$ can be represented using the gain entries $\boldsymbol{g}_{ijk}$ for the active triplet and the Kronecker-Delta symbols δ_{li} that are one if $l = i$ and zero otherwise,

$$g_l = g_i\, \delta_{li} + g_j\, \delta_{lj} + g_k\, \delta_{lk}. \tag{4}$$

The loudspeaker signals are found from the input s according to Eq. (2).

Known Issues

The number of active loudspeakers varies depending on where the virtual source lies. Note that a solution for the resulting modulation of the spread has been partly provided by Pulkki [33], however an extended method complying with Ambisonics will be presented in this paper. In the basic version of VBAP the cases depending on the virtual source direction are:

- Within loudspeaker triplet: 3 active loudspeakers,
- Between two loudspeakers: 2 active loudspeakers,
- On one loudspeaker: 1 active loudspeaker,
- $\boldsymbol{\theta}_s$ outside $\mathcal{H}_{\leq 90°}$: 0 active loudspeakers.

In the last case, the source signal vanishes entirely.

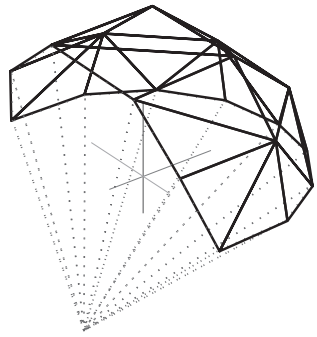
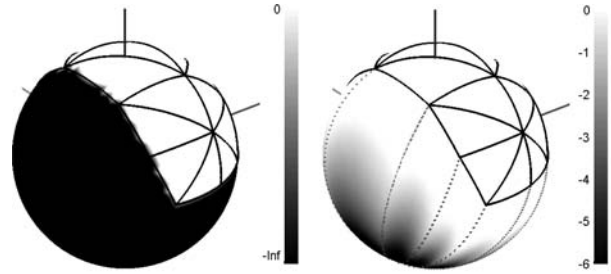


Fig. 2. Convex hull triangulation $\mathcal{H}_{\leq 90^\circ}$ for VBAP (solid lines) and the imaginary loudspeaker (dotted lines). This plot considers the exemplary arrangement of 19 loudspeakers specified in Table 3.



(a) standard VBAP-mapping (b) VBAP-mapping with imaginary loudspeaker

Fig. 4. Total signal power in dependence of virtual source position in dB (exemplary loudspeaker arrangement from Table 3).

1.1 Preserving Signals Outside the Panning Range of VBAP

To retain sounds of virtual sources that lie outside the admissible triangulation $\mathcal{H}_{\leq 90^\circ}$, while accepting localization mismatch in this case, one additional *imaginary* loudspeaker can be introduced, whose signal is omitted later, see Fig. 2. Roughly, its direction $\boldsymbol{\theta}_{\mathrm{L}+1}$ can be defined as opposing the *sum of weighted surface normals of the admissible triangulation*. If the admissible triangulation has more holes, an imaginary loudspeaker is positioned above each hole. For this purpose, an oriented rim is formulated by the edge vectors along the rim of the hole. Summing the cross-products of all pairwise neighboring vectors thereof yields a suitable position.

Due to the new imaginary loudspeaker(s), the convex hull triangulation encloses the listener in all thinkable cases with more than two loudspeakers. New virtual triangles are not excluded when their aperture is larger than 90°. Including the imaginary loudspeaker(s), VBAP is implemented according to the above description. Signals whose virtual source direction $\boldsymbol{\theta}_{\mathrm{s}}$ lies outside the admissible panning range are hereby better preserved and played back by the closest pair of loudspeakers. However, signals of virtual sources aligned with the imaginary loudspeaker(s) are entirely suppressed as they yield all-zero gains for the available, real loudspeakers, see Fig. 4(b).

2 PANNING USING AMBISONICS

Instead of discrete gains and vectors, Ambisonics uses a continuous *virtual panning function* of limited angular resolution. Ideally, this continuous function describes weights for panning on a continuous distribution of loudspeakers, cf. [34], which is, of course, not directly applicable in practice. Nevertheless, it allows for the adjustment of the angular resolution so that later discretization at optimal loudspeaker locations becomes possible. The continuous virtual panning function is based on the equivalence of a Dirac delta function at $\boldsymbol{\theta}_{\mathrm{s}}$ on the unit sphere $\mathbb{S}^2$ to a weighted sum of Legendre polynomials $P_n(\mu)$, see Table 1 and Appendix Eq. (33). The polynomials are evaluated at the scalar product between the direction $\boldsymbol{\theta}$ and the panning direction $\boldsymbol{\theta}_{\mathrm{s}}$

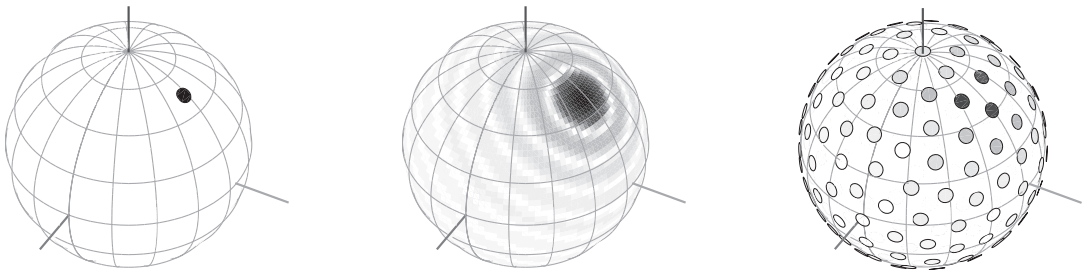
$$\delta(1 - \langle \boldsymbol{\theta}_{\mathrm{s}}, \boldsymbol{\theta} \rangle) = \sum_{n=0}^{\infty} \frac{2n+1}{4\pi} \, P_n(\langle \boldsymbol{\theta}_{\mathrm{s}}, \boldsymbol{\theta} \rangle). \tag{5}$$

The function represents an infinitesimally small dot at $\boldsymbol{\theta}_{\mathrm{s}}$, see Fig. 3(a). The Dirac delta function for the unit vectors $\boldsymbol{\theta}_{\mathrm{s}}$ and $\boldsymbol{\theta}$ is defined as

$$\delta(1 - \langle \boldsymbol{\theta}_{\mathrm{s}}, \boldsymbol{\theta} \rangle) = \begin{cases} \infty, & \text{at } \langle \boldsymbol{\theta}_{\mathrm{s}}, \boldsymbol{\theta} \rangle = 1, \\ 0, & \text{else,} \end{cases} \tag{6}$$

$$\text{and} \int_{\mathbb{S}^2} \delta(1 - \langle \boldsymbol{\theta}_{\mathrm{s}}, \boldsymbol{\theta} \rangle) \, \mathrm{d}\boldsymbol{\theta} = 1.$$

In the above representation in Legendre polynomials, the angular resolution of the dot can be ideally limited by truncating the sum in Eq. (5) to $n \leq \mathrm{N}$. Before discretizing the



(a) unlimited resolution (b) finite resolution (c) discretized distribution at 169 directions

Fig. 3. Virtual gain function in Ambisonics with ideally limited angular resolution.

Table 1. Legendre polynomials with the variable argument $\mu = \langle \boldsymbol{\theta}_s, \boldsymbol{\theta} \rangle$.

$P_n(\mu) = \sum_{k=0}^{n} c_k \mu^k$										
order $n =$	c_9	c_8	c_7	c_6	c_5	c_4	c_3	c_2	c_1	c_0
0										1
1									1	
2								$\frac{3}{2}$		$-\frac{1}{2}$
3							$\frac{5}{2}$		$-\frac{3}{2}$	
4						$\frac{35}{8}$		$-\frac{30}{8}$		$\frac{3}{8}$
5					$\frac{63}{8}$		$-\frac{70}{8}$		$\frac{15}{8}$	
6				$\frac{231}{16}$		$-\frac{315}{16}$		$\frac{105}{16}$		$-\frac{5}{16}$
7			$\frac{429}{16}$		$-\frac{693}{16}$		$\frac{315}{16}$		$-\frac{35}{16}$	
8		$\frac{6435}{128}$		$-\frac{12012}{128}$		$\frac{6930}{128}$		$-\frac{1260}{128}$		$\frac{35}{128}$
9	$\frac{12155}{128}$		$-\frac{25740}{128}$		$\frac{18018}{128}$		$-\frac{4620}{128}$		$\frac{315}{128}$	

virtual panning function at the loudspeaker directions, the adjustment of the truncation order N is utilized to minimize signal loss when panning between the loudspeakers, cf. Fig. 3(c). The resulting rotationally symmetric spread is shown in Fig. 3(b), and it contains typical but undesirable side-lobes. These are suppressed by inclusion of suitable weights a_n to obtain the virtual panning function

$$g(\boldsymbol{\theta}) = \sum_{n=0}^{\mathrm{N}} \frac{2n+1}{4\pi}\, a_n\, P_n(\langle \boldsymbol{\theta}_\mathrm{s}, \boldsymbol{\theta} \rangle). \tag{7}$$

Normally, Ambisonics is expressed by the spherical harmonics, see Section 6. For panning, the simplified polynomial expression based on the spherical harmonics addition theorem [36] is fully equivalent and has been chosen to maintain simplicity.

2.1 Panning-Invariant Energy and Spread

The virtual panning function $g(\boldsymbol{\theta})$ only changes its orientation depending on the panning direction $\boldsymbol{\theta}_\mathrm{s}$ but not its shape. Therefore its energy is panning-invariant, which becomes obvious after integrating over the squared virtual panning function $E = \int_{\mathbb{S}^2} |g(\langle \boldsymbol{\theta}, \boldsymbol{\theta}_\mathrm{s} \rangle)|^2 \, \mathrm{d}\boldsymbol{\theta}$ from Eq. (7), cf. Appendix Eq. (34),

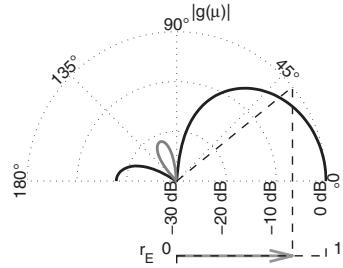
$$E = \sum_{n=0}^{\mathrm{N}} \frac{2n+1}{4\pi} |a_n|^2. \tag{8}$$

As descriptor of the spread of the rotationally symmetric virtual panning function, the $\boldsymbol{r}_\mathrm{E}$ vector can be used, cf. [10, 35, 11]. It can be interpreted as a Cartesian center of mass of the energetically weighted unit sphere using the weights $|g(\boldsymbol{\theta})|^2$ of Eq. (7), i.e., as the *energy centroid*. The corresponding $\boldsymbol{r}_\mathrm{E}$ vector is oriented toward $\boldsymbol{\theta}_\mathrm{s}$. Its derivation is given in the Appendix Eq. (35) and yields known results, cf. [11, p. 312],

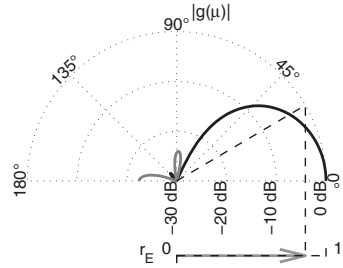
$$\boldsymbol{r}_\mathrm{E} = \boldsymbol{\theta}_\mathrm{s}\, \frac{\sum_{n=0}^{\mathrm{N}} (n+1)\, \Re\{a_n\, a_{n+1}^*\}}{2\pi\, E}. \tag{9}$$

The $\boldsymbol{r}_\mathrm{E}$ vector has a length between 0 and 1, which is indicated by the vector symbol in Fig. 5. If all spherical weights $|g(\boldsymbol{\theta})|^2$ are equal or balanced, the energy centroid lies in the center of the sphere, thus $\boldsymbol{r}_\mathrm{E}$ equals zero. On the other hand, if all the weight is sharply located $|g(\boldsymbol{\theta})|^2 = \delta(1 - \langle \boldsymbol{\theta}_\mathrm{s}, \boldsymbol{\theta} \rangle)^2$, the $\boldsymbol{r}_\mathrm{E}$ vector is unity and points to $\boldsymbol{\theta}_\mathrm{s}$.

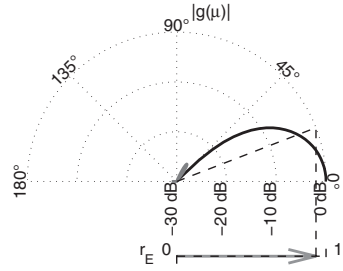
In general, the so-called max-$\boldsymbol{r}_\mathrm{E}$ weights are convenient, cf. [35, 11], because they optimize the energy concentration toward $\boldsymbol{\theta}_\mathrm{s}$ by maximizing the length $\|\boldsymbol{r}_\mathrm{E}\|$. Hereby, side-lobes are suppressed. The optimal solution, see Fig. 5, is



(a) N = 2



(b) N = 3



(c) N = 5

Fig. 5. Ambisonic panning functions, Eq. (7), of different truncation orders N with max-$\boldsymbol{r}_\mathrm{E}$ weights a_n, Eq. (10), cf. [35].

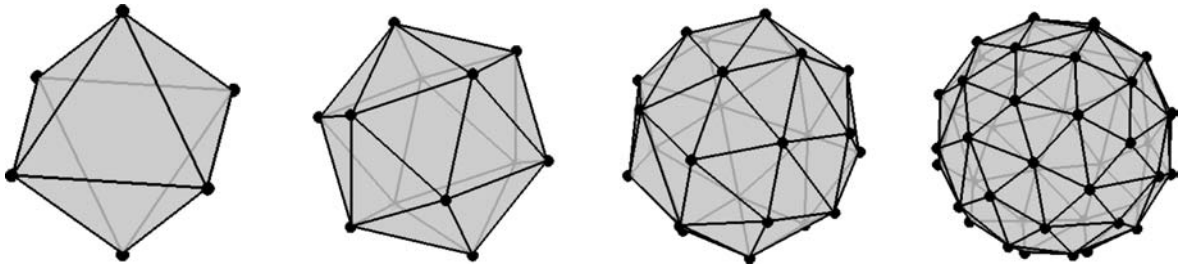


Fig. 6. t-designs: $t = 3$(octahedron), 5(icosahedron), 7, 9.

given in [35, 11] and derived in the Appendix Eq. (37), and it can be sufficiently well approximated by

$$a_n = P_n\left(\cos\left(\frac{137.9^\circ}{\mathrm{N}+1.51}\right)\right). \tag{10}$$

2.2 SPHERICAL T-DESIGN LOUDSPEAKER ARRANGEMENTS

The panning invariance of the measures E and $\|\boldsymbol{r}_\mathrm{E}\|$ is usually lost after the discretization of the virtual panning function by loudspeakers at the directions $\{\boldsymbol{\theta}_l\}$. Only special arrangements of loudspeakers permit direct sampling that retains both panning-invariant E and $\|\boldsymbol{r}_\mathrm{E}\|$: *spherical t-designs with* $t \geq 2\mathrm{N}+1$, cf. [37]. Examples of these special discretization schemes of the sphere are depicted in Fig. 6.

By definition, cf. [38], a t-design discretizes an arbitrary spherical polynomial $\mathcal{P}_n(\mu)$ of limited degree $n \leq t$ such that the integral and the discrete sum are equivalent

$$2\pi \int_{-1}^{1} \mathcal{P}_n(\mu)\,\mathrm{d}\mu = \frac{4\pi}{\mathrm{L}} \sum_{l=1}^{\mathrm{L}} \mathcal{P}_n(\mu_l). \tag{11}$$

Herein, the continuous variable is $\mu = \langle \boldsymbol{\theta}_\mathrm{s}, \boldsymbol{\theta} \rangle$ and the discrete one $\mu_l = \langle \boldsymbol{\theta}_\mathrm{s}, \boldsymbol{\theta}_l \rangle$; the direction $\boldsymbol{\theta}$ is a continuous variable, and $\{\boldsymbol{\theta}_l\}$ are the discrete L directions of a spherical t-design.

For a *panning-invariant energy* E, we desire that the energy integral, cf. Appendix Eq. (34), is equivalent to the sum over the discrete samples $\boldsymbol{\theta}_l$. To achieve that, the discretized Legendre polynomials must stay orthogonal. With a t-design, a discrete summation is equivalent to the integral expressing the orthogonality of two Legendre polynomials if $t \geq n + n'$

$$\begin{aligned} 2\pi \int_{-1}^{1} P_n(\mu) P_{n'}(\mu)\,\mathrm{d}\mu &= \frac{4\pi}{\mathrm{L}} \sum_{l} P_n(\mu_l) P_{n'}(\mu_l) \\ &= \frac{4\pi}{2n+1}\,\delta_{n'n}. \end{aligned} \tag{12}$$

A *panning-invariant energy centroid* $\boldsymbol{r}_\mathrm{E}$ is obtained for a t-design if the integral over one polynomial degree higher is equivalent to discrete summation, see $\mu P_{n'}(\mu)\, P_n(\mu)$ in Eq. (35), therefore $t \geq n + n' + 1$

$$2\pi \int_{-1}^{1} P_n(\mu) P_{n'+1}(\mu)\,\mathrm{d}\mu = \frac{4\pi}{\mathrm{L}} \sum_{l} P_n(\mu_l) P_{n'+1}(\mu_l). \tag{13}$$

For this reason, t-designs with $t \geq 2\mathrm{N}+1$ are the optimal loudspeaker arrangements for Ambisonic panning of the truncation order N as both energy and energy centroid match those of the virtual panning function.

2.3 t-Design Ambisonic Panning

Finally, we obtain signals of optimally placed loudspeakers by discretizing the virtual panning function Eq. (7).

$$g_l = \sum_{n=0}^{\mathrm{N}} a_n\, \frac{2n+1}{4\pi}\, P_n\left(\langle \boldsymbol{\theta}_\mathrm{s}, \boldsymbol{\theta}_l \rangle\right), \tag{14}$$

using a $(2\mathrm{N}+1)$-design $\{\boldsymbol{\theta}_l\}$. Essentially the signal processing structure is the same as in Fig. 1.

Known Issues

In playback situations, it will hardly ever be possible to mount speakers exactly at the locations of a t-design. In practice often compromises have to be made concerning the region covered by loudspeakers and the actual position of each. In such cases, careful Ambisonic decoder design becomes necessary with all its difficulties known from literature, cf. [39]. To obtain a solution whose behavior is easier to predict, a simple alternative solution is desirable.

3 PERFORMANCE MEASURES FOR DISCRETE PANNING FUNCTIONS

Similar to the continuous measures defined by E, Eq. (8), and $\boldsymbol{r}_\mathrm{E}$, Eq. (9), measures for evaluation and comparison of *discrete amplitude panning functions* are needed. In this case, Gerzon's energy and energy vector measures are directly applicable, cf. [10]. The first measure estimates the over-all loudness of the sound distributed by the loudspeakers as the energy of all gains (normalized using $\frac{4\pi}{\mathrm{L}}$)

$$\hat{E} = \frac{4\pi}{\mathrm{L}} \sum_{l=1}^{\mathrm{L}} g_l^2 = \frac{4\pi}{\mathrm{L}} \|\boldsymbol{g}\|^2, \tag{15}$$

which should ideally be panning-invariant. This energy estimator assumes an incoherent superposition of the loudspeaker signals at arbitrary listening positions and unknown acoustic properties of the playback room.

The second measure re-interprets the length of the $\hat{\boldsymbol{r}}_\mathrm{E}$ vector in terms of an angular spread, see Fig. 7. A unit sphere that is cropped at the distance $\|\hat{\boldsymbol{r}}_\mathrm{E}\|$ from its center

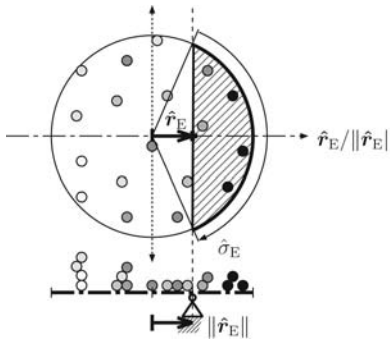


Fig. 7. The length of $\hat{\boldsymbol{r}}_\mathrm{E}$ describes how distant the centroid of the weighted sphere is from its center. Cropping the sphere at $\|\hat{\boldsymbol{r}}_\mathrm{E}\|$ yields the spread $\hat{\sigma}_E$ as cap size. The centroid refers to a point at which the weighted sphere is balanced under gravity and supported without torque, symbolized by the weighted beam below.

by a plane yields a spherical cap of the angular size

$$\hat{\sigma}_\mathrm{E} = 2\arccos\{\|\hat{\boldsymbol{r}}_\mathrm{E}\|\}\ 180^\circ/\pi,$$
$$\text{using } \hat{\boldsymbol{r}}_\mathrm{E} = \sum_{l=1}^{\mathrm{L}} \boldsymbol{\theta}_l\, g_l^2 \Big/ \sum_{l=1}^{\mathrm{L}} g_l^2. \tag{16}$$

The psychoacoustic relevance of this measure, i.e., the correlation between the computed angular spread and the perceived source width, does not seem to be fully established yet by psychoacoustical experimentation, which has just been started cf. [40]. Logically regarded, looking toward the panning direction, the calculated angular spread indicates the increase of lateral energy. From literature about room acoustics, it is well known that lateral energy is correlated with the perceived source width, cf. [41].

Both measures are applicable to any amplitude panning law that provides a set of gains $\{g_l\}$ for a virtual source direction $\boldsymbol{\theta}_s$ on the discrete loudspeaker arrangement $\{\theta_l\}$.

Performance Measures for *t*-Design Ambisonics

As described by Eqs. (12) and (13), using a t-design with $t \geq 2\mathrm{N}+1$ and max-$\boldsymbol{r}_\mathrm{E}$ weights, both performance measures are independent of the panning direction $\boldsymbol{\theta}_\mathrm{s}$ and equivalent to those of the virtual panning function:

$$\hat{E} = \sum_{n=0}^{\mathrm{N}} \frac{2n+1}{4\pi}\,|a_n|^2, \qquad \hat{\sigma}_\mathrm{E} = 2\,\frac{137.9^\circ}{\mathrm{N}+1.51}. \tag{17}$$

Performance Measures for VBAP

Due to its gain normalization in Eq. (3), VBAP ensures

$$\hat{E} = 1. \tag{18}$$

However, the angular spread $\hat{\sigma}_\mathrm{E}$ varies and rather tends to be as small as possible than as uniform as possible.

4 ALL-ROUND AMBISONIC PANNING: ALLRAP

Both above-mentioned panning approaches accomplish panning-invariant energy for arbitrary directions $\boldsymbol{\theta}_\mathrm{s}$. However, only VBAP supports arbitrary loudspeaker arrangements $\{\overset{\scriptscriptstyle\triangle}{\boldsymbol{\theta}}_l\}$, whereas only Ambisonic panning with t-design arrangements $\{\overset{\scriptscriptstyle\circ}{\boldsymbol{\theta}}_j\}$ provides a panning-invariant angular spread $\hat{\sigma}_\mathrm{E}$, see Table 2.

Table 2. Pros and Cons of presented panning approaches.

	VBAP	t-design Ambi $t \geq 2\mathrm{N}+1$	AllRAP
$\hat{E} = \text{const.}$	+	+	$\tilde{+}$
$\hat{\sigma}_\mathrm{E} = \text{const.}$	-	+	$\tilde{+}$
$\{\boldsymbol{\theta}_l\} = \text{arbitrary}$	+	-	+

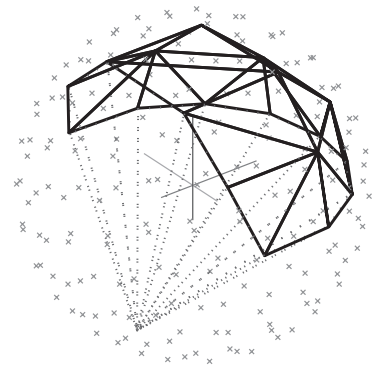


Fig. 8. AllRAP triangulation of the exemplary loudspeaker arrangement of Table 3, completed by one additional imaginary loudspeaker below; the crosses show the 180 virtual loudspeakers of an 18-design.

Batke and Keiler [42, 43] describe an approach to combine VBAP and Ambisonics. This approach uses VBAP for panning, but each real loudspeaker is replaced by a static, virtual loudspeaker at the same position. The virtual loudspeakers are rendered on the real ones using Ambisonics. In this way, per each virtual loudspeaker always several real loudspeakers are used. Hereby the tendency of VBAP to create the smallest spread for every virtual source is reduced and panning becomes uniform. However, still the approach relies on careful Ambisonic decoder design for rendering the virtual loudspeakers, which can be numerically ill-posed in some cases.

The AllRAP approach presented here uses a combination of Ambisonics and VBAP in the reverse order and with different arrangements. It takes into account that Ambisonic panning is ideally and safely achieved on a set of J t-design loudspeakers $\{\overset{\scriptscriptstyle\circ}{\boldsymbol{\theta}}_j\}$ as described above, see gray crosses in Fig. 8. These J directions are taken as virtual loudspeakers and rendered on the L real loudspeakers $\{\overset{\scriptscriptstyle\triangle}{\boldsymbol{\theta}}_l\}$ as J virtual sources using VBAP, see Fig. 8 and Eqs. (2) (3), which yields an L × J weight matrix

$$\overset{\scriptscriptstyle\triangle}{\mathbf{G}} = \left[\overset{\scriptscriptstyle\triangle}{\boldsymbol{g}}_1, \ldots, \overset{\scriptscriptstyle\triangle}{\boldsymbol{g}}_\mathrm{J}\right]. \tag{19}$$

For t-design Ambisonic panning on the J virtual loudspeakers, cf. Eq. (14), the following weight vector is used

$$\overset{\scriptscriptstyle\circ}{\boldsymbol{g}}(\boldsymbol{\theta}_\mathrm{s}) = \left[\overset{\scriptscriptstyle\circ}{g}_1(\boldsymbol{\theta}_\mathrm{s}), \ldots, \overset{\scriptscriptstyle\circ}{g}_\mathrm{J}(\boldsymbol{\theta}_\mathrm{s})\right]^\mathrm{T}. \tag{20}$$

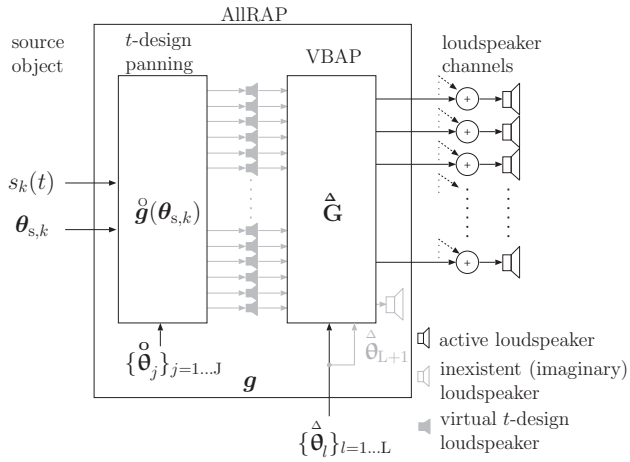


Fig. 9. In the block diagram of AllRAP, in which the panning function is composed of t-design Ambisonic panning and the multiple application of VBAP.

Putting together both equations, All-Round Ambisonic Panning (AllRAP) results in the weights

$$\boldsymbol{g} = \overset{\Delta}{\mathbf{G}}\ \overset{o}{\boldsymbol{g}}(\boldsymbol{\theta}_s). \tag{21}$$

The signal processing structure is the same as for general amplitude panning, cf. Fig. 1. The composition of the weights is described in the block diagram of Fig. 9. Note that the virtual loudspeaker signals it shows are not calculated in practice. The matrix-vector product in Eq. (21) only needs to be re-calculated when necessary due to source movement.

It is beneficial to choose a large number J of virtual loudspeakers that outnumbers the L given loudspeakers, $J \gg L$, in order to achieve uniform results. The triangulation and calculation of the static VBAP matrix in Eq. (19) is only done initially. Section 6.2 describes how the above result is used for Ambisonic decoding.

5 COMPARISON OF T-DESIGN AMBISONICS, VBAP, AND ALLRAP

The simulation results in Fig. 10 show the overall energy and angular spread using the exemplary loudspeaker coordinates from Table 3 for both VBAP and AllRAP. A 180 nodes 18-design available at [44] was employed. This 18-design is suited for Ambisonic orders $n < 9$. The optimal truncation order $N = 5$ for AllRAP was found by the method described in Section 7. The plane diagrams in Fig. 10 map the performance measures $\hat{E}(\boldsymbol{\theta}_s)$ and $\hat{\sigma}_E(\boldsymbol{\theta}_s)$, respectively, to grayscale pixels in the xy-plane, when applied to the exemplary loudspeaker arrangement. It shows only values for the panning directions $\boldsymbol{\theta}_s$ of the upper hemisphere that are covered by the loudspeakers. The horizontal panning directions map to a circle, the center of which refers to panning to the zenith.

Figs. 10 (a) and (b) show the known properties of VBAP: a panning-invariant overall energy $\hat{E}$ and a strong variation of the energy spread $\hat{\sigma}_E$. The new AllRAP technique, cf. Figs. 10 (c) and (d), offers nearly panning-invariant overall energy and spread.

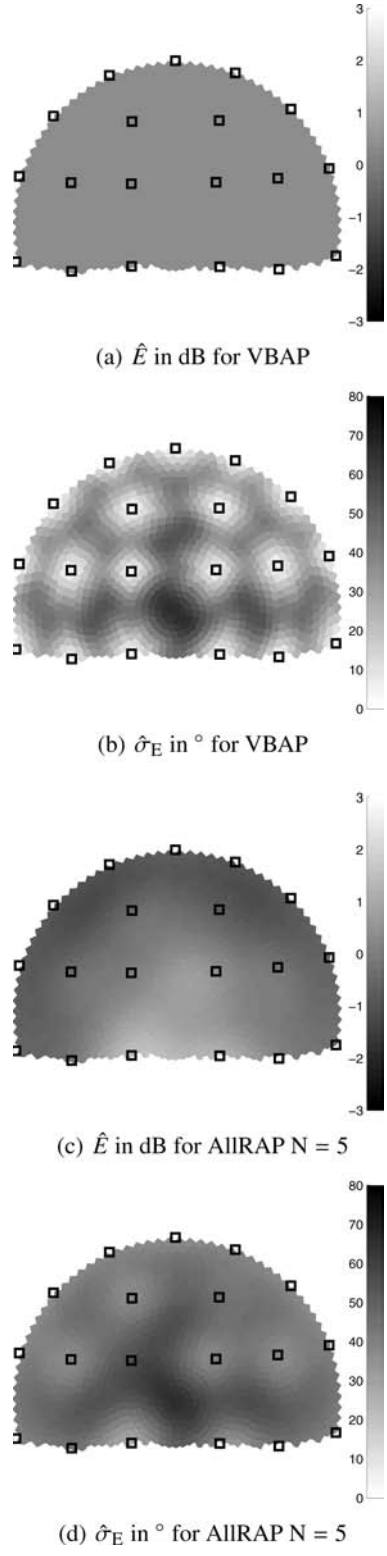


Fig. 10. Measures $\hat{E}$ (normalized to average) in dB and $\hat{\sigma}_E$ in ° for VBAP and AllRAP in gray scales after mapping the panning direction onto the plane. An exemplary loudspeaker arrangement is used (marked by squares; see Table 3); AllRAP uses a 180 nodes 18-design with $N = 5$.

Table 2 summarizes the pros and cons of VBAP compared to t-design Ambisonics. A trade-off is achieved with the proposed AllRAP approach.

Table 3. Azimuth and zenith angles of the exemplary arrangement of 19 loudspeakers that exists at the Institute of Electronic Music and Acoustics in Graz as a part of a 24 loudspeaker arrangement.

lspk. no.	azimuth angle φ in °	zenith angle ϑ in °	lspk. no.	azimuth angle φ in °	zenith angle ϑ in °	lspk. no.	azimuth angle φ in °	zenith angle ϑ in °
1	0.0	90.0	**10**	22.7	61.5	**16**	46.8	33.0
2	23.7	89.6	**11**	67.9	61.5	**17**	133.4	33.0
3	48.2	89.4	**12**	114.2	62.1	**18**	-133.4	33.4
4	72.6	89.3	**13**	-113.3	61.6	**19**	-43.4	32.3
5	103.1	89.4	**14**	-65.4	61.5			
6	-100.9	89.4	**15**	-22.7	62.0			
7	-69.8	89.6						
8	-44.8	89.5						
9	-21.4	89.5						

6 AMBISONIC ENCODING AND DECODING USING SPHERICAL HARMONICS

This section reviews some important features of Ambisonics. It hereby makes clear where the presented new method fits into the concept of Ambisonic decoding, i.e., when Ambisonics is not applied as a mere method of panning but also as a multichannel audio representation. Ambisonics with height is commonly formulated in spherical harmonics where it can be split up into *encoding* and *decoding*, similar as already proposed in an early work by Cooper [7] using azimuthal harmonics for horizontal-only playback. Ambisonic panning with height uses Legendre polynomials, Eq. (7), or equivalently the real-valued spherical harmonics, Appendix Eq. (41), that are related by the spherical harmonic addition theorem, cf. [45, Eq. (14.30.9)],

$$\frac{2n+1}{4\pi}\, P_n(\langle \boldsymbol{\theta}_\mathrm{s}, \boldsymbol{\theta}\rangle) = \sum_{m=-n}^{n} Y_n^m(\boldsymbol{\theta}_\mathrm{s})\, Y_n^m(\boldsymbol{\theta}), \tag{22}$$

and yield

$$g(\boldsymbol{\theta}) = \sum_{n=0}^{\mathrm{N}} \sum_{m=-n}^{n} a_n\, Y_n^m(\boldsymbol{\theta}_\mathrm{s})\, Y_n^m(\boldsymbol{\theta}). \tag{23}$$

In the spherical harmonic formulation, the variable direction $\boldsymbol{\theta}$ of the virtual panning function and the direction $\boldsymbol{\theta}_\mathrm{s}$ of the virtual source are split into two separate factors. The factor including $\boldsymbol{\theta}_\mathrm{s}$ defines the *encoder*, and a *decoder* can be defined that represents the dependency on the continuous direction variable $\boldsymbol{\theta}$ using loudspeakers at $\{\boldsymbol{\theta}_l\}_{l=1,\ldots,\mathrm{L}}$, cf. Fig. 11. Most importantly, encoder and decoder are distinct modules that can be tested and optimized individually, see, e.g., [46]. The concept of the intermediate multichannel Ambisonic representation allows to either add encoded sounds from object-based panning, or multichannel recordings from suitable microphone arrays, such as the Eigenmike for N = 3 or 4, cf. [21], or the classical Soundfield microphone with its B-format of N = 1, cf. [47]. By specifying playback directions, also audio formats based on standard loudspeaker arrangements, such as stereo, 5.1, or 22.2, cf. [48], can be embedded, cf. [46]. For playback, this representation can be decoded channel-based to more or less any given loudspeaker arrangement. Its spatial resolution is always limited to a certain order N for transmission and storage of audio scenes, cf. [46] [49].

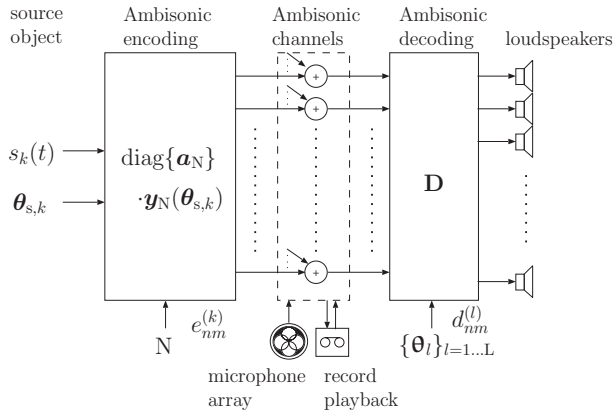


Fig. 11. Block diagram of Ambisonics with encoder and decoder: a virtual source is encoded and added to the Ambisonic channels. Ambisonic playback and Ambisonic microphone array signals can be mixed to these channels, or Ambisonic recordings taken from this mix. The decoder is a matrix that re-distributes Ambisonic channels to the loudspeakers.

6.1 Encoding and Decoding of Ambisonic Signals

Ambisonic signals $\chi_{nm}(t)$ obtained after encoding of K virtual sources are generally described as

$$\chi_{nm}(t) = \sum_{k=1}^{\mathrm{K}} e_{nm}^{(k)}\, s_k(t), \tag{24}$$

wherein $e_{nm}^{(k)}$ is the encoder that makes the k^{th} mono signal $s_k(t)$ appear at the direction $\boldsymbol{\theta}_{\mathrm{sk}}$. We define the encoder of one source as

$$e_{nm}^{(k)} = a_n\, Y_n^m(\boldsymbol{\theta}_{\mathrm{sk}}), \tag{25}$$

including the weights a_n, as we interpret them as a property of the virtual source. Typical literature, cf. [10] [11], employs them in the decoder instead, which is mathematically equivalent.

The Ambisonic representation can be employed as a production format for spatial audio material with many sources. It stores the sources efficiently in an uncompressed multi-

channel file as soon as they outnumber the spherical harmonics $K > (N + 1)^2$. Basically, the Ambisonic representation is not accompanied by restrictions to any specific microphone or loudspeaker arrangements during production or playback.

The loudspeaker signals $x_l(t)$ are obtained from $\chi_{nm}(t)$ by a decoder whose coefficients $d_{nm}^{(l)}$ can be defined in various ways, as described later,

$$x_l(t) = \sum_{n=0}^{N} \sum_{m=-n}^{n} d_{nm}^{(l)}\, \chi_{nm}(t). \tag{26}$$

Eqs. (24), (25), (26) for encoding and decoding altogether yield for one virtual source

$$g_l = \sum_{n=0}^{N} \sum_{m=-n}^{n} d_{nm}^{(l)}\, a_n\, Y_n^m(\boldsymbol{\theta}_s). \tag{27}$$

To simplify the notation, the following column vectors and matrices are useful

$$\begin{aligned} \boldsymbol{y}_N(\boldsymbol{\theta}) &= \left[Y_n^m(\boldsymbol{\theta})\right]_{n=0\ldots N, m=-n\ldots n}, \\ \boldsymbol{a}_N &= [a_n]_{n=0\ldots N, m=-n\ldots n}, \\ \mathbf{Y}_N &= [\boldsymbol{y}_N(\boldsymbol{\theta}_l)]^{l=1\ldots L}, \\ \mathbf{D} &= [d_{nm}^{(l)}]_{l=1\ldots L}^{n=0\ldots N, m=-n\ldots n}. \end{aligned}$$

Employing these definitions, encoding and decoding become, as shown in Fig. 11,

$$\boldsymbol{g} = \mathbf{D}\, \mathrm{diag}\{\boldsymbol{a}_N\}\, \boldsymbol{y}_N(\boldsymbol{\theta}_s), \tag{28}$$

and the open question is: *How to obtain a useful decoder matrix* $\mathbf{D}$*?*

Before defining All-Round Ambisonic Decoding (AllRAD), the interested reader is referred to the Appendix, in which the conventional decoder design approaches are summarized.

6.2 All-Round Ambisonic Decoding: AllRAD

Considering integration of conventional decoder design algorithms (Appendix) in professional sound equipment would make regularized matrix inversion or singular value decomposition necessary. As this might be too complicated for applied solutions, this section proposes decoder design avoiding matrix inversion exceeding 3×3. The notation in this section is based on Section 4 about AllRAP.

The spherical harmonics discretized with a suitable t-design can be used to define a simple decoder that is *sampling*, *mode matching*, and *energy preserving* at the same time, see Appendix. The required loudspeaker positions are not easy to set up in practice, but very easy to achieve if they are only virtual. Accordingly, the virtual Ambisonic decoder is simple:

$$\overset{\circ}{\mathbf{D}} = \frac{4\pi}{J}\, \overset{\circ}{\mathbf{Y}}{}_N^T. \tag{29}$$

Combining the virtual decoder with the VBAP rendering matrix Eq. (19) for the real loudspeakers, the following AllRAP-equivalent panning is achieved

$$\boldsymbol{g} = \frac{4\pi}{J}\, \overset{\triangle}{\mathbf{G}}\, \overset{\circ}{\mathbf{Y}}{}_N^T\, \mathrm{diag}\{\boldsymbol{a}_N\}\, \boldsymbol{y}_N(\boldsymbol{\theta}_s). \tag{30}$$

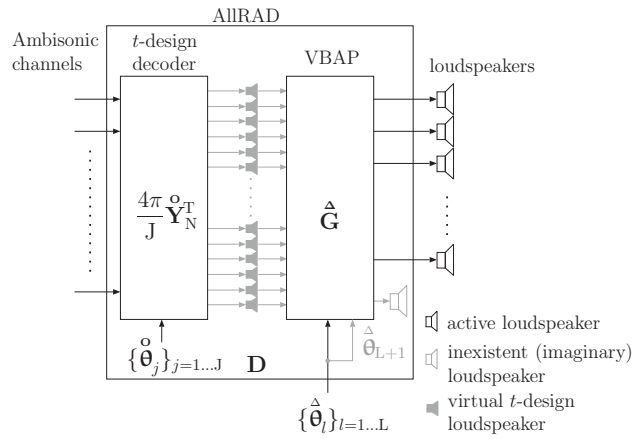


Fig. 12. Block diagram of the AllRAD: it decodes first with a simple t-design decoder to virtual loudspeakers that are rendered on the real loudspeaker arrangement by VBAP. In practice, by pre-multiplication of both the decoder and VBAP matrices, virtual signals need not be computed.

Comparison to Eq. (28) reveals the definition of *AllRAD*: it is the product of the $L \times J$ VBAP matrix with the $J \times (N + 1)^2$ virtual decoder as shown in Fig. 12

$$\mathbf{D} = \frac{4\pi}{J}\, \overset{\triangle}{\mathbf{G}}\, \overset{\circ}{\mathbf{Y}}{}_N^T. \tag{31}$$

The existence or numerical stability of AllRAD does not require uniformity of the loudspeaker arrangement or any specific relation between the maximum order N and the number of loudspeakers L. Nevertheless, the quality of the result depends on N and the geometry of the loudspeaker arrangement.

7 CHARACTERISTIC AMBISONIC TRUNCATION ORDER

The majority of conventional Ambisonic decoder designs, cf. Eqs. (41), (42), (43), for all orders $n \le N$ requires at least $(N + 1)^2$ loudspeakers uniformly covering a sphere. In principle, AllRAP and AllRAD are capable of Ambisonic playback using any N on any loudspeaker arrangement, even if the number of loudspeakers is small and they cover only a part of the surrounding directions, as in our exemplary loudspeaker arrangement. To ensure good results, the selection of a suitable truncation order N becomes necessary.

For this purpose the *equivalent max-*$\boldsymbol{r}_E$ *Ambisonic truncation order* is defined as a reformulation of the spread Eq. (17):

$$\hat{N}_e := 2\, \frac{137.9^\circ}{\hat{\sigma}_E} - 1.51 \tag{32}$$

to convert any spread $\hat{\sigma}_E$ of given discrete panning functions using Eq. (17), cf. Fig. 10.

Using the exemplary loudspeaker arrangement from Table 3, $\hat{\sigma}_E$ has been evaluated for a dense grid of panning directions on the loudspeaker arrangement, as in Fig. 10. After converting these to equivalent truncation orders

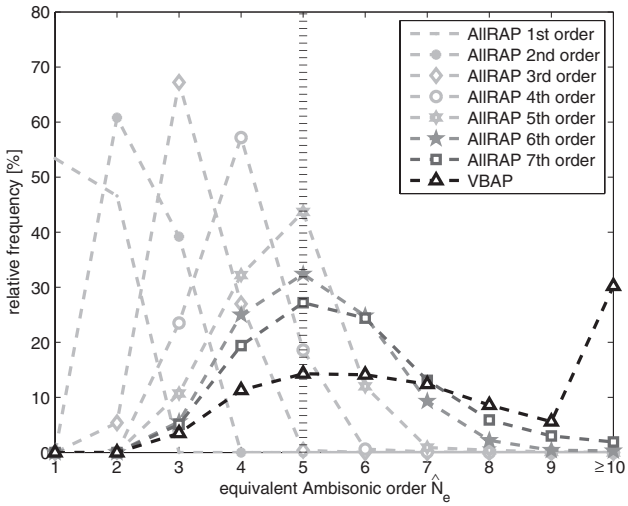


Fig. 13. Histograms of equivalent Ambisonic truncation orders $\hat{N}_e$ computed from energy spread for the exemplary loudspeaker arrangement of Table 3.

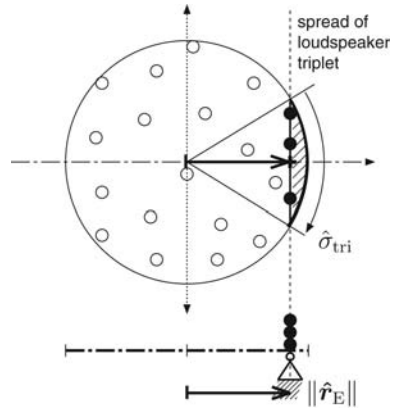


Fig. 14. Aperture angle $\hat{\sigma}_{\mathrm{tri}}$ of a loudspeaker triplet.

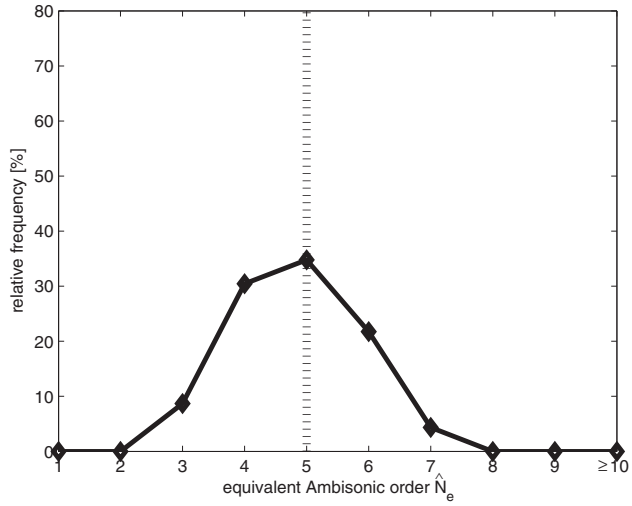


Fig. 15. Histogram of equivalent Ambisonic truncation orders $\hat{N}_e$ computed from triplet aperture angles $\hat{\sigma}_{\mathrm{tri}}$ of the exemplary loudspeaker arrangement of Table 3.

$\hat{N}_e$, Eq. (32), we calculate histograms showing the relative frequency of each truncation order for different settings. Fig. 13 shows histograms for AllRAP with N = 1, 2, 3, 4, 5, 6, 7 and for VBAP.

Interestingly, all virtual panning functions of the truncation orders $\mathrm{N} \geq 5$ peak at $\hat{N}_e = 5$, whereas all settings with $\mathrm{N} \leq 5$ yield most frequent values at the configured truncation order $\hat{N}_e = \mathrm{N}$ of the virtual panning function. Obviously the equivalent truncation order $\hat{N}_e = 5$ is a characteristic value for the exemplary loudspeaker arrangement.

Note that even VBAP yields the characteristic peak at $\hat{N}_e = 5$ (the second peak at $\hat{N}_e = 10$ is only due to fringing as the last interval collects all $\hat{N}_e \geq 9.5$ values in the histogram).

Summarizing, a virtual panning function of the truncation order $\mathrm{N} = 5$ seems to fit the exemplary loudspeaker arrangement as the equivalent spread occurs most often for several panning strategies. Therefore $\mathrm{N} = 5$ is considered as the *characteristic Ambisonic truncation order*.

As the characteristic Ambisonic truncation order only weakly depends on the panning strategy, there must be a simple way to find it based on the given loudspeaker arrangement. We can show that it is related to the average aperture angle of all loudspeaker triplets, cf. Fig. 14. The spread $\hat{\sigma}_{\mathrm{tri}}$ of one triplet corresponds to the spread obtained for VBAP with a virtual source centered in this triplet. The histogram containing this spread for all triplets in Fig. 15 indicates the same characteristic truncation order $\mathrm{N} = 5$ as Fig. 13 does.

Choosing the truncation order N of the panning functions correspondingly ensures the smoothest achievable Ambisonic panning at the highest achievable resolution.

8 CONCLUSION

This paper presented a novel and simple approach for panning of virtual sources with adjustable smoothness on any loudspeaker arrangement (All-Round Ambisonic Panning = AllRAP). It keeps the overall variation of the energy and the energy spread small. This is achieved by combining a slightly extended version of VBAP with Ambisonics. The approach is also applicable as decoder for the playback of Ambisonically recorded or encoded signals (All-Round Ambisonic Decoding = AllRAD). As our approach works also with loudspeaker arrangements that cover only parts of a sphere, it is suitable for upcoming surround with height formats such as, e.g., 22.2, cf. [48]. Complementary to the proposed AllRAP/AllRAD, we presented a method to determine the characteristic Ambisonic truncation order of a given loudspeaker arrangement. The use of this optimal order ensures the smoothest possible Ambisonic playback at the highest achievable directional resolution.

9 ACKNOWLEDGMENT

The authors gratefully thank Hannes Pomberger for proof-reading and fruitful discussions and the anonymous reviewers for comments that helped improving the paper.

Moreover, Alois Sontacchi encouraged the work on the paper and co-authored the first paper on this subject [37]. Our research is partly supported by the project AAP, which is funded by Austrian ministries BMVIT, BMWFJ, the Styrian Business Promotion Agency (SFG), and the departments 3 and 14 of the Styrian Government. The Austrian Research Promotion Agency (FFG) conducted the funding under the Competence Centers for Excellent Technologies (COMET, K-Project), a program of the above-mentioned institutions.

10 REFERENCES

[1] K. Wendt, *Das Richtungshören bei der Überlagerung zweier Schallfelder bei Intensitäts- und Laufzeitstereophonie*, Ph.D. thesis, Ph.D. thesis, RWTH Aachen, 1963.

[2] J. Blauert, *Spatial Hearing* (MIT Press, Cambridge, MA, 1983).

[3] G. Theile, *Über die Lokalisation im überlagerten Schallfeld*, Ph.D. thesis, Ph.D. thesis, Technische Universität Berlin, 1980.

[4] V. Pulkki, "Localization of Amplitude-Panned Virtual Sources II: Two- and Three-Dimensional Panning," *J. Audio Eng. Soc.*, vol. 49, no. 9, pp. 753-767 (2001 Sept.).

[5] V. Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning," *J. Audio Eng. Soc.*, vol. 45, no. 6, pp. 456-466 (1997 June).

[6] V. Pulkki and M. Karjalainen, "Localization of Amplitude-Panned Virtual Sources I: Stereophonic Panning," *J. Audio Eng. Soc.*, vol. 49, no. 9, pp. 739-752 (2001 Sept.).

[7] D. H. Cooper and T. Shiga, "Discrete-Matrix Multichannel Stereo," *J. Audio Eng. Soc.*, vol. 20, no. 5, pp. 346-360 (1972 June).

[8] M. A. Gerzon, "Ambisonics. Part Two: Studio techniques," *Studio Sound*, vol. 17, pp. 24-26 (1975).

[9] P. Fellget, "Ambisonics. Part One: General System Description," *Studio Sound*, vol. 17, pp. 20-22 (1975).

[10] M. A. Gerzon, "General Metatheory of Auditory Localization," presented at the 92nd Convention of the Audio Engineering Society (March, 1992), convention paper 3306.

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

[12] A. Sontacchi, *Dreidimensionale Schallfeldreproduktion für Lautsprecher- und Kopfhöreranwendungen*, Ph.D. thesis, Ph.D. thesis, TU Graz, 2003.

[13] M. A. Poletti, "Three-Dimensional Surround Sound Systems Based on Spherical Harmonics," *J. Audio Eng. Soc.*, vol. 53, no. 11, pp. 1004-1025 (2005 Nov.).

[14] D. Ward and T. Abhayapala, "Reproduction of a Plane-Wave Sound Field Using an Array of Loudspeakers," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 6, pp. 697-707 (2001 Sept.).

[15] J. Ahrens, *Analytic Methods of Sound Field Synthesis*(Springer, Berlin, 2011).

[16] D. Menzies and M. Al-akaidi, "Ambisonic Synthesis of Complex Sources," *J. Audio Eng. Soc.*, vol. 55, no. 10, pp. 864-876 (2007 Oct.).

[17] M. Kolundžija, C. Faller, and M. Vetterli, "Reproducing Sound Fields Using MIMO Acoustic Channel Inversion," *J. Audio Eng. Soc.*, vol. 59, no. 10, pp. 721-734 (2011 Oct.).

[18] P.-A. Gauthier and A. Berry, "Adaptive Wave Field Synthesis for Sound Field Reproduction: Theory, Experiments, and Future Perspectives," *J. Audio Eng. Soc.*, vol. 94, no. 12, pp. 1107-1124 (2007 Dec.).

[19] W. H. Chi, J. G. Ih, and M. M. Boone, "Holographic Design of a Source Array Achieving a Desired Sound Field," *J. Audio Eng. Soc.*, vol. 58, no. 4, pp. 282-298 (2010 Apr.).

[20] G. Cengarle, T. Mateos, and D. Bonsi, "A Second-Order Ambisonics Device Using Velocity Transducers," *J. Audio Eng. Soc.*, vol. 59, no. 9, pp. 656-668 (2011 Sept.).

[21] G. W. Elko, R. A. Kubli, and J. Meyer, "Audio System Based on at Least Second Order Eigenbeams," *Int Patent WO 03/061336 A1* (2003).

[22] V. Pulkki, "Spatial Sound Reproduction with Directional Audio Coding," *J. Audio Eng. Soc.*, vol. 56, no. 6, pp. 503-516 (2007 June).

[23] S. Bertet, J. Daniel, E. Parizet, and 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" in *Proc. Acoustics-08* (Joint ASA/EAA meeting, July, 2008).

[24] M. Frank and F. Zotter, "Localization Experiments Using Different 2D Ambisonics Decoders," in, *25. Tonmeistertagung, Leipzig* (VDT, Nov., 2008).

[25] M. Frank, *Perzeptiver Vergleich von Schallfeldreproduktionsverfahren unterschiedlicher räumlicher Bandbreite*, M thesis, M thesis, University of Music and Performing Arts, Institute of Electronic Music and Acoustics, Graz, and Deutsche Telekom Labs Berlin, 2009.

[26] A. Solvang, "Spectral Impairment of Two-Dimensional Higher Order Ambisonics," *J. Audio Eng. Soc.*, vol. 56, no. 4, pp. 267-279 (2008 April).

[27] G. Marentakis, N. Peters, and S. McAdams, "Auditory Resolution in Virtual Environments: Effects of Algorithms, Off-Centre Listening and Speaker Configuration," in *Proc. Acoustics-08, Paris* (Joint ASA/EAA meeting, July, 2008).

[28] M. Kratschmer and R. Rabenstein, "Implementing Ambisonics on a 48 Channel Circular Loudspeaker Array," in *Proc. 1st Ambisonics Symposium, Graz* (July, 2009).

[29] E. Bates, *The Composition and Performance of Spatial Music*, Ph.D. thesis, Ph.D. thesis, Trinity College Dublin, Dep. Electron. Electr. Eng., 2009.

[30] E. Scheirer, R. Vaananen, and J. Huopaniemi, "AudioBIFS: Describing Audio Scenes with the MPEG-4 Multimedia Standard," *IEEE Transactions on Multimedia*, vol. 1, no. 3, pp. 237-250 (1999 Sep.).

[31] C. B. Barber, D. P. Dobkin, and H. T. Huhdanpaa, "The Quickhull Algorithm for Convex Hulls," *ACM*

Transactions on Mathematical Software, vol. 22, no. 4, pp. 469–483 (1996).

[32] V. Pulkki, *Spatial Sound Generation and Perception by Amplitude Panning Techniques*, Ph.D. thesis, Ph.D. thesis, HUT, Finland, 2001.

[33] V. Pulkki, "Uniform Spreading of Amplitude Panned Virtual Sources," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, (Oct., 1999), pp. 187-190.

[34] Y. J. Wu and T. Abhayapala, "Soundfield Reproduction Using Theoretical Continuous Loudspeaker," in *International Conference on Acoustics, Speech and Signal Processing*, (IEEE, April, 2008), pp. 377-380.

[35] J. Daniel, J.-B. Rault, and J.-D. Polack, "Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions," presented at the 105th Convention of the Audio Engineering Society (1998 Sept.), convention paper 4795.

[36] E. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustic Holography* (Academic Press, London, UK , 1999).

[37] F. Zotter, M. Frank, and A. Sontacchi, "The Virtual t-Design Ambisonics-Rig Using VBAP," in *Proc. EAA EUROREGIO* (Sept., 2010).

[38] R. H. Hardin and N. J. A. Sloane, "McLaren's Improved Snub Cube and Other New Spherical Designs in Three Dimensions," *Discrete and Computational Geometry*, vol. 15, no. 4 (1996).

[39] M. A. Poletti, "Robust Two-Dimensional Surround Sound Reproduction for Nonuniform Loudspeaker Layouts," *J. Audio Eng. Soc.*, vol. 55, no. 7/8, pp. 568-610 (2007 Jul./Aug.).

[40] M. Frank, G. Marentakis, and A. Sontacchi, "A Simple Technical Measure for the Perceived Source Width," in, *Fortschritte der Akustik, DAGA*, (March, 2011).

[41] M. Barron and A. Marshall, "Spatial Impression due to Early Lateral Reflections in Concert Halls: The Derivation of a Physical Measure," *J. Sound and Vibration*, vol. 77, no. 2, pp. 211-232 (1981).

[42] J. M. Batke and F. Keiler, "Investigation of Robust Panning Functions for 3D Loudspeaker Setups," presented at the 128th Convention of the Audio Engineering Society (2010 May), convention paper 7979.

[43] J. M. Batke and F. Keiler, "Using VBAP-Derived Panning Functions for 3D Ambisonics Decoding," in *2nd Ambisonics Symposium* (May, 2010).

[44] R. H. Hardin and N. J. A. Sloane, "t-designs," http://www2.research.att.com/njas/sphdesigns/dim3/, accessed June 2012.

[45] F. W. J. Olver, R. F. Boisvert, and C. W. Clark (editors), *NIST Handbook of Mathematical Functions* (Cambridge University Press, Cambridge, UK, 2000), http://dlmf.nist.gov, accessed June 2012.

[46] M. A. Gerzon, "Hierarchical Transmission System for Multispeaker Stereo," *J. Audio Eng. Soc.*, vol. 40, no. 9, pp. 692-705 (1992 Sept).

[47] P. G. Craven and M. A. Gerzon, "Coincident Microphone Simulation Covering Three Dimensional Space and Yielding Various Directional Outputs," *US Patent* 4.042.779 (1975–1977).

[48] K. Hamasaki, T. Nishiguchi, K. Hiyama, and K. Ono, "Advanced Multichannel Audio Systems with Superior Impression of Presence and Reality," presented at the 116th Convention of the Audio Engineering Society (2004 May), convention paper 6053.

[49] C. Nachbar, F. Zotter, A. Sontacchi, and E. Deleflie, "Ambix – Suggesting an Ambisonic Format," in *3rd Ambisonics Symposium* (July, 2011).

[50] M. A. Poletti, "A Unified Theory of Horizontal Holographic Sound Systems," *J. Audio Eng. Soc.*, vol. 48, no. 12, pp. 1155-1182 (2000 Dec.).

[51] F. Zotter, H. Pomberger, and M. Noisternig, "Energy-Preserving Ambisonic Decoding," *Acta Acustica u. Acustica*, vol. 98, no. 1, pp. 37-47 (2012 Jan.).

APPENDIX

Dirac Delta in Legendre Polynomials

The Dirac delta function is equivalently expressed as [45, Eq. (1.17.22)]

$$\delta(\mu_0 - \mu) = \sum_{n=0}^{\infty} \frac{2n+1}{2} P_n(\mu_0) P_n(\mu). \tag{33}$$

By setting $\mu_0 = 1$, knowing $P_n(1) = 1$, and setting $\mu = \langle \boldsymbol{\theta}, \boldsymbol{\theta}_s \rangle$, it remains to re-normalize the above by 2π to obtain Eq. (5) with a unity surface integral as specified in Eq. (6).

Energy and Energy Centroid (r_E Vector) of a Virtual Panning Function

Inserting Eq. (7) into the energy integral and using the orthogonality of the Legendre polynomials, cf. [45, Eq. (14.17.6)], we obtain

$$\begin{aligned} E &= \int_{\mathbb{S}^2} |g(\langle \boldsymbol{\theta}, \boldsymbol{\theta}_s \rangle)|^2 \, \mathrm{d}\boldsymbol{\theta} = \int_0^{2\pi} \mathrm{d}\varphi \int_{-1}^{1} |g(\mu)|^2 \, \mathrm{d}\mu \qquad (34) \\ &= 2\pi \sum_{n=0}^{N} \sum_{n'=0}^{N} \frac{a_n \, a_{n'}^* \, (2n+1)(2n'+1)}{16\pi^2} \underbrace{\int_{-1}^{1} P_n(\mu) \, P_{n'}(\mu) \, \mathrm{d}\mu}_{=\frac{2}{2n'+1}\delta_{n'n}} \\ &= \sum_{n=0}^{N} \frac{2n+1}{4\pi} |a_n|^2. \end{aligned}$$

After inserting into the centroid integral over the squared weighting function, orthogonality, cf. [45, Eq. (14.17.6)], and the recurrence $\mu \, P_n(\mu) = \frac{n \, P_{n-1}(\mu) + (n+1) \, P_{n+1}(\mu)}{2n+1}$, cf. [45,

Eq. (14.10.3)], can be applied

$$\begin{aligned}\boldsymbol{r}_\mathrm{E} &= \frac{\int_{\mathbb{S}^2} \boldsymbol{\theta}\, |g(\boldsymbol{\theta})|^2\, \mathrm{d}\boldsymbol{\theta}}{\int_{\mathbb{S}^2} |g(\boldsymbol{\theta})|^2\, \mathrm{d}\boldsymbol{\theta}} = \boldsymbol{\theta}_\mathrm{s}\, \frac{\int_0^{2\pi} \mathrm{d}\varphi \int_{-1}^{1} \mu\, |g(\mu)|^2\, \mathrm{d}\mu}{E} \\ &= \boldsymbol{\theta}_\mathrm{s}\, \frac{2\pi \sum_{n,n'} \frac{a_n\, a_{n'}^*\, (2n+1)(2n'+1)}{16\pi^2} \int_{-1}^{1} \overbrace{\mu\, P_n(\mu)}\, P_{n'}(\mu)\, \mathrm{d}\mu}{E} \\ &= \boldsymbol{\theta}_\mathrm{s}\, \frac{\sum_{n,n'} \frac{a_n\, a_{n'}^*\, (2n'+1)}{8\pi} \int_{-1}^{1} \left[n\, P_{n-1}(\mu) + (n+1)\, P_{n+1}(\mu)\right] P_{n'}(\mu)\, \mathrm{d}\mu}{E} \\ &= \boldsymbol{\theta}_\mathrm{s}\, \frac{\sum_{n+1,n'} a_{n+1}\, a_{n'}^* \overbrace{\int_{-1}^{1} (2n'+1)(n+1)\, P_n(\mu)\, P_{n'}(\mu)\, \mathrm{d}\mu}^{=2(n+1)\delta_{n'n}}}{8\pi\, E} \\ &\quad + \frac{\sum_{n,n'+1} \overbrace{\int_{-1}^{1} a_n\, a_{n'+1}^*\, (2n'+3)(n+1)\, P_{n+1}(\mu)\, P_{n'+1}(\mu)\, \mathrm{d}\mu}^{=2(n+1)\delta_{n'n}}}{8\pi\, E} \\ &= \boldsymbol{\theta}_\mathrm{s}\, \frac{\sum_{n=0}^{\mathrm{N}} (n+1)\left[a_n\, a_{n+1}^* + a_n^*\, a_{n+1}\right]}{4\pi\, E} \\ \boldsymbol{r}_\mathrm{E} &= \boldsymbol{\theta}_\mathrm{s}\, \frac{\sum_{n=0}^{\mathrm{N}} (n+1)\, \Re\{a_n\, a_{n+1}^*\}}{2\pi\, E}. \end{aligned} \tag{35}$$

Maximization of the r_E Vector

In order to maximize the length of the $\boldsymbol{r}_\mathrm{E}$ vector for real-valued a_n

$$\|\boldsymbol{r}_\mathrm{E}\| = \frac{\tilde{r}_\mathrm{E}}{E}, \tag{36}$$

we differentiate it and equate the derivative to zero

$$\begin{aligned}\frac{\partial \|\boldsymbol{r}_\mathrm{E}\|}{\partial a_n} = 0 &= \frac{\frac{\partial}{\partial a_n}\tilde{r}_\mathrm{E} - \|\boldsymbol{r}_\mathrm{E}\|\, \frac{\partial}{\partial a_n} E}{E} \\ &= \frac{[n\, a_{n-1} + (n+1)\, a_{n+1}] - \|\boldsymbol{r}_\mathrm{E}\|\, (2n+1)\, a_n}{2\pi\, E} \\ \Rightarrow n\, a_{n-1} + (n+1)\, a_{n+1} &- \|\boldsymbol{r}_\mathrm{E}\|\, (2n+1)\, a_n = 0. \end{aligned} \tag{37}$$

The recurrence [45, Eq. (14.10.3)] of the Legendre polynomials $n\, P_{n-1}(\mu) + (n+1)\, P_{n+1}(\mu) - \mu\, (2n+1)\, P_n(\mu) = 0$ equals the recurrence of the optimum for the coefficients after setting $\mu = \|\boldsymbol{r}_\mathrm{E}\|$ and $a_n = P_n(\|\boldsymbol{r}_\mathrm{E}\|)$. Therefore the Legendre polynomials are optimal for $a_n = P_n(\|\boldsymbol{r}_\mathrm{E}\|)$. As a boundary condition, we desire that the weights should have a zero for the next higher order, i.e., $a_{\mathrm{N}+1} = P_{\mathrm{N}+1}(\|\boldsymbol{r}_\mathrm{E}\|) = 0$. In accordance with Daniel [35, 11] this means that the largest root of $P_{\mathrm{N}+1}(\max\|\boldsymbol{r}_\mathrm{E}\|)$ optimizes the length of the $\boldsymbol{r}_\mathrm{E}$ vector for the given order N. A simple approximation was found to model the location of this root for $\mathrm{N} \leq 1$. It was found by substituting $\vartheta_\mathrm{E}(N) = 180^\circ/\pi \arccos \|\max \boldsymbol{r}_\mathrm{E}(N)\|$ and linearly approximating its reciprocal for $N = 1, \ldots, 20$

$$\frac{1}{\vartheta_\mathrm{E}(N)} \approx k\, N + d = \frac{7.2536\, N + 10.9654}{1000}. \tag{38}$$

Eq. (10) only re-formulates this.

Addition theorem of the Legendre Functions and Spherical Harmonics

[45, Eq. (14.18.1)] becomes after replacing $P_n^{-m}(\cos\vartheta_1)$ by $(-1)^m \frac{(n-m)!}{(n+m)!} P_n^m(\cos\vartheta_1)$, cf. [45, Eq. (14.9.3)], noticing that $\cos(\alpha-\beta) = \cos(\alpha)\cos(\beta) + \sin(\alpha)\sin(\beta)$, $P_n^m = 0$ for $|m| > n$, re-definition of the left hand side argument $\cos\vartheta_1 \cos\vartheta_2 + \sin\vartheta_1 \sin\vartheta_2\, [\cos(\vartheta_1)\cos(\vartheta_2) + \sin(\vartheta_1)\sin(\vartheta)] =: \langle\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\rangle$, and gathering the terms under one sum by using $(2-\delta_{m,0})$ with the Kronecker delta $\delta_{m,0}$, which is 1 for $m = 0$ and zero else,

$$\begin{aligned}&P_n(\langle\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\rangle) = \\ &\sum_{m=0}^{n} \tfrac{(2-\delta_{m,0})(n-m)!}{(n+m)!}\, P_n^m(\cos\vartheta_1)\cos(m\varphi_1) \\ &\times P_n^m(\cos\vartheta_2)\cos(m\varphi_2) + \sum_{m=1}^{n} \tfrac{(2-\delta_{m,0})(n-m)!}{(n+m)!} \\ &P_n^m(\cos\vartheta_1)\sin(m\varphi_1) P_n^m(\cos\vartheta_2)\sin(m\varphi_2). \end{aligned} \tag{39}$$

Real-valued spherical harmonics $Y_n^m(\boldsymbol{\theta})$ fulfilling the spherical harmonic addition theorem, cf. Eq. (22), are defined after splitting the common factor $\frac{(2-\delta_{m,0})(n-m)!}{(n+m)!}$ from above and $\frac{2n+1}{4\pi}$ from the theorem into two square root factors and suppressing negative values by $|m|$ to let the sign of m select between sine or cosine with

$$\begin{aligned}&Y_n^m(\boldsymbol{\theta}) = \\ &\sqrt{\tfrac{(2n+1)\,(2-\delta_{m,0})\,(n-|m|)!}{4\pi\,(n+|m|)!}} \\ &\times P_n^{|m|}(\cos(\vartheta)) \begin{cases} \cos(m\varphi), & \text{for } m \geq 0 \\ \sin(|m|\varphi), & \text{otherwise.} \end{cases} \end{aligned} \tag{40}$$

Conventional Ambisonic Decoder Design Approaches

The *sampling decoder* directly discretizes the virtual panning function of Eq. (7) and applies a normalization $\frac{4\pi}{\mathrm{L}}$. The loudspeaker arrangement can be arbitrary but affects the result of the sampling decoder

$$\mathbf{D} = \frac{4\pi}{\mathrm{L}}\, \mathbf{Y}_\mathrm{N}^\mathrm{T}. \tag{41}$$

Alternatively, the so-called *mode-matching* equations have been considered, whose solution yield an indirect discretization of the virtual panning function, cf. [50]. Each loudspeaker is represented by the spherical harmonic spectrum $\boldsymbol{y}_\mathrm{N}(\boldsymbol{\theta}_l)$ of a Dirac delta. To re-produce the spherical harmonic spectrum $\boldsymbol{y}(\boldsymbol{\theta}_\mathrm{s})$ of a virtual source, the spectra of the loudspeakers are superimposed by weights $\boldsymbol{g}$. Considering weights of the form $\boldsymbol{g} = \mathbf{D}\, \boldsymbol{y}(\boldsymbol{\theta}_\mathrm{s})$ and using these in a weighted sum of the loudspeaker spectra $\mathbf{Y}_\mathrm{N}\, \boldsymbol{g}$, the reproduction task can be expressed as a system of mode-matching equations $\boldsymbol{y}_\mathrm{N}(\boldsymbol{\theta}_s) = \mathbf{Y}_\mathrm{N}\, \boldsymbol{g}$. The matrix $\mathbf{D}$ is *mode matching* if

$$\mathbf{D} = \mathbf{Y}_\mathrm{N}^\mathrm{T}\left(\mathbf{Y}_\mathrm{N}\mathbf{Y}_\mathrm{N}^\mathrm{T}\right)^{-1} \tag{42}$$

exists. Despite mode-matching works in a central listening spot and anechoic spaces, it requires at least $\mathrm{L} \geq (\mathrm{N}+1)^2$ uniformly distributed loudspeakers to make the inversion well-posed. As further improvement, *energy preserving decoding* [51] has been proposed recently. It requires the same number of loudspeakers $\mathrm{L} \geq (\mathrm{N}+1)^2$ but imposes no uniformity constraint. It is based on the truncated singular value decomposition of the matrix $\mathbf{Y}_\mathrm{N} = \mathbf{U}\,\mathbf{S}\,\mathbf{V}^\mathrm{T}$. This yields an orthogonal square matrix $\mathbf{U}$, a rectangular matrix $\mathbf{V}$ with orthonormal columns, and a diagonal matrix

S, which define a decoder

$$\mathbf{D} = \frac{4\pi}{\mathrm{L}}\mathbf{V}\,\mathbf{U}^{\mathrm{T}}. \tag{43}$$

t-design Ambisonic decoding has special properties: due to Eqs. (12) and (22), the spherical harmonics discretized with a suitable *t*-design exhibit orthogonal rows. Utilizing these optimal loudspeaker arrangements, decoding by *sampling*, *mode matching*, and *energy preservation* yields the same result

$$\mathbf{D} = \mathbf{Y}^{\mathrm{T}}(\mathbf{Y}\mathbf{Y}^{\mathrm{T}})^{-1} \equiv \frac{4\pi}{\mathrm{L}}\mathbf{V}\,\mathbf{U}^{\mathrm{T}} \equiv \frac{4\pi}{\mathrm{L}}\,\mathbf{Y}^{\mathrm{T}}, \tag{44}$$

as the involved matrices simplify to $\mathbf{Y}\mathbf{Y}^{\mathrm{T}} = \frac{\mathrm{L}}{4\pi}\,\mathbf{I}$, $\mathbf{U} = \mathbf{I}$, $\mathbf{S} = \frac{\mathrm{L}}{4\pi}\mathbf{I}$, and $\mathbf{V} = \frac{4\pi}{\mathrm{L}}\mathbf{Y}$.

THE AUTHORS

Franz Zotter

Matthias Frank

Dr. Franz Zotter finished his doctoral degree with a thesis on "Analysis and Synthesis of Sound-Radiation with Spherical Arrays" in 2009 at the Institute of Music and Electronic Music (IEM), University of Music and Performing Arts (KUG), which he has joined in 2004 while locally finishing his diploma degree in electrical engineering - audio engineering. For intensifying work on his Ph.D. thesis he was awarded a research scholarship to pursue scientific works at CNMAT, UC Berkeley in the fall of 2007. With expertise in theory and experimental practice on holographic analysis and holophonic resynthesis of sound, he worked with exchange scholars and supervised various master students. Franz Zotter was awarded the DEGA Lothar-Cremer prize of the German Acoustical Society (DEGA) in 2012. Working as a senior scientist, he is involved in research projects, and he teaches acoustic holography and holophony, acoustic measurement techniques, and leads a seminar on algorithms in acoustics and computer music. Moreover, Franz is a member of the DEGA, co-organized the first Ambisonics Symposium in 2009 and the DAFx-10 as scientific chair, which were held in Graz, and he became a DAFx board member in 2011.

●

Matthias Frank received his Diploma in Electrical Engineering - Audio Engineering in 2009, for his thesis on perceptual comparison of spatial sound field reproduction techniques at the Institute of Electronic Music and Acoustics, University of Music and Performing Arts (KUG) in Graz. For the practical work on his thesis, he was a visiting research scholar at the Deutsche Telekom Labs, Technical University of Berlin, Germany. In 2009 he joined IEM as a University Assistant who teaches analytic hearing for the analysis of recordings, leads a recording technology seminar and a lab course for algorithms in computer music and acoustics. Matthias Frank is an expert on psychoacoustics with a focus on loudspeaker-based spatial audio and on product sound quality evaluation and design. His PhD research deals with spatial and timbral attributes of phantom sources created by multiple loudspeakers, such as in Ambisonics. He was involved in the organization of DAFx-10 and is proud to have Prof. Jens Blauert as the external supervisor of his PhD studies. He regularly plays drums in a band and percussion in various wind and symphony orchestras. Matthias is member of the AES Student Section Graz.