Bibliographical reference

Ambisonic Decoding With Constant Angular Spread, N. Epain, C. T. Jin and F. Zotter, Acta Acustica **vol. 100** (Number 5), 2014, pp. 928-936

DOI
https://doi.org/10.3813/AAA.918772

ACTA ACUSTICA UNITED WITH ACUSTICA
Vol. 100 (2014) 928 – 936

DOI 10.3813/AAA.918772

Ambisonic Decoding With Constant Angular Spread

N. Epain[1)], C. T. Jin[1)], F. Zotter[2)]

[1)] Computing and Audio Research Laboratory, The University of Sydney, Sydney, Australia.
[nicolas.epain, craig.jin]@sydney.edu.au

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

Summary

Ambisonic decoding refers to the process of reconstructing a sound field represented by spherical harmonic modes up to a given order. The issue with the spherical harmonic representation is that perfect reconstruction of the sound field is typically possible only within an area whose size is inversely proportional to the frequency. Therefore, in order to decode ambisonic signals for high-frequency sounds or a wide listening area, one has to rely on other criteria than the sound field reconstruction error. Classic criteria for the derivation of ambisonic decoding matrices are the total energy of the loudspeaker signals and the direction and norm of the so-called energy vector, which corresponds to the energy-weighted sum of the unit vectors pointing to the directions of the loudspeakers. The underlying idea behind using such criteria is that they are somewhat related to the perceptual attributes of the reproduced sound field. In particular, the norm of the energy vector can be interpreted in terms of the angular spread of energy across loudspeakers, which was recently shown to be correlated to the perceived width of virtual sources. In previous works, ambisonic decoding methods have been presented which yield a constant loudspeaker energy and minimal energy-vector direction mismatch across virtual source directions. However, in the case of irregular speaker layouts, these methods result in a varying angular spread across directions. In this paper we present a method for calculating ambisonic decoding matrices providing a nearly constant angular spread across source directions while maintaining a constant energy and very low energy-vector direction mismatch.

PACS no. 43.60.Fg, 43.38.Md

1. Introduction

The ambisonic or Higher-Order Ambisonic sound field reproduction technique [1, 2] relies on the decomposition of the sound field into spherical harmonic modes up to a given order, L. When playback is intended for only one listener, the loudspeaker signals are classically calculated such that the spherical harmonic modes of the sound field reconstructed by the loudspeakers are identical to that of the desired, virtual or recorded, sound field. This method of ambisonic playback is often referred to as *mode-matching* decoding. The objective of mode-matching decoding is to physically reproduce the desired sound field around the listener. However, limiting the summation of the spherical harmonic modes up to order L only provides an accurate approximation of the sound field inside a small region of space, the size of which is inversely proportional to the frequency [3]. For instance using an order-4 ambisonic system at 2 kHz, the sound field is accurately reproduced within a region of space roughly the size of the human head. At higher frequencies, the size of the accurate spatial region shrinks to less than the size of the human head. Of course, when playback is intended for more than one listener, the area in which the sound field is to be accurately reproduced increases and the maximum frequency for physically accurate sound field reconstruction decreases.

In order to improve the perceptual quality of ambisonic playback for high frequency sounds or for spaces accommodating a large audience, various techniques have been proposed. A common feature of these techniques is the use of some form of amplitude panning based on a vector model of the sound field. The inventor of ambisonics, Gerzon, used various vector models to characterize the properties of perceived sound objects generated using amplitude panning [4]. Optimization criteria based on the so-called *energy vector*[1] or on the loudspeaker energy appear especially useful for the design of ambisonic decoding matrices. This approach was used in recent work [5, 6, 7, 8] to gain control over, e.g., loudness in the situation where the loudspeakers are irregularly distributed in space. Other

Received 21 May 2014,
accepted 14 July 2014.

[1] The energy-weighted sum of the unit vectors pointing to the loudspeaker directions.

works with similar approaches were recently pre sented for first-order Ambisonics by Arteaga, Heller, Benjamin, and Lee [9, 10].

The use of heuristic rules such as Gerzon's energy vector for the calculation of panning gains or ambisonic decoding matrices is somewhat controversial, as it is not fully supported by rigorous perceptual experiments. For instance, the direction in which a virtual source is perceived may depend on a number of factors, such as the source signal, the positions of the loudspeakers relative to the listener, etc. Nevertheless, recent works [11, 12] show that in the context of amplitude panning over a horizontal loudspeaker array: a) the direction of the energy vector could be used to predict the perceived source direction of broadband stimuli with a reasonable accuracy; and b) for these same stimuli, there is a correlation between the perceived source width and the norm of the energy vector. As well, when considering a wide listening area, the energy of the loudspeaker signals provides a good criterion for predicting the loudness of a virtual source.

In this work, our aim is to develop an ambisonic decoding method that provides control of the relevant energy-vector measures and the loudspeaker energy. In particular, this includes controlling the norm of the energy vector so that it is constant across source directions. Our approach can be decomposed in two steps: first, design amplitude panning rules yielding the desired energy vector and energy level properties; second, calculate an ambisonic decoding matrix which matches the properties of the panning as accurately as possible. In the first step, we employ a method derived from Multiple Direction Amplitude Panning (MDAP) [13], whereby loudspeaker panning gains corresponding to numerous virtual source directions are combined together to result in a wider auditory object. This virtual source combination is then iteratively optimized until the energy vector has the desired properties. In the second step, we derive the ambisonic decoding matrix that best fits the panning rules in the least-square sense. This is similar to the methods presented in [5, 14, 15]. The advantage of this approach is that it may be applied to decoders with any ambisonic order, and to arbitrary, irregular loudspeaker arrays. Note that the performance of the decoder is naturally limited by the physical characteristics of the loudspeaker setup, as well as by the inherent resolution of the ambisonic representation. For example, it is impossible to obt ain narrow virtual sources across every direction when the loudspeakers are very sparsely distributed. Likewise, low-order ambisonic decoders generally result in sound energy that is widely spread across loudspeakers.

The outline of the paper is as follows. In Section 2, we briefly review the concept of ambisonic decoding and introduce performance criteria based on the characteristics of the energy vector. In Section 3, we introduce Vector-Base Intensity Panning (VBIP), which is derived from VBAP. We then show that VBIP yields a vector $\mathbf{r}_\mathrm{E}$ pointing to the direction of the desired virtual source, but does not control the norm of this vector. In Section 4, we introduce Multiple-Direction Intensity Panning (MDIP), whereby VBIP gains corresponding to several virtual source directions are combined. We then demonstrate that carefully designed MDIP gains provide $\mathbf{r}_\mathrm{E}$ vectors that point in the desired direction while having a constant norm. Lastly, in Section 5, we detail the calculation of HOA decoding matrices that optimally fit the MDIP gains. The performance of the decoder is then illustrated in a practical example.

2. Ambisonic decoding

In this section we briefly review the concept of ambisonic decoding and define quality measures for a given decoder or panning method. In the frequency domain, the ambisonic decoding equation relates the vector of the loudspeaker signals, $\mathbf{q}$, to the vector of ambisonic signals, $\mathbf{b}$, and is given by [2]

$$\mathbf{q} = \mathbf{D}\,\mathbf{b}\,, \tag{1}$$

where $\mathbf{D}$ is referred to as the ambisonic *decoding matrix* and $\mathbf{q}$ and $\mathbf{b}$ are given by

$$\mathbf{q} = \left[q_1,\ q_2,\ \ldots,\ q_N\right]^\mathsf{T},$$
$$\mathbf{b} = \left[b_{(0,0)},\ b_{(1,-1)},\ \ldots,\ b_{(L,L)}\right]^\mathsf{T}.$$

In the presence of a single, unit-amplitude plane wave incoming from direction $\Omega_S = (\theta_S, \phi_S)$, where θ_S and ϕ_S denote the elevation and azimuth, respectively, the ambisonic signals are given by [2]

$$\mathbf{b}(\Omega_S) = \mathbf{y}(\Omega_S)\, a_S, \tag{2}$$

where a_S denotes the complex signal corresponding to the plane-wave source and $\mathbf{y}(\Omega_S)$ is the vector of the spherical harmonic function values for the direction Ω_S, *i.e.*,

$$\mathbf{y}(\Omega_S) = \left[Y_0^0(\Omega_S),\ Y_1^{-1}(\Omega_S),\ \ldots,\ Y_L^L(\Omega_S)\right]^\mathsf{T}. \tag{3}$$

Substituting the expression for $\mathbf{b}$ into equation (1), the loudspeakers signals are then given by

$$\mathbf{q}(\Omega_S) = \mathbf{D}\ \mathbf{y}(\Omega_S)\, a_S. \tag{4}$$

Therefore, *encoding* a plane-wave source signal into ambisonic signals and *decoding* these signals to the loudspeakers is equivalent to *panning* the source with the gains $\mathbf{g}(\Omega_S)$, given by

$$\mathbf{g}(\Omega_S) = \mathbf{D}\ \mathbf{y}(\Omega_S)$$
$$\mathbf{g}(\Omega_S) = \left[g_1(\Omega_S),\ g_2(\Omega_S),\ \ldots,\ g_N(\Omega_S)\right]^\mathsf{T}. \tag{5}$$

There are many different ways to derive ambisonic decoding matrices or, more generally, amplitude panning gains. An important mathematical quantity characterizing the amplitude panning gains is simply the sum of the squared panning gains, which can be interpreted as the total energy of the loudspeaker signals relative to the energy of the plane-wave source signal. In the case where ambisonic encoding and decoding are used to pan a single plane-wave source, this quantity is given by

$$E(\Omega_S) = \sum_{i=1}^{N} g_i^2(\Omega_S). \tag{6}$$

For the sake of brevity, we herein refer to E as the *loudspeaker energy*. In general it is desirable for the loudspeaker energy to be constant across panning directions. This helps to ensure that if a sound source, such as a musical instrument, is moved to another direction at a constant distance from the listener, the perceived loudness remains reasonably constant. This can be understood as follows. One can interpret $E(\Omega_S)$ as providing a crude estimate of the energy of the sound field produced by the loudspeakers, relative to the energy of the source signal. Therefore, when it increases for a particular panning direction compared to another panning direction, it is likely the perceived loudness of a source also increases for this panning direction. Note that the loudspeaker energy only provides a rough estimate of the actual acoustic energy of the sound field around a listener. Estimating the perceived source loudness accurately requires taking into account many factors, including the position and orientation of the listener relative to the speakers, the frequency, the acoustic properties of the listening room, etc. Nevertheless, as a first approximation, the loudspeaker energy is a reasonable quantity to use for equalization of sound level across directions.

Another common mathematical quantity used to derive ambisonic decoding matrices is the so-called *energy vector*, $\mathbf{r}_\mathrm{E}$, first introduced by Gerzon. For example, *max-rE* decoders attempt to maximize the length of the vector $\mathbf{r}_\mathrm{E}$ across every source direction. In the presence of a single plane wave in direction, Ω_S, the energy vector is given by

$$\mathbf{r}_\mathrm{E}(\Omega_S) = \frac{\mathbf{X}\,\mathbf{e}(\Omega_S)}{E(\Omega_S)}, \tag{7}$$

where $\mathbf{e}(\Omega_S)$ denotes the vector of the loudspeaker signal energies,

$$\mathbf{e}(\Omega_S) = \left[g_1^2(\Omega_S),\ g_2^2(\Omega_S),\ \ldots,\ g_N^2(\Omega_S)\right]^\mathsf{T}, \tag{8}$$

and $\mathbf{X}$ is the matrix whose columns comprise the unit vectors pointing in the direction of the loudspeakers,

$$\mathbf{X} = [\mathbf{x}_1,\ \mathbf{x}_2,\ \ldots,\ \mathbf{x}_N],$$
$$\text{where}\ \mathbf{x}_i = [x_i, y_i, z_i]^\mathsf{T},\quad \begin{cases} x_i = \cos(\theta_i)\cos(\phi_i), \\ y_i = \cos(\theta_i)\sin(\phi_i), \\ z_i = \sin(\theta_i). \end{cases} \tag{9}$$

In a recent work, Frank [12] showed that the direction of the energy vector predicts the perceived source direction with reasonable accuracy when using Vector-Base Amplitude Panning (VBAP) [16] or when using ambisonics to pan sources across a horizontal loudspeaker array. Note that the phase of the individual loudspeaker gains is not taken into account in the calculation of the energy vector, so it may not be precise. In other words, there may be perceptual consequences related to these phase differences that are not captured by the energy vector. Nevertheless, for the context of this paper, the loudspeaker signals are generally assumed to be mostly in phase. Thus, it follows from these considerations that it is desirable for the ambisonic decoding matrices to result in $\mathbf{r}_\mathrm{E}$ vectors pointing in the direction of the desired virtual sources. We define the $\mathbf{r}_\mathrm{E}$ direction mismatch, $\Delta\Omega(\Omega_S)$, which measures the angular distance between the direction of the energy vector and the direction of the desired virtual source, as

$$\Delta\Omega(\Omega_S) = \arccos\left(\frac{\mathbf{r}_\mathrm{E}(\Omega_S)^\mathsf{T}\mathbf{x}_S}{\|\mathbf{r}_\mathrm{E}(\Omega_S)\|}\right), \tag{10}$$

where $\mathbf{x}_S$ is the unit vector pointing to the direction Ω_S.

Another recent work by Frank [11] shows a correlation between the norm of the energy vector and the source width perceived by a listener in the context of amplitude panning using a horizontal loudspeaker array. More specifically, the shorter the energy vector, the wider the perceived source. The intuition behind this result is that, for a given loudspeaker energy value and a given energy vector direction, a smaller norm implies that the energy is more widely distributed across the loudspeakers. On the other hand, a unit norm can be obtained only if a single speaker is used, in which case the source should be perceived as very narrow. We characterize this property of the panning via a mathematical quantity that we refer to as the angular spread, $\sigma(\Omega_S)$, which corresponds to the angular width of the area over which the sound energy is distributed, and is defined by

$$\sigma(\Omega_S) = 2\arccos\left(2\,\|\mathbf{r}_\mathrm{E}(\Omega_S)\| - 1\right). \tag{11}$$

With this definition, the angular spread is equal to the angular aperture of the equivalent continuous sound source distribution that would result in the observed energy vector norm. For instance, the spread is equal to 0 when the norm of the energy vector is one, which corresponds to a source distribution with an aperture equal to 0 (only one active speaker). On the other hand, the spread is equal to 2π when the norm of the energy vector is 0, which corresponds to a situation where the sound sources are evenly distributed over the sphere, *i.e.* the angular aperture of the equivalent source distribution is 2π. The derivation of equation (11) above is provided in the Appendix. Note that this definition of the angular spread is slightly different from that given in [11].

The angular spread is naturally only a very rough predictor of the perceived source width. Methods for controlling the width of virtual sources typically applyy decorrelation to the loudspeaker signals. As we have defined it, the angular spread value is mathematically independent of the correlation between the loudspeaker signals. Nevertheless, it may be desirable that the angular spread remain relatively constant across virtual source directions. This issue has, to the best of our knowledge, not been addressed in the existing literature. In the following, our aim is to design ambisonic decoders providing: a) an optimally constant loudspeaker energy; b) an optimally small $\mathbf{r}_E$ direction mismatch; and c) an optimally constant angular spread. Our approach to this problem is to first design panning gains with the desired properties and then calculate the decoding matrix which optimally matches these gains.

3. Vector-Base Intensity Panning

In this section we briefly review Vector-Base Intensity Panning (VBIP) [17] and show that this panning technique provides optimal performances in terms of loudspeaker energy and $\mathbf{r}_E$ vector direction. In the Vector-Base Amplitude Panning (VBAP) [16], the loudspeaker gains are calculated such that the *velocity vector* [2] points to the direction of the source. Likewise, in VBIP, the gains are calculated such that: 1) the $\mathbf{r}_E$ vector points to the desired direction ; 2) only the three loudspeakers that are the most closely located to the virtual source direction are employed for any given source direction ; 3) the loudspeaker energy is equal to one. For a virtual source located in direction Ω_S, the vector of VBIP loudspeaker energies, $\tilde{\mathbf{e}}(\Omega_S)$, is given by

$$\begin{aligned}\tilde{\mathbf{e}}(\Omega_S) &= [\tilde{e}_1(\Omega_S),\ \tilde{e}_2(\Omega_S),\ \ldots,\ \tilde{e}_N(\Omega_S)]^\mathsf{T} \\ &= [0\ldots0,\ \tilde{e}_u,\ 0\ldots0,\ \tilde{e}_v,\ 0\ldots0,\ \tilde{e}_w,\ 0\ldots0]^\mathsf{T},\end{aligned} \tag{12}$$

where u, v and w are the indices of the three loudspeakers that are the most closely located to the virtual source direction, the energies of which are denoted $\tilde{e}_u$, $\tilde{e}_v$ and $\tilde{e}_w$. These energies are given by

$$\begin{aligned}[\tilde{e}_u,\ \tilde{e}_v,\ \tilde{e}_w] &= \frac{[\rho_u,\ \rho_v,\ \rho_w]}{\rho_u+\rho_v+\rho_w}, \\ [\rho_u,\ \rho_v,\ \rho_w]^\mathsf{T} &= \mathbf{X}_{uvw}^{-1}\,\mathbf{x}_S, \\ \mathbf{X}_{uvw} &= [\mathbf{x}_u,\ \mathbf{x}_v,\ \mathbf{x}_w],\end{aligned} \tag{13}$$

where ρ_u, ρ_v and ρ_w can be referred to as the un-normalized loudspeaker energies and $\mathbf{x}_u$, $\mathbf{x}_v$ and $\mathbf{x}_w$ are defined as in equation (9). Note that, when the virtual source is exactly located in the direction of a loudspeaker, only this loudspeaker is used.

The loudspeaker energy resulting from VBIP, $\tilde{E}(\Omega_S)$, is equal to 1, as we have

$$\begin{aligned}\tilde{E}(\Omega_S) &= \sum_{i=1}^{N}\tilde{e}_i(\Omega_S) = \tilde{e}_u+\tilde{e}_v+\tilde{e}_w, \\ &= \frac{\rho_u+\rho_v+\rho_w}{\rho_u+\rho_v+\rho_w} = 1.\end{aligned} \tag{14}$$

As well, it can easily be verified that the energy vector resulting from VBIP, $\tilde{\mathbf{r}}_E(\Omega_S)$, points to the desired direction. From equations (7) and (13), we have

$$\begin{aligned}\tilde{\mathbf{r}}_E(\Omega_S) &= \frac{\mathbf{X}\,\tilde{\mathbf{e}}(\Omega_S)}{\tilde{E}(\Omega_S)} = \mathbf{X}\,\tilde{\mathbf{e}}(\Omega_S) = \mathbf{X}_{uvw}\left[\tilde{e}_u,\ \tilde{e}_v,\ \tilde{e}_w\right]^\mathsf{T}, \\ &= \frac{\mathbf{X}_{uvw}\,\mathbf{X}_{uvw}^{-1}\,\mathbf{x}_S}{\rho_u+\rho_v+\rho_w} = \frac{\mathbf{x}_S}{\rho_u+\rho_v+\rho_w}.\end{aligned} \tag{15}$$

Figure 1. CARLab's 32-channel loudspeaker setup.

Although the VBIP energy vector points to the desired direction, the norm of this vector is $(\rho_u+\rho_v+\rho_w)^{-1}$, which means that the angular spread varies as a function of the virtual source direction. In particular the norm is maximal and equal to one when the virtual source is located in the exact direction of a loudspeaker. Only in this condition is a single loudspeaker used. On the other hand, the norm is smaller when the three closest loudspeakers are relatively far from the virtual source direction, yielding a larger angular spread.

In order to illustrate the properties of VBIP, we calculate the VBIP gains for the setup of 32 loudspeakers located in the Computing and Audio Research Laboratory (CARLab) facilities, at the University of Sydney (see Figure 1). The resulting loudspeaker energy, $\mathbf{r}_E$ direction mismatch and spread are shown in Figure 2. While the energy level and $\mathbf{r}_E$ direction mismatch are optimal for every source direction, the spread varies significantly from zero (in the loudspeaker directions) to approximately 92° (at the north and south poles). This property is undesirable because it may result in a varying perceived source width as a constant virtual source moves around the sphere.

4. Constant-spread Multiple-Direction Intensity Panning

We now show that it is possible to derive panning gains achieving the same loudspeaker energy and direction performances as that obtained with VBIP while controlling the angular spread. The approach taken here is similar to that of the Multiple-Direction Amplitude Panning [13], whereby VBAP gains corresponding to a number, P, of auxiliary virtual source directions are combined to increase the perceived source width. Because we combine VBIP gains, we refer to this technique as the Multiple-Direction Intensity Panning (MDIP). It is important to note here that it is impossible to obtain angular spread values that are smaller than that obtained using VBIP. This is because VBIP already employs optimally few loudspeakers. Combining several VBIP gain vectors results in using loudspeakers located farther from the virtual source and thus can only yield a larger angular spread value.

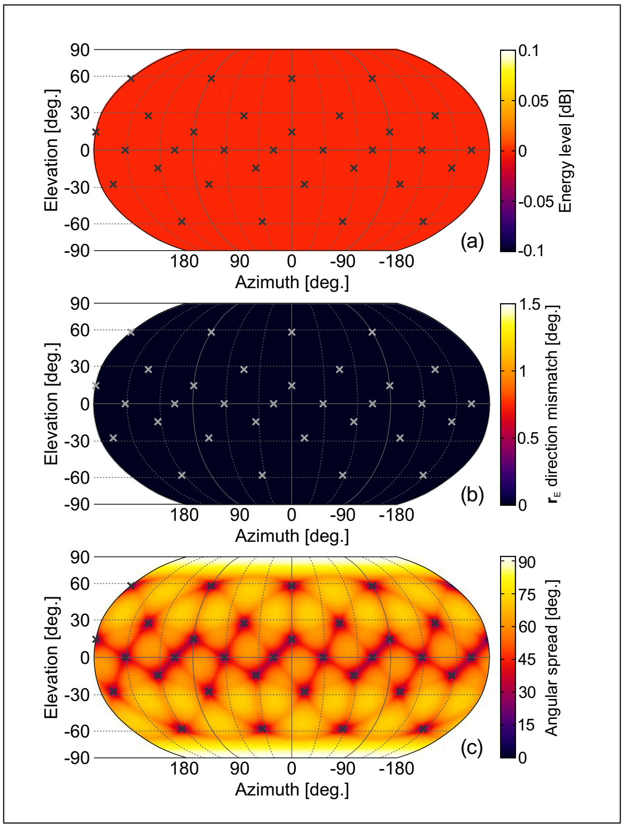

Figure 2. Performance of VBIP using CARLab's 32 loudspeaker setup: a) Loudspeaker energy; b) $\mathbf{r}_\mathrm{E}$ direction mismatch; c) Angular spread, as a function of the virtual source direction. The crosses indicate the position of the 32 loudspeakers in CARLab's 3D audio facility.

The un-normalised energies, $\mathring{e}_i(\Omega_S)$, of the MDIP loudspeaker signals for source direction Ω_S consist of a weighted sum of the VBIP loudspeaker signal energies for P auxiliary directions Ω_{Sj}, *i.e.*

$$\mathring{e}_i(\Omega_S) = \sum_j^P \gamma_j\, \tilde{e}_i(\Omega_{Sj}), \quad (16)$$

where γ_j denotes the weight associated with the j-th auxiliary source and $\tilde{e}_i(\Omega_{Sj})$ is the energy of the i-th loudspeaker signal when using VBIP for a virtual source located in direction Ω_{Sj}, as defined in equation (12). As with VBIP we impose that the total energy of the loudspeaker signals is equal to one, therefore we calculate the MDIP panning gains as

$$\kappa_i(\Omega_S) = \sqrt{\frac{\mathring{e}_i(\Omega_S)}{\sum_i^N \mathring{e}_i(\Omega_S)}}\,. \quad (17)$$

It is easy to show that the energy vector, $\mathring{\mathbf{r}}_E(\Omega_S)$, corresponding to the MDIP panning is a weighted sum of the VBIP energy vectors corresponding to the auxiliary sources. Using equation (16) and recalling the definition of the energy vector, we have

$$\mathring{\mathbf{r}}_E(\Omega_S) = \frac{\sum_i^N \mathring{e}_i(\Omega_S)\,\mathbf{x}_i}{\sum_i^N \mathring{e}_i(\Omega_S)} = \frac{\sum_i^N \sum_j^P \gamma_j\, \tilde{e}_i(\Omega_{Sj})\,\mathbf{x}_i}{\sum_i^N \mathring{e}_i(\Omega_S)} = \frac{\sum_j^P \gamma_j \sum_i^N \tilde{e}_i(\Omega_{Sj})\,\mathbf{x}_i}{\sum_i^N \mathring{e}_i(\Omega_S)} = \frac{\sum_j^P \gamma_j\, \tilde{\mathbf{r}}_E(\Omega_{Sj})}{\sum_i^N \mathring{e}_i(\Omega_S)}. \quad (18)$$

Thus, we see that the vector $\mathring{\mathbf{r}}_E(\Omega_S)$ points to the direction Ω_S if the following conditions are met: a) the auxiliary source directions Ω_{Sj} are symmetrically distributed around the target source directions Ω_S; b) the weights γ_j only depend on the angular distance between directions Ω_S and Ω_{Sj}; and c) the vectors $\tilde{\mathbf{r}}_E(\Omega_{Sj})$ have equal norms.

As shown in section 3, the VBIP panning does not result in energy vectors with equal norm, therefore we need to compensate for this discrepancy to obtain MDIP energy vectors pointing in the desired direction. This can be done by defining the weights γ_j as

$$\gamma_j = \frac{w_\alpha(\Omega_S, \Omega_{Sj})}{\|\tilde{\mathbf{r}}_\mathrm{E}(\Omega_{Sj})\|}, \quad (19)$$

where $w_\alpha(\Omega_S, \Omega_{Sj})$ is a spatial windowing function which depends only on the angular distance, $\angle(\Omega_S, \Omega_{Sj})$, between directions Ω_S and Ω_{Sj}, and α denotes the aperture angle of the window. In this work we use a Tukey window, given by

$$w_\alpha(\Omega_S, \Omega_{Sj}) = \begin{cases} 1 & \text{if } \angle(\Omega_S, \Omega_{Sj}) < \alpha/2, \\ 0 & \text{if } \angle(\Omega_S, \Omega_{Sj}) > \alpha, \\ \cos^2\left[\pi\left(\frac{\angle(\Omega_S, \Omega_{Sj})}{\alpha} - \frac{1}{2}\right)\right] & \text{otherwise}. \end{cases} \quad (20)$$

Note that other choices are possible for the spatial windowing function. It is easy to show that, with the weights given in equation (19), the MDIP energy vector is guaranteed to point to the target direction Ω_S, provided that the auxiliary source directions are symmetrically distributed around Ω_S. Substituting the expression of equation (19) in equation (18) and recalling equation (15) we obtain

$$\mathring{\mathbf{r}}_E(\Omega_S) = \frac{1}{\sum_i^N \mathring{e}_i(\Omega_S)} \sum_j^P w_\alpha(\Omega_S, \Omega_{Sj}) \frac{\tilde{\mathbf{r}}_E(\Omega_{Sj})}{\|\tilde{\mathbf{r}}_\mathrm{E}(\Omega_{Sj})\|} = \frac{\sum_j^P w_\alpha(\Omega_S, \Omega_{Sj})\,\mathbf{x}_{Sj}}{\sum_i^N \mathring{e}_i(\Omega_S)}. \quad (21)$$

Within this framework, calculating the panning gains which yield a given spread value thus simply consists in finding the appropriate window aperture value α for each virtual source direction Ω_S: a larger aperture activates auxiliary sources located farther away from the target source direction, therefore yielding a greater spread value.

The procedure for the calculation of constant-spread MDIP gains for a given target source direction Ω_S then consists in three steps: 1) Generate a dense set of P auxiliary source positions Ω_{Sj} distributed evenly along concentric rings around Ω_S; 2) Estimate the value of the angular

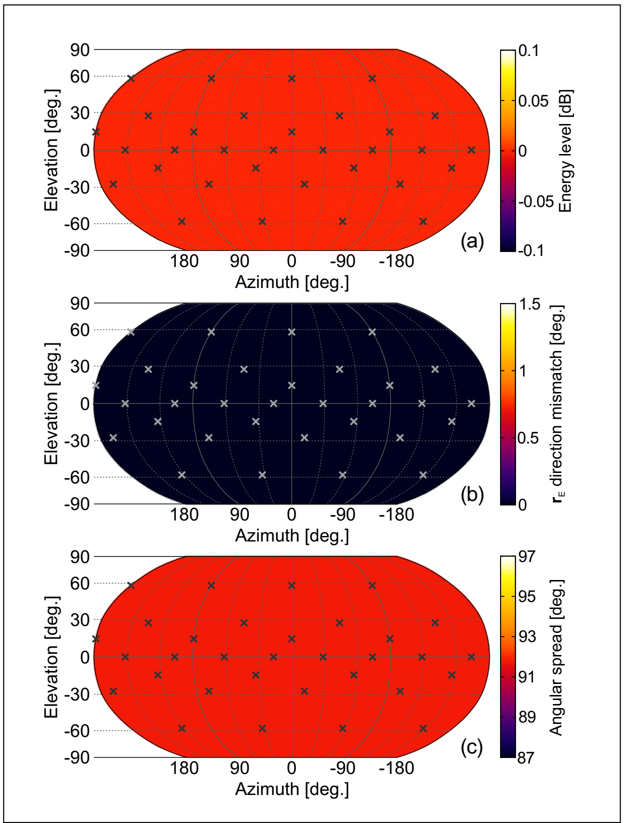

Figure 3. Performance of the constant-spread MDIP using CARLab's 32 loudspeaker setup: a) Loudspeaker energy; b) $\mathbf{r}_E$ direction mismatch; c) Angular spread, as a function of the virtual source direction.

aperture $\alpha(\Omega_S)$ such that the norm of the energy vector $\mathring{\mathbf{r}}_E(\Omega_S)$ matches the desired value. This can be done using a non-linear minimization algorithm, such as that used in MATLAB's *fminsearch* or *fzero* functions; 3) Normalise the loudspeaker energies such that the total energy is equal to one and calculate the loudspeaker gains as the square roots of these energies.

We used the procedure described above to calculate panning gains for the CARLab 32 loudspeaker setup. The gains were calculated for 2562 virtual source directions, relatively evenly distributed over the sphere. In the calculations, the target angular spread was kept constant and equal to 92°, which corresponds to the maximum value obtained using the VBIP gains (north and south pole directions). This is because, as mentioned above, the angular spread can only be larger than that obtained with VBIP. The results are shown in Figure 3. Clearly, the MDIP gains result in a constant loudspeaker energy, zero $\mathbf{r}_E$ direction mismatch, and a 92° angular spread for every virtual source direction.

5. Constant-spread Ambisonic decoding

We now demonstrate the design of ambisonic decoders yielding properties similar to that of the constant-spread MDIP. In order to derive an ambisonic decoding matrix for a given loudspeaker setup, we first calculate constant-spread MDIP gains for a large number, Q, of virtual source directions evenly distributed over the sphere. We then define a matrix, $\mathbf{K}$, containing the collection of computed MDIP gains

$$\mathbf{K} = \left[\boldsymbol{\kappa}(\Omega_{S1}),\ \boldsymbol{\kappa}(\Omega_{S2}),\ \ldots,\ \boldsymbol{\kappa}(\Omega_{SQ})\right], \tag{22}$$

where $\boldsymbol{\kappa}(\Omega_{Sq})$ denotes the vector of the MDIP panning gains for target direction Ω_{Sq}. Our approach to the design of a constant-spread ambisonic decoder is to calculate the decoding matrix which results in panning gains that match $\mathbf{K}$ optimally, in the least-square sense. Recalling equation (5), the matrix, $\mathbf{G}$, of the panning gains corresponding to the HOA encoding and decoding of plane waves located in the directions Ω_{Sq} is given by

$$\mathbf{G} = \mathbf{D}\,\mathbf{Y}, \tag{23}$$

where $\mathbf{D}$ is the decoding matrix and $\mathbf{G}$ and $\mathbf{Y}$ are given by

$$\begin{aligned} \mathbf{G} &= \left[\mathbf{g}(\Omega_{S1}),\ \mathbf{g}(\Omega_{S2}),\ \ldots,\ \mathbf{g}\left(\Omega_{SQ}\right)\right], \\ \mathbf{Y} &= \left[\mathbf{y}(\Omega_{S1}),\ \mathbf{y}(\Omega_{S2}),\ \ldots,\ \mathbf{y}(\Omega_{SQ})\right], \\ \mathbf{y}(\Omega_{Sq}) &= \left[Y_0^0(\Omega_{Sq}),\ Y_1^{-1}(\Omega_{Sq}),\ \ldots,\ Y_L^L(\Omega_{Sq})\right]^{\mathsf{T}}. \end{aligned} \tag{24}$$

The optimal constant-spread decoding matrix, $\mathbf{D}_{\mathrm{CS}}$, is the matrix which minimizes the mismatch between $\mathbf{K}$ and $\mathbf{G}$, hence it is given by

$$\mathbf{D}_{\mathrm{CS}} = \arg\min_{\mathbf{D}} \|\mathbf{K} - \mathbf{D}\mathbf{Y}\|_{\mathrm{F}}, \tag{25}$$

where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm. equation (25) admits a closed-form solution, which is simply the product of $\mathbf{K}$ with the Moore-Penrose pseudo-inverse of $\mathbf{Y}$, *i.e.*,

$$\mathbf{D}_{\mathrm{CS}} = \mathbf{K}\,\mathbf{Y}^{\mathsf{T}}\left(\mathbf{Y}\mathbf{Y}^{\mathsf{T}}\right)^{-1}. \tag{26}$$

Note that the number of virtual source directions, Q, should be strictly greater than the total number of spherical harmonic components for $\mathbf{D}$ to be a least-square solution. As well, the position of the virtual sources may affect the conditioning of the matrix $\mathbf{Y}\mathbf{Y}^{\mathsf{T}}$. As a rule of thumb, using about twice as many virtual sources as there are spherical harmonics should prevent the problem from being ill-posed. Possible choices for the virtual source positions are the Lebedev grids or spherical-T designs [18].

We now demonstrate the performance of the constant-spread HOA decoder. We again consider CARLab's 32 loudspeaker setup. We used the method described in Section 4 to compute constant-spread panning gains for 2562 virtual source directions. These gains were calculated to provide a constant spread value of 92° for every virtual source direction. The virtual source directions were obtained by successively refining the triangular faces of an icosahedron and are relatively evenly distributed over the sphere. We then used equation (26) to derive HOA decoding matrices for orders 1 to 10 and calculated the loudspeaker energy, $\mathbf{r}_E$ direction mismatch, and angular spread as a function of the virtual source directions.

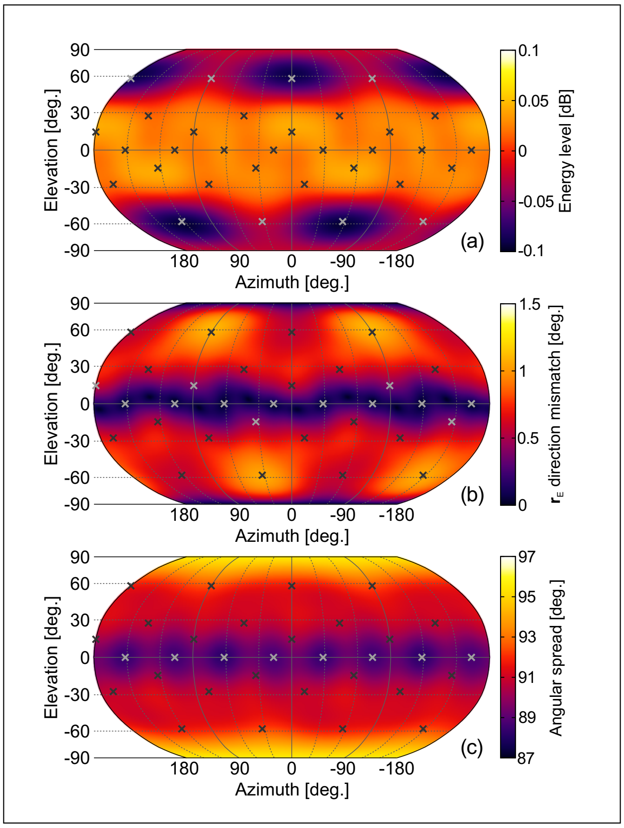

Figure 4. Performance of the constant-spread order-4 HOA decoder: a) Loudspeaker energy; b) $\mathbf{r}_\mathrm{E}$ direction mismatch; c) Angular spread, as a function of the virtual source direction.

Figure 4 shows the performance of the order-4 constant-spread decoder. The loudspeaker energy is practically constant over source directions, with variations on the order of 0.1 dB. As well, the mismatch between the energy vector direction and the virtual source direction is very small. The $\mathbf{r}_\mathrm{E}$ direction mismatch is less than 1.5° for every virtual source direction, which is probably imperceptible to the listener. Note that the $\mathbf{r}_\mathrm{E}$ direction mismatch is minimal for source directions located close to the equator, which is were the loudspeakers are the most densely distributed. Lastly, the angular spread varies about ± 3° around the target value of 92°. Again, slightly lesser spread values are obtained around the equator, where loudspeakers are more densely packed, than for sources located close to the poles, where loudspeakers are more sparsely distributed.

In order to observe the effect of the HOA order on the performance of the decoder, we have summarized the performances of the constant-spread decoders for orders 1 to 10 in Figure 5. The order has a negligible impact on the loudspeaker energy, the values being comprised between −0.1 and 0.15 dB for every direction at every order. The difference between the maximum and minimum loudspeaker energies is thus well below the just-noticeable difference (JND) of 1 dB [19, p. 174]. Unlike the loudspeaker energy, the $\mathbf{r}_\mathrm{E}$ direction mismatch and spread vary as a function of the order. In particular, the $\mathbf{r}_\mathrm{E}$ direction mismatch decreases with increasing order, which can be

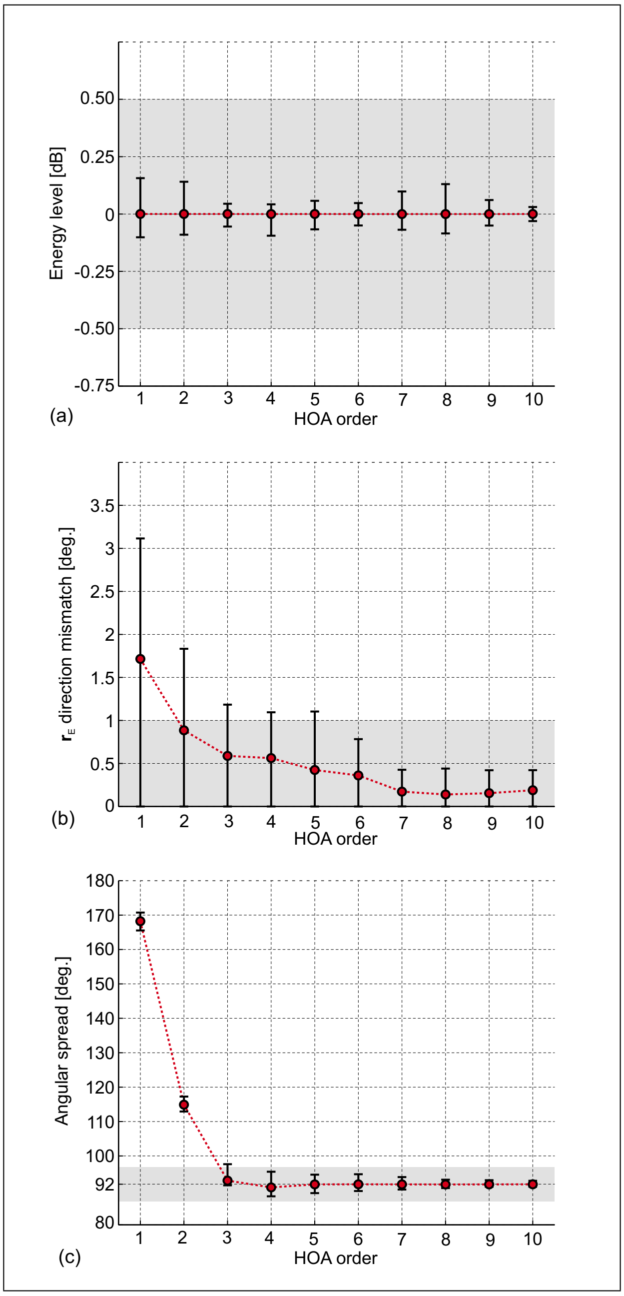

Figure 5. Global performance of the constant-spread HOA decoder as a function of the HOA order: a) Loudspeaker energy; b) $\mathbf{r}_\mathrm{E}$ direction mismatch; c) Angular spread. The round markers represent the average values over every virtual source direction and the bars indicate the minimum and maximum values. The grey areas represent the regions where values are located below the JNDs: ±0.5 dB for the loudspeaker energy (JND = 1 dB); 0–1° for the $\mathbf{r}_\mathrm{E}$ direction mismatch (JND = 1°); and 92 ± 5° for the angular spread (JND ≈ 10°).

explained by the fact that more degrees of freedoms are available to match the performance of the MDIP. Nevertheless, the $\mathbf{r}_\mathrm{E}$ direction mismatch is on the order of a few degrees even at order 1. As well, for orders higher than 3, the largest $\mathbf{r}_\mathrm{E}$ direction mismatch lies in the range of the minimum JND (localization blur) for directional sound localization in free field of 1° [20, p. 38]. Regarding the angular spread, lower orders, especially 1 and 2, result in spread values that are considerably larger than the target 92° value. This is simply due to the fact that, at these orders, the inherent resolution of the ambisonic representation is too coarse to achieve the target angular spread. Note that the spread is nonetheless approximately constant as a function of the source direction for orders 1 and 2. For higher orders, the angular spread values are approximately within the range of the JND, which we estimate to be about 6° based on the results of perceptual experiments presented in [11, p. 50]. In summary, the performance of the constant-spread decoders is good for orders 1 and 2, considering the limitation in spatial resolution, and excellent for orders higher than 3.

6. Conclusions

We have presented a method for designing ambisonic decoding matrices which yield excellent performance in terms of $\mathbf{r}_\mathrm{E}$ vector direction and loudspeaker energy, while providing an almost constant angular spread across directions. The constant-spread decoder is particularly useful for setups where the loudspeakers are irregularly distributed, as classic decoders typically yield varying angular spreads with setups of this kind. Because the minimum achievable angular spread ultimately depends on the angular distance between the loudspeakers, making the spread value constant for every virtual source direction comes at the cost of an increased spread value in the regions where the loudspeakers are the most densely distributed. Therefore, setups whereby the loudspeakers are evenly distributed across directions are desirable because they allow for a greater range of constant spread values.

In our experiments, the constant-spread decoding matrices were found to match the MDIP gains with a surprising accuracy for orders higher than 3. This is because the directivity patterns corresponding to the MDIP gains are relatively wide and there is no major difficulty reproducing these with spherical harmonic functions. By contrast, it is much more difficult to derive ambisonic decoding matrices that match the VBIP gains accurately, as these correspond to narrower directive beams.

Our formulation of constant-spread MDIP requires that a windowing function is optimized for a large number of virtual source directions. As a result the panning gains for a given target angular spread must be calculated offline. This raises the question of how to efficiently control the angular spread during playback. In the context of ambisonics, a possible solution could be to interpolate between several pre-calculated decoding matrices corresponding to different spread values.

Lastly, note that the criteria used for the calculation of the decoding matrix are mostly based on the properties of the energy vector. Although experiments conducted by Frank [11, 12] showed a relation between these criteria and the perceptual attributes of virtual sources, a proper perceptual evaluation of the methods described in this paper is required. This evaluation will be conducted in future work.

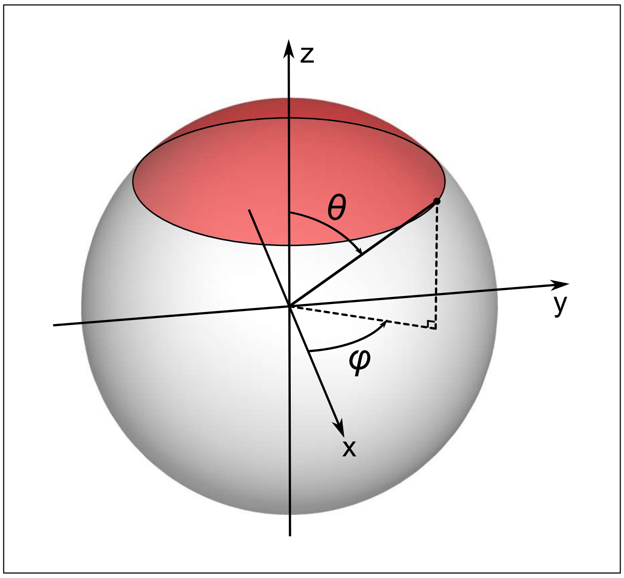

Figure A1. The notations used for describing the geometry in the Appendix are shown.

Appendix

In this appendix we derive the definition of the angular spread as given in equation (11). Assume the sound field is emitted by a continuous distribution of loudspeakers forming a cone with angle σ, as illustrated by the red cap in Figure A1. Further, assume that the energy density of the speaker distribution is constant and equal to 1. The energy vector is then given by

$$\mathbf{r}_\mathrm{E} = \frac{\int_0^{\frac{\sigma}{2}}\int_0^{2\pi} (\sin\theta\cos\varphi\mathbf{i} + \sin\theta\sin\varphi\mathbf{j} + \cos\theta\mathbf{k})\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\varphi}{\int_0^{\frac{\sigma}{2}}\int_0^{2\pi}\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\varphi}, \tag{A1}$$

where $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ are the unit vector pointing to the directions x, y and z, respectively. Because the speaker distribution is symmetric around the z axis, the expression of the energy vector simplifies to

$$\begin{aligned}\mathbf{r}_\mathrm{E} &= \frac{\int_0^{\frac{\sigma}{2}}\int_0^{2\pi}\cos\theta\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\varphi}{\int_0^{\frac{\sigma}{2}}\int_0^{2\pi}\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\varphi}\mathbf{k} \\ &= \frac{\int_0^{\frac{\sigma}{2}}\cos\theta\sin\theta\,\mathrm{d}\theta}{\int_0^{\frac{\sigma}{2}}\sin\theta\,\mathrm{d}\theta}\mathbf{k} = \frac{\int_0^{\frac{\sigma}{2}}\sin 2\theta\,\mathrm{d}\theta}{2\int_0^{\frac{\sigma}{2}}\sin\theta\,\mathrm{d}\theta}\mathbf{k} \\ &= \frac{1-\cos\sigma}{4\left(1-\cos\frac{\sigma}{2}\right)}\mathbf{k} = \frac{1+\sin^2\frac{\sigma}{2}-\cos^2\frac{\sigma}{2}}{4\left(1-\cos\frac{\sigma}{2}\right)}\mathbf{k} \\ &= \frac{1-\cos^2\frac{\sigma}{2}}{2\left(1-\cos\frac{\sigma}{2}\right)}\mathbf{k} = \frac{1+\cos\frac{\sigma}{2}}{2}\mathbf{k}.\end{aligned} \tag{A2}$$

Therefore, the norm of the energy vector is given by

$$\|\mathbf{r}_\mathrm{E}\| = \frac{1+\cos\frac{\sigma}{2}}{2}. \tag{A3}$$

Lastly, we can express the aperture angle σ as a function of the norm of the energy vector, *i.e.*,

$$\sigma = 2 \arccos \left(2 \| \mathbf{r}_{\mathrm{E}} \| - 1 \right). \tag{A4}$$

References

[1] M. Gerzon: Periphony: With-height sound reproduction. The Journal of the Audio Engineering Soceity **21** (1973) 2–10.

[2] J. Daniel: Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia. Dissertation. Université Paris 6, Paris, France, 2000.

[3] D. B. Ward, T. D. Abhayapala: Reproduction of a plane-wave sound field using an array of loudspeakers. IEEE Transactions on Speech and Audio Processing **9** (2001) 697–707. 1063-6676.

[4] M. Gerzon: General metatheory of auditory localisation. 92nd Convention of the Audio Engineering Society, Vienna, Austria, March 1992.

[5] J.-M. Batke, F. Keiler: Using vbap-derived panning functions for 3D Ambisonic decoding. Proc. 2nd Int. Symp. Ambisonics and Spherical Acoustics, Paris, 5 2010.

[6] F. Keiler, J.-M. Batke: Evaluation of virtual source localization using 3-D loudspeaker setups. Audio Engineering Society Convention 128, 5 2010.

[7] F. Zotter, H. Pomberger, M. Noisternig: Energy-preserving Ambisonic decoding. Acta Acustica united with Acustica **98** (2012) 37–47.

[8] F. Zotter, M. Frank: All-round Ambisonic panning and decoding. Journal of the Audio Engineering Society **60** (2012) 807–820.

[9] D. Arteaga: An ambisonic decoder for irregular 3d loudspeaker arrays. prepr. 8918, 134th AES Conv., Rome, 2013.

[10] A. Heller, E. Benjamin, R. Lee: A toolkit for the design of ambisonic decoders. Proc. Linux Audio Conference, CCRMA, Stanford, 2012.

[11] M. Frank: Phantom sources using multiple loudspeakers in the horizontal plane. Dissertation. University of Music and Performing Arts, Graz, 2013.

[12] M. Frank: Localization using different amplitude-panning methods in the frontal horizontal plane. Proceedings of the EAA Joint Symposium on Auralization and Ambisonics, Berlin, Germany, April 3-5 2014, 41–47.

[13] V. Pulkki: Uniform spreading of amplitude panned virtual sources. Applications of Signal Processing to Audio and Acoustics, 1999 IEEE Workshop on, 1999, 187–190.

[14] N. Epain, C. Jin: Considering VBAP's role - an HOA and compressed sensing perspective. Presented at the AES 131th Convention, New York, October 2011, available online at: www.ee.usyd.edu.au/carlab/presentations.htm.

[15] F. Zotter, M. Frank, H. Pomberger: Comparison of energy-preserving and all-round ambisonic decoders. AIA-DAGA 2013, Meran, 2013.

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

[17] J.-M. Pernaux, P. Boussard, J.-M. Jot: Virtual sound source positioning and mixing in 5.1 implementation on the real-time system genesis. Proc. Conf. Digital Audio Effects (DAFx-98), Barcelona, November 1998, 76–80.

[18] N. Sloane, R. Hardin, W. Smith: Tables of spherical codes. neilsloane.com/packings/.

[19] E. Zwicker, H. Fastl: Psychoacoustics: facts and models. 2 ed. Springer, Berlin, 1999.

[20] J. Blauert: Spatial hearing: the psychophysics of human sound localization. MIT Press, Cambridge, MA, 1983.