Spectral Impairment for Two-Dimensional Higher Order Ambisonics*

AUDUN SOLVANG, *AES Student Member*
(auduso@q2s.ntnu.no)

Centre for Quantifiable Quality of Service in Communication Systems, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway

The representation of sound fields by spherical or cylindrical harmonics, also known as higher order Ambisonics (HOA), is a flexible format where the accuracy of the representation depends on the order N. It is shown that reproducing 2D HOA with uniformly distributed loudspeakers on a circle radiating plane waves in a nonreverberant environment will lead to spectral impairment if the number of loudspeakers M is higher than $2N + 1$. The impairments vary with both the angle, that is, the angular difference between source and receiver relative to the center of the circle, and the product kr, where k is the wavenumber and r the distance from the center of the circle. It is therefore not possible to correct the impairments with a filter for multiple radii. For classical first-order Ambisonics the near perfect reconstruction area is so small that it must be regarded solely as a sweet-spot technique and filtering should be considered when using more than three loudspeakers. Furthermore, for HOA the number of loudspeakers is a tradeoff between the reproduction error for $kr < N$ and spectral impairments for $kr > N$.

0 INTRODUCTION

The decomposition of sound fields has been formulated in several ways for both cylindrical harmonics [1]–[5] and plane waves [6] in the two-dimensional case and for spherical harmonics in the three-dimensional case [7]–[9]. This decomposition allows for near perfect representation over a larger area than the traditional sweet spot and is maybe best known as higher order Ambisonics (HOA). It is layered, that is, by including more channels the accuracy of the representation in terms of kr will increase, where k is the wavenumber and r the distance from the center of the loudspeaker array.

Another way of representing sound fields is by a spatial sampling of the pressure and particle velocity over a closed surface, which is used to recreate the wavefield, also known as holophony or wavefield synthesis [10]. This technique is by nature not as flexible in loudspeaker positioning and scalability.

If a frequency limit is defined for near perfect reproduction of the sound field, spatial harmonics decomposition would be scalable in terms of reproduction area and would be well suited for the storage, consumption, and transmission of spatial audio, such as universal multimedia access (UMA) [11] and transmission over the internet.

The reproduction of a spatial harmonics decomposed sound field is flexible in terms of loudspeaker configurations [12]–[15], which is advantageous for UMA. This is certainly true for the near perfect reconstruction region, but as shown in [3], there will be impairments in both spectrum and phase for the nonperfect reproduction region when more loudspeakers are used than the minimum required. A compensation technique is also presented in [3], which is based on filtering, but since the accuracy of the representation depends on kr, a compensation by filtering could be expected to work well only for sweet-spot listening where the listener's head is centered in $r = 0$.

This can clearly be observed if a diffuse field consisting of pink noise is reproduced over a large number of loudspeakers in the horizontal plane (2D HOA) with different HOA orders. The spectral coloration can be observed to change when changing the HOA order or moving away from the sweet spot. Compensation by filtering would not be suitable in this scenario. This calls for an elucidation of the impairment. Loudspeakers radiating plane waves and uniformly distributed on a circle and a nonreverberant environment are considered for simplicity.

*Manuscript received 2007 July 17; revised 2008 March 3 and 31.

1 HIGHER ORDER AMBISONICS

1.1 Basic Relations

A monochromatic wavefield can be expressed in the frequency domain by HOA signals $B_n^{\pm 1}$ [7] by the relation

$$p(r, k, \theta) = B_0^{+1} J_0(kr) + \sum_{n=1}^{\infty} \left(B_n^{+1} \sqrt{2} \cos n\theta + B_n^{-1} \sqrt{2} \sin n\theta J_n(kr) \right) \quad (1)$$

or by cylindrical harmonics written in complex notation [2],

$$p(r, k, \theta) = \sum_{n=-\infty}^{\infty} \mathrm{j}^n q_n \mathrm{e}^{\mathrm{j}n\theta} J_n(kr) \quad (2)$$

where $J_n(z)$ is the Bessel function of order n of the first kind, r is the distance from the point of origin, and $k = 2\pi/\lambda = 2\pi f/c$ is the wavenumber.

Eq. (1) is related to Eq. (2) by

$$B_0^{+1} = q_0, \qquad B_n^{+1} = \frac{\mathrm{j}^n (q_n + q_{-n})}{\sqrt{2}}, \qquad B_n^{-1} = \frac{\mathrm{j}^{n+1} (q_n - q_{-n})}{\sqrt{2}} \quad (3)$$

and the derivations and results achieved for the cylindrical harmonic signals in this paper apply to the HOA signals as well.

In practice, the series in Eqs. (1) and (2) must be truncated due to limited bandwidth or microphone technique, which results in a degradation of the sound field, except at $r = 0$. The truncation limit is denoted by the HOA order that corresponds to the highest azimuthal frequency represented. For HOA order N, Eq. (2) will turn into

$$p(r, k, \theta) = \sum_{n=-N}^{N} \mathrm{j}^n q_n \mathrm{e}^{\mathrm{j}n\theta} J_n(kr). \quad (4)$$

Since Eq. (4) is the inverse relation of a Fourier series with respect to the angle θ of the sound pressure p at a radius r, a transformation into the azimuthal frequency domain is defined by

$$q_n = \frac{1}{2\pi \mathrm{j}^n J_n(kr)} \int_0^{2\pi} p(r, k, \theta) \mathrm{e}^{-\mathrm{j}n\theta} \, \mathrm{d}\theta. \quad (5)$$

In practice q_n can be found by utilizing microphone arrays [2], [4]–[10].

Modeling the wavefield as a sum of S monochromatic plane waves leads to

$$p(r, k, \theta) = \sum_{s=1}^{S} A_s \mathrm{e}^{\mathrm{j}kr \cos(\theta - \theta_s)} \quad (6)$$

and using the relation in Eq. (36) in the Appendix leads to Eq. (4) becoming

$$q(k)_n = \sum_{s=1}^{S} A_s \mathrm{e}^{-\mathrm{j}n\theta_s} \quad (7)$$

which means that the HOA B-format encoding can be formulated as an angular weighting of the pressure in one position.

Using matrix notation for the plane wave model results in

$$\boldsymbol{q} = \boldsymbol{E}\boldsymbol{p} \quad (8)$$

where

$$\boldsymbol{q} = \begin{bmatrix} q(k)_{-N} \\ \cdots \\ \cdots \\ q(k)_N \end{bmatrix}, \qquad \boldsymbol{E} = \begin{bmatrix} \mathrm{e}^{-\mathrm{j}N\theta_1} & \cdots & \mathrm{e}^{-\mathrm{j}N\theta_S} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ 1 & \cdots & 1 \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \mathrm{e}^{\mathrm{j}N\theta_1} & \cdots & \mathrm{e}^{\mathrm{j}N\theta_S} \end{bmatrix},$$

$$\boldsymbol{p} = \begin{bmatrix} A_1 \\ \cdots \\ \cdots \\ A_S \end{bmatrix}. \quad (9)$$

The decoding $\boldsymbol{D}$ of the HOA B-format to the D-format, that is, the loudspeaker signals, can then be expressed as

$$\boldsymbol{p}_l = \boldsymbol{D}\boldsymbol{q} = \boldsymbol{D}\boldsymbol{E}\boldsymbol{p} \quad (10)$$

where by the "reencoding principle" [7],

$$\boldsymbol{D} = \mathrm{pinv}(\boldsymbol{E}_\mathrm{d}) = \boldsymbol{E}_\mathrm{d}^\mathrm{H} (\boldsymbol{E}_\mathrm{d} \boldsymbol{E}_\mathrm{d}^\mathrm{H})^{-1}. \quad (11)$$

$\boldsymbol{E}_\mathrm{d}$ is determined by the angles of the loudspeakers and $(.)^\mathrm{H}$ is the Hermitian operator. The encoding–decoding process can be viewed as a plane-wave decomposition of the original wavefield via a cylindrical representation of the wavefield. In the case of a regular layout of M loudspeakers then

$$\boldsymbol{E}_\mathrm{d} = \begin{bmatrix} \mathrm{e}^{-\mathrm{j}N\frac{2\pi}{M}} & \cdots & \mathrm{e}^{-\mathrm{j}N\frac{2\pi M}{M}} \\ \cdots & \cdots & \cdots \\ 1 & \cdots & 1 \\ \cdots & \cdots & \cdots \\ \mathrm{e}^{\mathrm{j}N\frac{2\pi}{M}} & \cdots & \mathrm{e}^{\mathrm{j}N\frac{2\pi M}{M}} \end{bmatrix} \quad (12)$$

and

$$\boldsymbol{D} = \frac{1}{M} \boldsymbol{E}_\mathrm{d}^\mathrm{H} \quad (13)$$

which implies that generally $M \geq 2N + 1$ for the existence of an exact solution. It is shown in [2], [16] that $M \geq 2N$, but the limit $M \geq 2N + 1$ is used in this work for simplicity.

The loudspeaker signal in direction θ_m at $r = 0$ can then be found by using the definition in Eq. (38) in the Appendix,

$$\begin{aligned} p_l(0, k, \theta_m) &= \frac{1}{M} \sum_{n=-N}^{N} \mathrm{e}^{-\mathrm{j}n\theta_m} q(k)_n^* \\ &= \frac{1}{M} \sum_{n=-N}^{N} \mathrm{e}^{-\mathrm{j}n\theta_m} \sum_{s=1}^{S} A_s^* \mathrm{e}^{(\mathrm{j}n/\theta_s)} \\ &= \frac{1}{M} \sum_{s=1}^{S} A_s^* \sum_{n=-N}^{N} \mathrm{e}^{-\mathrm{j}n(\theta_m - \theta_s)} \\ &= \frac{1}{M} \sum_{s=1}^{S} A_s^* \mathrm{csinc}_{2N+1}(\theta_m - \theta_s). \end{aligned} \quad (14)$$

This is also known as basic decoding [3].

Truly perfect reconstruction is not feasible except for the point of origin, but the mean reconstruction error can be found to be below −14 dB (0.04%) [6], [8], [15] as long as $kr < N$. Since this typically does not give near perfect reproduction up to 20 kHz even at $r = 100$ mm (corresponding to the human head in the sweet spot), it is important to study what happens for $kr > N$.

1.2 Influence of the Number of Loudspeakers

For simplicity, the distance to the loudspeakers is considered to be so long that the radiated waves are close to plane for all frequencies and the theory presented in the previous section can be used.

The common error measure in previous literature is the normalized mean square pressure difference [1], [3], [8], [9], [15]. This measure is very good for small errors but becomes quickly large for $kr > N$ and gives little elucidation on the error in this region. The auditory system is insensitive to phase at high frequencies ($kr > N$) [17], [18] and utilizes interaural level differences (ILDs) for localization. Together with the possibility for spectral coloration this motivates for employing impairments in the sound intensity as an error measurement in the nonperfect reproduction area. The pressure magnitude squared, which is proportional to the sound intensity, in HOA reproduction relative to the original sound field is therefore used as an error measure in this work. This measure can, however, not quantize in any possible deviation in the angle, but we know that the angle of incidence generally will be wrong for $kr > N$.

The case of only one source in direction θ_1 and M uniformly distributed loudspeakers will be used throughout the rest of this paper. The relation in Eq. (36) then leads to the sound pressure at a receiver in the HOA reconstructed wavefield as being expressed as

$$\begin{aligned} p(r, k, \theta)_A &= \sum_{m=0}^{M-1} \mathrm{e}^{\mathrm{j}kr\cos\left(\frac{2\pi m}{M}-\theta\right)} p_l\left(0, k, \frac{2\pi m}{M}\right) \\ &= \frac{1}{M}\sum_{m=0}^{M-1} A_1^* \mathrm{e}^{\mathrm{j}kr\cos\left(\frac{2\pi m}{M}-\theta\right)} \mathrm{csinc}_{2N+1}\left(\frac{2\pi m}{M}-\theta_1\right) \\ &= \frac{A_1^*}{M}\sum_{m=0}^{M-1}\sum_{n=-\infty}^{\infty} \mathrm{j}^n J_n(kr)\mathrm{e}^{\mathrm{j}n\left(\frac{2\pi m}{M}-\theta\right)} \sum_{l=-N}^{N} \mathrm{e}^{-\mathrm{j}l\left(\frac{2\pi m}{M}-\theta_1\right)} \\ &= \frac{A_1^*}{M}\sum_{l=-N}^{N} \mathrm{e}^{\mathrm{j}l\theta_1} \sum_{n=-\infty}^{\infty} \mathrm{j}^n J_n(kr)\mathrm{e}^{-\mathrm{j}n\theta} \sum_{m=0}^{M-1} \mathrm{e}^{-\mathrm{j}(l-n)\frac{2\pi m}{M}} \\ &= A_1^* \sum_{l=-N}^{N} \mathrm{e}^{\mathrm{j}l(\theta_1-\theta)} \sum_{n=-\infty}^{\infty} \mathrm{j}^{nM+l} J_{nM+l}(kr)\mathrm{e}^{-\mathrm{j}nM\theta}. \end{aligned} \tag{15}$$

For $kr < M - N$ Eq. (15) can be simplified to

$$p(r, k, \theta)_A \approx A_1 \sum_{l=-N}^{N} \mathrm{e}^{\mathrm{j}l(\theta_1-\theta)} \mathrm{j}^l J_l(kr) \tag{16}$$

since the Bessel terms have small values when the argument is smaller than the order.

Following the same argument, the truncation error becomes small for $kr < N$ so that

$$\begin{aligned} p(r, k, \theta)_A &\approx A_1 \sum_{l=-\infty}^{\infty} \mathrm{e}^{\mathrm{j}l(\theta_1-\theta)} \mathrm{j}^l J_l(kr) \\ &= p(r, k, \theta) \qquad kr < N \end{aligned} \tag{17}$$

that is, near perfect reproduction.

1.2.1 Relative Intensity

The squared pressure of the original sound field in Eq. (6) can now be expressed as

$$p(r, k, \theta)p(r, k, \theta)^* = |A_1|^2 \tag{18}$$

and the sound field resulting from HOA as

$$\begin{aligned} &p(r, k, \theta)_A p(r, k, \theta)_A^* \\ &\quad = \frac{|A_1|^2}{M^2} \sum_{m=0}^{M-1}\sum_{p=0}^{M-1} \mathrm{e}^{\mathrm{j}kr\left[\cos\left(\frac{2\pi m}{M}-\theta\right)-\cos\left(\frac{2\pi p}{M}-\theta\right)\right]} \\ &\qquad \times \mathrm{csinc}_{2N+1}\left(\frac{2\pi m}{M}-\theta_1\right) \mathrm{csinc}_{2N+1}^*\left(\frac{2\pi p}{M}-\theta_1\right). \end{aligned} \tag{19}$$

The relative sound intensity of the HOA wavefield is therefore

$$\begin{aligned} I_{\mathrm{rel}}(kr, \theta, \theta_1) &= \frac{|p(r, k, \theta)_A|^2}{|p(r, k, \theta)|^2} \\ &= \frac{1}{M^2} \sum_{m=0}^{M-1}\sum_{p=0}^{M-1} \mathrm{e}^{\mathrm{j}kr\left[\cos\left(\frac{2\pi m}{M}-\theta\right)-\cos\left(\frac{2\pi p}{M}-\theta\right)\right]} \\ &\qquad \times \mathrm{csinc}_{2N+1}\left(\frac{2\pi m}{M}-\theta_1\right) \mathrm{csinc}_{2N+1}\left(\frac{2\pi p}{M}-\theta_1\right) \end{aligned} \tag{20}$$

where the conjugated notation of csinc_{2N+1} is omitted since it has a real value.

1.2.2 Mean Relative Intensity

The expression for the expected relative intensity in Eq. (20) over the angle is derived in Eq. (40) in the Appendix and is given by

$$\bar{I}_{\mathrm{rel}}(kr, \theta_1) = \frac{1}{M}\sum_{m=0}^{M-1} J_0\left[2kr\sin\left(\frac{\pi m}{M}\right)\right] \mathrm{csinc}_{2N+1}\left(\frac{2\pi m}{M}\right). \tag{21}$$

In other words, the mean relative intensity is the value of J_0 [$2kr \sin(\pi m/M)$] for $m = 0$ when the discrete Fourier transform (DFT) with respect to m is truncated at $(2N + 1)/2M$, which follows from the second to last line in Eq. (40), or recognizing Eq. (21) as the expression for a circular convolution [19]. Interestingly the mean relative intensity is independent of the angle of incidence given the assumptions in this work.

When $kr \to \infty$, then J_0 [$2kr$ sin ($\pi m/M$)] will have a nonzero value only for $m = 0$ and hence

$$\lim_{kr\to\infty} \bar{I}_{\text{rel}} = \frac{\text{csinc}_{2N+1}(0)}{M} = \frac{2N+1}{M}. \tag{22}$$

1.3 Coloration and Interaural Level and Time Differences

In order to evaluate the perceptual consequences, the resulting ear signals need to be evaluated.

The head-related impulse responses (HRIRs) for a Neumann KU8li dummy head were measured in an anechoic chamber with 1° resolution at 2-m distance between the loudspeaker and the dummy head. Symmetry was assumed so that the HRIR for the left ear was also used for the right ear, and simulations of a virtual source in between two loudspeakers with both the loudspeakers and the virtual source radiating plane waves were performed in MATLAB [20]. The head was facing the virtual source at all times, motivated by the fact that the minimum audible angle is smallest in this case [21]. The resulting ear signals should theoretically have no interaural level or time differences. The HRIRs from the loudspeakers were time shifted according to the different positions of the head when it was moved gradually from the center of the loudspeaker array to the right. The ERBFilterbank in the Auditory Toolbox for MATLAB [22] was employed on the binaural impulse response for both HOA reproduction and the real source.

The ILD for each band was calculated as the relative energy in the left ear to the right and the interaural time differences (ITDs) were determined by locating the position of the maximum of the cross correlation of the half-wave-rectified right and left ear signals. The spectral coloration was calculated as the mean energy in each band for both ears in the HOA reproduction relative to the mean energy that would be produced by the real source. The results are discussed in Section 2.3 (see Figs. 9–11).

1.4 Alternative HOA Decoding

Filtering the HOA B-format channels leads to a different decoding technique [3]. Weighting each B-format channel with a frequency-dependent gain $g(2\pi f/c)_n = g(k)_n$ leads to the plane-wave model phase mode spectrum of

$$\hat{q}(k)_n = g(k)_n q(k)_n = \sum_{s=1}^{S} g(k)_n p_s(0, k, \theta_s) e^{-jn\theta_s}. \tag{23}$$

The desired loudspeaker pressure in Eq. (14) now becomes

$$\begin{aligned} p_l(k, \theta_m) &= \frac{1}{M} \sum_{s=1}^{S} p_s(0, k, \theta_s) \sum_{n=-N}^{N} g(k)_n e^{-jn(\theta_m-\theta_s)} \\ &= \frac{1}{M} \sum_{s=1}^{S} p_s(0, k, \theta_s) f_{2N+1}(k, \theta_m - \theta_s). \end{aligned} \tag{24}$$

Inserting Eq. (24) instead of Eq. (14) into Eq. (15), assuming a regular loudspeaker layout and one source, leads to

$$\begin{aligned} p_r(r, k, \theta)_A = \frac{p_1(0, k, \theta_1)}{M} \sum_{m=0}^{M-1} \sum_{n=-N}^{N} g(k)_n \\ e^{jkr\cos\left(\frac{2\pi m}{M}-\theta\right)-jn\left(\frac{2\pi m}{M}-\theta_1\right)} \end{aligned} \tag{25}$$

and

$$\begin{aligned} I_{\text{rel}}(kr, \theta, \theta_1) = \frac{1}{M^2} \sum_{m=0}^{M-1} \sum_{p=0}^{M-1} e^{jkr\left[\cos\left(\frac{2\pi m}{M}-\theta\right)-\cos\left(\frac{2\pi p}{M}-\theta\right)\right]} \\ \times f_{2N+1}\left(k, \frac{2\pi m}{M} - \theta_1\right) f^*_{2N+1}\left(k, \frac{2\pi p}{M} - \theta_1\right). \end{aligned} \tag{26}$$

The difference is that the $\text{csinc}_{2N+1}(\theta)$ panning function has been replaced by a new frequency-dependent panning function,

$$f_{2N+1}(k, \theta) = \sum_{n=-N}^{N} g(k)_n e^{-jn\theta}. \tag{27}$$

This is analogous to window shaping in the time–frequency domain [19]. Using the notation in Eq. (39), the mean relative intensity can be expressed as

$$\begin{aligned} \bar{I}_{\text{rel}} &= \frac{1}{M} \sum_{n=-N}^{N} g(k)_n g^*(k)_{-n} \mathcal{D}_{kr,M}(n) \\ &= \frac{1}{M} \sum_{m=0}^{M-1} J_0\left[2kr \sin\left(\frac{\pi m}{M}\right)\right] \sum_{n=-N}^{N} g(k)_n g^*(k)_{-n} e^{-jn\frac{2\pi m}{M}} \\ &= \frac{1}{M} \sum_{n=-N}^{N} g(k)_n g^*(k)_{-n} \mathcal{D}_{kr,M}(n). \end{aligned} \tag{28}$$

The B-format filtering coefficients $g(k)_n$ have been proposed to be used in order to compensate for the frequency-dependent spectral coloration [3]. A frequency-independent scaling g_n can also be utilized to reproduce the HOA B format over $M = 2N$ loudspeakers [2], [16].

2 DISCUSSION

2.1 Spectral Impairment in Different *kr* Regions

The theoretical expressions in Section 1.2 are illustrated in Figs. 1–13. The figures illustrate the mean spectral impairment as a function of *kr* (averaged over receiver angle, for a fixed source position) as well as the detailed spectral impairment as a function of *kr* and receiver angle.

The mean relative intensity as a function of *kr* is calculated for a fixed number of loudspeakers, *M* and varying order *N*, as given by Eq. (21). Figs. 1–3 show that above the high-frequency (high *kr*) limit a lower level results, as pointed out in [3] and given by Eq. (22). It can be noted that when the number of loudspeakers is matched to the order, that is, when $M = 2N+1$, the mean level stays correct. This can, for example, be seen in Fig. 1 for order 7.

The effect of a perfectly matched number of loudspeakers is further illustrated in Fig. 4. For order $N = 1$, that is, basic Ambisonics, and three loudspeakers, the mean spectral impairment (averaged over receiver angle) is zero, as seen by the leftmost diagrams in Fig. 4. However, when the source is positioned in the middle between two loudspeakers (the worst case), Fig. 4(b) shows that even if the mean level is correct [as is also seen by Eq. (21)], the spectral impairment will vary with the receiver angle above the correct reproduction frequency range. This implies that if a listener rotates the head, a varying spectral impairment will result. Also, if a source is rotating around the listener, a varying spectral impairment will result.

As soon as the number of loudspeakers is larger than the minimum required, $M > 2N + 1$, the mean level will not be correct (as was also shown in Figs. 1 and 2). Figs. 5 and 6 illustrate this for order $N = 1$ (basic Ambisonics) and an increasing number of loudspeakers. Interestingly the mean spectral impairment is independent of the source angle as given by Eq. (21), which can be seen by the leftmost curves in each figure. For example, the two leftmost curves in Fig. 5 are identical, as are the ones in Fig. 6. Still, the detailed spectral impairment will depend on the source angle, as is seen by comparing the rightmost plots in Figs. 5 and 6.

By increasing the number of loudspeakers an intermediate frequency range is emerging. In Figs. 6 and 7, it is clear that a transition appears around $kr = M - N$. When the number of loudspeakers is kept at 15 but the order is increased from $N = 1$ to $N = 4$ (see Fig. 7), the near perfect reproduction range ($kr < N$) is increasing so that the intermediate frequency range is getting smaller. Finally, when the order is high enough to match the number of loudspeakers ($N = 7$, $M = 15$), Fig. 8 illustrates in the same way as Fig.

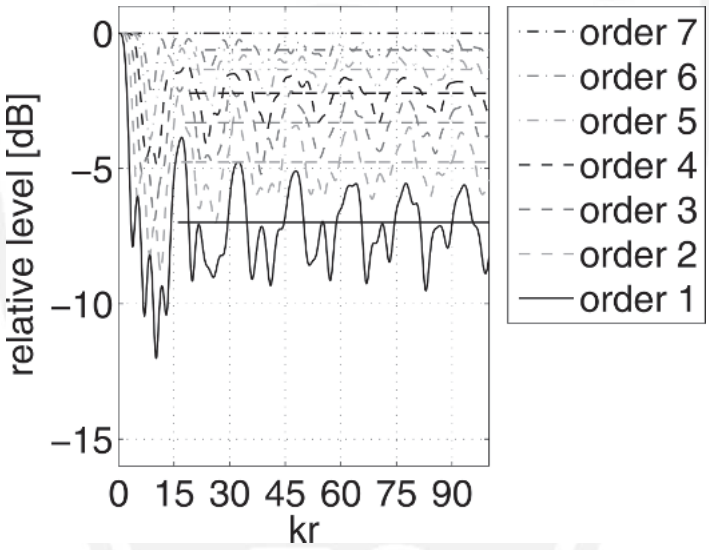

Fig. 1. Mean relative level averaged over receiver angle for HOA reproduction using basic decoding, 15 loudspeakers. Straight lines—asymptotic values.

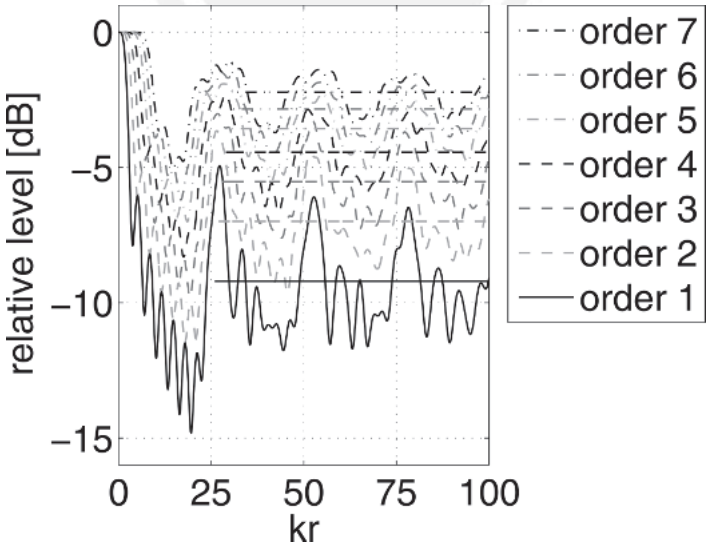

Fig. 2. Mean relative level averaged over receiver angle for HOA reproduction using basic decoding, 25 loudspeakers. Straight lines—asymptotic values.

4 that the mean level is correct, but there will be spectral impairment that will depend on the source angle.

- $kr < N$ As long as the Bessel terms of order higher than N have little significance, the impairment will be small, as stated in Eq. (17). It corresponds well to the criterion $kr \leq N$ [6], [8], [15]. The number of loudspeakers M has little influence as long as $M \geq 2N + 1$, as pointed out in Section 1.1.
- $N < kr < M - N$ Truncation of the plane-wave model phase mode spectrum at N results in an azimuthal fluctuation pattern given by the circular convolution

$$|e^{jkr\cos(\theta)} * \mathrm{csinc}_{2N+1}(\theta - \theta_1)|^2.$$

This relation results from recognizing Eq. (16) as being of the form of a DFT and using the properties of the DFT [19]. Whether the source is positioned in the direction of a loudspeaker or between loudspeakers will have no impact on the pattern; the source position determines only the orientation of the pattern. Aliasing effects will occur when $M/2 < kr$, analogous to a violation of the sampling theorem. The effects will, however, not be noticeable because of the truncation of the phase mode spectrum, analogous to low-pass filtering. The truncation leads to the observable kr-dependent relative intensity of about N/kr deduced from the amount of energy the truncation has "filtered out,"

$$I_{\mathrm{rel}} = \frac{N}{kr}|e^{jkr\cos(\theta)} * \mathrm{csinc}_{2N+1}(\theta - \theta_1)|^2. \quad (29)$$

- $M - N < kr$ Bessel terms with arguments higher than $M - N$ will begin to contribute significantly to Eq. (15),

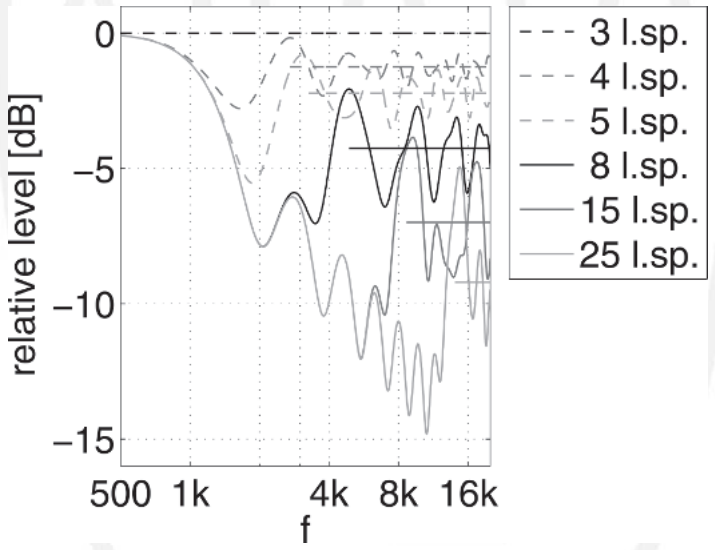

Fig. 3. Mean relative power density spectrum for Ambisonics (HOA order 1) at 0.1-m radius using basic decoding. Straight lines—asymptotic values.

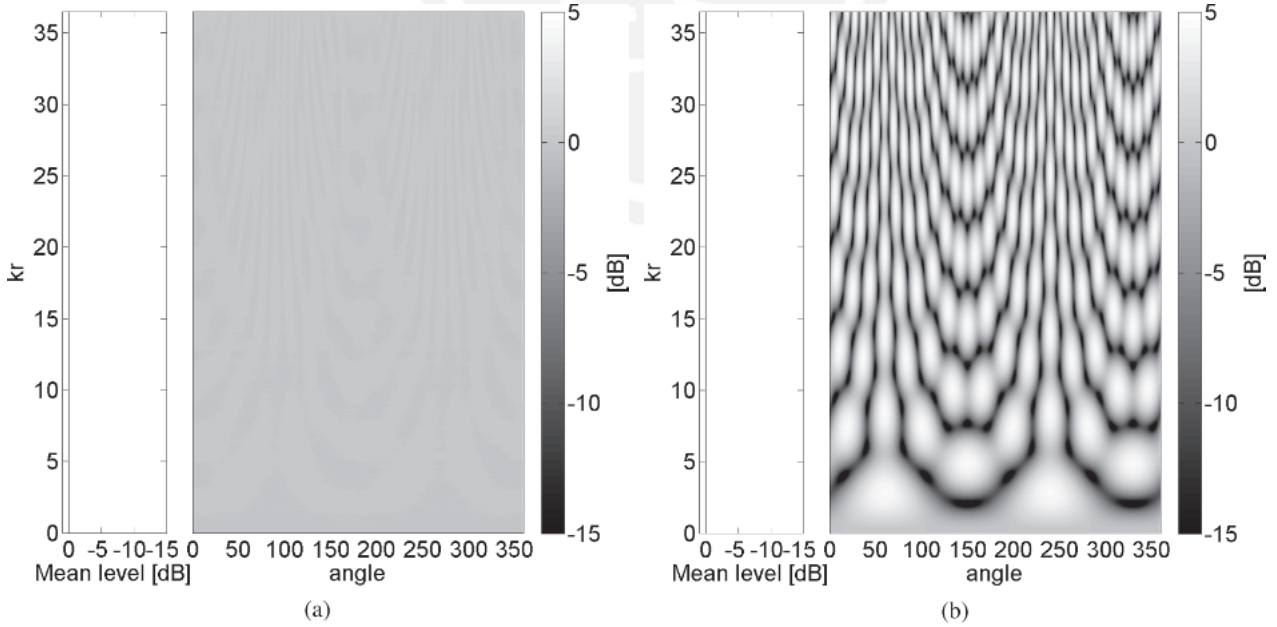

Fig. 4. Level of Ambisonics (HOA order 1) reproduction relative to original sound field; basic decoding, 3 loudspeakers. (a) Source at 0° (in direction of one loudspeaker). (b) Source at 60° (between two loudspeakers).

and aliasing will be introduced. The angular fluctuation will therefore change due to the extra angular phase contribution $e^{nM\theta}$ in the last line of Eq. (15), that is, the angular fluctuation pattern is not only dependent on the position of the source but also an M, the number of loudspeakers. When the source is positioned between loudspeakers large impacts on the pattern can be observed in Figs. 4–8.

From Figs. 1–3 it can be seen that the relative level ripples around the asymptotic value in Eq. (22). In Eq. (21) the term $J_0[2kr \sin(\pi m/M)]$ is weighted with $\mathrm{csinc}_{2N+1}(2\pi m/M)$, both being periodic and symmetric in m with the period M. Since the Bessel terms within the main lobe of $\mathrm{csinc}_{2N+1}(2\pi m/M)$ will have the largest weight, the number of contributing terms should be about proportional to M/N. The minima and maxima of the Bessel function for large arguments in Eq. (37) in the Appendix leads to expressions for peaks and dips in intensity.

$$\left.\begin{aligned} kr_{\max} &= \frac{2\pi n + \pi/4}{2\sin(\pi m/M)} \\ kr_{\min} &= \frac{2\pi n - \pi/4}{2\sin(\pi m/M)} \end{aligned}\right\} \qquad n \in \mathbb{N} \wedge n \gg N.$$

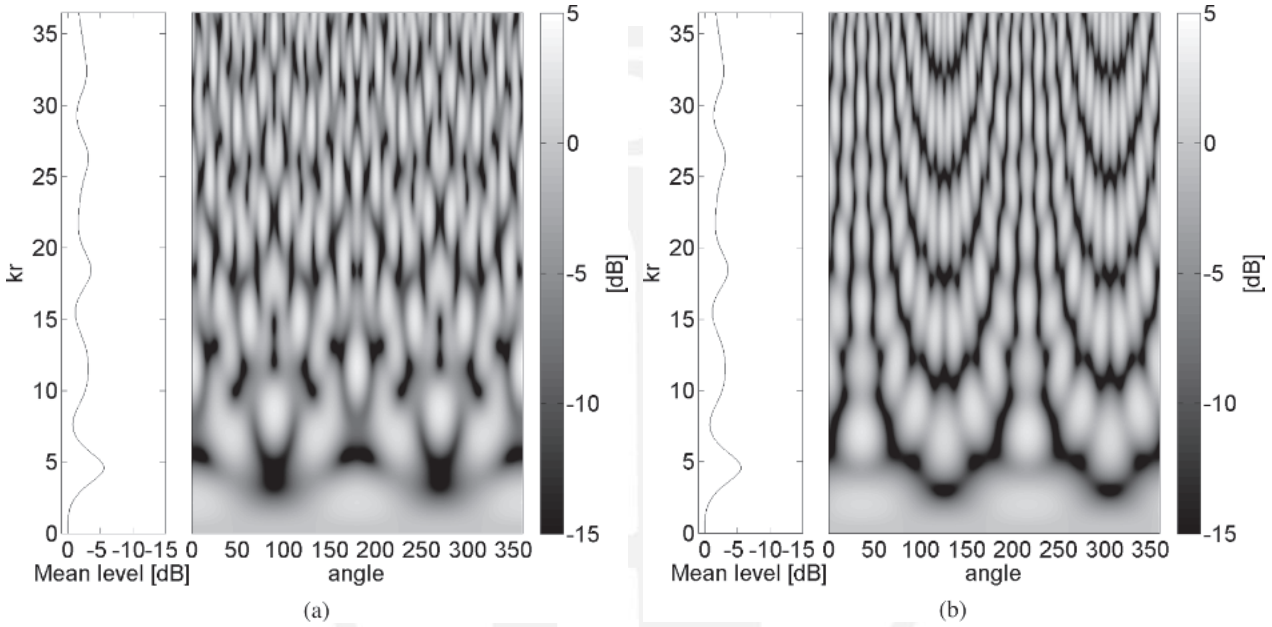

Fig. 5. Level of Ambisonics (HOA order 1) reproduction relative to original sound field; basic decoding, 5 loudspeakers. (a) Source at 0° (in direction of one loudspeaker). (b) Source at 36° (between two loudspeakers).

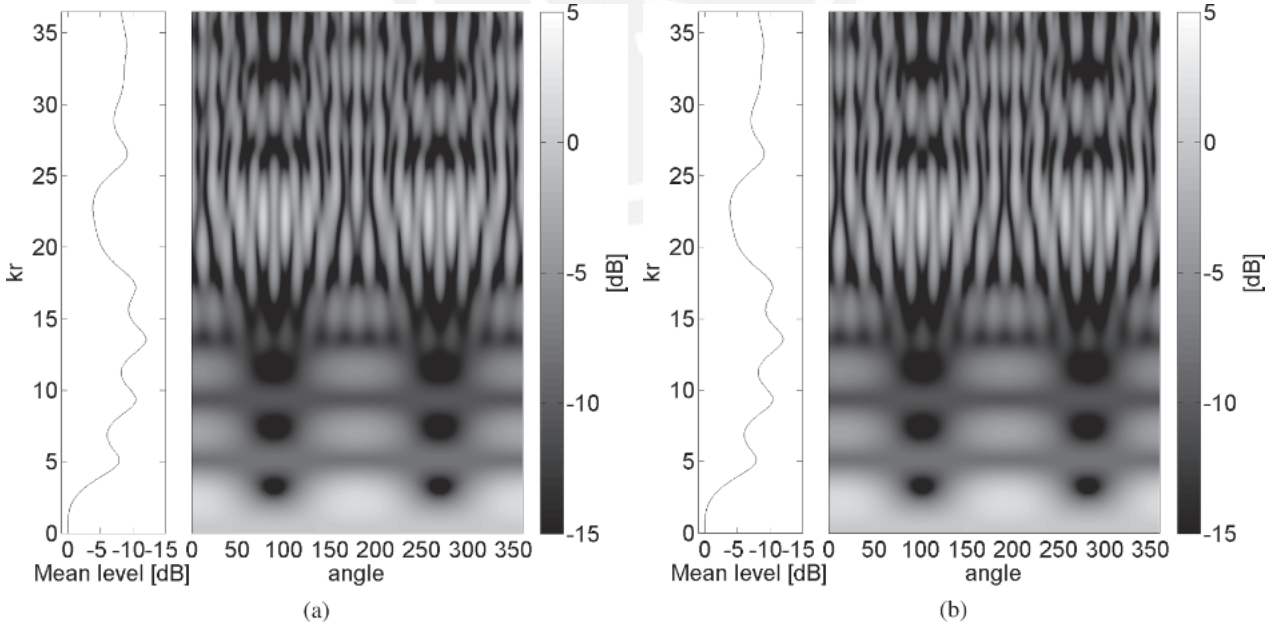

Fig. 6. Level of Ambisonics (HOA order 1) reproduction relative to original sound field; basic decoding, 15 loudspeakers. (a) Source at 0° (in direction of one loudspeaker). (b) Source at 12° (between two loudspeakers).

The ripple will therefore be a series of about M/N almost harmonic terms. If $M \gg \pi$ then $\sin(\pi/M) \approx \pi/M$ and the main ripple will have a periodicity of about $kr = M$, which can be observed in Fig. 2. It can also be observed that the number of harmonics increases when M/N increases, but that J_0 [2 kr sin (π/M)] will be the term contributing the most to the impairment since $\mathrm{csinc}_{2N+1}(2\pi m/M)$ is largest for $m = M - 1$ and $m = 1$.

2.2 Advantages of Utilizing More Loudspeakers than 2*N*+1

As discussed in section 2.1, Bessel terms of order higher than $M - N$ will introduce aliasing in the sound pressure in Eq. (15). While not that prominent, this will also have an effect on the near perfect reproduction region, especially at the boundary $kr \approx N$. Increasing the number of loudspeakers leads to the order of the destructive Bessel terms relative to the argument $kr \approx N$ becoming higher. From the nature of the Bessel function of the first kind,

$$|J_n(x)| < |J_m(x)|, \qquad |x| < |m| < |n|$$

this implies that the magnitude will be smaller and hence the destructive effect will be lower, especially at the $kr = N$ boundary.

This means that by using more loudspeakers the spatial sound reproduction has higher fidelity within the near perfect reproduction region, and especially at $kr \approx N$.

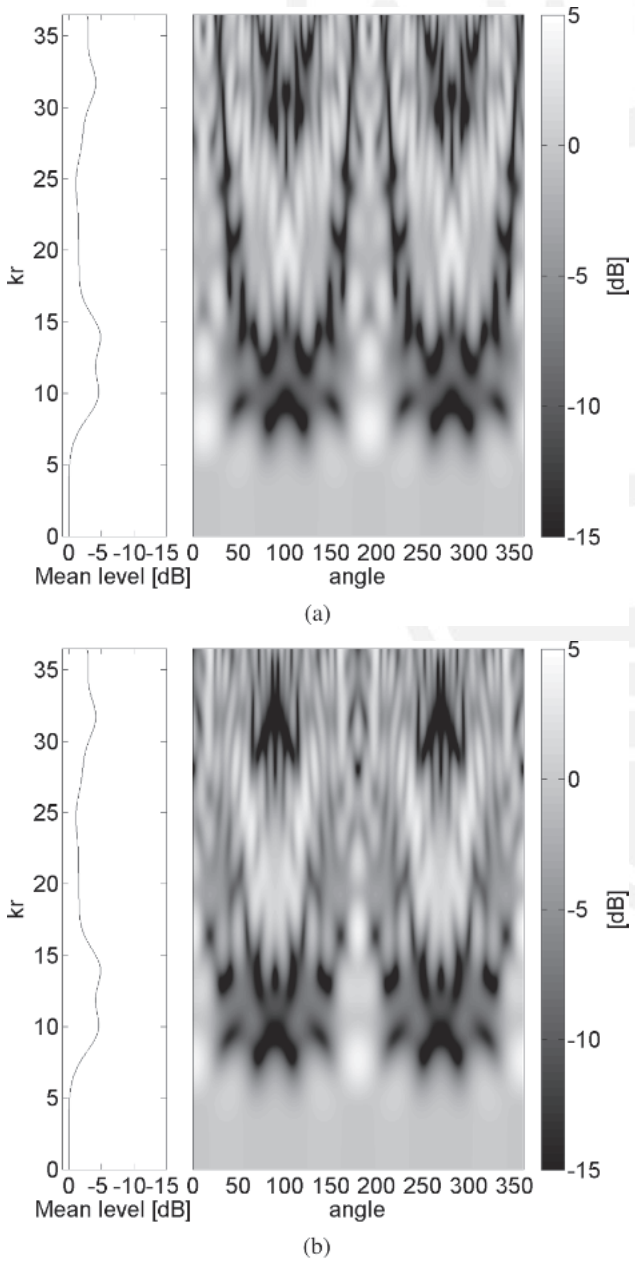

Fig. 7. Level of Ambisonics (HOA order 4) reproduction relative to original sound field; basic decoding, 15 loudspeakers. (a) Source at 0° (in direction of one loudspeaker). (b) Source at 12° (between two loudspeakers).

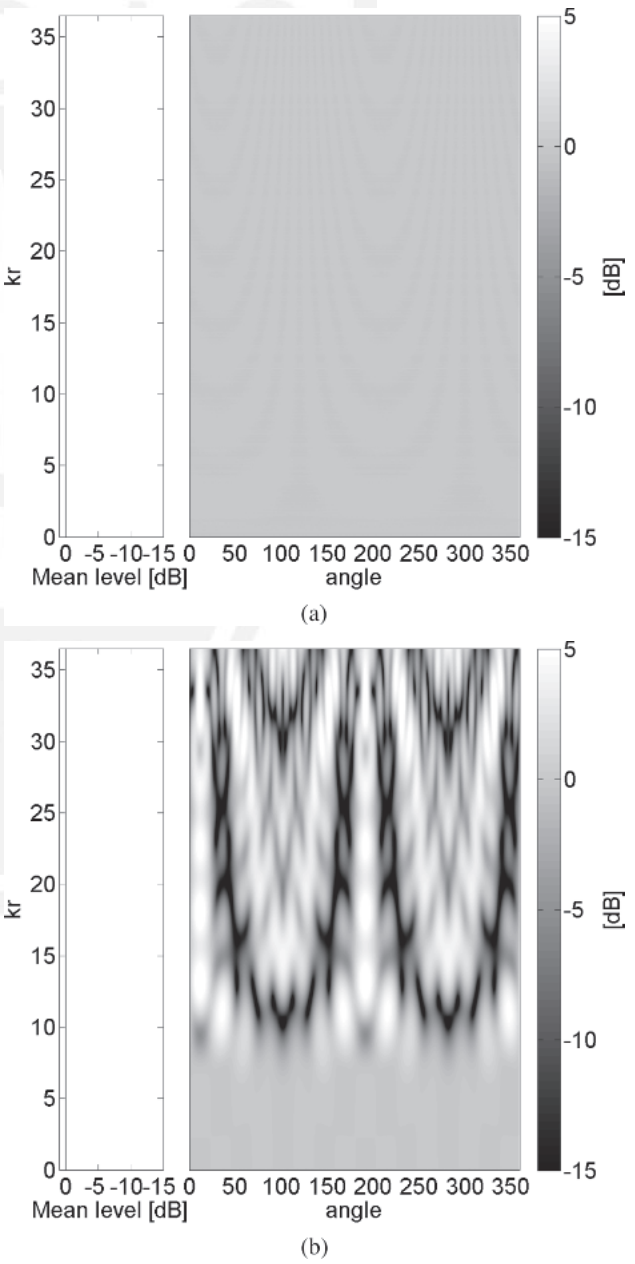

Fig. 8. Level of Ambisonics (HOA order 7) reproduction relative to original sound field; basic decoding, 15 loudspeakers. (a) Source at 0° (in direction of one loudspeaker). (b) Source at 12° (between two loudspeakers).

This suggests that the number of loudspeakers chosen is a tradeoff between lower reproduction errors within the near perfect reproduction region and spectral impairment for $kr > N$.

2.3 Perceptual Consequences

The interaural level difference, time difference, and spectral coloration are shown in Figs. 9–11. Since the distance between the ears of a person is about 19.5 mm [23] the intensity impairment can lead to ILD for a noncentered listener. Since the head is not acoustically transparent above about 700 Hz, the diffraction around the subject's head will also affect the ILD for $kr > N$. An example of the shift in ILD for the reproduced sound field relative the original sound field is shown in Fig. 9 together with the ITD in Fig. 10 and the spectral coloration in Fig. 11. We note the following.

- $kr < N$ ILD and ITD are correct, no spectral coloration.
- $N < kr < M - N$ ILD and ITD are biased in a regular way across frequencies that can result in a localization shift. The cues are, however, partially conflicting, suggesting that the auditory object can be perceived as diffuse [18], [24], as reported in [3]. The degree of spectral coloration is dependent on the relative number of loudspeakers.
- $kr > M - N$ Pseudorandom fluctuations in ILD and ITD, which will probably lead to a more diffuse auditory object. This is, however, independent of whether or not the number of loudspeakers is matched to the order. The dips in the spectrum are not that severe, but there is a generally lower level if the number of loudspeakers is not matched.

Since a human's most effective method for resolving localization ambiguities is head movements [17], the subject would probably utilize this in HOA. However, this would not result in more consistent localization cues that might reinforce the sensation of a diffuse auditory object.

Since the low-frequency localization cues are the dominating ones [25], it would be expected that localization would be possible as long as the listener is seated so close

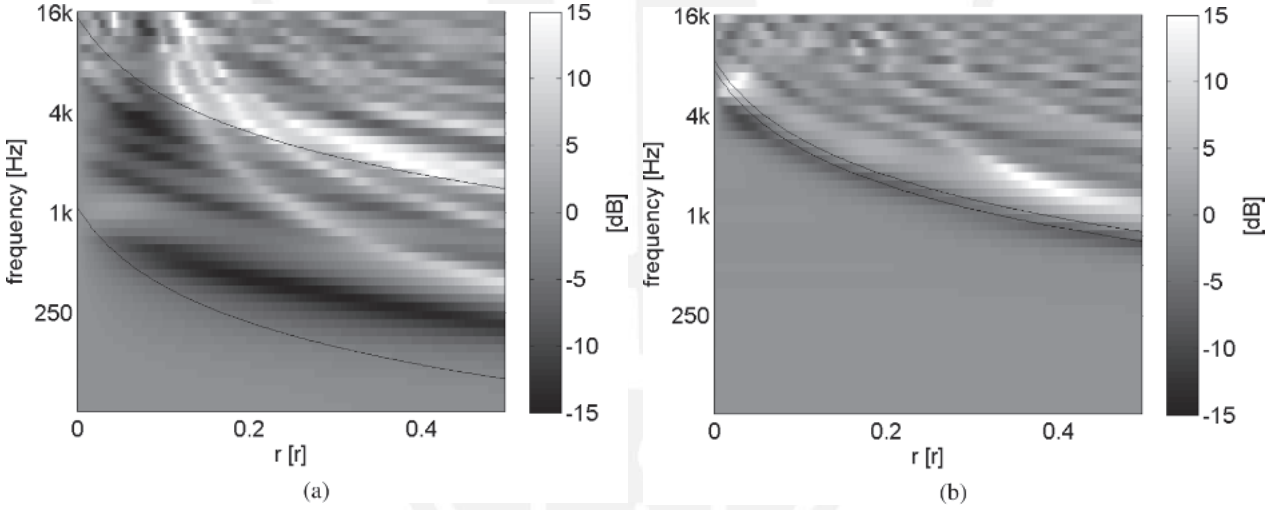

Fig. 9. ILD for 15 loudspeakers; $kr = N$ and $kr = N - M$ limits plotted in black. (a) Order 1. (b) Order 7.

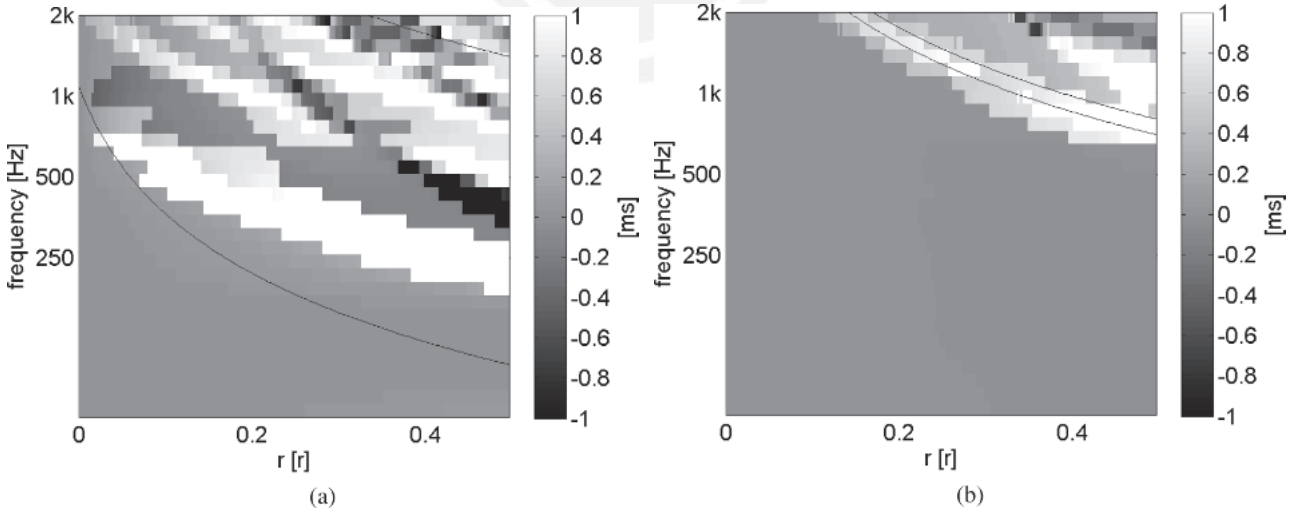

Fig. 10. ITD for 15 loudspeakers; $kr = N$ and $kr = N - M$ limits plotted in black. (a) Order 1. (b) Order 7.

to the center of the loudspeakers that sufficient low-frequency information is present. The distance will be dependent on the order. Motivated by stereo reproduction that is capable of reproducing consistent cues up to 700 Hz this distance might be $r \approx cN/2\pi 700 = 0.078N$.

2.4 Compensating for Level Impairments by Alternative HOA Decoding

Deriving the filter coefficients described in [3] and Section 1.4 for alternative HOA decoding must be performed for a radius r_{comp}. One disturbing effect is that when compensating for the spectral differences by filtering, the frequency limit for near perfect reconstruction inside the compensation radius will be scaled down, and a higher sound level can be observed above this frequency. That is, the compensation will destroy near perfect reconstruction regions. Another weakness is that filtering cannot compensate for the angular variations in sound level; it will only alter the pattern, for example, for $N < kr < M - N$ the fluctuation pattern in Eq. (29) will become

$$I_{rel} = |e^{jkr\cos(\theta)} * f_{2N+1}(k, \theta - \theta_1)|^2. \quad (30)$$

2.5 Importance for Classical Ambisonics Reproduction

Considering classical Ambisonics, that is, HOA order 1, and sweet-spot listening ($r \approx 0.1$ m), the sound field can be almost perfectly reconstructed up to about 550 Hz as long as one uses a minimum of three loudspeakers ($M \geq 3$). It follows from the discussion in Section 2.1 and Eq. (29) that when utilizing more loudspeakers the spectrum will gradually drop from $f = c/2\pi \times 0.1 \approx 550$ Hz to about $f = (M - 1)c/2\pi \times 0.1 \approx 550(M - 1)$ Hz, where the levels are reduced by approximately $-10 \log (M - 1)$ dB. Above this frequency the spectrum will fluctuate around 4.8–10 log (M) dB, which follows from the discussion in Section 2.1. These effects can be observed explicitly in Fig. 3 and implicitly in Figs. 1, 2, 4–6, and 11(a). They could be compensated for at the sweet spot by filtering the Ambisonics B-format channels with filters optimized for reproduction at $r \approx 0.1$ m [3]. The fidelity of the reproduction will be lower when the ears are not positioned at $r \approx 0.1$ m [Figs. 9(a), (10a), and 11(a)], and since the head is not a perfect sphere and the azimuthal variation in intensity cannot be compensated for, one must expect spectral changes for frequencies above at least 550 Hz when rotating the head at sweet-spot listening. Furthermore, filtering cannot compensate for the regular radial fluctuation pattern, which could lead to either a shift or an inconsistency in localization above 2 kHz, as observed in [3].

However, the discussion in Section 2.2 and the fact that low-frequency localization cues dominate over high-frequency cues [25] suggest that in order to achieve better reproduction at about 550 Hz some small spectral coloration can be sacrificed by utilizing a few more loudspeakers than the minimum required.

3 CONCLUSION

When representing two-dimensional sound fields by HOA and reproducing them over uniformly distributed loudspeakers radiating plane waves on a circle and in a nonreverberant environment, there is near perfect reconstruction as long as kr is smaller than the HOA order N. As long as the number of loudspeakers is larger than the minimum number required, the number used is insignificant for the spectral coloration.

When $kr > N$, angle- and kr-dependent intensity impairments increase when the number of loudspeakers increases above $2N + 1$. The impairments will lead to perceived spectral coloration and reduced localization accuracy that cannot be compensated for by filtering. It is worth mentioning that when using $2N + 1$ loudspeakers spectral impairments can still be observed if the virtual sound source is positioned between two loudspeakers.

In order to avoid spectral impairments the number of loudspeakers should be kept as low as possible. However, utilizing more loudspeakers can decrease the reproduction error in the near perfect reproduction region, $kr < N$, especially at $kr \approx N$. The number of loudspeakers must be a

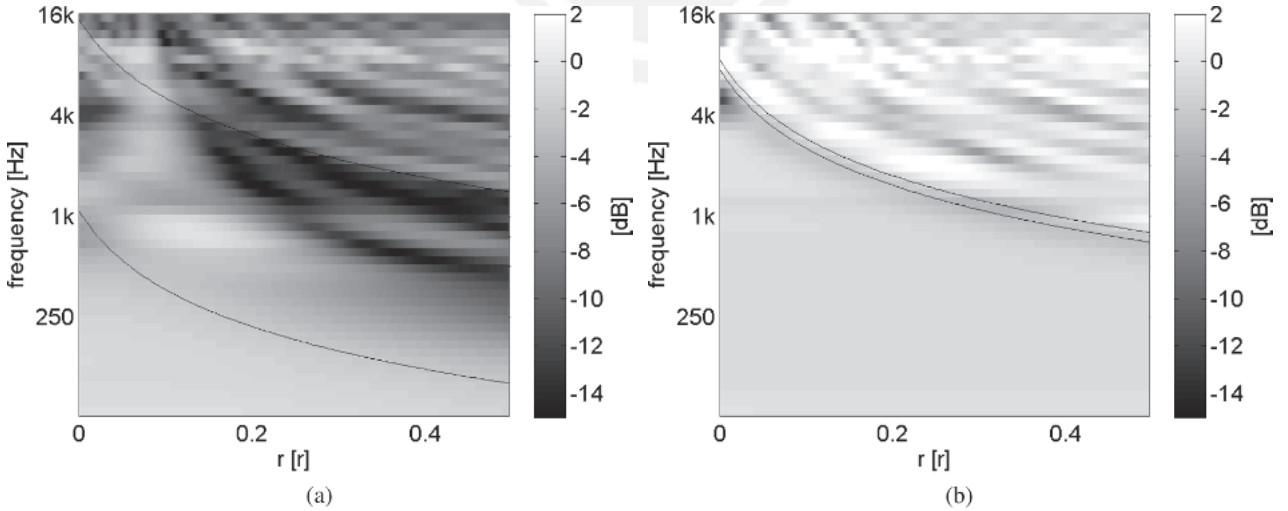

Fig. 11. Spectral coloration for 15 loudspeakers; $kr = N$ and $kr = N - M$ limits plotted in black. (a) Order 1. (b) Order 7.

tradeoff between a lower reproduction error at $kr \approx N$ and spectral impairments for $kr > N$.

This implies that classical Ambisonics can only be used in sweet-spot listening, and compensation must be performed if using more than $2N + 1$ loudspeakers. However, filtering cannot compensate for a regular radial fluctuation pattern, which could lead to a shift in localization or a diffuse sound image over 2 kHz.

When reproducing a sound field over a larger listening area than the sweet spot, HOA should be employed. The number of loudspeakers should be matched to the order that is determined by the radius of the listening area.

4 ACKNOWLEDGMENT

The author would like to thank U. Peter Svensson and Mark Poletti for guidance and support in this work.

5 REFERENCES

[1] D. H. Cooper and T. Shiga, "Discrete-Matrix Multichannel Stereo," *J. Audio Eng. Soc.*, vol. 20, pp. 346–360 (1972 June).

[2] M. A. Poletti, "A Unified Theory of Horizontal Holographic Sound Systems," *J. Audio Eng. Soc.*, vol. 48, pp. 1155–1182 (2000 Dec.).

[3] 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, J. Audio Eng. Soc.* (*Abstracts*), vol. 46, pp. 1034, 1035 (1998 Nov.), preprint 4795.

[4] M. A. Poletti, "Effect of Noise and Transducer Variability on the Performance of Circular Microphone Arrays," *J. Audio Eng. Soc*, vol. 53, pp. 371–384 (2005 May).

[5] E. Hulsebos, T. Schuurmans, D. de Vries, and R. Boone, "Circular Microphone Array for Discrete Multichannel Audio Recording," presented at the 114th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 51, p. 405 (2003 May), preprint 5716.

[6] B. Støfringsdal and U. P. Svensson, "Conversion of Discretely Sampled Sound Field Data to Auralization Formats," *J. Audio Eng. Soc.*, vol. 54, pp. 380–400 (2006 May).

[7] J. Daniel, R. Nicol, and S. Moreau, "Further Investigations of High-Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging," presented at the 114th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 51, p. 425 (2003 May), preprint 5788.

[8] S. Moreau, J. Daniel, and S. Bertet, "3D Sound Field Recording with Higher Order Ambisonics—Objective Measurements and Validation of a 4th-Order Spherical Microphone," presented at the 120th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 54, p. 740 (2006 July/Aug.), convention paper 6857.

[9] M. A. Poletti, "Three-Dimensional Surround Sound Systems Based on Spherical Harmonics," *J. Audio Eng. Soc.*, vol. 53, pp. 1004–1025 (2005 Nov.).

[10] E. Hulsebos, D. de Vries, and E. Bourdillat, "Improved Microphone Array Configurations for Auralization of Sound Fields by Wave-Field Synthesis," *J. Audio Eng. Soc.*, vol. 50, pp. 779–790 (2002 Oct.)

[11] A. Perkis, Y. Abdeljaoued, C. Christopoulos, T. Ebrahimi, and J. F. Chicharo, "Universal Multimedia Access from Wired and Wireless Systems," *Circuits, Sys., Signal Process., Special Issue on Multimedia Communications*, vol. 20, pp. 387–402 (2001).

[12] M. Neukom, "Decoding Second Order Ambisonics to 5.1 Surround Systems," presented at the 121st Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 54, p. 1287 (2006 Dec.), convention paper 6980.

[13] E. Benjamin, R. Lee, and A. Heller, "Localization in Horizontal-Only Ambisonic Systems," presented at the 121st Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 54, p. 283 (2006 Dec.), convention paper 6967.

[14] M. A. Gerzon, "Ambisonics in Multichannel Broadcasting and Video," *J. Audio Eng. Soc*, vol. 33, pp. 859–871 (1985 Nov.).

[15] D. B. Ward and T. D. Abhayapala, "Reproduction of a Plane-Wave Sound Field Using an Array of Loudspeakers," *IEEE Trans. Speech Audio Process.*, vol. 9, pp. 697–707 (2001 Sept.).

[16] M. Poletti, "The Design of Encoding Functions for Stereophonic and Polyphonic Sound Systems," *J. Audio Eng. Soc.*, vol. 44, pp. 948–963 (1996 Nov.).

[17] J. Blauert, *Spatial Hearing*—revised ed.: *The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA, London, UK, 1997).

[18] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, 5th ed. (Academic Press, Cambridge, UK, 2003).

[19] J. G. Proakis and D. G. Manolakis, "*Digital Signal Processing: Principles, Algorithms, and Applications*, 3rd ed. (Prentice-Hall, Upper Saddle River, NJ, 1996).

[20] MATLAB, http://www.mathworks.com/products/matlab/ (2007 May 1).

[21] A. W. Mills, "On the Minimum Audible Angle," *J. Acoust. Soc. Am.*, vol. 30, pp. 237–246 (1958).

[22] M. Slaney, "Auditory Toolbox," version 2, Tech. Rep. 1998–10, Interval Research Corp., Palo Alto, CA, (1998).

[23] E. A. G. Shaw, "Acoustical Features of the Human External Ear," in *Binaural and Spatial Hearing in Real and Virtual Environments*, ch. 2, pp. 25–48 (Lawrence Erlbaum Associates, Mahwah, NJ, 1997).

[24] D. Griesinger, "The Psychoacoustics of Apparent Source Width, Spaciousness and Envelopment in Performance Spaces," *Acta Acustica*, vol. 83, pp. 721–731 (1997).

[25] F. L. Wightman and D. J. Kistler, "The Dominant Role of Low-Frequency Interaural Time Differences in Sound Localization," *J. Acoust. Soc. Am.*, vol. 91, pp. 1648–1661 (1992 Mar.).

APPENDIX

Fourier Series and Its Inverse

$$W(n) = \mathcal{F}[w(\theta)] = \frac{1}{2\pi}\int_{-\pi}^{\pi} w(\theta)e^{-jn\theta}\,d\theta \tag{31}$$

$$w(\theta) = \mathcal{F}^{-1}[W(n)] = \sum_{n=-\infty}^{\infty} W(n)e^{jn\theta}. \tag{32}$$

Discrete Fourier Transform (DFT) and Its Inverse

$$W(n) = \sum_{m=0}^{M-1} w(m)e^{-j\frac{2\pi mn}{M}} \tag{33}$$

$$w(m) = \frac{1}{M}\sum_{n=0}^{M-1} W(n)e^{\frac{2\pi mn}{M}}. \tag{34}$$

Bessel Function of the First Kind

$$J_n(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-jn\theta+jz\sin(\theta)}\,d\theta = \mathcal{F}[e^{jz\sin(\theta)}] \tag{35}$$

$$e^{jz\cos\theta} = \sum_{n=-\infty}^{\infty} j^n J_n(z)e^{jn\theta} = \mathcal{F}^{-1}[j^n J_n(z)] \tag{36}$$

$$\lim_{|z|\to\infty} J_n(z) \approx \sqrt{\frac{2}{\pi z}}\cos\left(z - \frac{n\pi}{2} - \frac{\pi}{4}\right). \tag{37}$$

Circular Sinc, csinc

$$\mathrm{csinc}_{2N+1}(\theta) \triangleq \frac{\sin[(N+\tfrac{1}{2})\theta]}{\sin(\theta/2)} = \sum_{n=-N}^{N} e^{-jn\theta}. \tag{38}$$

Deriving the Mean Relative Intensity

Using the relation in Eqs. (35) and (38) and denoting the DFT of $J_0\,[2kr\,\sin\,(\pi m/M)]$ by

$$\mathcal{D}_{kr,M}(l) \triangleq \sum_{m=0}^{M-1} J_0\left[2kr\sin\left(\frac{\pi m}{M}\right)\right]e^{-j\frac{2\pi ml}{M}} \tag{39}$$

the expression for the mean over the azimuth angle of the relative intensity in Eq. (20) can be expressed as

$$\begin{aligned}
\bar{I}_{\mathrm{rel}}^{\tau}(kr,\theta_1) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} I_{\mathrm{rel}}(kr,\theta,\theta_1)\,d\theta \\
&= \frac{1}{M^2}\sum_{m=0}^{M-1}\sum_{p=0}^{M-1}\mathrm{csinc}_{2N+1}\left(\frac{2\pi m}{M}-\theta_1\right)\mathrm{csinc}_{2N+1}\left(\frac{2\pi p}{M}-\theta_1\right) \\
&\quad\times\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{jkr\left[\cos\left(\frac{2\pi m}{M}-\theta\right)-\cos\left(\frac{2\pi p}{M}-\theta\right)\right]}\,d\theta \\
&= \frac{1}{M^2}\sum_{m=0}^{M-1}\sum_{p=0}^{M-1}\mathrm{csinc}_{2N+1}\left(\frac{2\pi m}{M}-\theta_1\right)\mathrm{csinc}_{2N+1}\left(\frac{2\pi p}{M}-\theta_1\right) \\
&\quad\times\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{j2kr\sin\left(\frac{2\pi m/M-2\pi p/M}{2}\right)\sin\left(\theta-\frac{2\pi m/M+2\pi p/M}{2}\right)}\,d\theta \\
&= \frac{1}{M^2}\sum_{m=0}^{M-1}\sum_{p=0}^{M-1}\mathrm{csinc}_{2N+1}\left(\frac{2\pi m}{M}-\theta_1\right)\mathrm{csinc}_{2N+1}\left(\frac{2\pi p}{M}-\theta_1\right) \\
&\quad\times J_0\left[2kr\sin\frac{\pi(m-p)}{M}\right] \\
&= \frac{1}{M^2}\sum_{p=0}^{M-1}\sum_{n=-N}^{N} e^{jn\left(\theta_1-\frac{2\pi p}{M}\right)}\sum_{l=-N}^{N} e^{jl\theta_1} \\
&\quad\times\sum_{m=0}^{M-1} e^{-jl\frac{2\pi m}{M}}J_0\left[2kr\sin\frac{\pi(m-p)}{M}\right] \\
&= \frac{1}{M^2}\sum_{p=0}^{M-1}\sum_{n=-N}^{N} e^{jn\theta_1}\,e^{-jn\frac{2\pi p}{M}}\sum_{l=-N}^{N} e^{jl\theta_1}[e^{-jl\frac{2\pi p}{M}}\mathcal{D}_{kr,M}(l)] \\
&= \frac{1}{M^2}\sum_{n=-N}^{N}\sum_{l=-N}^{N} e^{j\theta_1(n+l)}\mathcal{D}_{kr,M}(l)M\delta(n+l) \\
&= \frac{1}{M}\sum_{n=-N}^{N}\mathcal{D}_{kr,M}(n) \\
&= \frac{1}{M}\sum_{m=0}^{M-1} J_0\left[2kr\sin\left(\frac{\pi m}{M}\right)\right]\mathrm{csinc}_{2N+1}\left(\frac{2\pi m}{M}\right).
\end{aligned} \tag{40}$$

THE AUTHOR

Audun Solvang was born in Norway in 1977. He received an M.Sc. degree in 2005 from the Norwegian University of Science and Technology (NTNU), majoring in acoustics. Since 2003 he has been following the integrated Ph.D. education at Q2S (Center for Quantifiable Quality of Service in Communication Systems) at NTNU. His research topic has been three-dimensional audio, focusing on representation, reproduction, and perception.