



Audio Engineering Society Convention Paper

Presented at the 131st Convention
2011 October 20–23 New York, USA

This Convention paper was selected based on a submitted abstract and 750-word precis that have been peer reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This convention paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Mixed-order Ambisonics recording and playback for improving horizontal directionality

Sylvain Favrot¹, Marton Marschall¹, Johannes Käsbach¹, Jörg Buchholz², Tobias Weller¹

¹Centre for Applied Hearing Research, Department of Electrical Engineering, Technical University of Denmark, Lyngby, Denmark

²National Acoustic Laboratories, Australian Hearing and Audiology group, Department of Linguistics, Macquarie University, Australia

Correspondence should be addressed to Sylvain Favrot (sf@elektro.dtu.dk)

ABSTRACT

Planar (2D) and periphonic (3D) higher-order Ambisonics (HOA) systems are widely used to reproduce spatial properties of acoustic scenarios. Mixed-order Ambisonics (MOA) systems combine the benefit of higher order 2D systems, i.e. a high spatial resolution over a larger usable frequency bandwidth, with a lower order 3D system to reproduce elevated sound sources. In order to record MOA signals, the location of the microphones on a hard sphere were optimized to provide a robust MOA encoding. A detailed analysis of the encoding and decoding process showed that MOA can improve both the spatial resolution in the horizontal plane and the usable frequency bandwidth for playback as well as recording. Hence the described MOA scheme provides a promising method for improving the performance of current 3D sound reproduction systems.

1. INTRODUCTION

The use of virtual auditory environments in applications such as psychoacoustic or hearing instrument technology research requires a very precise spatial resolution over a large frequency bandwidth for a single listener. Higher-order Ambisonics (HOA) is

used in this context to either (i) auralize simulated virtual auditory environments (e.g., [6]) or to (ii) capture and reproduce acoustic environments (e.g., [5, 15, 1]). In both cases, planar (2D) Ambisonics as well as periphonic (3D) Ambisonics can be used. For a given Ambisonics order M , 2D systems require far fewer loudspeakers and/or microphones ($2M + 1$

versus $(M + 1)^2$) and provide a higher spatial resolution over a larger usable frequency bandwidth. However, 2D systems are restricted to the horizontal plane and thus can not naturally reproduce elevated sound sources. In order to combine the benefits of both 2D and 3D systems, the mixed-order Ambisonics (MOA) scheme has been introduced which combines a higher order 2D system with a lower order 3D system [3, 17]. Since human auditory localization is much more accurate in the horizontal plane than in the vertical plane [2], such a mixed-order approach is also highly encouraged from an auditory perception perspective. Moreover, most sound sources that are of interest in hearing or hearing aid research are located in the horizontal plane. Mixed-order Ambisonics has initially been referred to as “hybrid representation” [3].

For HOA recordings, microphone arrays typically consist of a large number of pressure microphones that are homogeneously distributed on the surface of a rigid sphere. 3D HOA signals are encoded from all the microphone outputs (e.g., [5, 15, 1]). Thereby, the usable frequency bandwidth and spatial resolution are limited by the number of microphones and their distance to each other. Both limitations are of major concern when the recorded acoustic scenes are used for psychoacoustic experiments or the evaluation of hearing instruments. For the same practical number of microphones, but with a higher density on the equator, the MOA approach can theoretically improve the spatial resolution of the array (in the horizontal plane) and increase the usable frequency bandwidth while retaining some vertical spatial information. For the playback, MOA is relevant for arrays with a higher density of loudspeakers in the horizontal plane. For example, installations designed to be used for both 2D and 3D reproduction are particularly suited.

Although the general idea of mixed-order Ambisonics has already been mentioned in the literature [3, 13, 17], no detailed implementations have been proposed so far. This manuscript provides a general framework for both the recording and playback of MOA signals with arbitrary 2D/3D order combination. The technique’s inherent limitations are discussed through the simulation of the whole signal chain for different example recording layouts. An objective evaluation of MOA playback as well as a

quantitative evaluation of the benefit of MOA over standard 3D HOA for microphone arrays are presented.

2. BACKGROUND

First, the principle of using flush mounted microphones on a solid sphere for recording higher-order Ambisonics (HOA) signals is shortly summarized. The notations and nomenclature of the manuscript follow the ones from [12] using spherical coordinates in which a point is described by its radius r , azimuth θ ($0 \leq \theta \leq 2\pi$), and elevation δ ($-\pi/2 \leq \delta \leq \pi/2$) according to the origin O (in the following either the center of the solid sphere or the center of the loudspeaker array).

2.1. Pressure over a sphere

The pressure at the surface of a solid sphere of radius R for a point (R, θ, δ) can be expressed as [12]:

$$p(kR, \theta, \delta) = \sum_{m=0}^{\infty} W_m(kR) \sum_{n=0}^m \sum_{\sigma=\pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta, \delta) \quad (1)$$

with k being the wave number, the weighting factor $W_m(kr)$

$$W_m(kR) = i^m \left(j_m(kR) - \frac{j'_m(kR)}{h'_m(kR)} h_m(kR) \right) \quad (2)$$

and B_{mn}^{σ} the Fourier-Bessel series coefficient or Ambisonics components and $Y_{mn}^{\sigma}(\theta, \delta)$ the (“real-value”) spherical harmonic functions [11] (SHF) described as N3D normalized by [12]:

$$\begin{aligned} Y_{mn}^{\sigma}(\theta, \delta) &= a_{mn}(\delta) \cdot b_n(\theta) \quad \text{with} \\ a_{mn}(\delta) &= \sqrt{2m+1} P_{mn}(\sin \delta) \\ b_n(\theta) &= \begin{cases} \cos n\theta & \text{if } \sigma = +1 \\ \sin n\theta & \text{if } \sigma = -1 \end{cases} \end{aligned} \quad (3)$$

with P_{mn} are the “Schmidt seminormalized” associated Legendre functions.

2.2. Spherical higher-order Ambisonics microphone

Ambisonics components B_{mn}^{σ} can be estimated with a spherical HOA microphone array consisting of Q microphones flush mounted on a rigid sphere. In

this case, Eq. 1 can be approximated in a discrete matrix form truncated at order M such that $K = (M + 1)^2 \leq Q$:

$$\mathbf{T} \cdot \mathbf{b} = \mathbf{s} \quad (4)$$

where \mathbf{b} is the vector of the K Ambisonics components B_{mn}^σ up to order M , \mathbf{s} is the Q -long vector of the q microphone pressure signals $p_q(j\omega)$ and \mathbf{T} represents the transfer matrix of size $Q \times K$

$$\begin{aligned} \mathbf{T} &= \mathbf{Y} \cdot \text{diag}(W_m(kR)) \quad \text{with} \\ \mathbf{Y} &= [\mathbf{c}_0 \cdots \mathbf{c}_q \cdots \mathbf{c}_Q]^T \\ \mathbf{c}_q &= [Y_{00}^{+1}(\theta_q, \delta_q) \cdots Y_{mn}^\sigma(\theta_q, \delta_q) \cdots Y_{M0}^1(\theta_q, \delta_q)] \end{aligned} \quad (5)$$

HOA signals \mathbf{b} are estimated by solving Eq. 4 with a least-squares solution [12, 14]

$$\mathbf{b} = \text{diag} \left[\frac{F_m(kR)}{W_m(kR)} \right] \mathbf{E} \cdot \mathbf{s} \quad (6)$$

where \mathbf{E} is the pseudo-inverse of matrix \mathbf{Y} , which is given by $\mathbf{E} = (\mathbf{Y}^t \mathbf{Y})^{-1} \mathbf{Y}^t$ and $F_m(kR)$ are regularization filters intended to avoid excessive microphone amplification at low frequencies and described by:

$$F_m(kR) = \frac{|W_m(kR)|^2}{|W_m(kR)|^2 + \lambda^2} \quad (7)$$

where λ is the parameter that controls the regularization. The least squares solution minimizes the residual error which gives good results when the microphone layout is regular in the sense that the SHF evaluated at the microphone positions form an orthonormal basis.

Recently, an adaption of this method has been proposed to extract 2D HOA signals from a circular microphone array mounted on a hard sphere [18].

3. MIXED-ORDER AMBISONICS

The mixed-order Ambisonics approach relies on the combination of 3D HOA at a periphonic order M_{3D} with 2D HOA at a planar order of M_{2D} where

$M_{3D} < M_{2D}$. More specifically, it implies the combination of their harmonic functions, cylindrical for 2D HOA and spherical for 3D HOA. A smooth transition of the source characteristics (spatial and energetic) is desired from horizontal to elevated sources. However, the cylindrical and spherical harmonic functions are not compatible with each other due to different weighting functions. In order to retain the flexibility of HOA encoding and decoding, the approach followed in this paper consists of using spherical harmonics functions up to the periphonic order M_{3D} extended with the horizontal functions of spherical harmonics (i.e., $m = n$) until the planar order M_{2D} . With reference to Eq. 1, the pressure of the surface of a solid sphere is then approximated by:

$$\begin{aligned} p(r, \theta, \delta) &\simeq \sum_{m=0}^{M_{3D}} W_m(kR) \sum_{n=0}^m \sum_{\sigma=\pm 1} B_{mn}^\sigma Y_{mn}^\sigma(\theta, \delta) \\ &+ \sum_{m=M_{3D}+1}^{M_{2D}} W_m(kR) \sum_{\sigma=\pm 1} B_{mm}^\sigma Y_{mm}^\sigma(\theta, \delta), \end{aligned} \quad (8)$$

This means that MOA encoding and decoding is similar to HOA encoding (see section 2.2) and decoding (e.g. [4]), but using a mixed-order SHF matrix $\tilde{\mathbf{Y}}$ with rows $\tilde{\mathbf{c}}$ such as

$$\tilde{\mathbf{c}} = [Y_{00}^{+1} \cdots Y_{mn}^\sigma \cdots Y_{M_{3D}0}^1 \cdots Y_{mm}^\sigma \cdots Y_{M_{2D}M_{2D}}^1] \quad (9)$$

with $\tilde{\mathbf{c}}$ of length $K = (M_{3D} + 1)^2 + 2(M_{2D} - M_{3D})$ [9]. Although the realization of alternative MOA schemes is possible [17], the described approach seems to require the least number of microphones.

3.1. Transducer layout

Since the encoding and the decoding of the described MOA scheme rely on the pseudo-inverse of the SHF matrix \mathbf{Y} (least squares error minimization), this matrix needs to be well-conditioned in order to provide a robust encoding and decoding solution. The 2-norm condition number $\kappa(\mathbf{Y})$ of matrix \mathbf{Y} has been shown to be a suitable metric for evaluating the robustness of a transducer layout [16]. It is calculated as the ratio of the largest singular value to the smallest non-zero one and a large value indicates an ill-conditioned problem. In a first attempt at deriving microphone layouts that are suitable for

MOA encoding, a number of Q microphones were randomly placed on a rigid sphere and then their positions were iteratively varied to minimize the conditioning number. Thereby the search routine from Li et al. [10] was applied. Unfortunately this algorithm was always trapped in local minima, providing non satisfactory solutions. In order to overcome this problem, possible microphone layouts were restricted to structures where microphones were distributed along a limited number of latitudes. Since, the use of MOA is intended to provide a higher directionality in the horizontal plane, a higher density of transducers is required on the equator. The total number of transducers Q is required to be larger than the number of SHFs K and is the limiting cost factor in most applications.

First, for MOA of order M_{3D} , M_{2D} , a ring of equiangle $2(M_{2D} + 1)$ transducers at elevation $\delta = 0$ was considered. Then, additional rings with q equiangle transducers at elevation $\pm\delta_i$ completed the layout. For different numbers of elevated transducers and different number of rings, the elevation angle δ_i of the additional rings was optimized to minimize the condition number $\kappa(\mathbf{Y})$ of the corresponding matrix \mathbf{Y} . An example MOA configuration of $M_{3D} = 3$ and $M_{2D} = 7$ ($K = 24$) was considered and, with this method, a number of transducers of at least $Q = 28$ and 5 rings was necessary to obtain a condition number lower than 2. In the following, two layouts were selected, one with $Q = 28$ and the other with $Q = 36$ (see Table 1 for details).

Q	$\kappa(\mathbf{Y})$	θ	δ	q
28	1.72	0	$\pm\pi/2$	2×1
		$2\pi k/5, k = 0..4$	± 0.737	2×5
		$2\pi k/16, k = 0..15$	0	16
36	1.61	0	$\pm\pi/2$	2×1
		$2\pi k/9, k = 0..8$	± 0.727	2×9
		$2\pi k/16, k = 0..15$	0	16

Table 1: Two transducer layouts for $Q = 28$ and $Q = 36$ for MOA $M_{3D} = 3$, $M_{2D} = 7$. For each layout, each line corresponds to a ring.

4. OBJECTIVE EVALUATION

A preliminary objective evaluation was performed to investigate the performance of MOA encoding and

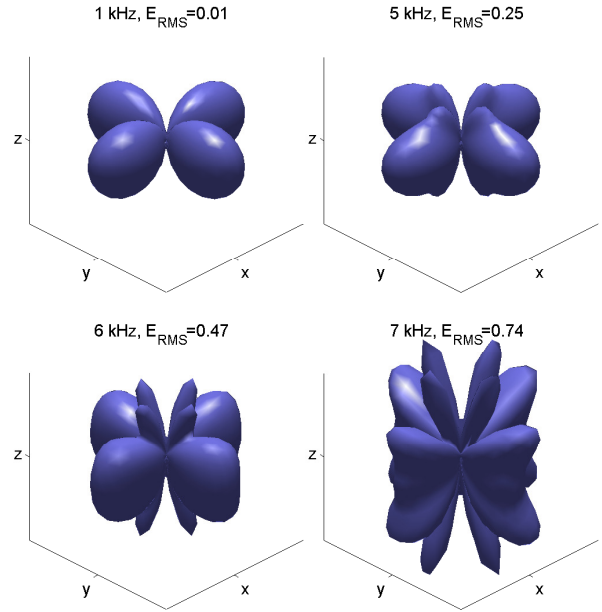


Fig. 1: Directivity characteristic of the SHF Y_{22}^{+1} for the microphone layout $Q = 36$.

decoding for an example configuration of $M_{3D} = 3$ and $M_{2D} = 7$.

4.1. MOA Encoding

For MOA encoding, the directivity characteristics of the encoded SHFs were analyzed. A hard sphere of radius of $R = 5$ cm was considered with either the $Q = 28$ or $Q = 36$ layouts described in the previous section. In order to account for sensor noise, white noise was simulated for each microphone such that the obtained signal-to-noise ratio was +65 dB.

The pressure at the sensors was simulated according to Eq. 1 for 1002 plane waves incoming from full 3D sound incidence (i.e., placing plane wave sources on a regular spherical grid around the microphone array [8]). For each of these plane wave sources, the MOA signals \mathbf{b} were derived according to Eq. 6 with the regularization parameter $\lambda = -40$ dB. Figure 1 shows the resulting directivity pattern of the spherical harmonic function Y_{22}^{+1} for the microphone layout $Q = 36$ for frequencies 1, 5, 6 and 7 kHz.

The encoded directivity patterns match the SHF of Y_{22}^{+1} at 1 kHz and start showing aliasing effects at

the higher frequencies. For an overview of the performance of this encoding, the root mean square (RMS) error E_{RMS} between the obtained and the theoretical SHF patterns was calculated for all SHFs and plotted in Fig. 2 and Fig. 3 against frequency for the layout $Q = 28$ and $Q = 36$ respectively. The y-axis on these figures represent the SHF index ordered from $m = 0..M$ and for each order m , from $n = m..0$ (for the 3D orders) and for each n , $\sigma = 1..-1$.

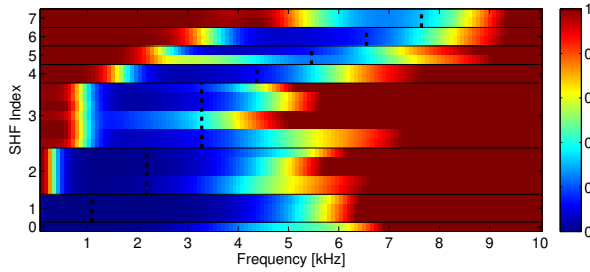


Fig. 2: RMS error (E_{RMS}) of SHFs for MOA $M_{3D} = 3$, $M_{2D} = 7$ and $Q = 28$.

The effect of regularization can be seen at low frequencies where $E_{RMS} = 1$ until a frequency that increases with the order m . The regularization effectively attenuates SHFs of high order to avoid large gains in the same way as in HOA [12]. At high frequency, E_{RMS} increases to large value showing that the obtained SHF patterns become aliased. This aliasing effect depends on the order m of the SHFs and leads to high gains in MOA signals.

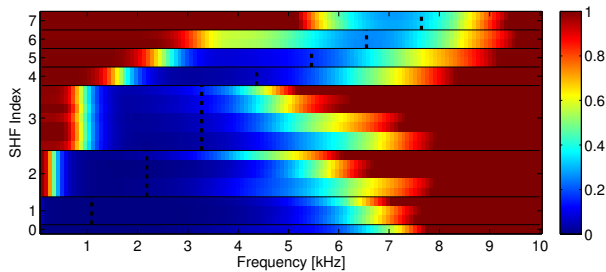


Fig. 3: RMS error (E_{RMS}) of SHFs for MOA $M_{3D} = 3$, $M_{2D} = 7$ and $Q = 36$.

As an indication, for each order m , vertical dashed lines represent the theoretical frequency f_a from where the aliasing error is dominant $2\pi f_a/cR = m$ [15] with c the sound velocity. For the microphone

layout $Q = 36$ (Fig. 3), this aliasing frequency f_a matches the increase of RMS error from $m = 2$. In addition, within each 3D order m , the RMS error increases faster with frequency as n decreases. This means that the non-horizontal SHFs are aliased at lower frequencies than the horizontal SHFs. For the microphone layout $Q = 28$ (Fig. 2), a similar behavior can be observed except for the 3D order $m = 0..3$, where the increase of RMS error does not seem to depend on the order of the SHF. In addition it should be noted that some SHFs ($Y_{32}^{\pm 1}$ and Y_{55}^1) show higher RMS errors than the others of equal order m . The addition of microphones, i.e., from the layout $Q = 28$ to $Q = 36$, provides a larger usable bandwidth of the MOA signals.

4.2. MOA Decoding

The energy vector concept [7] was applied to objectively evaluate the elevation dependency of the energy and directionality of MOA reproduced sound sources.

The MOA encoded signals \mathbf{b} with the microphone layout $Q = 36$ presented in the previous section were simulated for 3 sources with constant azimuth $\theta = 0$ and elevation $\delta = 0, \pi/4$ and $\pi/2$. These simulated MOA signals were decoded onto a loudspeaker layouts with $L = 92$ regularly distributed loudspeakers [8] around a 2 m radius sphere. From the derived gain of the loudspeakers G_i , the norm of the energy vector r_E , the total energy $E = \sum |G_i|^2$ and the sum of all loudspeakers $G = \sum G_i$ are plotted in Fig. 4 (solid line). For comparison, these parameters (dashed line) were also plotted for a 3D HOA microphone array of order $M = 3$ with $Q = 18$ and using the same loudspeaker array.

Below 1 kHz, for the MOA case, the 3 sources produce the same r_E since the higher order horizontal (2D) SHFs are not activated. The norm r_E increases as the SHFs are being activated due to the regularization as seen in Fig. 3. Between 1 and 3.5 kHz, as the 2D SHFs ($M_{3D} \leq m \leq M_{2D}$) are activated, r_E slightly increases to reach 0.85 for the horizontal source $\delta = 0$ and remains constant (approx. 0.77) for the elevated sources. For 3rd order HOA, the horizontal source produces significantly lower r_E at 0.73. At frequencies above 3.5 kHz, r_E for the horizontal source with MOA decreases until 0.5 at 8 kHz, whereas for elevated sources, r_E drops below 0.5 from 4 kHz upwards and shows similar values as for

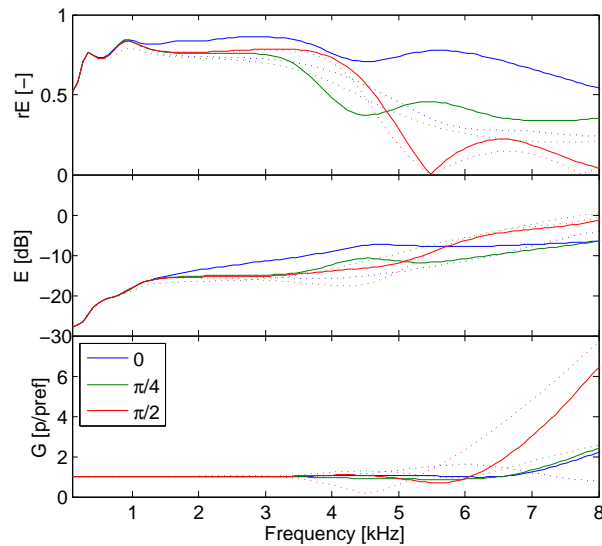


Fig. 4: Energy vector norm r_E , total energy E and total gain G for 3 elevation angles for the microphone layout $Q = 36$ MOA $M_{3D} = 3$, $M_{2D} = 7$ (solid line) and for a microphone layout $Q = 18$ for HOA 3rd order (dashed line).

HOA. This indicates that for horizontal sources the MOA scheme provides both better directivity and a larger bandwidth, and for elevated sources similar directivity as M_{3D} HOA.

It should be noted that the sum of all loudspeaker gains increases from 6 kHz and, for the source at elevation $\delta = \pi/2$, reaches very high values. This is due to the aliasing of the lower order components as shown in Fig. 3. A specific regularization scheme of the encoding is needed in order to attenuate these low order components at high frequency.

5. CONCLUSION

This study has introduced a mixed-order Ambisonics scheme and provides an example that shows that MOA can improve both spatial resolution and usable frequency bandwidth in the horizontal plane for playback as well as recording. Hence, MOA provides a promising method for improving the performance of microphone arrays for psychoacoustic research and hearing aid evaluation when a limited number of loudspeakers are available.

More effort is needed to provide a better regularization scheme (frequency dependent) in order to avoid

very large gains at high frequencies.

6. ACKNOWLEDGMENT

The project was supported by the Oticon foundation and the Hearing CRC, Australia.

7. REFERENCES

- [1] Abhayapala, T.D., and Ward D.B., “Theory and design of high order sound field microphones using spherical microphone array,” presented at the int. conf. on acoustics, speech and signal processing, Orlando, USA, 2002.
- [2] Blauert, J., “Spatial Hearing: The psychophysics of human sound localization,” Revised edition, MIT Press, 1997.
- [3] Daniel, J. (2000), “Representation de champs acoustiques, application a la transmission et a la reproduction de scenes sonores complexes dans un contexte multimedia,” Ph.D. thesis, 1996-2000 Universite Paris 6.
- [4] Daniel, J., “Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new Ambisonic format,” presented at the AES 23rd conference, 2003.
- [5] Elko, G.W. and Meyer, J., (2004) “A scalable spherical microphone array for spatial sound capture,” J. Acoust. Soc. Am. 116, 2525.
- [6] Favrot, S. and Buchholz, J. M., (2010). “LoRA - A loudspeaker-based room auralization system,” Acta Acustica united with Acustica, 96(2), 364-375.
- [7] Gerzon M.A., “General metatheory of auditory localisation,” presented at the AES 92nd convention, Vienna, Austria, 1992.
- [8] Hollerweger, F., “Periphonic sound spatialization in multi-user virtual environments,” Master’s thesis, Institute of Electronic Music and Acoustics, 2005.
- [9] Käsbaach, J., Favrot, S., Buchholz, J., “Evaluation of a mixed-order planar and periphonic Ambisonics playback implementation,” proceedings of the Forum Acusticum 2011, Aalborg, Denmark, 2011.

- [10] Li, Z., Duraiswami, R., Grassi, E., and Davis, L. S., "Flexible layout and optimal cancellation of the orthonormality error for spherical microphone arrays," presented at the int. conf. on acoustics, speech and signal processing, Montreal, Canada, 2004.
- [11] P.M. Morse and H. Feshbach, "Methods of theoretical physics," McGraw-Hill, 1953.
- [12] Moreau, S., Daniel, J. and Bertet, S., "3D sound field recording with higher order ambisonics," presented at the AES120th convention, Paris, France, preprint 6857, 2006.
- [13] Noisternig, M, "a 3D real time rendering engine for binaural sound reproduction," presented at the int. conf. on auditory display, Boston, USA, 2003.
- [14] Poletti, M.A. "Three-dimensional surround sound systems based on spherical harmonics," J. Audio Eng. Soc., 53(11):1004-1025, 2005.
- [15] Rafaely, B., "Analysis and Design of Spherical Microphone Arrays," IEEE, transactions on speech and audio processing, 13(1), 135-143, 2005.
- [16] Rafaely, B., "The spherical shell microphone array," IEEE, transactions on speech and audio processing, 16(4), 740-747, 2008.
- [17] Travis, C. "A new mixed-order scheme for Ambisonics signals," presented at the 1st Ambisonics Symposium, Graz, Austria, 2009.
- [18] Weller, T., Favrot, S. and Buchholz, J. M., "Application of a circular 2D hard-sphere microphone array for higher-order Ambisonics auralization," proceedings of the Forum Acusticum 2011, Aalborg, Denmark, 2011.