Technical University of Denmark

Johannes Käsbach

Development and evaluation of a mixed-order Ambisonics playback system

Master's thesis, November 2010

DTU Electrical Engineering
Department of Electrical Engineering

Development and evaluation of a mixed-order Ambisonics playback system

Report written by:
Johannes Käsbach – S081558

Supervisors:
Jörg Buchholz, Associate Professor
Sylvain Favrot, Post Doc

J. Käsbach

DTU Elektro
Danmarks Tekniske Universitet
2800 Kgs. Lyngby
Denmark

studieadministration@elektro.dtu.dk

Project period:	26-04-2010 – 15-11-2010
ECTS:	35
Education:	M.Sc. Engineering Acoustics
Class:	Public
Edition:	1. edition
Remarks:	This report is submitted as partial fulfillment of the requirements for graduation in the above education at the Technical University of Denmark.

Acknowledgements

First of all, I would like to thank my two supervisors, Jörg Buchholz and Sylvain Favrot, for their great and patient supervision. I enjoyed very much working with them. They had always a free minute for me and taught me a lot of things.

Furthermore, I am happy about the education in our cozy department where I got so much help of all the teachers at any time and who were always open for questions and discussions.

Moreover, I appreciated the support by Jörgen Rasmussen and Tom Arent Petersen concerning any technical issue.

A thankful word has to be said to Jesper Andersen who supplied me with nice and proper recordings for my investigations of a concert listening scenario.

I am also thankful for all the people that participated in my listening experiments and offered their free time for me.

I want to give a big hug to all my friends I got to know during my studies that always kept my motivation alive and discussed with me about life and science.

Finally, I want to thank my parents and my brothers, without them my education would not have been possible and who were always there for me when I needed them.

Thanks a lot to all of you, einen lieben Dank og mange tak.

Johannes Käsbach

Abstract

In Higher Order Ambisonics (HOA) playback systems, which can be used to auralise Virtual Sound Environments (VSE), an authentic reproduction according to human perception is of interest. One approach to this goal is to combine the benefits of conventional 3D and 2D Ambisonic reproduction systems in an optimal way, which in principle can be achieved by following a mixed-order Ambisonics approach (e.g., Travis, 2008).

In this thesis, a mixed-order Ambisonic playback system was developed by extending the spherical harmonics decomposition of the three-dimensional sound field with additional horizontal components. Thereby, consideration of the orthonormality properties of the spherical harmonic functions were necessary to determine the maximal 2D and 3D orders for a given loudspeaker array. Based on this analysis, an alternative mixed-order implementation, that required a truncated order of the inherent Legendre functions, was proposed.

Throughout this study it was shown with means of objective evaluations that both approaches "effectively" improved the directional focus and spatial horizontal resolution of periphonic (3D) systems and approached the high spatial resolution of surround (2D) systems in case of horizontal reproduced sources. The importance of a regular horizontal loudspeaker ring integrated into the setup was thereby highlighted. The implemented algorithms provided a smooth transition in spatial resolution with increasing elevation of sound sources from the horizontal plane.

The advantages of the two mixed-order Ambisonics systems observed in the objective evaluation were confirmed by two listening tests. In the first experiment, 12 normal hearing listeners evaluated the apparent source width of an anechoic pulsed noise signal. In the second, pilot experiment, 8 normal hearing listeners rated different spatial qualities for two complex listening situations, consisting of multiple instruments playing in a highly reverberant environment. Simultaneously, the present implementations led to artifacts, such as spectral coloration effects and an overall power that is dependent on the elevation angle. Both artifacts were more complicated to treat as in a conventional 2D and 3D system. In addition, elevated sound source locations (apart from the horizontal plane) led to different (downshifted) apparent source positions. Possible solutions to the different problems were suggested.

Contents

1 Introduction 9

2 Theory 12

2.1 Human sound localisation 12

2.2 Higher order Ambisonics - 3D 17

2.2.1 Deriving the Bessel-Fourier series for the pressure field 17

2.2.2 Spherical harmonic functions 19

2.2.3 Encoding, re-encoding and decoding 21

2.2.4 First-Order Ambisonic (B-Format) 24

2.3 Higher order Ambisonics - 2D 25

3 Objective analysis tools 28

3.1 Loudspeaker arrays 28

3.2 Directivity plots 30

3.3 Monochromatic soundfield plots 32

3.4 Velocity and energy vector 33

3.5 Orthonormality Matrix 39

4 Implementation of a mixed-order Ambisonics playback system 42

5 Objective evaluation 46

5.1 Directivity analysis 46

5.2 Quasi-regular system (92 LS-array) 48

5.3 Non-regular and symmetric system (30 LS-array) 52

5.4 Frequency spectrum considerations 56

6 Subjective evaluation 59

6.1 Experiment A 59

6.1.1 Methods 60

6.1.2 Results 65

6.1.3 Discussion 68

6.2 Experiment B 70

6.2.1 Methods 70

6.2.2 Results and discussion 72

7 Further and future work 75

7.1 Transition frequency and energy normalisation 75

7.2 Optimisation decoding methods 78

7.3 Future experiments 79

8 Conclusions 80

9 Appendix 82

List of Figures 93

List of Tables 93

List of Abbreviations 94

List of Symbols 94

References 97

1 Introduction

Higher Order Ambisonics (HOA) is a spatial sound encoding and decoding approach dealing with the physical reconstruction of sound fields. In these terms it refers to the field of science of holophony, which is often considered as being the acoustical equivalent to holography.

For the reconstruction, periphonic (3D) as well as surround (2D) playback systems are conventionally available that both reveal advantages and disadvantages. While systems of the latter category have higher spatial resolution in the horizontal domain, they are simultaneously limited to this domain and are not able to naturally represent elevated sound sources, which is the most advantageous feature of periphonic systems. In turn, 3D systems require more loudspeakers in order to achieve a comparable resolution. When considering human abilities in spatial localisation, the importance of the horizontal plane, where human accuracy is much better than apart from this plane, is highlighted ([1]). Moreover, many natural sound sources appear in that plane in all day life situations, such as for example conversation, traffic noise or music performances.

The before mentioned systems reach an ideal performance for regular setups, which refers to equally distributed loudspeakers in the according spatial domain, i.e. on a ring in a 2D and on the surface of a sphere in a 3D representation. While this is an easy task in case of a loudspeaker ring, uniformly distributed loudspeakers on a sphere are limited to five specific cases linked to the vertices of Plato's polyhedrons or approximated ideal solutions have to be used[1]. In addition, practical solutions of 3D Ambisonics playback systems are often given with a horizontal loudspeaker ring with additional loudspeakers in elevated positions.

The here described facts make it reasonable to develop 3D systems with an improved horizontal resolution or 2D systems with added 3D information in order to fit the technical to the human auditory system and thereby using technical resources in an optimal way. Such systems will be referred to as mixed-order Ambisonics in the following.

Ambisonics was originally developed in the 1960s and 70s, where the research work of Michael Gerzon contributed mostly to its progress. Thereafter, its extensions to higher-order systems known as HOA were invented. An elaborated study is given in Jérôme

[1] Plato's polyhedrons are: Tetrahedron, hexahedron, octahedron, dodecahedron and icosahedron [14]. Approximated ideal solutions are provided by [10].

Daniel's thesis [5] and comprehensive results are presented in [6], where also a comparison to another common holophonic approach, the Wave Field Synthesis (WFS) technique based on the Huygens-Fresnel principle and the Kirchhoff-Helmholtz Integral (KHI), is described. Mixed-order approaches have been mentioned in the literature e.g. in [20] with the suggestion of an alternative two-parameter scheme (#H#V) and also in [10], where the terminology hybrid system is used. In the before mentioned work of Daniel, the term mixed-order system is used in context with the combination of Ambisonic signals encoded with different orders.

At the research center CAHR at the Department of Electrical Engineering at the Technical University of Denmark (DTU) a virtual auditory environment called loudspeaker-based room auralization (LoRA) Toolbox that uses HOA coding strategies was developed by Favrot in [7]. The system, implemented at the facilities of DTU (Spacelab), consists of 29 loudspeakers in an irregular setup and is a practical solution to a 3D Ambisonics playback system. Complex real-life sound environments can be tested on a listener in such a setup and allow for (1) the performance of basic research on the signal processing and perception of the normal, impaired, and aided-impaired auditory system and (2) for evaluation, testing, and fitting of binaural (two interconnected) hearing aids. The improvement of the LoRA system by a mixed-order implementation is desired in order to achieve a maximal authenticity of such scenes. In the future, applications of recordings of complex listening scenarios, as for example in cafeterias or train stations, by a novel microphone array technique such as in [22] that support a mixed-order Ambisonics system are of interest. However, before such microphone array can be applied, first the corresponding loudspeaker playback technique needs to be investigated.

The aim of this thesis is to develop a mixed-order Ambisonics playback system and to investigate and evaluate its properties. In this sense, fundamental background is given in chapter 2. First, human sound localisation is summarised in section 2.1. Adjacently, the basic theory behind Higher Order Ambisonics will be explained in section 2.2 and 2.3 and common analysis tools that are necessary to investigate an Ambisonic system by simulation studies based on this theory are presented in chapter 3. The theory is explained for the case of a periphonic (3D) as well as for a surround (2D) Ambisonic system. In the 3D case an ideal reconstruction system with a (approximated) regular 92 loudspeaker array and an irregular but symmetric system with 30 loudspeakers, which is an idealised adapted array to the Spacelab, are used for simulation studies. A regular horizontal loudspeaker ring consisting of 16 loudspeakers, which is part of the latter system, will be used in the 2D case. Practical limitations will be highlighted throughout the analysis.

An algorithm for a possible solution of a mixed-order Ambisonic system that bases on the extension of horizontal components in the spherical harmonic functions is developed and described in chapter 4. The importance of the orthonormality properties of the spherical harmonic functions as a measure of the reproduction quality is highlighted and these properties will be used to determine the combined horizontal and periphonic order of a certain loudspeaker setup.

The performance of such an implementation is studied in an objective analysis in chapter 5 making use of the before derived common analysis methods. The investigations focus thereby on the high frequency domain, where reproduction is most sensitive and energy unbalances are introduced to the system. Throughout the entire thesis the developed mixed-order system is compared to the state-of-the-art 3D and 2D implementation.

In chapter 6 these systems are evaluated in two listening experiments and the results are compared to the simulation studies. The listening tests are performed in the Spacelab.

Further analysis in respect to spectral changes of the different systems under investigation, that go in hand with directional dependent energy contributions, and suggestions about their treatment are presented and discussed in section 5.4 and chapter 7.

2 Theory

2.1 Human sound localisation

As mentioned in the introduction the main motivation for investigating a mixed-order Ambisonic system is to adapt a sound reproduction system to the limitations of the hearing system and thereby using resources of the system in an optimal way. Before going into details about technical aspects, human spatial hearing properties should be outlined. This chapter is thought of giving a brief summary of this topic referring to [1] by focussing on information that is important for the system design as well as for the experimental design for the subjective evaluation procedure.
Following the designation and definitions in [1], the terminology sound event is used to describe physical properties of a source and an auditory event as the perceptional aspects of this source. Regarding Figure 1, where the human head is supposed to be placed in the middle of the coordinate system facing the positive x-axis, spatial hearing can be considered in different planes, i.e. in the horizontal plane (xy plane), the median plane (xz plane) and the frontal plane (yz plane). Perception of distance is linked to the radius r.

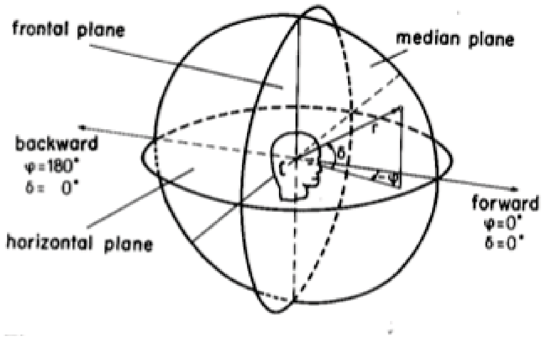


Figure 1: Head-related spherical coordinate system [1]. Note that φ denotes the azimuthal angle in contrast to the in Figure 6 introduced spherical coordinate system.

Being concerned with spatial hearing, absolute localization of an auditory event regarding the introduced coordinate system and the ability of just audible discrimination between two presentations, called the localization blur, are measures of investigation. The horizontal plane is thereby of special interest since it is the most accurate one in human spatial perception. For frontal sources the just noticeable difference (JND) takes a resolution of up to $\Delta(\varphi = 0) = 1°$. This value is dependent on the type of stimulus. An overview is given in [1], p.39, Table 2.1. The localization blur in the horizontal plane for 4 different source positions is shown in Figure 2 where pulsed white noise has been used as stimulus.

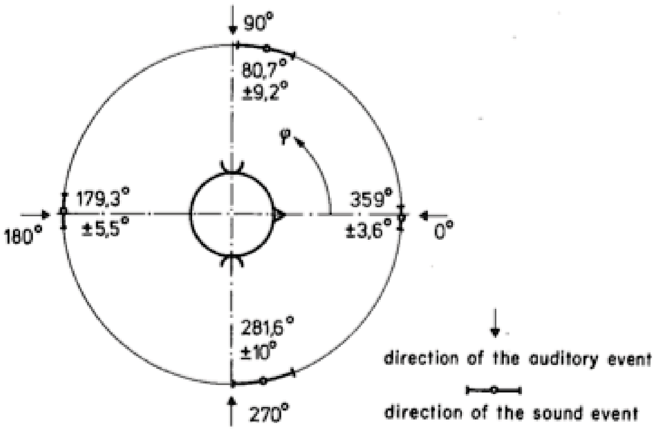


Figure 2: Localisation blur in the horizontal plane for pulsed white noise at 70 phon with a pulse width of 100ms [1]. Here φ denotes the azimuthal angle.

In case of different ear signals (dichotic condition), i.e. for a sound source positioned to the right or the left of the median plane, two spatial cues mainly provide human's for localising the source's direction: Interaural time differences (ITDs) and interaural level differences (ILDs), where the former ones are mainly used to localise low frequency sounds and the latter ones for high frequencies. Sound waves diffract around the head, which leads to time delays between the two ears (due to the different path length) providing ITDs. Interaural time differences range from 0 μs for a frontal source position at 0° to 690 μs for a source position at the opposite ear (±90°) [13]. The hearing system can use the signal's fine structure itself as a cue, also referred to as interaural phase differences (IPDs), up to an approximate frequency limit of 1500 Hz or using the signal's envelope structure, which requires non-stationary signals and works also for higher frequencies. With increasing frequency the head's dimensions become comparable with the wavelength. Reflections occur at the head, causing a shadow that results into different sound pressure levels at the ears and thereby provides ILDs. Whereas level differences do not occur naturally below 500 Hz, they can take values of more than 20dB at high frequencies (around 6 kHz) ([13], pp.235-238), which is true for sources in the far field. The situation is different for close sources, where ILDs also can deal as a cue for evaluating the distance of the source and occur even at lower frequencies.

The transition frequency between low and high frequencies depends on the considered cues and tasks and therefore cannot be given with a single accurate value, but is often given as a transition area of 1.5 to 2 kHz (according to [5], p.35). The distinction between low and high frequencies is an issue for models used for objective evaluations of Ambisonic reproduction systems (see section 3.4) and is crucial in the Ambisonic system design itself (see chapter 7).

An effect, referred to as cone of confusion, describes the fact that an auditory event can be localised in the front or in the rear in areas of source positions that cause identical ILDs and ITDs. Such ambiguities can be resolved by head rotation. Very crucial in spatial detection is the frequency content of sound sources. In general broad band signals provide additional cues and help to resolve ambiguities. Such cues are given by interferences between reflections from the pinna and the torso. This results into characteristic filters for different source positions, referred to as head related transfer functions (HRTFs). For broadband signals differences in magnitude and phase between single frequency components are provided cues, whereas for narrowband signals the phase information is lost. In the measurements of acoustical sound fields generated by an Ambisonic reproduction system often dummy heads are used (e.g. in [19]) in order to provide appropriate spectral information for objective evaluations as performed in section 5.4.

Since in the median plane binaural cues like ITDs and ILDs cannot be used due to identical ear signals (diotic condition), auditory localisation only relies on monaural cues such as the spectral content. By nature, that results into a localization performance that is worse than in the horizontal plane: The JND in that plane takes the smallest value $\Delta(\delta = 0) = 4°$ for white noise. The localization blur in the median plane is illustrated in Figure 3 for a continuous speech signal.

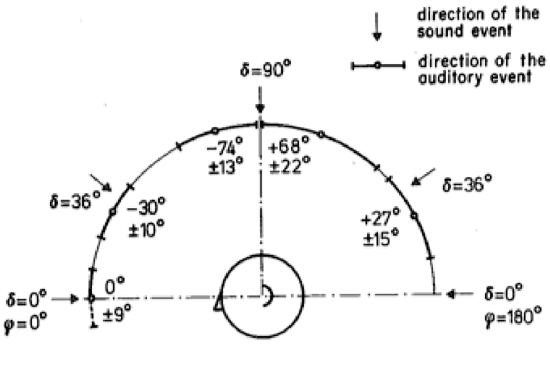


Figure 3: Localization blur in the median plane for continuous speech [1]. Here φ denotes the azimuthal angle and δ the elevation angle.

It can be deduced from the figure that there are disagreements between source positions and the auditory events indicated by front-back reversals and that the localization blur takes higher values compared to the horizontal plane. It should just be mentioned that there is a trend for short pulsed signals that they are most likely perceived in the backward plane.

In the following, for the design of listening tests such as Experiment A (section 6.1) and its evaluation, it is important to distinguish between limits in human performance such as a difference limen (JNDs) and technical limitations of the investigated Ambisonic system. The here made considerations make it reasonable to use broad band signals in localisation experiments in order to provide sufficient frequency information to the listener. In addition, the familiarity with a certain signal is very important. For white or pink pulsed noise the localisation is correct, i.e. the sound source and auditory event are congruent, in 90 % without familiarizing the test-subject with the signal, but training further improves performance.

In the special case of narrowband signals, a source is localised depending on the signal's center frequency independent from the real source location. By increasing the center frequency of a narrowband signal the auditory event follows frequency dependent paths as illustrated in Figure 4. In the design of Ambisonic reproduction systems, spectral changes (see section 5.4) are introduced due to system specific filter characteristics and can be a reason for incorrect apparent sound source locations in case of pronounced (narrowband) frequency regions (see section 6.1.3).

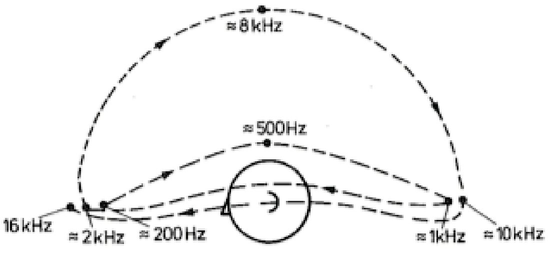


Figure 4: Simplified illustration of the path of an auditory event in the median plane depending on the signal's center frequency for narrow-band signals incident anywhere in that plane [1].

In Figure 5 ITD thresholds as a function of signal duration are plotted for various literature data of different stimulus types (see [9]). It illustrates the fact that localisation improves with the signal duration, where a saturation can be reached even at 50 ms for a 1 kHz sinusoid (crosses, data from Ricard and Hafter (1973)) or at maximum 700 ms for broadband noise bursts (squares, data from Tobias and Zerlin (1959)). The signal length is another aspect that has to be considered in the design of listening experiments. It has to be long enough to provide sufficient information (here described in terms of ITDs) for the auditory system of a human listener.

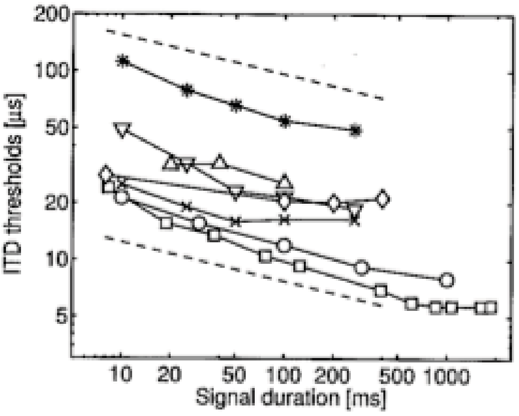


Figure 5: ITD thresholds as a function of signal duration. Summary of experimental results from various literature [9]. The average slope of all literature data is shown by the dashed lines for thresholds measured for stimulus durations between about 10 and 400 ms.

When two or multiple coherent sources are present, the according auditory event is localised in the direction of the sound that arrives earlier. This is called the precedence effect and refers to the law of the first wavefront. This effect explains the importance of direct sound compared to reflections for localisation issues in a room for example. Therefore direct sound needs highest reproduction precision in applications of virtual sound environments as it is implemented in the LoRA toolbox used in Experiment B (section 6.2).

2.2 Higher order Ambisonics - 3D

The Ambisonic principle is derived from solving the wave equation in spherical coordinates for a central listening spot resulting into a limited reconstructed sound field area called the sweet spot. It bases on the decomposition of the physical sound field into spherical harmonic functions, where the reproduction precision and the dimensions of the sweet area are dependent on the order of these functions. In the coding process, the assumption is made that loudspeakers as well as sources emit plane waves and therefore no information about the distance of the source is provided. The encoding of finite distant sources and encoding near field characteristics are described in [6], but are not part of this thesis.

2.2.1 Deriving the Bessel-Fourier series for the pressure field

A sound field is generally described in the frequency domain by the well-known homogeneous Helmholtz equation, assuming wave propagation in a linear and lossless medium and assuming time-invariance, and is in terms of the pressure

$$\nabla^2 p + k^2 p = 0. \tag{1}$$

The wavenumber is given by $k = \omega / c$ where $\omega = 2\pi f$ is the angular frequency and c is the speed of sound. The homogeneous equation indicates the lack of sources inside the considered field. In order to get to the mathematical principle of Ambisonics, the Helmholtz equation is solved in the spherical coordinate system.

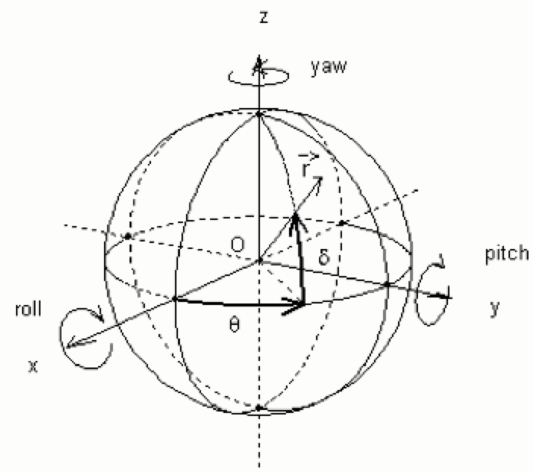


Figure 6: Spherical coordinate system with the three elementary rotation degrees [6].

The methodology of finding the solution to the homogeneous Helmholtz equation is shortly summarised by following [23] and the conventions concerned with Ambisonics used in this thesis follow those of [6]. In the spherical coordinate system under consideration (with origin at $\vec{r} = 0$) each point is described by the two angles azimuth θ and elevation δ and the radius r as indicated in Figure 6. As can be seen from the figure the azimuth angle θ is measured counterclockwise from the positive x-axis in the xy plane and the angle δ describes the elevation from exactly this plane. The transformation from spherical coordinates to cartesian coordinates is given by

$$\begin{aligned} x &= rcos\delta cos\theta \\ y &= rcos\delta sin\theta \\ z &= rsin\delta \end{aligned} \tag{2}$$

and the transformation from cartesian to spherical coordinates by

$$\begin{aligned} \theta &= arctan\frac{y}{x} \\ \delta &= arctan\frac{z}{\sqrt{x^2+y^2}} \\ r &= \sqrt{x^2+y^2+z^2}. \end{aligned} \tag{3}$$

Writing the homogeneous Helmholtz equation as suggested in the spherical coordinate system the Laplace operator ∇^2 becomes

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\cos\delta}\frac{\partial}{\partial\delta}\left(\cos\delta\frac{\partial}{\partial\delta}\right) + \frac{1}{r^2\cos^2\delta}\frac{\partial^2}{\partial\theta^2}. \tag{4}$$

Solving eq. (1) by means of the method of separation of variables the solution can be expressed as a product of functions depending on only one coordinate (time considerations are neglected here)

$$p(r,\theta,\delta) = p_r(r)p_\theta(\theta)p_\delta(\delta). \tag{5}$$

Including the solutions to each individual homogenous ordinary differential equation the final solution for the pressure field is then given by the spherical Fourier-Bessel series

$$\begin{aligned} p(r,\theta,\delta) = &\sum_{m=0}^{\infty} j^m \underbrace{j_m(kr)}_{p_r(r)} \sum_{0\leqslant n\leqslant m,\sigma=\pm1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta,\delta) \\ &+ \sum_{m=0}^{\infty} j^m \overbrace{h_m^{-}(kr)} \sum_{0\leqslant n\leqslant m,\sigma=\pm1} A_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta,\delta). \end{aligned} \tag{6}$$

The nature of this equation is shortly discussed in the following. Considering the general solution for $p_r(r)$ it is described by an arbitrary combination of a spherical Bessel function $j_m(kr)$ and a spherical Neumann function $n_m(kr)$ both of order m ([11])

$$p_r(r) = c_m j_m(kr) + d_m n_m(kr). \tag{7}$$

In terms of the used definitions the pressure field can be split into a "through-going" and an "outgoing" field. In the latter case that is caused by inside sources it can be shown that $d_m = -jc_m$ in order to satisfy the boundary condition at infinity which is also known as the Sommerfield radiation condition. This is resulting into the divergent spherical Hankel functions $h_m^- = j_m(kr) - jn_m(kr)$ which is the term found in the second series of eq. (6) and is associated with the weighting coefficients A_{mn}^{σ}. Since Ambisonics basically assumes a "centered listening area that is free of virtual sources" [6], meaning that there are no inside sources, it follows that $A_{mn}^{\sigma} = 0$. Ambisonics can rather be understood by the description of a plane wave field in terms of the spherical coordinate system, meaning that the considered listening area describes the "through-going" sound field caused by a source far away and outside this area. This is only possible when the coefficient d_m in eq. (7) is zero since the Neumann functions diverge at $kr = 0$ and a finite representation which is given by the spherical Bessel functions is needed. This refers to the first series in eq. (6) where B_{mn}^{σ} are the associated weighting coefficients.

2.2.2 Spherical harmonic functions

In the Ambisonic approach, taking the before made considerations into account, the pressure field is decomposed into a series of radial functions and directional functions Y_{mn}^{σ} called spherical harmonics that are weighted by the weighting coefficients B_{mn}^{σ} (which are described in the next section)

$$p(r,\theta,\delta) = \sum_{m=0}^{\infty} j^m j_m(kr) \sum_{0\leqslant n\leqslant m,\sigma=\pm 1} B^{\sigma}_{mn} Y^{\sigma}_{mn}(\theta,\delta). \tag{8}$$

The spherical harmonic functions[2]are here defined as

$$Y^{\sigma(N3D)}_{mn}(\theta,\delta) = \sqrt{2m+1} N_{mn} \underbrace{P_{mn}(\sin\delta)}_{p_\delta(\delta)} \cdot \begin{cases} \cos n\theta \text{ if } \sigma = +1 \\ \underbrace{\sin n\theta \text{ if } \sigma = -1 \text{ (ignored if n=0)}}_{p_\theta(\theta)} \end{cases} \tag{9}$$

where

$$N_{mn} = \sqrt{(2-\delta_{0,n})\frac{(m-n)!}{(m+n)!}}. \tag{10}$$

is a normalisation factor in the version of the common Schmidt semi-normalisation. The eigenfunctions $P_{mn}(x)$ corresponding to the solution of $p_\delta(\delta)$ are the associated Legendre functions of degree m and order n = 0,1,...,m, evaluated for each element of $x = \sin\delta$. The solution for $p_\theta(\theta)$ is given by one of the two linear independent trigonometric functions $\cos n\theta$ and $\sin n\theta$ depending on wether σ is +1 or −1. The Kronecker symbol given by δ_{pq} equals unity only for $p = q$ and is naught otherwise. The spherical harmonic functions represent a product of linear independent eigenfunctions and therefore form an orthonormal base

$$\left\langle Y^{\sigma}_{mn} | Y^{\sigma'}_{m'n'} \right\rangle = \delta_{mm'}\delta_{nn'}\delta_{\sigma\sigma'} \tag{11}$$

meaning that at least one index must be different in order to get two linear independent eigenvectors, where the spherical scalar product is defined as

$$\langle F|G\rangle_{4\pi} = \frac{1}{4\pi} \oint F(\theta,\delta)G(\theta,\delta)d\Omega. \tag{12}$$

In Figure 7 the spherical harmonic functions with the usual designation are illustrated in three dimensional plots. The functions Y^{-1}_{m0} are zero for any m and are therefore ignored. For each order m there are $(2m+1)$ components available, including 2 horizontal components (those with n=m) for $m \geqslant 1$. Here the first 16 components corresponding to

[2] The given definition of the spherical harmonic functions refers also to the full normalisation 3D (N3D) convention. Without the factor $\sqrt{2m+1}$ the semi-normalisation 3D (SN3D) convention is represented. Further conventions and their conversions are presented in [5] (pp.156-157).

$m = 0...3$ are shown. These patterns describe the directional selectivity, both in encoding and reproduction of a sound source.

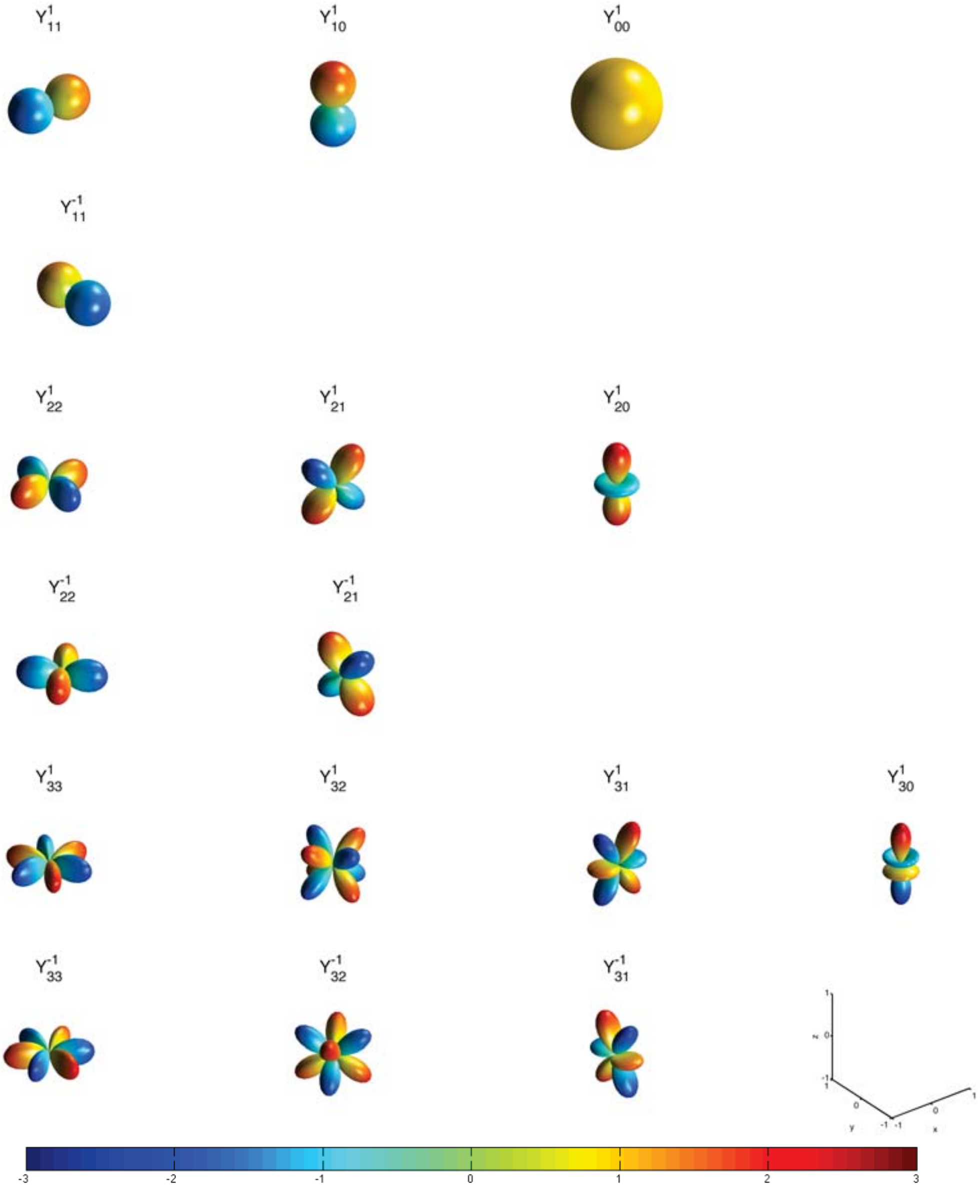


Figure 7: 3D spherical harmonics for $m = 0$ to $m = 3$.

2.2.3 Encoding, re-encoding and decoding

As has been stated before the directional functions $Y^{\sigma}_{mn}(\theta, \delta)$ are weighted with the coefficients B^{σ}_{mn} which are determined in the following. Describing the pressure that is generated by a plane wave, originating from a source direction $(\theta_{src}, \delta_{src})$ and conveying the signal s_{src}, in a spherical coordinate system (with origin at $\vec{r} = 0$) results into a similar series as in eq. (8) by following [10]

$$p(r,\theta,\delta) = s_{src}\sum_{m=0}^{\infty} j^m \sum_{0\leqslant n\leqslant m,\sigma=\pm1} Y_{mn}^{\sigma}(\theta_{src},\delta_{src})Y_{mn}^{\sigma}(\theta,\delta)j_m(kr). \tag{13}$$

The two pressure field descriptions in eq. (8) and eq. (13) should give the same result for a certain position (r,θ,δ) and therefore allow to compare them. It is apparent that the weighting coefficients B_{mn}^{σ} for a plane wave are determined by

$$B_{mn}^{\sigma} = s_{src}Y_{mn}^{\sigma}(\theta_{src},\delta_{src}). \tag{14}$$

This equation represents the encoding process of sound sources in the approach of Ambisonics. The information that is encoded for a single sound source is its signal s_{src} multiplied with the value of the respective spherical harmonic function Y_{mn}^{σ} evaluated at the direction $(\theta_{src},\delta_{src})$. Depending on the total number of Ambisonic components a single source is encoded in a set of signals B_{mn}^{σ} each of which are also referred to as Ambisonic channels [5].

So far, an infinite Fourier-Bessel series has been considered. In reality this series has to be truncated by the Ambisonic order M due to practical limitations. Doing so leads to the truncated Bessel-Fourier series

$$p(r,\theta,\delta) = \sum_{m=0}^{M} j^m j_m(kr) \sum_{0\leqslant n\leqslant m,\sigma=\pm1} B_{mn}^{\sigma}Y_{mn}^{\sigma}(\theta,\delta). \tag{15}$$

Thereby, also the maximum order for the spherical harmonic functions is determined which leads to the total number of Ambisonic components

$$K_{3D} = (M+1)^2. \tag{16}$$

The following remark should be made at this point: Equations (8) and (13) have been used to derive the important encoding equation (14), but do not play any further role themselves in the Ambisonic encoding and decoding algorithm, at least not in a playback system where simulated sources are used. The Bessel functions $j_m(kr)$ are not part of the algorithm's implementation and therefore no considerations of the radial distance is provided. These functions get important for example when dealing with encoding based on sound field recording techniques using a proper microphone array design such as described in [14].

Further, the aim is to derive appropriate loudspeaker gains, where the loudspeakers are considered to be regularly distributed on the surface of a sphere. As stated before the loudspeakers are assumed to emit plane waves allowing for re-encoding each loudspeaker at its specified direction $(\theta_{ls}, \delta_{ls})$ in a representation comparable to eq. (14)

$$\tilde{B}^{\sigma}_{mn} = \sum_{j=1}^{L} s_{ls_j} Y^{\sigma}_{mn}(\theta_{ls_j}, \delta_{ls_j}). \tag{17}$$

This leads to the re-encoding principle in matrix notation

$$\vec{\tilde{B}} = C\vec{s}_{ls}$$

$$\vec{\tilde{B}} = \begin{pmatrix} \tilde{B}^1_{00} \\ \tilde{B}^1_{11} \\ \tilde{B}^{-1}_{11} \\ \vdots \\ \tilde{B}^{\sigma}_{mn} \end{pmatrix} C = \begin{pmatrix} Y^1_{00}(\theta_{ls_1}, \delta_{ls_1}) & Y^1_{00}(\theta_{ls_2}, \delta_{ls_2}) & \cdots & Y^1_{00}(\theta_{ls_L}, \delta_{ls_L}) \\ Y^1_{11}(\theta_{ls_1}, \delta_{ls_1}) & Y^1_{11}(\theta_{ls_2}, \delta_{ls_2}) & \cdots & Y^1_{11}(\theta_{ls_L}, \delta_{ls_L}) \\ Y^{-1}_{11}(\theta_{ls_1}, \delta_{ls_1}) & Y^{-1}_{11}(\theta_{ls_2}, \delta_{ls_2}) & \cdots & Y^{-1}_{11}(\theta_{ls_L}, \delta_{ls_L}) \\ \vdots & \vdots & \ddots & \vdots \\ Y^{\sigma}_{mn}(\theta_{ls_1}, \delta_{ls_1}) & Y^{\sigma}_{mn}(\theta_{ls_2}, \delta_{ls_2}) & \cdots & Y^{\sigma}_{mn}(\theta_{ls_L}, \delta_{ls_L}) \end{pmatrix} \vec{s}_{ls} = \begin{pmatrix} s_{ls_1} \\ s_{ls_2} \\ \vdots \\ s_{ls_L} \end{pmatrix} \tag{18}$$

The associated spherical harmonic functions Y^{σ}_{mn} evaluated for each loudspeaker position are written in a matrix C, called the re-encoding matrix, where each column contains the sampled spherical harmonic functions for one loudspeaker. Due to a limited amount of in total L loudspeakers the formalism becomes discrete[3]. Taking the practical limitations of a discrete loudspeaker array and a finite number of Ambisonic components into account matrix C is of dimensions K x L.

In order to derive the necessary loudspeaker gains s_{ls} the following boundary condition is applied, ensuring that the encoded soundfield of a single source equals the resynthesized soundfield

$$\begin{aligned} B^{\sigma}_{mn} &= \tilde{B}^{\sigma}_{mn} \\ s_{src} Y^{\sigma}_{mn}(\theta_{src}, \delta_{src}) &= \sum_{j=1}^{L} s_{ls_j} Y^{\sigma}_{mn}(\theta_{ls_j}, \delta_{ls_j}). \end{aligned} \tag{19}$$

Note that in Ambisonic systems all loudspeakers are always contributing to the resynthesized sound field, no matter of which direction a sound source is reproduced. The orthonormality property of the spherical harmonics allows thereby for the reconstruction

[3] meaning that the continuous integral of the scalar product in eq. (12) becomes discrete as well.

where the integration over a solid angle of 4π in the definition of the scalar product in eq. (12) is satisfied. Equation (19) makes it possible to solve eq. (18) for the loudspeaker gains by inverting matrix C resulting into the decoding process

$$\vec{s}_{ls} = C^{-1}\vec{B} = D\vec{B} \tag{20}$$

The resulting matrix D is called decoding matrix[4] and has the dimensions L x K. To ensure that eq. (18) is not an undetermined system of equations (undersampling), which would result into spatial aliasing, there should be at least as many loudspeakers as encoded Ambisonic channels

$$L \geqslant K \tag{21}$$

It is common practice to use more loudspeakers ($L > K_{3D}$) resulting into an overdetermined system. There is an upper limit though since otherwise perceptual artefacts such as coloration are introduced into the reproduction due to coherent loudspeaker signals [19]. The inverse matrix of C in eq. (20) can just be determined for a square matrix. In case of the non-square matrix the inversion is given by the pseudo-inverse defined as

$$D = pinv(C) = C^T(CC^T)^{-1}, \tag{22}$$

where CC^T is always a square matrix that can be inverted. In case of regular loudspeaker layouts the decoding matrix reduces to [6]

$$D = \frac{1}{L}C^T. \tag{23}$$

2.2.4 First-Order Ambisonic (B-Format)

As an example of the foregoing considerations the encoding and decoding process is illustrated for an first order ($M = 1$) 3D Ambisonic system, referred to as B-format developed by Gerzon. With $M = 1$ at least 4 loudspeakers ($L = 4$) are needed according to eq. (16) and (21). Following eq. (14) a single sound source is encoded by the first four Ambisonic channels known as

[4] The here derived formula refers to a basic decoder according to [5] that will be used in terms of this thesis. Optimisation decoding methods are mentioned in section 7.2.

$$\begin{aligned} B^1_{00} &= W = s_{src}\frac{1}{\sqrt{3}} \\ B^1_{11} &= X = s_{src}\cos\theta_{src}\sin\delta_{src} \\ B^{-1}_{11} &= Y = s_{src}\sin\theta_{src}\sin\delta_{src} \\ B^1_{10} &= Z = s_{src}\sin\delta_{src}, \end{aligned} \tag{24}$$

where the first channel reflects the pressure and the three following channels define its gradient, which is the acoustic velocity, in the indicated direction. Note the normalisation by the factor $\frac{1}{\sqrt{3}}$. The re-encoding matrix (17) with dimensions 4 x 4 contains the spherical harmonic functions

$$\begin{aligned} Y^1_{00} &= 1 \\ Y^1_{11} &= \sqrt{3}\cos\theta_{ls}\sin\delta_{ls} \\ Y^{-1}_{11} &= \sqrt{3}\sin\theta_{ls}\sin\delta_{ls} \\ Y^1_{10} &= \sqrt{3}\sin\delta_{ls}. \end{aligned} \tag{25}$$

Making use of decoding equation (20) the driving signal s_{ls_j} for the j-th loudspeaker is then given by

$$\begin{aligned} s_{ls_j} &= \frac{1}{L}\left(\frac{1}{\sqrt{3}}WY^1_{00} + XY^1_{11} + YY^{-1}_{11} + ZY^1_{10}\right) \\ &= \frac{1}{L}\left(\frac{1}{\sqrt{3}}W + X\cos\theta_{ls_j}\sin\delta_{ls_j} + Y\sin\theta_{ls_j}\sin\delta_{ls_j} + Z\sin\delta_{ls_j}\right), \end{aligned} \tag{26}$$

where also the same normalisation factor has been applied to the re-encoded signals.

2.3 Higher order Ambisonics - 2D

In this section the Ambisonic formalism is shown for a 2D (horizontal-only) reproduction system by highlighting the changes to a 3D system.

Since no elevation is considered any longer the coordinate system becomes a cylindrical one with azimuth angle θ and radius r, where $z = 0$. The Fourier-Bessel series becomes

$$p(r,\theta) = B_{00}^{+1(N2D)} J_0(kr) + \sum_{m=1,\sigma=\pm1}^{\infty} B_{mm}^{\sigma(N2D)} Y_{mm}^{\sigma(N2D)}(\theta,0) J_m(kr), \tag{27}$$

where $J_m(kr)$ are the cylindrical Bessel functions, B_{mm}^{σ} denote circular Ambisonic channels and the circular harmonic functions $Y_{mm}^{\sigma(N2D)}(\theta)$ [5]are given by

$$Y_{mm}^{\sigma(N2D)}(\theta,0) = \sqrt{2} \begin{cases} \cos m\theta \text{ if } \sigma = +1 \\ \sin m\theta \text{ if } \sigma = -1 \text{ (ignored if m=0).} \end{cases} \tag{28}$$

The circular harmonics can be linked to the spherical harmonics $Y_{mm}^{\sigma(N3D)}(\theta,\delta)$ from eq. (9) by introducing a weighting factor, resulting in

$$Y_{mm}^{\sigma(N2D)}(\theta,\delta) = \sqrt{\frac{2^{2m} m!^2}{(2m+1)!}} Y_{mm}^{\sigma(N3D)}(\theta,\delta). \tag{29}$$

The circular harmonic functions are represented by the eigenfunctions for angle θ and therefore form an orthonormal basis with

$$\left\langle Y_{mm}^{\sigma} | Y_{m'm'}^{\sigma'} \right\rangle = \delta_{mm'} \delta_{\sigma\sigma'}, \tag{30}$$

where the circular scalar product is given by

$$\langle F|G \rangle_{2\pi} = \frac{1}{2\pi} \int_0^{2\pi} F(\theta) G(\theta) d\theta. \tag{31}$$

By truncating the Fourier-Bessel series in eq. (27) to the Ambisonic order M the total number of horizontal Ambisonic components is determined by

$$K_{2D} = 2M + 1, \tag{32}$$

indicating that for the same order M less loudspeakers (for $L = K$) are necessary compared to a 3D system, since the number of loudspeakers is linear proportional to the order in contrast to a quadratic relation in the 3D case (eq. (16)).

The circular harmonic functions are illustrated in Figure 8 for components $m = 0$ to $m = 3$. Comparing them to the spherical ones in Figure 7 it is prominent that $Y_{mm}^{\sigma(2D)}(\theta)$ components equal (when applying the scaling-factor) components $Y_{mn}^{\sigma(3D)}(\theta,\delta)$ with n=m

[5] For reasons of consistency with the spherical harmonic functions the full normalisation 2D (N2D) convention is used here.

seen for $\delta = 0°$.

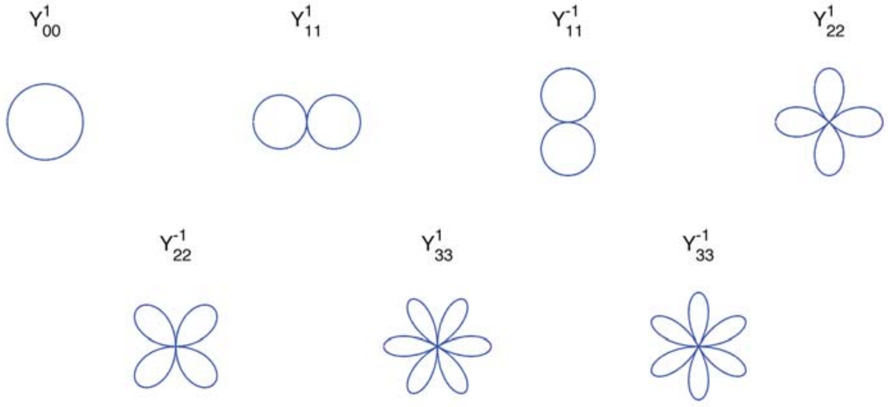


Figure 8: 2D circular harmonics for $m = 0$ to $m = 3$.

3 Objective analysis tools

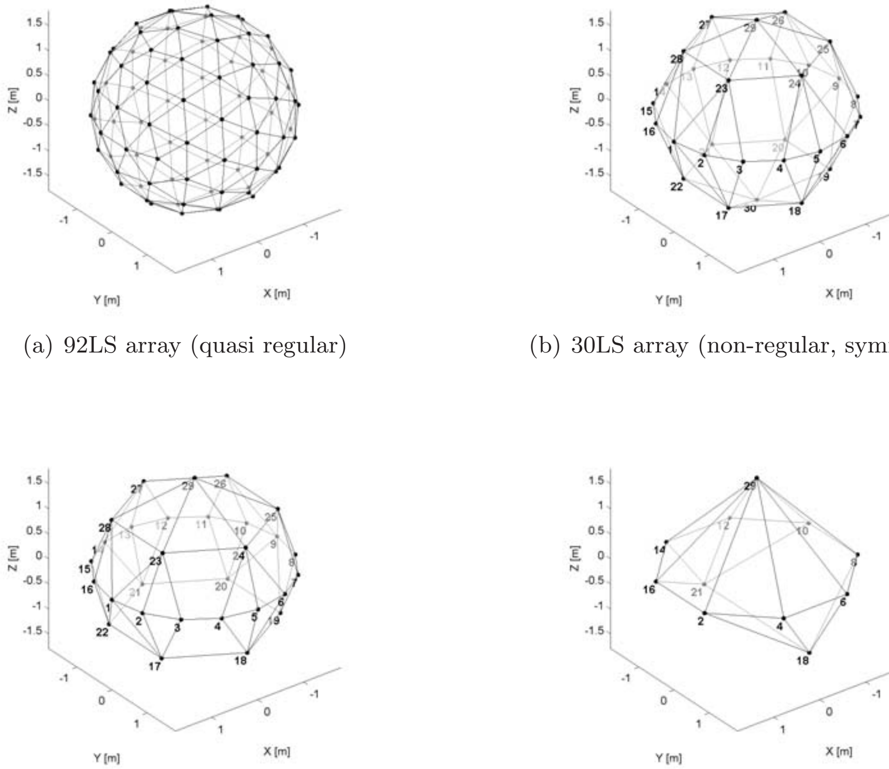
In this chapter common analysis tools that are used for the objective evaluation of 3D and 2D Ambisonic reproduction systems are introduced according to the underlying theory presented in the previous chapter. The chosen setups for the simulation studies are a quasi regular 92 and a non-regular 30 loudspeaker system (section 3.1). While the former one is used to illustrate conventional ideal 3D Ambisonic reproduction, the latter one highlights the effects of having a non-regular design as commonly used in practice. Horizontal-only Ambisonic reproduction is presented in terms of a regular 16 loudspeaker rig, which is part of the non-regular system. The properties of the different configurations are graphically illustrated in terms of directivity patterns (section 3.2), sound field simulations (section 3.3) and measures of localisation properties (section 3.4). Since each setup leads to a discretisation of the Ambisonic principle (according to the re-encoding principle in eq. (18)), the quality of this process is evaluated by investigating orthonormality properties (section 3.5).

3.1 Loudspeaker arrays

For the following simulation studies two 3D loudspeaker arrays are used. The first one is to be considered as an ideal loudspeaker array consisting of 92 loudspeakers in a quasi regular layout (see Figure 9 (a)), i.e. an approximated equal distribution of loudspeakers (spatial sampling) on the surface of a sphere.[6]Its maximum Ambisonic order is $M = 8$ (acc. to eq. (16)) and it will be denoted as 92LS array in the following.

The second layout consists of 30 loudspeakers (Figure 9 (b)) that are irregularly, but symmetricly distributed and is an idealised version of the Spacelab system, implemented at the facilities of DTU. It contains 3 loudspeaker rings with equidistant azimuth spacing between the loudspeakers, a horizontal one ($\delta = 0°$) with 16 loudspeakers and a spacing of 22.5° and 2 elevated rings with each 6 loudspeakers, resulting into a spacing of 60° between individual loudspeakers on each ring, at $\delta = 45°$ and $\delta = -45°$. In addition there is a single loudspeaker on top and at the bottom ($\delta = 90°$ and $\delta = -90°$). The system is limited to an maximum order of $M = 4$ in a periphonic (3D) reproduction and $M = 7$ in a horizontal-only (2D) reproduction (see chapter 2.3) when just making use of the 16 loudspeakers in the horizontal plane (denoted as 16LS array).

[6] This is not a trivial problem. For creating such an array the function geosphere.m from the 3LD library provided by [10] is used.



(a) 92LS array (quasi regular)

(b) 30LS array (non-regular, symmetric)

(c) 29LS array/Spacelab (non-regular)

(d) 11LS array (non-regular)

Figure 9: Illustration of the loudspeaker arrays. Setup (a) and (b) are used for simulation studies, (c) and (d) for listening tests presented in chapter 6.

Already at this point, two further systems that will be used for the subjective evaluations in chapter 6 are introduced. The Spacelab (see Figure 9 (c)) which will be denoted as 29LS array and a reduced version of that setup (see Figure 9 (d)), which has 11 loudspeakers in total (11LS array). The only differences of the 29LS array to the 30LS array are the missing loudspeaker at the bottom and shifted elevation angles of the two loudspeaker rings, i.e. $\delta = 36,5°$ and $\delta = -34°$. The before mentioned orders are the same.

The 11LS array contains 8 equidistant loudspeakers in the horizontal plane (every second from the 16LS array starting with loudspeaker no.2 at $\theta = 22.5°$) and three additional loudspeakers (no. 18, 21 and 29) for elevated sound sources. This array is asymmetric and can be more understood as a surround array with 3 additional speakers in order to represent elevation. It thereby represents a practical solution to a low-order system realisable in the Spacelab. The maximum orders for this system are $M = 3$ in a horizontal-only and $M = 2$ in a periphonic reproduction.

3.2 Directivity plots

According to eq. (20) Ambisonic decoding results in all loudspeakers playing the same source signal but weighted with source and loudspeaker direction dependent gains. Plotting these gains for a single encoded source with $s_{src} = 1$ originating from direction $(\theta_{src} = 0°, \delta_{src} = 0°)$ in respect to the corresponding loudspeaker positions $(\theta_{ls}, \delta_{ls})$ a three-dimensional directivity pattern is created. The patterns are shown in Figure 10 up to an order $M = 5$ and indicate that loudspeaker gains take positive as well as negative values. These figures highlight the fact that the energy emitted by the loudspeakers is more and more focussed towards the direction of the source the higher the Ambisonic order M. Accordingly to the source position, the directivity pattern changes direction by maintaining its shape.

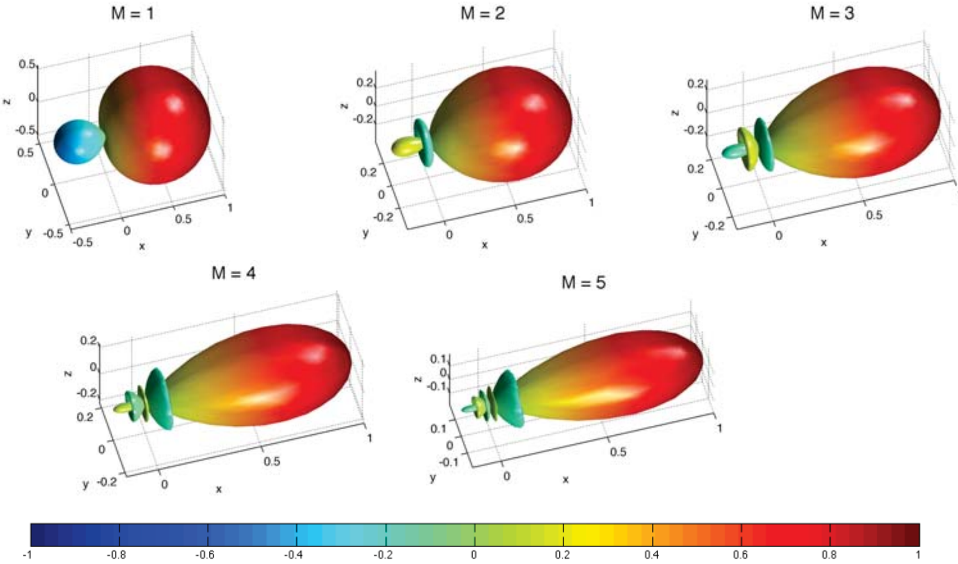


Figure 10: Directivity plots for $M = 1$ to $M = 5$.

The directivity plots for horizontal-only reproduction are shown in Figure 11. In order to compare the directivity plots of 2D and 3D, a horizontal representation of the 3D directivity plots from Figure 10 is also shown. The horizontal-only plots are generated by making use of the equivalent panning functions $G(\gamma)$ for 2D and 3D, which are derived in [6] by making use of decoding equation (20), and finally lead to the loudspeaker driving signals with $s_j = s_{src}.G(\gamma)$. The panning functions are

$$
\begin{aligned}
G(\gamma) &= \frac{1}{N}\left(1 + 2\sum_{m=1}^{M}\cos m\gamma\right) \text{ for 2D,}\\
G(\gamma) &= \frac{1}{N}\sum_{m=0}^{M}(2m+1)P_m(\cos\gamma) \text{ for 3D,}
\end{aligned}
\tag{33}
$$

where $\gamma = \theta_{src} - \theta_j$ denotes the angle between the source direction θ_{src} and the loudspeaker position θ_j in the $\delta = 0$ plane and equals $-\theta_j$ in case of a frontal source at $\theta_{src} = 0°$, which is the case for the following considerations.

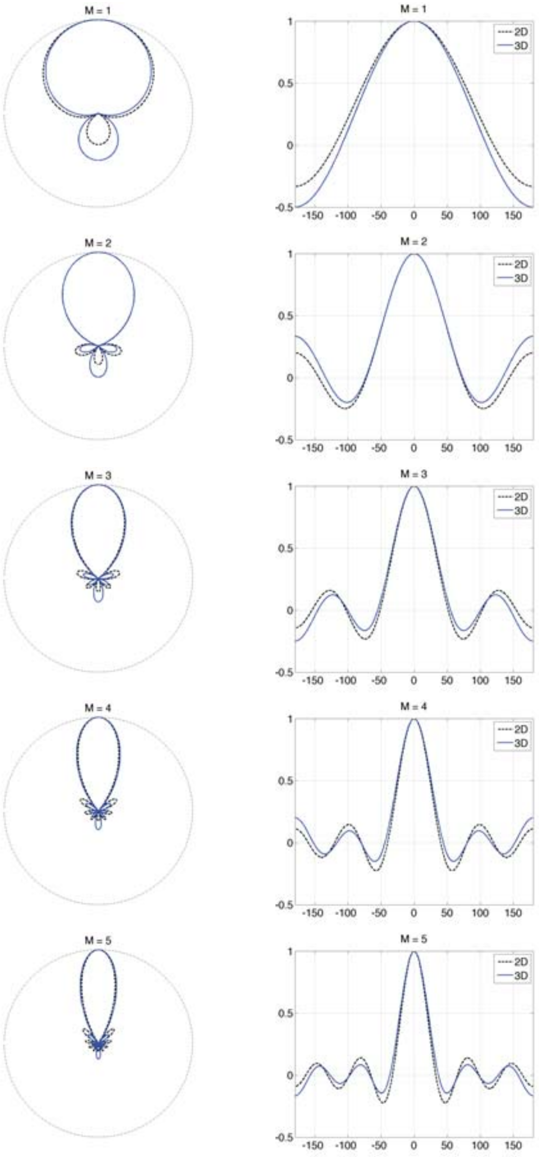


Figure 11: Equivalent panning laws for 2D and 3D in polar and linear representation for $M = 1$ to $M = 5$ (normalised to their maximum value).

Note that $P_m(\cos\gamma)$ denote the unassociated Legendre functions. In Figure 11 the panning laws are shown in a polar and in a linear representation. When comparing the 2D with the 3D directivity pattern the following remarks can be made: The main lobes are of similar shape, but in the 2D representation the energy is slightly more focussed into the main lobe, i.e. into a slightly narrower angular range, for orders $M > 1$. The side lobes appear in different strength for both cases. The lobe to the back is always stronger in the 3D representation, while this is the case in the 2D representation for the other side lobes. Therefore the 2D representation tends more to focus energy towards the source direction than in the 3D representation for the same order M. From the linear graphs it is apparent that the zero crossings in both representations are not identical.

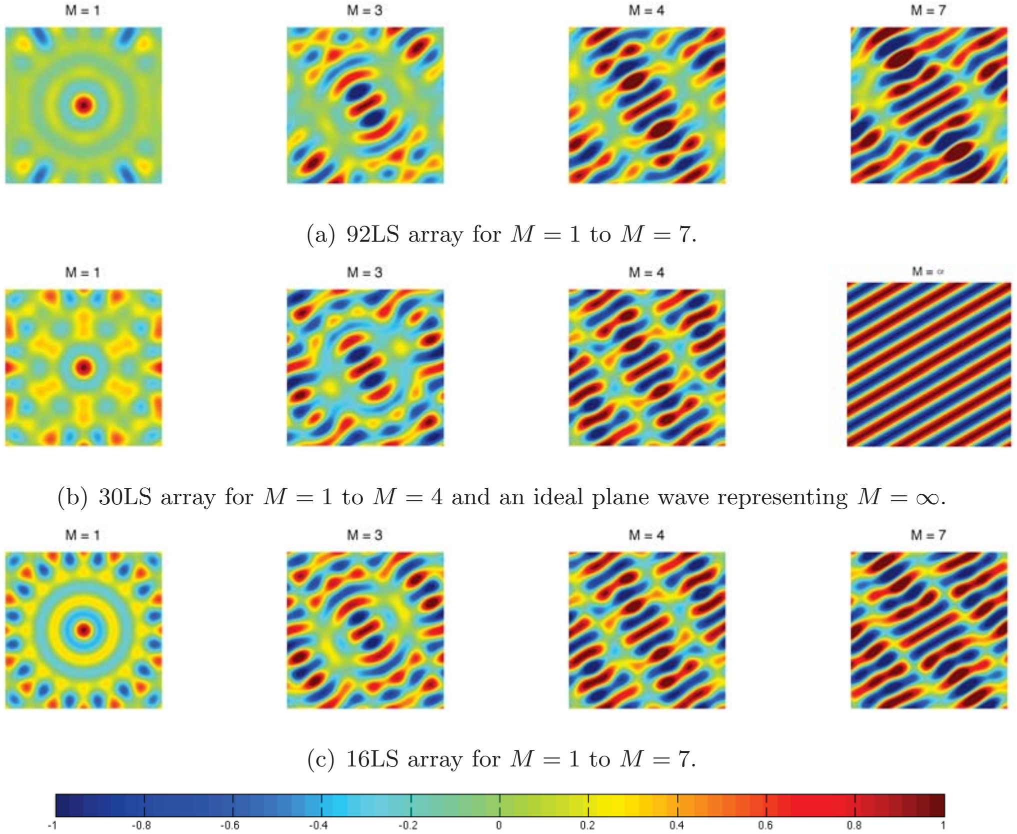
3.3 Monochromatic soundfield plots

One of the limitations of Ambisonic is the restriction of an exact sound field reproduction to a limited area. This sweet spot is basically determined by the following relation

$$M = kr, \tag{34}$$

which follows a rule of thump saying that the reproduction error is below 4% by fulfilling this relation [7]. When increasing either the wavenumber k or the distance r from the center, the error increases as well. The reproduction is also downgraded when decreasing the Ambisonic order M.

In Figure 12 monochromatic soundfields in the xy plane for the two different 3D loudspeaker arrays, 92LS and 30LS, and for the 2D surround system (16LS array) for a single encoded source with amplitude $s_{src} = 1$, frequency $f = 1000$ Hz and direction $(\theta_{src} = 30°, \delta_{src} = 0°)$ are shown for different truncation orders M. The plots are generated by superimposing plane waves emitted from the loudspeaker positions with amplitudes scaled by their corresponding calculated gains (following eq. (20)). According to eq. 8 the aim of Ambisonics is to reconstruct an ideal plane wave (corresponding to $M = \infty$) shown in the right plot of case (b). By increasing the order M the ideal case is approached. Note that for the same order M the reproduction by the 3D and 2D systems are similar. While for a first order case ($M = 1$) correct reproduction is limited to a very small sweet spot, its radius increases with the order. In the case of the 92LS and 16LS array for $M = 7$ the radius of the sweet spot is $r \approx 44$ cm (by making use of eq.(34)) and for the 30LS array the radius is limited by its maximum order of $M = 4$ resulting into a smaller sweet spot of $r \approx 22$ cm. Note that for orders $M \geqslant 3$ the plane field is also reproduced at areas outside the sweet spot.



(a) 92LS array for $M = 1$ to $M = 7$.

(b) 30LS array for $M = 1$ to $M = 4$ and an ideal plane wave representing $M = \infty$.

(c) 16LS array for $M = 1$ to $M = 7$.

Figure 12: Monochromatic soundfield plots for 3D and 2D reproduction systems.

3.4 Velocity and energy vector

In Gerzon's "General Metatheory of Auditory Localisation" from 1992 [8] a mathematical theory about human auditory perception is developed. It considers the human auditory system as "black box" responding to an incident sound field. Its purpose is to support the systematic development of complete surround systems and the author's aim is to keep this process simple and mathematically tractable. One of these simplifications is the plane wave assumption. Two models, the velocity and the energy model, are formulated for a first Ambisonic order system (B-format) referring to as models of first and second degree, respectively. The models provide two vector definitions that are useful in order to predict the localisation of sound in a reproduced sound field. Their prediction can be interpreted as physical measures of localisation, but cannot exactly be translated to human perception (compare to intensity as physical and loudness as perceptional quantity). According to the models, the velocity vector $\vec{V}$ is defined as

$$\vec{V} = r_V \vec{u}_V = \operatorname{Re} \frac{\sum\limits_{j=1}^{L} g_j \vec{u}_j}{W_V} = \operatorname{Re} \frac{\sum\limits_{j=1}^{L} g_j \vec{u}_j}{\sum\limits_{j=1}^{L} g_j} \tag{35}$$

and the energy vector $\vec{E}$ as

$$\vec{E} = r_E \vec{u}_E = \frac{\sum\limits_{j=1}^{L} |g_j|^2 \vec{u}_j}{W_E} = \frac{\sum\limits_{j=1}^{L} |g_j|^2 \vec{u}_j}{\sum\limits_{j=1}^{L} |g_j|^2}, \tag{36}$$

where $\vec{u_j}$ is a unity vector that represents the direction of the j-th loudspeaker in cartesian coordinates and g_j is the corresponding gain, that in general is complex. Assuming plane waves in Ambisonics results into real gains derived by eq. (20), so that $\vec{g} = \vec{s_{ls}}$. While the first vector is supposed to be valid for frequencies < 700 Hz the second one applies for higher frequencies in the region from 500 to 5000 Hz. Both measures are normalised, in case of the velocity vector by the sum of all loudspeaker gains (W_V) and in case of the energy vector by the total energy (W_E) reflected by the sum of the squared loudspeaker gains, in order to be independent of the overall signal level. The two vectors contain the calculated direction in u_V, u_E that is weighted by the corresponding magnitude value r_V, r_E. For an ideally reproduced single sound source the condition $u_V = u_{src}$ with $r_V = 1$ or $u_E = u_{src}$ with $r_E \rightarrow 1$ should be fulfilled in the sweet spot. As an example, these conditions are strictly fulfilled in case of a single loudspeaker representation. It is possible that r_V takes values > 1 in case of a very small sum of loudspeaker gains (see the example of the 29LS array in the Appendix), but r_E is strictly less than unity, as proofed in [8] p.17. In case that the sum of loudspeaker gains becomes very small, which is possible due to negative and positive gains, destructive interference is indicated. This cannot happen when calculating the total energy since the individual gains are squared.

In [5] p.179 it is stated that for a regular loudspeaker setup and assuming $L > K$ the total energy W_E is independent of the source direction and takes the constant values

$$W_E = \frac{K}{L} = \begin{cases} \frac{2M+1}{L} \text{ (pure 2D coding)} \\ \frac{(M+1)^2}{L} \text{ (pure 3D coding)} \end{cases} \tag{37}$$

Accordingly, r_E takes a constant value

$$r_E = \begin{cases} \frac{2M}{2M+1} \text{ (pure 2D coding)} \\ \frac{M}{M+1} \text{ (pure 3D coding)} \end{cases} \tag{38}$$

As an example, r_E equals 0.5 for a first order 3D reproduction system and takes a value of 0.67 for a horizontal-only reproduction system of the same order.

In terms of the predicted direction u_V or u_E an error from the real sound source position u_{src} can be formulated and will be used in the following in the form of spherical coordinates (by using transformation eq. (3)) as

$$\begin{aligned} \delta_{V_{err}} &= \delta_{src} - \delta_V & \delta_{E_{err}} &= \delta_{src} - \delta_E \\ \theta_{V_{err}} &= \theta_{src} - \theta_V & \theta_{E_{err}} &= \theta_{src} - \theta_E . \end{aligned} \tag{39}$$

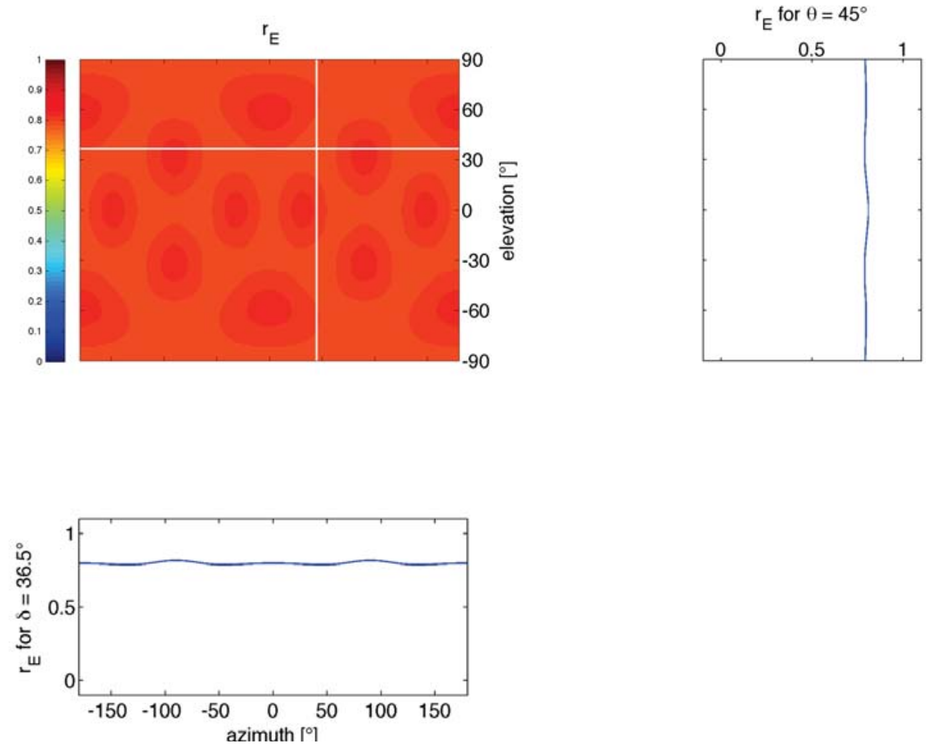


Figure 13: Energy vector magnitude r_E for $M = 4$ for the 92 LS array. A constant magnitude of ~ 0.8 independent of the source direction as expected from eq. (38) is indicated in the map plot as well as for a selected elevation and azimuth angle as displayed below and beside, respectively.

In the following the two systems, 92LS and 30LS array, are analysed in terms of the velocity and energy vector for the same Ambisonic order of $M = 4$ (which is the maximum order for the latter mentioned array) in order to make them comparable. Starting with the ideal system, a homogeneous reproduced sound field in terms of directional perception is expected. Indeed, the velocity vector magnitude r_V is unity and the two errors $\delta_{V_{err}}$ and $\theta_{V_{err}}$ are zero for all directions (not illustrated), indicating an ideal reproduction for low frequencies.

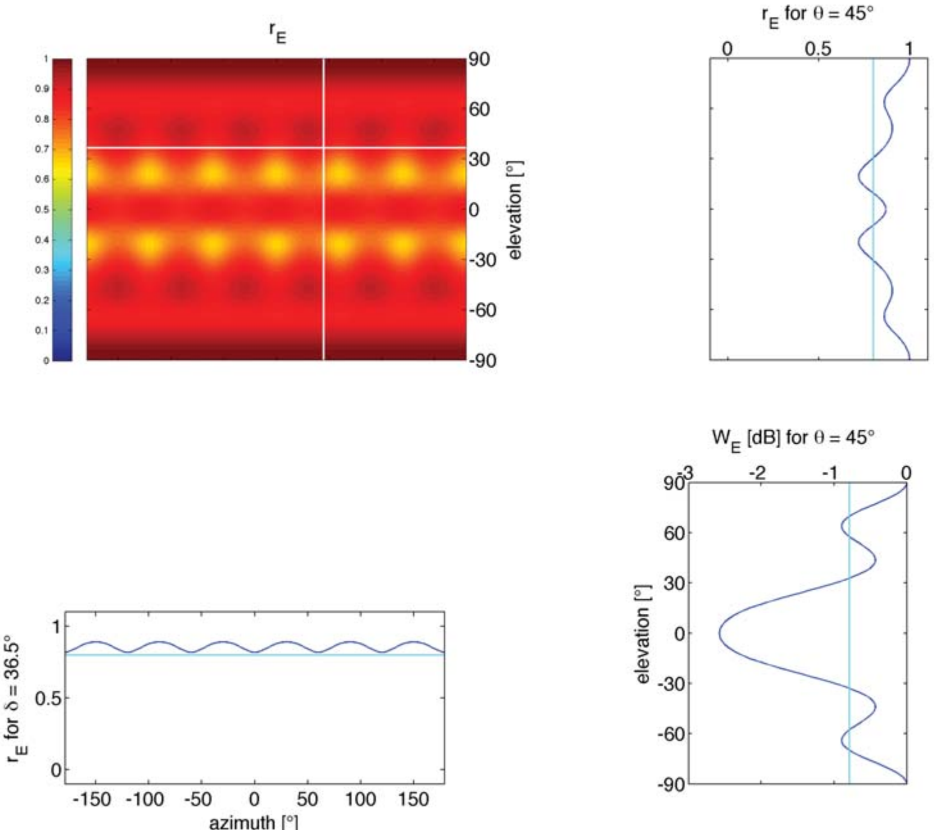


Figure 14: Energy vector magnitude for $M = 4$ for the 30 LS array and total energy of the system. The magnitude r_E deviates from a constant value (acc. to eq. 38) as in case of a regular setup of this order (light blue line) and is maximised around loudspeaker positions. The total energy W_E regarding changes in elevation is not constant either compared to its estimate in eq. (37) and has a minimum for non-elevated sources.

The results for the energy vector magnitude are shown in Figure 13. The figure is created by plotting the magnitude value over the surface of the sphere in 2D, i.e. over all possible source directions with θ on the abscissa and δ on the ordinate. The plot reflects a homogeneous magnitude value of $r_E \approx 0.8$ as expected from eq. (38). The graphs below and to the right indicate the variation of the magnitude for a specific elevation and azimuth angle, respectively which are highlighted by the two lines in the map plot. The two errors $\delta_{E_{err}}$ and $\theta_{E_{err}}$ are both naught and therefore not illustrated. A level change of 0.6dB in the distribution of energy is present for this system for all angles. According to [25] this value is below the JND in sound pressure level of 1 dB and would therefore not be detectable by a human listener.

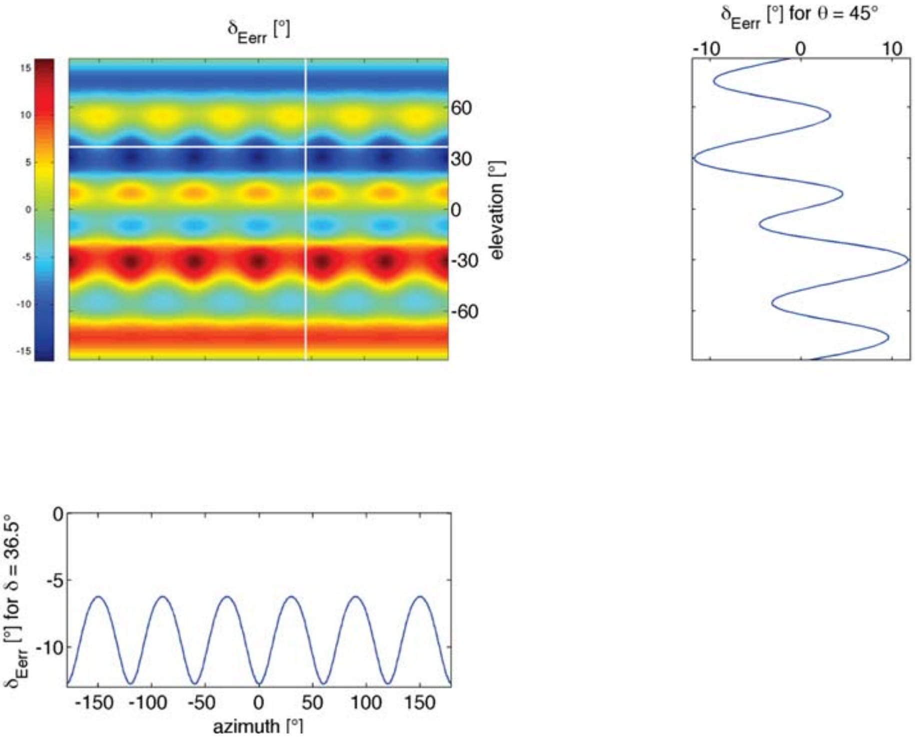


Figure 15: Elevation error of the energy vector for $M = 4$ for the 30 LS array. The error is highly dependent on the loudspeaker positions. Significant variations of the error are especially prominent in dependence of the elevation angle. A systematical change of the error value is present in dependence of the azimuth.

The 30LS array is as well ideal when analysed in terms of the velocity vector: The magnitude r_V is unity, independent of the source position, and the reproduction is errorless. Limitations of the system become prominent at higher frequencies as investigated by the energy vector shown in Figures 14 to 16. The magnitude (Figure 14) is maximised around loudspeaker positions with $r_E = 0.9$ and takes minimum values in between. Due to loudspeakers at top and bottom of the array a value of unity is achieved. The loudspeaker positions become also visible when looking at the map plot. Changes with elevation reveal more drastic variations in the magnitude than changes with azimuth though. The irregular loudspeaker setup causes an unequal distribution of energy and takes a minimum value of -2.6dB in the horizontal plane. Due to the system's irregularity, errors are introduced into the reproduction. This is shown in Figures 15 and 16 where the elevation and azimuth error are plotted, respectively. Coinciding with the before made statements about the magnitude errors in elevation angle become more prominent than errors in azimuthal changes of source position. The error $\delta_{E_{err}}$ ranges from $-10°$ to $+10°$ by changing elevation and varies systematically from $-6°$ to $-12°$ over changes in azimuth. Regarding $\theta_{E_{err}}$, systematical deviations of $\pm6°$ occur with changes over azimuth and peaks of $+10°$ are observable for elevation angles of $\pm28°$.

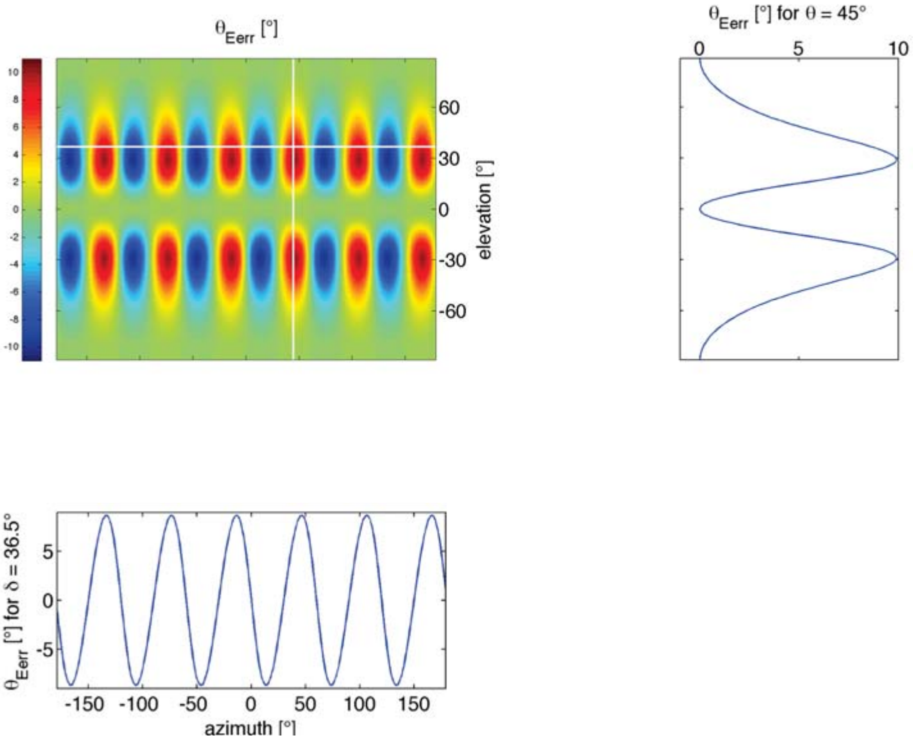


Figure 16: Azimuth error of the energy vector for $M = 4$ for the 30 LS array. Two maxima are prominent at elevation angles of $\pm28°$ and systematical variations of the error are obtained along the azimuthal angle. The error is influenced by the loudspeaker positions.

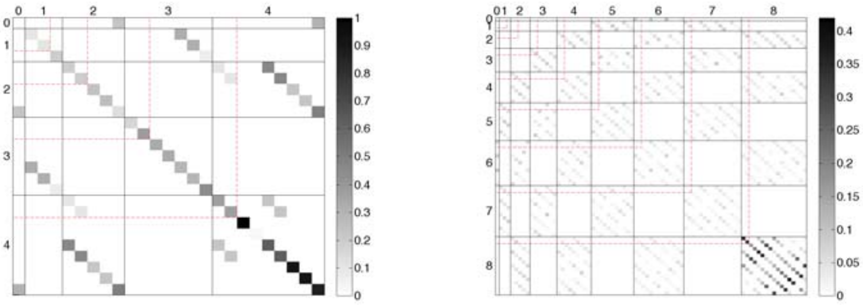
The horizontal-only system works ideally in terms of the velocity and energy vector considerations, which is due to its regular setup. Graphical illustrations are therefore omitted here. The magnitude of the velocity vector r_V is unity for all azimuth angles and no errors in localisation occur. The magnitude of the energy vector is constant, as it is expected from eq. (38), with $r_E = 0.9$ and likewise u_E equals u_{src} for all angles. Also the energy is equally distributed with $W_E = -2.5$dB.

3.5 Orthonormality Matrix

Practical limitations of a certain loudspeaker array occur due to the sampling of the spherical harmonic functions performed in the re-encoding. This causes reproduction errors which are to investigate. The discretisation can lead to violations of the definition of the orthonormal basis in eq. (40), meaning that the spherical scalar product $\left\langle Y_{mn}^{\sigma} | Y_{m'n'}^{\sigma'} \right\rangle$ is deviating from zero. This deviation can be expressed as an error and is illustrated by the orthonormality matrix defined as [14]

$$U = I_k - \frac{1}{L} C C^T, \tag{40}$$

where I_k is the K x K identity matrix, L the number of loudspeakers and C the re-encoding matrix from eq. (18). The matrix is symmetric and its graphical illustration indicates the orthonormality error between two sampled spherical harmonics by small black squares evaluated on a scale from 0 to 1. The order M of the spherical harmonic functions is shown on both axes, highlighting horizontal components by red dashed lines.

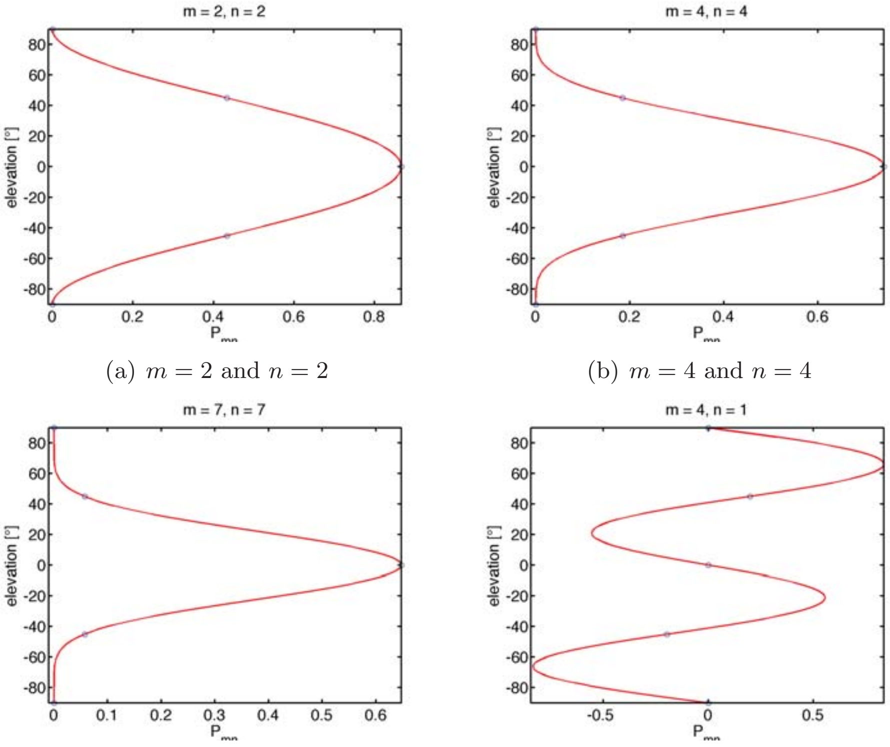


(a) 30 LS array

(b) 92 LS array

Figure 17: Orthonormality matrix for the 92LS and 30LS array. Black squares indicate orthonormality errors between spherical harmonics.

In Figure 17 the orthonormality matrix for the loudspeaker arrays is shown for a pure 3D reproduction. For the 30LS array errors occur mainly on the diagonal (which is between harmonic functions of same order), but also for harmonic functions of different orders. It is conspicuous that the highest errors of 100% occur on the diagonal for the maximum order of $M = 4$ which indicates clearly the system's limitations. As a conclusion, the actual recommended maximum order is $M = 3$ and is therefore lower than the one derived in eq. (16). By applying the new limit to the system the maximal error is reduced to 45% - results into a more homogeneous map plot in the simulation studies of the energy vector's magnitude r_E (Figure 14) - with the trade-off of having an overall slightly reduced magnitude. The azimuth error $\theta_{E_{err}}$ is totally removed (Figure 16), which is also the case for the elevation error $\delta_{E_{err}}$ for changes with the azimuthal angle. Even though the number of loudspeakers is high for the 92LS array a maximum error of about 40% arises for the maximum order of $M = 8$. To ensure an even more homogeneous reproduction with such a system the order should not exceed $M = 7$.



(a) $m = 2$ and $n = 2$

(b) $m = 4$ and $n = 4$

(c) $m = 7$ and $n = 7$

(d) $m = 4$ and $n = 1$

Figure 18: Discretisation (circles) of continuous associated Legendre functions by the 30LS array. The graph is tilted by 90° for a better illustration of the elevation angle. Insufficient discretisation is present in cases (b) to (d).

It can be shown that there is no error for a horizontal-only reproduction with the regular 16LS array. From this it can be concluded that the error in a periphonic representation in case of a non-uniform sampling arises from the discretisation of the continuous associated Legendre functions that are representing the discrimination of elevation angles. Four Legendre functions and their discretisation by the 30LS array are illustrated in Figure 18 as an example. The discretised points represent the elevation positions of the loudspeakers. While Figure 18 (a) can be considered as sufficient sampling, Figure 18 (d) shows a case for insufficient discretisation that results into a high orthonormality error. The Legendre functions for $m = n$ get narrower with increasing order (compare (a) to (c)) which leads as well to insufficient sampling by the 30LS array in case (b). Case (c) will be referred to in the next chapter.

4 Implementation of a mixed-order Ambisonics playback system

The goals of a mixed-order playback system are here defined as:

- Improvement of localisation in the horizontal plane (compared to a pure 3D representation)
- Maintaining the properties of a periphonic system for elevated sources
- Smooth perceptional transition between representations of horizontal and elevated sources
- The flexibility of Ambisonic systems - such as portability which is the independence of the encoding and decoding stage - should be maintained.
- Coloration effects should be minimised.

In order to implement a mixed-order Ambisonic system, two general global strategies can be thought of in the author's point of view: In the first strategy loudspeaker gains s_{ls} are calculated separately for a horizontal-only and a periphonic reproduction. They are then weighted appropriately. This method presumes a loudspeaker ring for horizontal only reproduction in addition to or included into the entire reproduction system in order to be able to reproduce horizontal and elevated sources separately. The second and chosen strategy is simple and effective as will be shown in the following analysis. It is based on the spherical harmonic functions $Y_{mn}^{\sigma(3D)}$ for pure periphonic representations. Since as mentioned earlier horizontal components (those with m=n) are already included in the spherical harmonic functions, they can be used to represent (i.e. to encode and to decode) horizontal sources. In the implementation further horizontal components have simply to be added to existing 3D components. The definition of separate orders, the horizontal order M_{2D} and the periphonic order M_{3D} is hereby necessary. Considering Figure 7, a mixed-order system of order $M_{2D} = 3$ and $M_{3D} = 2$ for example is achieved by taking the first nine components (first 4 rows), representing pure 3D components of order 2, and adding the components Y_{33}^{1} and Y_{33}^{-1} as additional 2D components. The algorithm calculates sufficient 3D components and selects the right components according to the specified orders. In the given example, this is by omitting the last five components $Y_{32}^{1}, Y_{32}^{-1}, Y_{31}^{1}, Y_{31}^{-1}$ and Y_{30}^{1}. As a result, there are $(2m + 1)$ 3D components per order $1 \leqslant m \leqslant M_{3D}$ including 2 horizontal components as usual and in addition 2 horizontal components (with just using $n = m$) are added per order $M_{3D} < m \leqslant M_{2D}$.

The truncated Bessel-Fourier series from eq. (15) is modified accordingly to

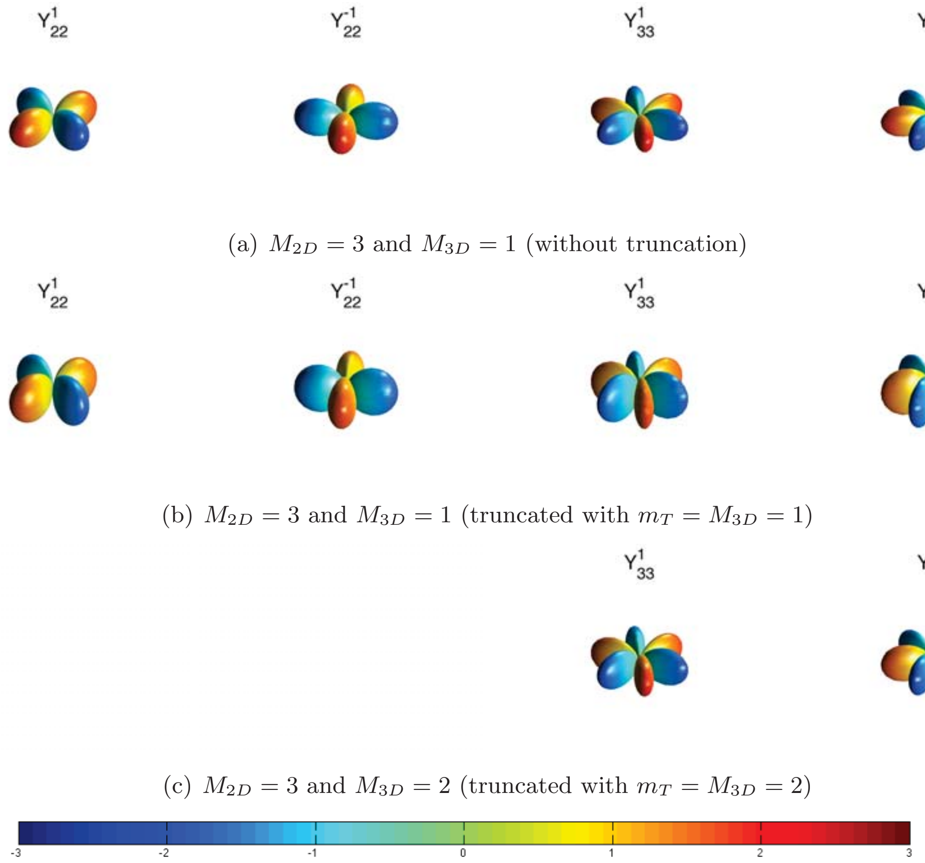
$$\begin{aligned} p(r, \theta, \delta) = &\sum_{m=0}^{M_{3D}} j^m j_m(kr) \sum_{0 \leqslant n \leqslant m, \sigma=\pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta, \delta) \\ &+ \sum_{m=M_{3D}}^{M_{2D}} j^m j_m(kr) \sum_{\sigma=\pm 1} B_{mm}^{\sigma} Y_{mm}^{\sigma}(\theta, \delta). \end{aligned} \tag{41}$$

A relation similar to eq. (29) is simply given by

$$Y_{mm}^{\sigma(2D)}(\theta, \delta) = Y_{mm}^{\sigma(3D)}(\theta, \delta). \tag{42}$$

Commenting on eq. (29), this relation sustains the portability and flexibility of Ambisonic systems since it allows playback of elevated sources on a horizontal-only playback system, which can be interpreted as projection of elevated sources into the horizontal plane as also mentioned by [20]. This feature of projection provides the mixed-order system with a 'inherent smoothing' in the transition from horizontal to elevated source image reproduction: The directional weighting and thereby the influence of horizontal components decreases with the elevation angle. The functions that describe the smoothing are the according Legendre functions P_{mm} (with a maximum at $\delta = 0$ and zero at the poles), which are included in the horizontal components, since they are derived from the original spherical harmonic functions themselves (eq. (42)).

Before implementing a mixed order system the system's limitations - for example investigated by the orthonormality matrix (see section 3.5) - should be known, so that the maximum order $M_{3D_{max}}$ of a certain loudspeaker system keeps the reproduction error below a certain limit that is that spherical harmonic components which lead to errors of 100% should definitely be excluded. The maximal order $M_{2D_{max}}$ depends either on the number of loudspeakers in the regular horizontal loudspeaker ring as in the case of the 30LS array or $M_{2D_{max}} = M_{3D_{max}}$ as in case of a regular loudspeaker array such as the 92LS array that does not have a regular horizontal loudspeaker ring.



(a) $M_{2D} = 3$ and $M_{3D} = 1$ (without truncation)

(b) $M_{2D} = 3$ and $M_{3D} = 1$ (truncated with $m_T = M_{3D} = 1$)

(c) $M_{2D} = 3$ and $M_{3D} = 2$ (truncated with $m_T = M_{3D} = 2$)

Figure 19: Truncated spherical harmonics in comparison with the non-truncated ones. Only the additional horizontal components (those with $m = n$) are displayed. The shape of the spherical harmonics (vertical expansion) changes accordingly to the applied truncated order of the Legendre functions.

Especially in the first mentioned case it is important not to exceed the limits of $M_{3D_{max}}$, but since the spherical harmonics are used to derive the horizontal components a contradiction arises: The system is meant to have $M_{2D} > M_{3D}$ to improve the horizontal reproduction, but this violates the system's limitations. Taking the 30LS array as an example, its maximal recommended order had been determined as $M_{3D} = 3$. So the system's limitations in terms of the orthonormality properties are violated, when this order is exceeded (see Figures 18 (a) to (c) and 26). To resolve this contradiction the maximum order of the Legendre functions used in the calculations of the spherical harmonic functions in eq. (9) must be truncated to $m \leqslant m_T = M_{3D_{max}}$ in order to stay inside the limits:

$$Y_{mn}^{\sigma}(\theta,\delta) = \begin{cases} Y_{mn}^{\sigma(3D)}(\theta,\delta) & \text{if } m \leqslant M_{3D} \\ \\ \sqrt{2m_T+1}N_{m_Tm_T}P_{m_Tm_T}(\sin\delta)\begin{cases}\cos m\theta \text{ if } \sigma=+1 \\ \\ \sin m\theta \text{ if } \sigma=-1\end{cases} & \text{if } M_{3D} < m = n \leqslant M_{2D}. \end{cases} \tag{43}$$

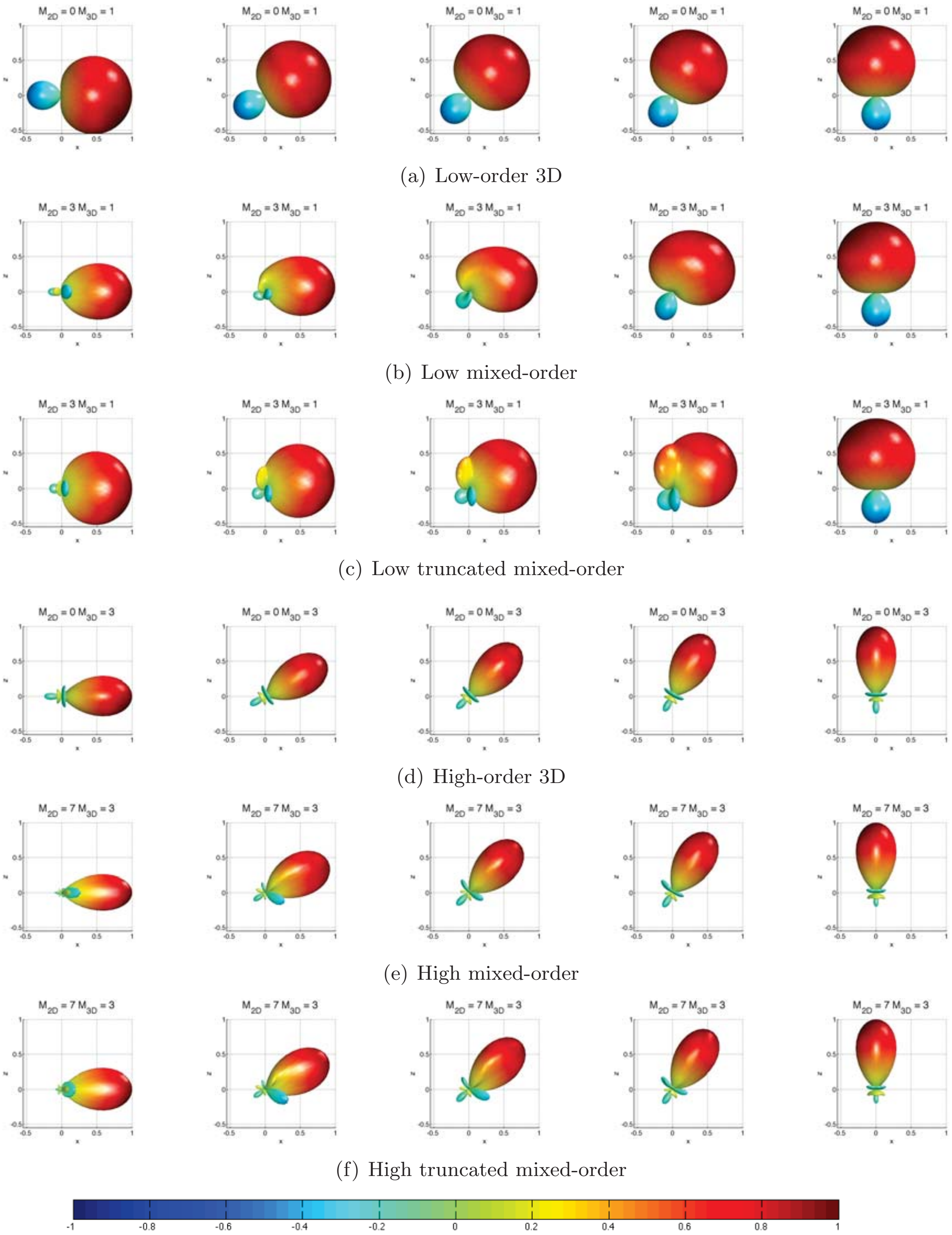
The effect of truncation on the spherical harmonic functions is illustrated in Figure 19. For example for a truncation of $m_T = M_{3D} = 1$ as shown in the Figure 19(c) the maintained sinusoidal shape in the spherical harmonic functions considering elevation is prominent due to that the Legendre function $P_{11}(\sin\delta) = \sin(\delta)$. In the following both, the truncated and the non-truncated mixed-order system will be investigated.

5 Objective evaluation

This chapter deals with the objective evaluation of a mixed-order Ambisonic implementation as described in the previous chapter. The two derived approaches - non-truncated and truncated mixed-order - are investigated and compared to a conventional 3D implementation in terms of the analysis tools that have been introduced in chapter 3. The influence on the directivity properties is outlined in section 5.1. As before, the two setups, 92LS and 30LS array, are used for simulation studies in section 5.2 and 5.3. Furthermore, the influence of a mixed-order implementation on the frequency spectrum is presented in section 5.4. Such investigations are just reasonable for existing loudspeaker setups and are therefore carried on for the Spacelab.

5.1 Directivity analysis

Regarding directivity, differences between a pure 3D and a mixed-order representation is outlined in the following, taking also the influence of truncation into account. In Figure 20 distributions of loudspeaker gains in the elevation plane (xz plane) for five different elevation angles (from left to right: $0°, 30°, 45°, 60°$ and $90°$) and for two different combinations of orders, representing a low-order (panels (a) to (c)) and a high-order system (panels (d) to (f)), are shown. With increasing elevation, the shape of a mixed-order plot approaches the one of a pure 3D representation due to the vanishing influence of the horizontal spherical harmonic components, so that source representations at the zenith (shown on the right side of the figure) result into equal directivity plots. It had been observed though that minor differences to the pure 3D coding can be present for certain combinations of orders. In case of a pure horizontal source (left side of the figure), it becomes obvious that the mainlobe's vertical dimensions are determined by the specified order M_{3D}. The directivity pattern in a pure 3D system is rotational symmetric, whereas in a mixed-order system energy is focussed into the horizontal plane which is seen at both, the mainlobe and the sidelobes. The mainlobe of the truncated mixed-order system takes a similar vertical expansion as the pure 3D system though, as the operation of truncation implies. While the directivity properties of a pure 3D system are maintained in the transition area between horizontal and elevated sound sources, the symmetric balloon plot gets deformed for the mixed-order systems indicating the focus of energy towards the horizontal plane. The deformation artifacts and asymmetries are prominent for the mainlobe as well as for the sidelobes, especially in case of the low-order system.



(a) Low-order 3D

(b) Low mixed-order

(c) Low truncated mixed-order

(d) High-order 3D

(e) High mixed-order

(f) High truncated mixed-order

Figure 20: Directivity plots for a low- and a high mixed-order system (non-truncated and truncated) in comparison with a pure 3D system (with the specified order M_{3D}) for different elevated source positions (from left to right: $\delta = 0°, 30°, 45°, 60°$ and $90°$) illustrated in the xz plane.

In case of the high-order system, mainly in-symmetries in the sidelobes are remarkable and are intensified when truncation is applied. This can be explained by the truncated order of the legendre functions which allows for a higher intensity of positive and negative elevated loudspeakers due to a reduced directional selectivity.

Considering changes in the horizontal plane (xy plane) as illustrated in Figure 21 the shape of horizontal-only reproduction systems is approached, so that the order M_{2D} is the determining factor as desired: The mainlobe is identical in shape and the sidelobes are identical in quantity to a pure 2D system, whereas their stronger intensity is first reduced when truncation is applied to the system.

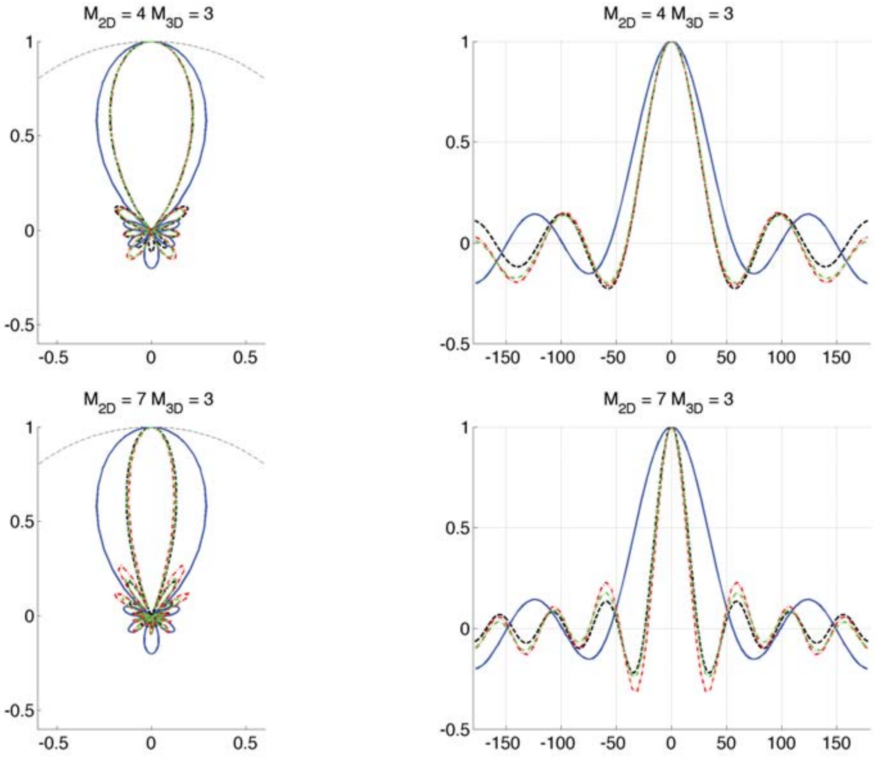


Figure 21: Directivity plots for the mixed-order coding techniques in comparison with pure 2D and 3D coding illustrated in the xy plane. 2D case: black; 3D case: blue solid line; mixed case: red dash-dotted line; truncated mixed case: green dash-dotted line.

5.2 Quasi-regular system (92 LS-array)

In the following the mixed-order algorithm is tested in simulation studies by using the 92LS and 30LS array as in the previous sections in order to stress differences between a regular and a non-regular setup. Even though the question of how meaningful an improvement in the horizontal plane in a system that is designed for a uniform 3D representation such as the 92LS array arises the analysis is carried on to highlight some interesting features.

First of all the functionality of a mixed-order algorithm can be well explained on such a regular array and at the same time its disadvantage for horizontal reproductions will be outlined.

The analysis is best illustrated in terms of the energy vector. In Figure 22 the dependence of the energy vector magnitude r_E on the order M is shown for two fixed source positions. All cases, that is a pure 2D system - using the 16LS array - the pure 3D case, mixed case and truncated mixed case by making use of the 92LS array are in comparison. The abscissa shows for the pure 2D and 3D coding the appropriate order M and for the mixed-order coding techniques the variation of order M_{2D}. The periphonic order M_{3D}, as indicated at the top of each graph and illustrated by the horizontal blue line, is fixed. Considering first a horizontal source it becomes obvious that r_E is improved with increasing order as mentioned earlier (eq. (38)). At the same time, the representation by the horizontal-only system is better than compared to the pure periphonic system as expected. Note that in both cases the improvement is getting less with increasing M, where in the horizontal case the function is even more compressed. (This can be explained by the declining maximum value of the legendre functions at $\delta = 0°$ referring to eq. (29).)

As desired the mixed-order case improves r_E compared to the pure 3D case (with specified order M_{3D}), is as well improving with increasing order, but is not as good as the pure 2D representation. This is due to the lacking regular horizontal loudspeaker array in that specific array. The 92LS array is just supplied with 12 loudspeakers in the horizontal plane that are non-regularly distributed. Furthermore, no difference occurs between the non-truncated and truncated mixed case.

When elevating the sound source to $\delta = 45°$ it is remarkable that the horizontal reproduction system saturates towards a certain r_E for $M \geqslant 2$. This highlights the effect of projection described earlier in section 4 (Regarding eq. (29) the legendre functions with m=n take small values for higher elevation angles and therefore have a vanishing influence). For a low order $M_{3D} = 1$ the mixed case supplies the system with a smaller improvement compared to a horizontal source, but further improvement depending on order M is gained from applying truncation. For a higher order M_{3D} no further improvement can be achieved, so that the system is as good as a pure periphonic system independent of increasing order M.

In Figure 23 the effect on the energy vector's magnitude r_E, the energy distribution W_E and the elevation error $\delta_{E_{err}}$ are highlighted for two specific combinations of orders

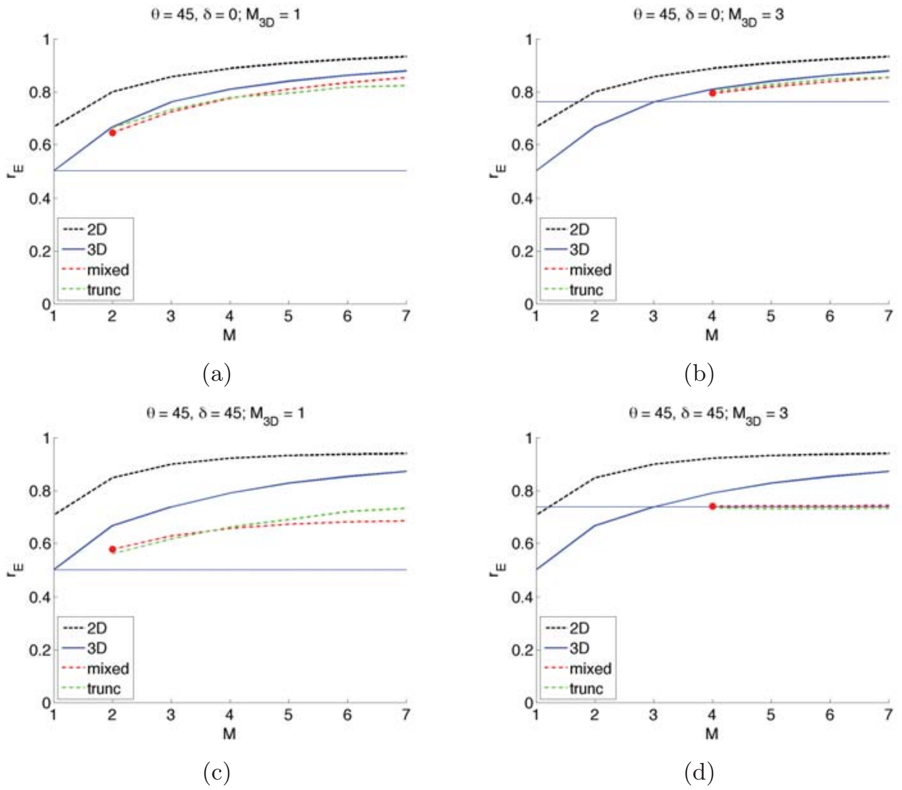
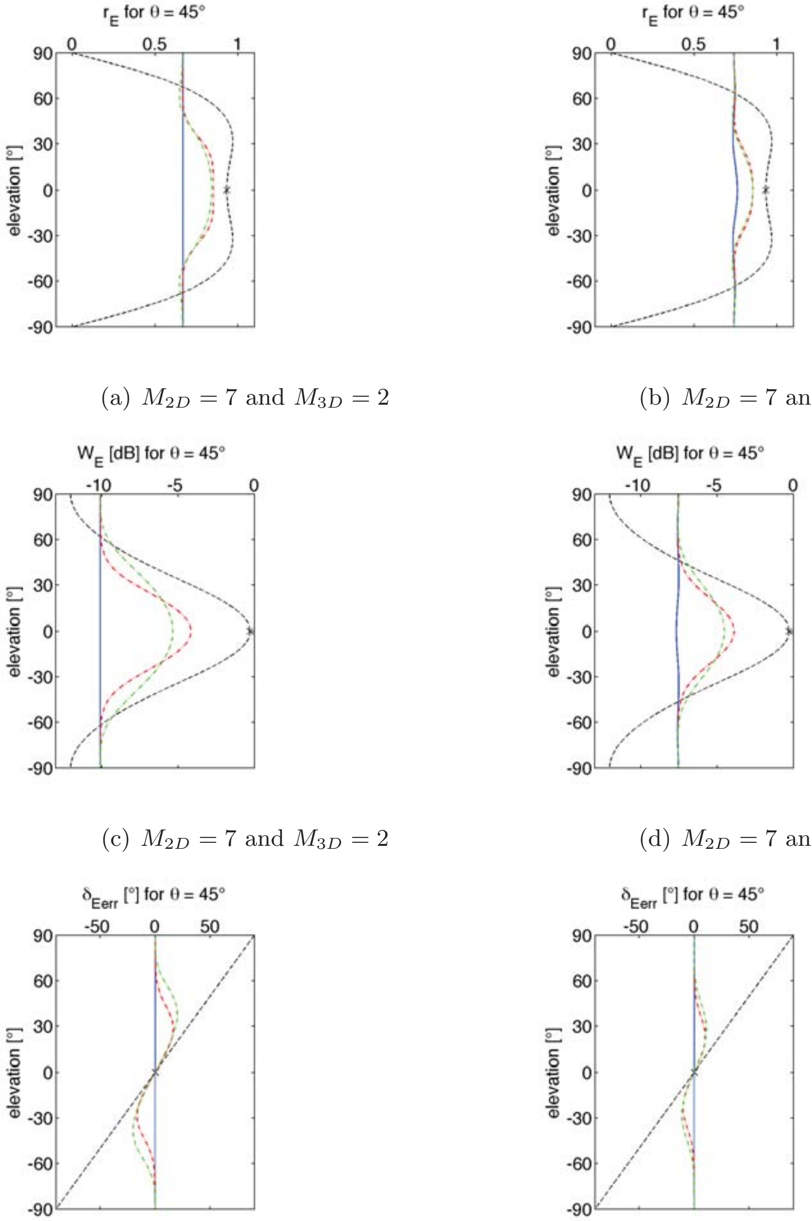


Figure 22: Comparison of the different coding techniques in terms of the energy vector magnitude r_E in dependence of order M for the 92LS array for two specified source positions. Note that the 2D case is plotted for the 16LS array (elevated sources can be reproduced by this array by making use of eq. (29)). The order M refers to the order of a conventional 2D representation (16LS array) and 3D representation (92LS array). The mixed-order implementation applied to the 92LS array has the specified order M_{3D} and $M = M_{2D}$ is the varying parameter. The mixed-order case is to compare with the straight line as this indicates the improvement to a pure 3D coding of this order.

$M_{2D} = 7$ in combination with $M_{3D} = 2$ or $M_{3D} = 3$ by varying the elevation angle of the sound source. Again the improvement by the mixed-order compared to a pure periphonic system is obvious, but is less for a smaller difference in orders M_{2D} and M_{3D}. Additionally, the truncation has the effect of smoothing the transition area between horizontal and elevated sources. The improvement of r_E in the horizontal plane goes in hand with a focussing of energy towards this plane, which is the trade-off condition of a mixed-order system. However, applying truncation smoothes the energy distribution likewise. Regarding the elevation error, a mixed-order system increases this error significantly compared to an errorless pure 3D reproduction, but is resolved by lowering the difference between the orders M_{2D} and M_{3D}. Error $\theta_{E_{err}}$ with changes over δ is smaller than 1° for all combinations. For comparison, the behaviour of a pure 2D system is illustrated as well in all plots.

Regarding the velocity vector, the sum of loudspeaker gains W_V and the velocity vector magnitude r_V are always unity in all cases disregarding which combination of orders is chosen. At the same time reproduction errors $\theta_{V_{err}}$ and $\delta_{V_{err}}$ are zero in all cases.



(a) $M_{2D} = 7$ and $M_{3D} = 2$

(b) $M_{2D} = 7$ and $M_{3D} = 3$

(c) $M_{2D} = 7$ and $M_{3D} = 2$

(d) $M_{2D} = 7$ and $M_{3D} = 3$

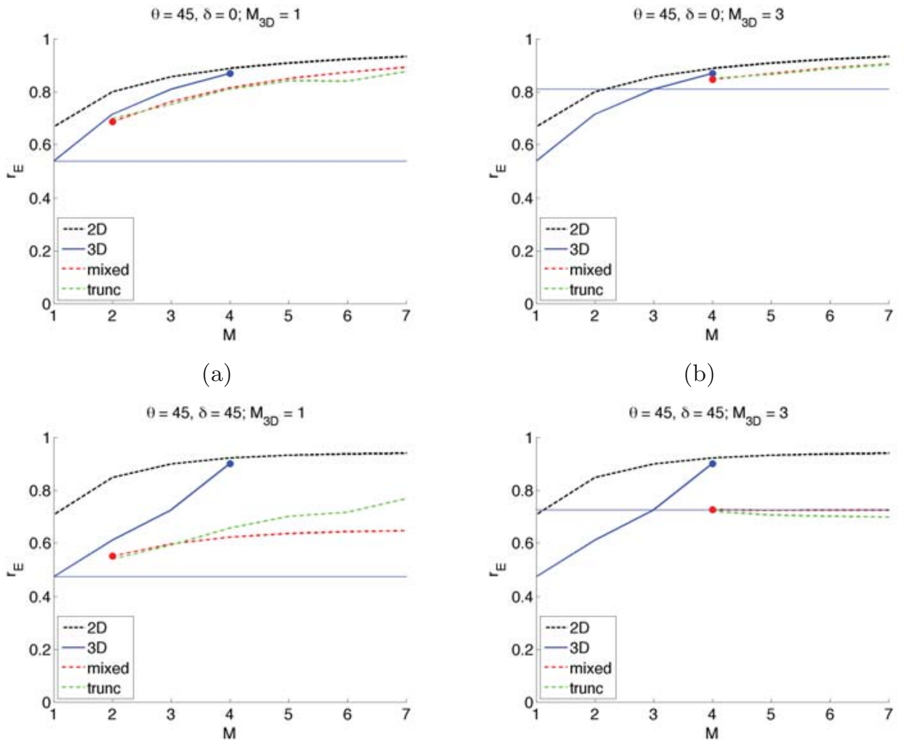
(e) $M_{2D} = 7$ and $M_{3D} = 2$

(f) $M_{2D} = 7$ and $M_{3D} = 3$

Figure 23: Investigation of the energy vector $\vec{E}$ (r_E, W_E and $\delta_{E_{err}}$) for the 92 LS array in case of the different coding techniques for a constant horizontal order $M_{2D} = 7$ and two different periphonic orders $M_{3D} = 2$ and 3. 3D case: blue solid line; mixed case: red dash-dotted line; truncated mixed case: green dash-dotted line. For comparison the 2D case for the 16LS array (black cross) and its behaviour for sources with $\delta \neq 0°$ by making use of eq. (29) (black dashed line) is also shown.

5.3 Non-regular and symmetric system (30 LS-array)

In case of the non-regular, but symmetric 30LS array the pure horizontal reproduction system is approached by a mixed-order system with increasing order M_{2D} and is further improved for a higher order M_{3D} as shown in Figure 24. An additional small improvement is gained by applying truncation for a horizontal source reproduction. For an elevated source the same effects as for the 92LS array are shown, whereas the application of truncation lowers r_E slightly compared to a pure 3D reproduction in case of a higher order M_{3D}. Note that since this array is not independent of the source position (as shown in section 3.4), slightly varying results are obtained for other source positions.



(a)

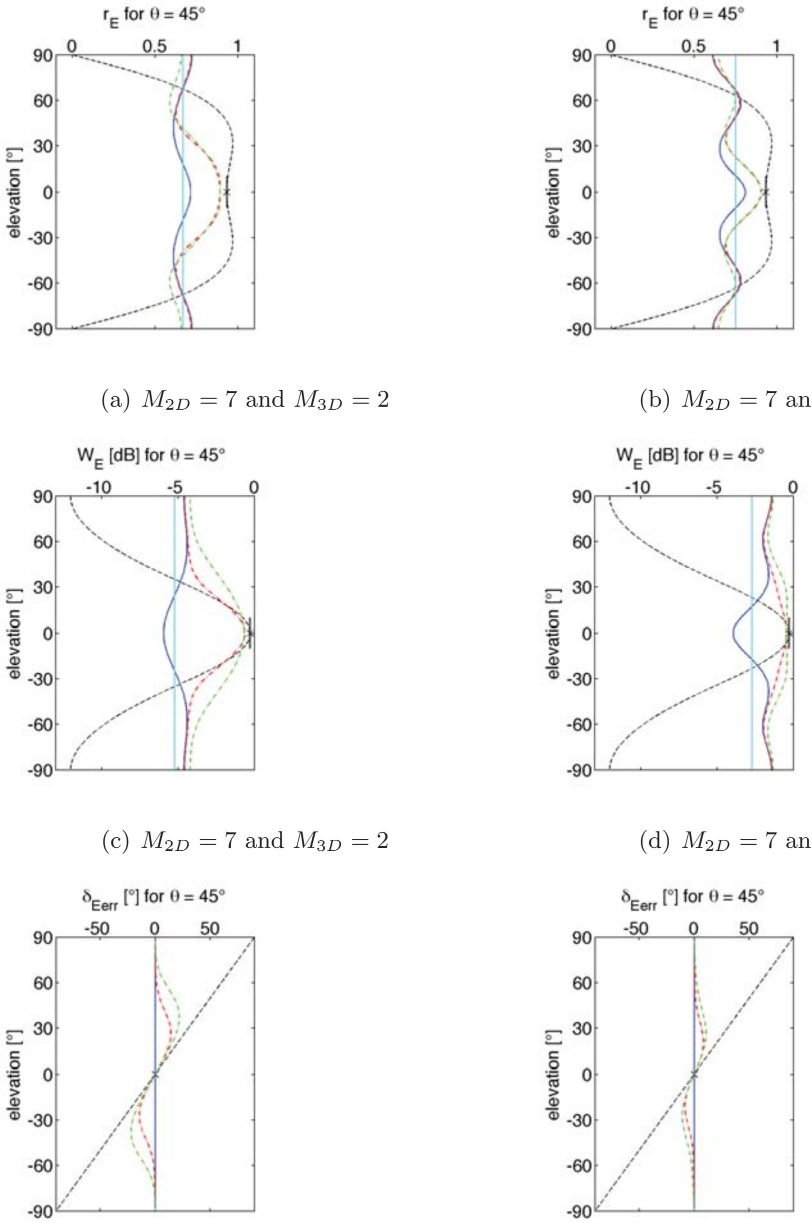
(b)

(c)

(d)

Figure 24: Comparison of the different coding techniques in terms of the energy vector magnitude r_E in dependence of order M for the 30LS array for two specified source positions. The order M refers to the order of a conventional 2D (sources with $\delta \neq 0°$ are plotted by making use of eq. (29)) or 3D representation. The mixed-order implementation has the specified order M_{3D} and $M = M_{2D}$ is the varying parameter. The mixed-order case is to compare with the straight line as this indicates the improvement to a pure 3D coding of this order.

These observations coincide with Figure 25. In addtion the constant values of r_E and W_E as computed with eq. (38) and (37), respectively are indicated. They represent exactly the values for the 2D coding and an averaged value for the 3D coding as mentioned earlier. The



(a) $M_{2D} = 7$ and $M_{3D} = 2$

(b) $M_{2D} = 7$ and $M_{3D} = 3$

(c) $M_{2D} = 7$ and $M_{3D} = 2$

(d) $M_{2D} = 7$ and $M_{3D} = 3$

(e) $M_{2D} = 7$ and $M_{3D} = 2$

(f) $M_{2D} = 7$ and $M_{3D} = 3$

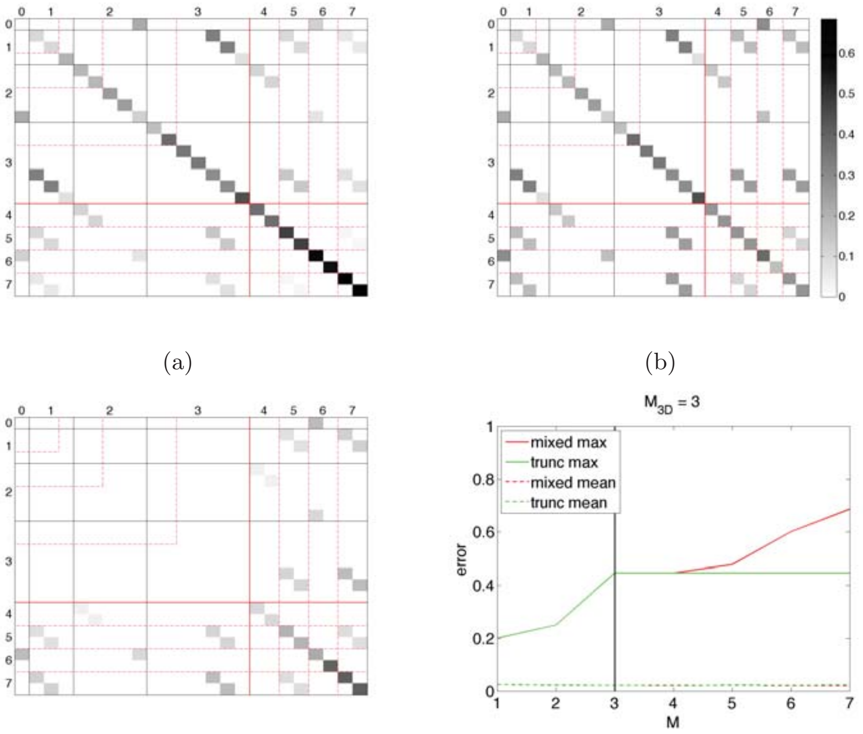
Figure 25: Investigation of the energy vector $\vec{E}$ (r_E, W_E and $\delta_{E_{err}}$) for the 30LS array in case of the different coding techniques for a constant horizontal order $M_{2D} = 7$ and two different periphonic orders $M_{3D} = 2$ and 3. 2D case: black cross ($\delta = 0$) and dashed-line ($\delta \neq 0°$ acc. to eq. (29)); 3D case: blue solid line; mixed case: red dash-dotted line; truncated mixed case: green dash-dotted line. Estimates of r_E and W_E in case of a regular 3D (cyan solid line) and 2D (short black solid line) loudspeaker setup according to eq. (38) and (37), respectively, are shown additionally.

estimations of the energy vector for the two mixed-order systems lie in between the two other systems, where the desired improvement in the horizontal plane and the transition towards a pure 3D system is well indicated. Considering the energy distribution, this array

has a gap in the horizontal plane (pure 3D). The application of the mixed-order algorithm leads to an energy contribution that is similar to one of the pure 2D coding.
A smoothing again is achieved with truncation. Observations about the elevation error are unchanged compared to the 92LS array. The error $\theta_{E_{err}}$ with change over δ (not illustrated) is deviating from the errorless pure 2D and 3D case in the mixed-order systems, but is smaller than $\pm 2^\circ$ in the worst case.
Analysis of the velocity vector result into a magnitude value r_V of unity and an errorless reproduction in all cases.

At this point the importance of a regular horizontal loudspeaker ring should be highlighted as can be seen from the comparison between the here given systems, the 92LS and the 30LS array. Even though the former system is an ideal system in terms of a pure periphonic reproduction, it lacks precision in the horizontal plane, where loudspeakers are non-regularly placed and missing. In contrast, a horizontal regular loudspeaker ring integrated into the setup further improves the spatial resolution in that plane when applying a mixed-order coding technique as it has been illustrated in terms of the 30LS array.

In order to investigate the orthonormality properties of the adapted spherical harmonic functions of a mixed-order system, the orthonormality matrix for the 30LS array is considered (Figure 26). The periphonic order is kept fixed with $M_{3D} = 3$ and the horizontal order M_{2D} is varied. By doing so, it is prominent that the maximal error increases with the horizontal order, from 45% for a pure 3D system ($M_{2D} = M_{3D} = 3$) to 69% for a mixed-order system of order $M_{2D} = 7$ and $M_{3D} = 3$ as it is shown in panel (d). When truncation is applied the maximal error is constant for $M_{3D} \geqslant 3$ disregarding of the horizontal order. The total mean error is thereby kept constant at 2,1% disregarding of order or wether truncation is applied. The other 3 plots ((a) to (c))illustrate the orthonormality matrix for the maximal determined order for this loudspeaker array when used as a mixed-order system. Looking at the difference between a truncated and a non-truncated system, high errors in the additional horizontal spherical harmonic components, especially occurring on the diagonal of the matrix, are reduced, but more equally distributed among components, which makes the constant mean error reasonable. Note that the color scale is chosen according to the maximal error.



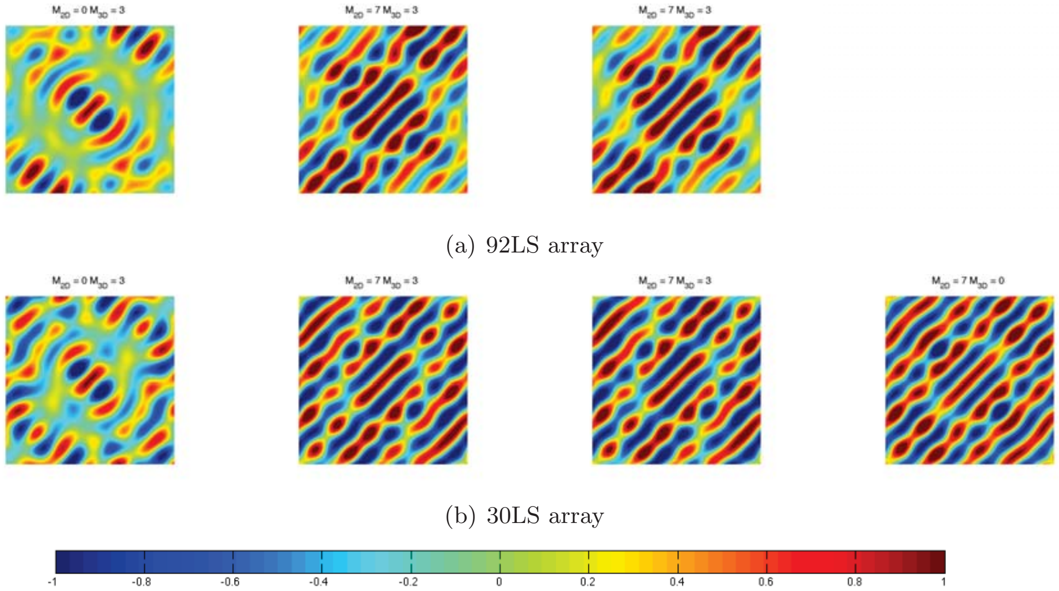
(a)

(b)

(c)

(d)

Figure 26: Orthonormality considerations for the 30LS array in a mixed-order application. (a) to (c) Orthonormality matrix for the combined order $M_{2D} = 7$ and $M_{3D} = 3$ where the scaling of colors refers to the maximal error: (a) Mixed-order case, (b) truncated mixed-order case and (c) absolute difference between both. (d) Errors in respect to M_{2D} with a fixed periphonic order $M_{3D} = 3$.



(a) 92LS array

(b) 30LS array

Figure 27: Comparison of monochromatic soundfield plots ($f = 1000$Hz, $\theta_{src} = 45°$, $\delta_{src} = 0°$) for the different coding techniques applied to the (a) 92LS and (b) 30LS array. From left to right: Pure 3D, mixed, truncated mixed and pure 2D case (for the horizontal ring of the 30LS array).

In order to complete the analysis of a mixed-order system, the synthesied wavefield is considered for the 92LS and 30LS array as illustrated in Figure 27. The pure 3D coding ($M_{3D} = 3$) for both setups is shown to the left of the figure. By applying mixed-order coding to the systems ($M_{2D} = 7$ and $M_{3D} = 3$), the sweet spot area increases in radius and approaches the pure 2D case (lower right plot). There is no visible difference between the truncated and non-truncated coding for each setup individually. The two setups deviate from each other though in the reproduced soundfield outside the sweet area, where an improvement in case of the 30LS array is present.

5.4 Frequency spectrum considerations

Spectral effects are introduced into the reproduction of any Ambisonic system since coherent loudspeaker signals are used and are especially audible when moving the head away from the sweet area. The purpose of this section is to investigate the changes of the frequency spectrum when mixed-order coding is applied. Therefore binaural room impulse responses (BRIR) measured for a B&K dummy head are used to simulate the power spectrum at the two ears of a human listener. Since such measurements can just be performed on existing loudspeaker setups, the Spacelab has been used for this investigation. Its loudspeaker configuration has already been introduced in section 3.1. The obtained power spectra will also be helpful for the interpretation of the results from the listening tests presented in the next chapter and are therefore presented already at this point. The subjective evaluation is performed on (1) a high mixed-order system ($M_{2D} = 7$ and $M_{3D} = 3$), which corresponds to the maximal defined order of the Spacelab (29LS array), and (2) a low mixed-order system ($M_{2D} = 3$ and $M_{3D} = 1$). A reduced order requires at the same time a physically reduced number of loudspeakers, since otherwise this leads to pronounced spectral impairment as described in [19]. For the low-order system, a reduced version of the Spacelab, the 11LS array (see section 3.1), is therefore used.

Two modifications are necessary for a listening test which are described in the following: As had been shown earlier in this chapter the different Ambisonic coding techniques result into different total energy levels depending on the reproduced sound source position. In order to achieve equal levels a normalisation is required, where each loudspeaker gain is normalised by the total energy of all loudspeaker signals

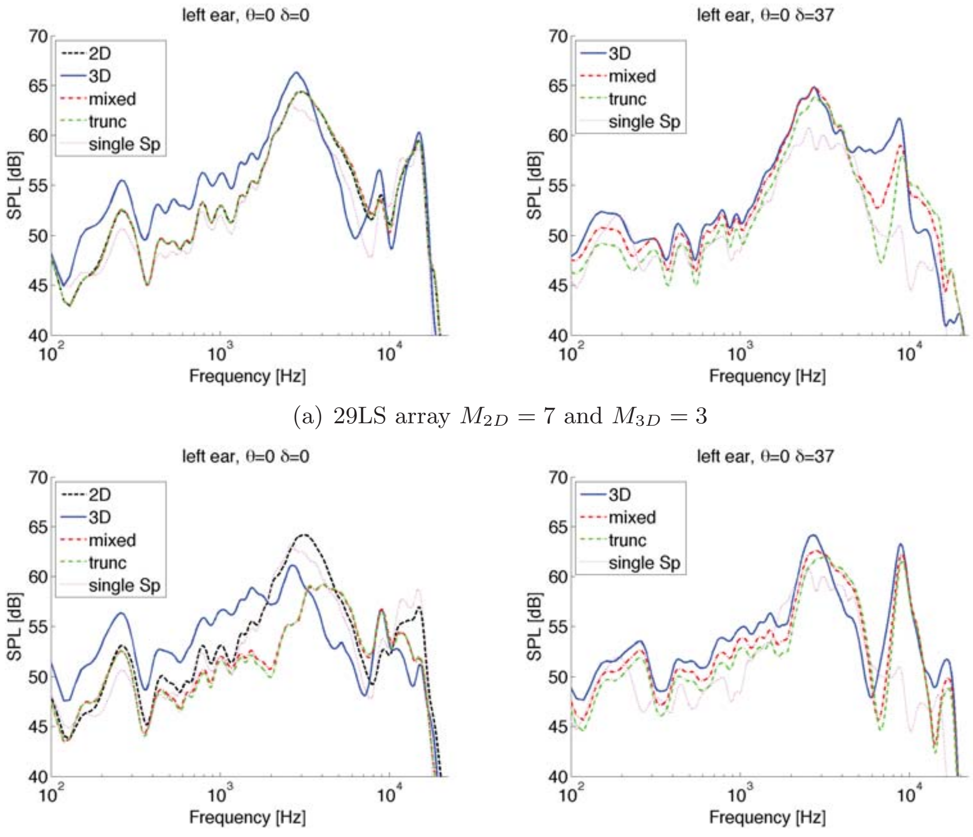
$$\vec{s}_{ls_{norm}} = \frac{\vec{s_{ls}}}{\sqrt{W_E}} = \frac{\vec{s_{ls}}}{\sqrt{\sum\limits_{j=1}^{L} |s_{ls_j}|^2}}. \tag{44}$$

Note that this normalisation accounts only for the level difference obtained for high frequencies (according to the definition of the energy vector given in section 3) and therefore the normalisation is incorrect for low frequencies, where a pressure normalisation should be applied. Since energy normalisation is not a trivial task in a mixed-order system due to the elevation angle dependent energy contribution further details are left for the discussion in section 7.1.

In addition, equalisation filters, that perform a time-alignment, sound pressure level and amplitude frequency response equalisation for each loudspeaker with regard to the center point of the loudspeaker array, are applied as described in [7] Appendix B. These modifications are accordingly applied to the measured BRIRs. The results for the different coding techniques are presented in Figure 28 on a logarithmic frequency scale in the range of 100 Hz to 22 kHz. For better illustration the obtained spectra are smoothened with a gammatone filterbank of order 4. Two different source positions are shown, a horizontal one at $\theta = \delta = 0°$ and an elevated position at $\theta = 0°$ and $\delta = 37°$, which correspond to those used in Experiment A of chapter 6. In order to compare the multiple loudspeaker responses (Ambisonics) to one of a single loudspeaker for the two specified source positions, its spectrum is shown in the same plots.

The spectra are relatively symmetric for the left and the right ear as expected and therefore just the results for the left ear are plotted. For the horizontal listening position, in case of the 29LS array (panel (a) left), the equal spectra of a pure 2D and the mixed-order coding techniques are close to the response of a single loudspeaker at this position. A deviating spectrum is obtained for the pure 3D coding, which is upward shifted about 4 dB at low and mid frequencies (up to around 2 kHz). For high frequencies, an attenuation of around 5 dB is present in a frequency interval of 2 to 7 kHz and peaks and dips at around 10 kHz are more pronounced and shifted in frequency. In case of the 11LS array (panel (b) left) the 2D coding technique does not significantly change the frequency content either, compared to the single loudspeaker, but in contrast stronger deviations occur for the mixed-order coding techniques. Observations about the 3D coding are similar as described for the 29LS array.

When elevating the sound source, the spectra of the two mixed-order approaches get comparable to the one of the 3D coding for both systems with differences in the high frequency region though. In case of the high order system (panel (a) right), the dip at around 7kHz gets more pronounced when using the non-truncated mixed-order technique (approaching the single loudspeaker response for that specific frequency region) and even more when truncation is applied to the system. The frequency content above 10 kHz is slightly boosted for the mixed-order techniques. The obtained differences between the Ambisonic coding techniques are less prominent in case of the low-order system (panel (b) right). Dips at 7kHz and above 10kHz are rather attenuated by the mixed-order compared the 3D coding. Compared to the single loudspeaker response the spectra are mainly amplified at high frequencies for both systems.



(a) 29LS array $M_{2D} = 7$ and $M_{3D} = 3$

(b) 11LS array $M_{2D} = 3$ and $M_{3D} = 1$

Figure 28: Simulated power spectrum for BRIRs of a B&K dummy head applying different coding techniques to the 29LS (Spacelab) and 11LS array (reduced version of the Spacelab) for two source positions. A single loudspeaker response for the specified positions is shown for comparison.

From a perceptional point of view, a slight low frequency boost and high frequency attenuation is expected for a horizontal listening position in case of the 3D coding compared to the other coding techniques. For an elevated sound source, the differences are expected to vanish or being less prominent.

6 Subjective evaluation

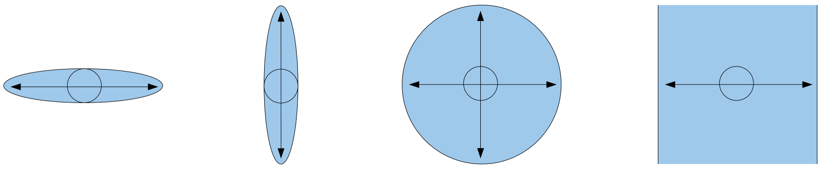
The objective evaluation presented in the previous chapter promises significant improvement of spatial resolution in the horizontal plane on auditory perception when mixed-order coding is applied: Energy is focused towards the desired plane (directivity patterns) and a better localisation is provided as r_E is improved. This chapter is to validate the simulations by means of two subjective evaluation studies, where a proper choice of attributes is necessary. The following list summarises possible perceptual attributes that are worth to investigate for any Ambisonic playback system:

- Apparent sound source focus
- Coloration
- Localisation
- Loudness
- Distance

It is the aim of the first experiment (Experiment A) to estimate a perceptional attribute that can be linked to the 'physical' quantities, directional focus and localisability (r_E) of an Ambisonic signal. This link is given by the attribute *apparent sound source focus* and is more specifically described by *apparent source width* for the experimental investigation. The purpose of the second experiment (Experiment B) is (1) to reveal tendencies of preference in a complex listening scenario and (2) to investigate the multi-dimensional field of perceptual involved attributes. For both studies a high mixed-order system as provided by the Spacelab (29LS array) and a low mixed-order system (11LS array) are used.

6.1 Experiment A

In this experiment the attribute *apparent source width* was evaluated in a listening test procedure in order to get subjective results that are comparable to the objective evaluation. This attribute is shortly explained in the following in order to avoid misunderstandings. The explanation was also given to each test subject. Possible expansions from a certain reference position (a single loudspeaker) are indicated schematically in Figure 29. The attribute *apparent source width* refers to the indicated horizontal expansion as shown in case d) where limits in source height are left unspecified. This means that case d) covers case a) as well as case c) whereas vertical expansions as shown in case b) are not specifically considered.



(a) horizontal (b) vertical (c) two-dimensional (d) app. source width

Figure 29: Possible source expansions from a reference position.

On the basis of the objective evaluation the following null and alternative hypotheses are assumed for the listening experiment. A mixed-order system

H_{0_1}: behaves as a pure 3D system in the horizontal plane,

H_{0_2}: behaves as a pure 3D system for elevated sources (i.e. anywhere apart from the horizontal plane),

H_{A_1}: improves the properties of a pure 3D system in the horizontal plane (furthermore approaches thereby the behaviour of a pure 2D system),

H_{A_2}: its performance is worse than that of a pure 3D system for elevated sources

for a specified combination of orders (horizontal and periphonic order) regarding the chosen attribute *apparent source width* (dependent variable).

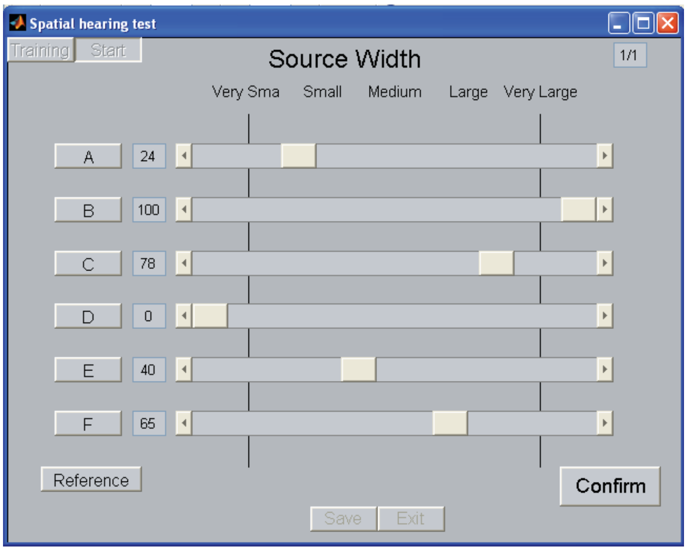
The desired goal of the subjective evaluation is to reject the first null hypothesis H_{0_1}, thereby proving H_{A_1} and at the same time to validate the second null hypothesis H_{0_2}.

6.1.1 Methods

Procedure

The experimental procedure bases on the MUSHRA test, which stands for "MUlti Stimulus test with Hidden Reference and Anchor" and is described in [18]. This test was originally developed to evaluate different audio coding techniques (e.g. MPEG) by rating their quality on a scale between 0 and 100. Several stimuli are presented (in random order) on one screen at the same time and therefore allow for direct comparison between them.

For the present listening test a similar testing procedure was used where the task of evaluating quality was replaced by rating *apparent source width* on a scale from *'very small'* ($\cong 0$) to *'very large'* ($\cong 100$) for the different coding techniques, which are in this case the different Ambisonic coding techniques that have been described and objectively evaluated in course of this thesis: pure 3D, pure 2D, mixed and truncated mixed coding. The reference was a single loudspeaker, which is considered as having the smallest apparent source width ($\cong 0$) and the anchor was produced by using Vector Based Panning (VBP) as described in [17], which is basically the stereo format in this case. The anchor was not designed in the way as being the widest source (according to *'bad'* quality in the original MUSHRA test). It should be more considered as additional anchor position in the used scale in order to make results reproducible and to ensure the exploitation of the entire scale by the test-subjects. Although the use of an anchor can result into bias problems as investigated in [24], it had been nevertheless decided to integrate one into the procedure. It was the aim of the present study to obtain a relative ranking between the different Ambisonic coding techniques. The absolute assigned values were thereby of minor interest, contrarily to the original MUSHRA test, where the results are linked to absolute quality ranking. The graphical user interface presented on a touchscreen is shown in Figure 30. The MUSHRA test environment in MATLAB was provided by Jens Brehm Nielsen and adapted accordingly.



A 24

B 100

C 78

D 0

E 40

F 65

Figure 30: Graphical User Interface (GUI) for the experimental procedure.

The experiment was performed in the Spacelab at the facilities of DTU for a high order and a low order system. It was decided to present only frontal sources to the test-subjects, since human localisation abilities are the most accurate ones for such positions (see section

2.1) and it was the technical system that was to investigate. A photo of the experimental setup is shown in Figure 31. This view represents the front of the loudspeaker array referring to the positive x-direction of the specified coordinate system in Figure 6.

The following two source positions were chosen: One in the horizontal plane ($\delta = 0°$) and the other one at an elevated position of $\delta = 37°$, at which the pure 2D coding was excluded from the evaluation procedure. It was decided to test the system in its worst condition, which is placing a source in between loudspeakers. As earlier simulation studies concerning the magnitude of the energy vector r_E have shown, this quantity takes lower values at these positions compared to sources that are congruent with loudspeaker positions. Therefore it was necessary to install 3 additional loudspeakers - two at $\delta = 0°$ with $\theta = +11°$ and $-11°$ and one at $\delta = 37°$ with $\theta = 0°$ - that are indicated by blue circles in the picture and dealt as reference positions for the experimental procedure. The reason for two reference positions in the horizontal plane was simply to avoid any bias in the results caused by a presentation to just one side of the median plane. Note that these three speakers are not part of the actual Ambisonic playback system.

The two systems (high and low order) were investigated separately, where a break of approximately 10 minutes were enforced in between the listening procedure before presenting the second system. The system to start with was in randomised sequence. The 2 possible source positions (horizontal and elevated) were also presented randomly and repeated once for each subject. The presentation of the two possible horizontal positions ($\pm 11°$) were given in alternating order (one at the initial presentation and the other at the repetition) and their results are averaged and designated as $\theta = 11°$ in the following. The randomisation took into account that there were an equal number of subjects starting with the high or low order system, horizontal or elevated source position and $+11°$ or $-11°$ in case of the horizontal source direction. This resulted into 4 listening conditions per system.

Some subjects changed their mind from the initial to the repeated evaluation of a certain source position, so that not a constant trend in his or her response was obtained. A constant trend hereby refers to the ranking order between the presented coding techniques in the task of determining the *apparent source width* and does not refer to a change in the actual assigned numbers. In such a case the presentation of that specific source position was repeated again for the subject before proceeding with the second system. The last two 'stable' answers (with a constant trend) were then taken into account for the results, whereas all previous presentations were considered as extended training conditions. In average this happened for one to two conditions per subject. The experiment satisfied a balanced design for the statistical analysis.

Each test-subject was instructed with a description in written form (see Appendix), where he or she was always asked to identify the reference by rating one of the stimuli as 0 and encouraged to use the entire scale for the evaluation of the other stimuli. The instructed listening position is shown on a photo in Figure 31 (right), that was looking to the front, avoiding head movement and not closing the eyes. Of course, the subjects had too look down to the touchscreen in order to rate the stimuli, but were asked to keep an upright, frontal head position when evaluating a sound. Proceeding the experiment itself, a short training session including one repetition was held in order to familiarize the subject with the task and the stimuli. The total listening procedure lasted around one hour per subject.

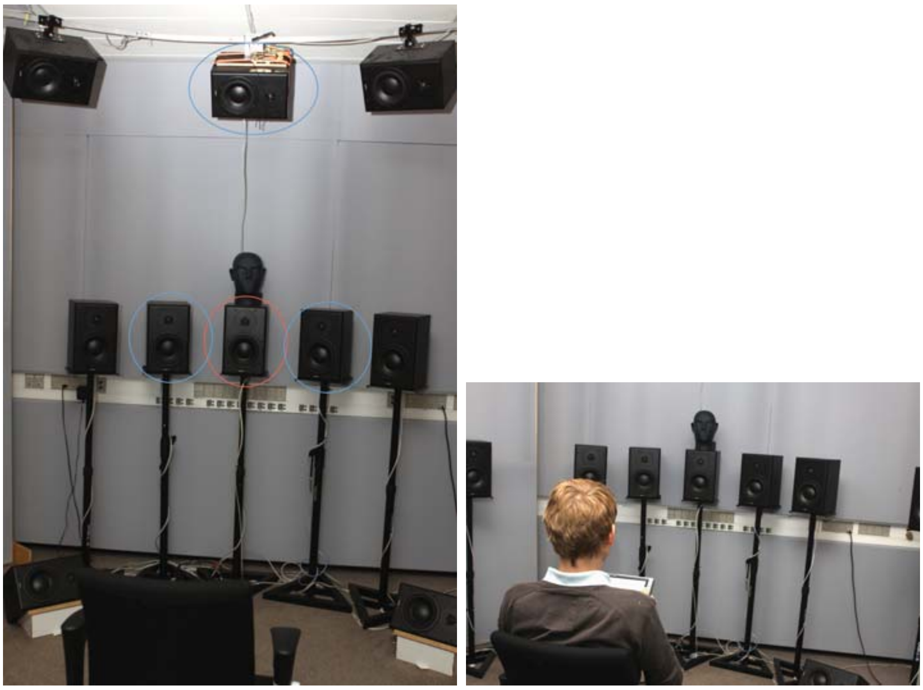


Figure 31: Left: Experimental setup with 3 added loudspeakers (not part of the Ambisonic system) as reference positions (marked blue) and the training reference position (marked red). Right: Instructed listening position.

Stimuli

For the purpose of this experiment the encoded stimulus was chosen as pulsed white noise with a pulse width of 100ms. It is a common stimulus used in localisation experiments, e.g. conducted by Blauert [1], and justifies its use since it provides sufficient spectral information and signal length as described in section 2.1. The pronounced onset slopes of the non-stationary signal provide additional cues (envelope ITDs) that help in localising sources. The first and last 5ms of each pulse were rounded off by a hanning window shaped weighting in order to avoid sharp transitions, so that the sound was more pleasant to listen to. The presentation level was at 50 dB SPL.

One of the two systems under test was the 29LS array (Spacelab) with its maximum determined order $M_{2D} = 7$ and $M_{3D} = 3$ (high mixed-order). The objective properties of this array are very similar to the one of the 30LS system, except for reproduced sources significantly below the hoizontal plane ($\delta < 40°$)). The entire analysis can be found in the Appendix. The second system was the 11LS array which represents one possible solution for a low mixed-order system. The number of loudspeakers is physically reduced in order to minimise coloration effects as described in [19]. This results into a maximum order of $M_{2D} = 3$ and $M_{3D} = 1$ after orthonormality considerations (see Appendix). Both arrays are shown in section 3.1.

In order to minimise additional cues caused by loudness differences, the calculated loudspeaker gains were power-normalised as described by eq. (44). In addition the individual loudspeakers were calibrated with equalisation filters as described in [7]. After calibrating and normalising the system, equal loudness was approximately achieved for the different stimuli, as was confirmed by three trained listeners from the department. Note that the different coding techniques result into different coloration effects as analysed in section 5.4.

Due to varying source positions, appropriate anchors had to be implemented. The loudspeaker base angle (see Figure 32) for the anchors differed slightly for the two listening positions because existing loudspeakers in the array were used for the loudspeaker pairs. This led to a base angle of $\varphi_0 = \pm 34°$ for the two horizontal reference positions and $\varphi_0 = \pm 30°$ for the elevated listening position. The according loudspeaker gains were calculated as described in [17] for a two-dimensional loudspeaker base and were normalised using a high frequency local panning method (for $f > 700Hz$) called Vector Base Intensity Panning (VBIP) provided by [15] resulting into

$$\vec{s}_{VBIP} = \sqrt{\frac{\vec{g}}{\sum_{j=1}^{L} g_j}}. \tag{45}$$

This normalisation method was chosen in contrast to Vector Base Amplitude Panning (for $f < 700Hz$) to be consistent with the chosen power normalisation used for the Ambisonic systems (see eq. (36) and eq. (44)).

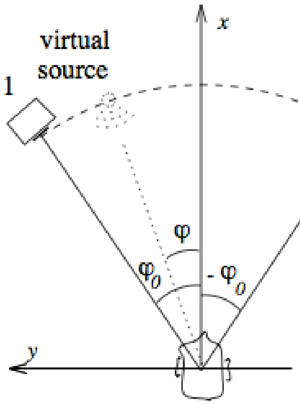


Figure 32: Illustration of the loudspeaker base angle φ_0 [17].

For the training condition a reference position at $\theta = 0$ and $\delta = 0$ (as indicated in Figure 31) was chosen. Conventional 2D and 3D coding of different orders plus an anchor with a base angle of $\varphi_0 = 45°$ were chosen in the way of describing best differences in *apparent source width*. This resulted into 6 stimuli in total: 3D ($M = 3$), 3D ($M = 1$), 2D ($M = 5$), 2D ($M = 2$), anchor and reference. The results from the training session verified that the task of rating the attribute *apparent source width* were understood by a subject as low-order signals were rated as having a large source expansion, whereas high-order signals were rated as small.

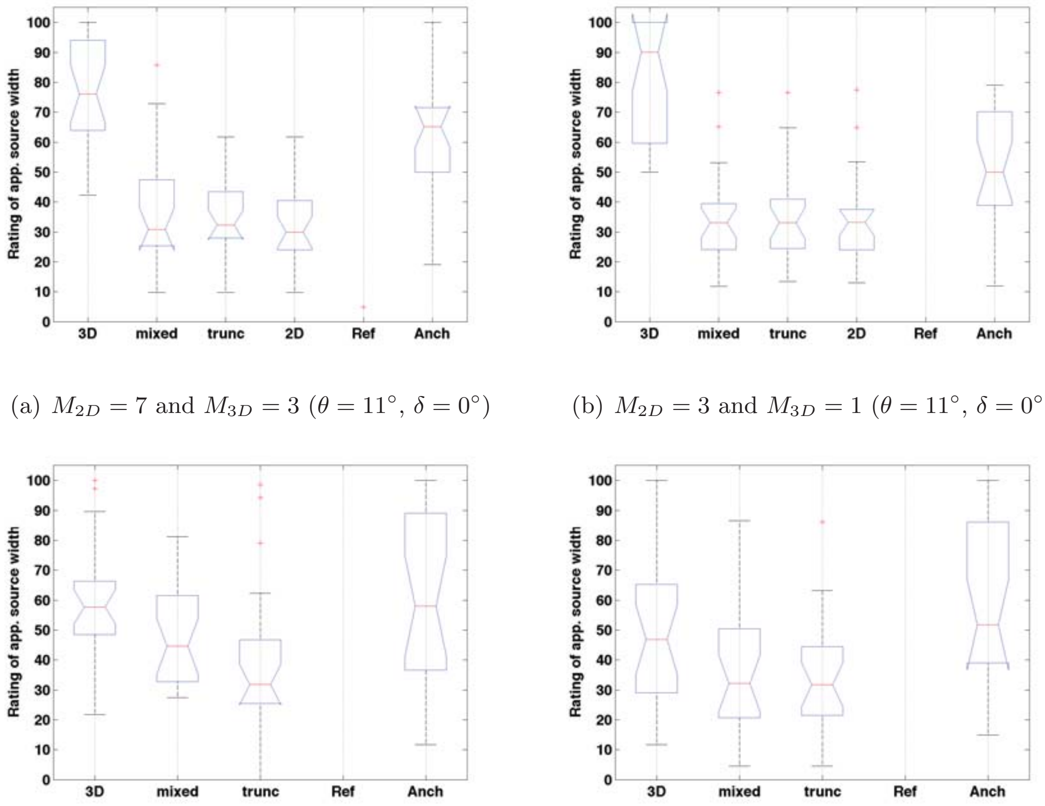
Subjects

In total 12 test-subjects (10 male, 2 female) with normal hearing in the age of between 24 and 30 years were tested. All subjects were experienced in participating in psychoacoustic tests. Their individual results were averaged and are presented by following the recommendations specified in [18].

6.1.2 Results

The graphical illustration of the results represent a box-and-whisker plot as described in [21] pp. 39-43, obtained by using the *boxplot* method in MATLAB. The median of the data for each treatment (coding) is indicated by a red line. The box around the median value marks the 25th and 75th percentiles and thereby describes the interquartile range. The whiskers contain points lying inside the interval of 1.5 times the interquartile range. Values outside that range are considered as outliers and are represented by red pluses. In order to statistically validate the results a balanced two-way ANOVA (analysis of variance) test was performed by using the MATLAB function *anova2*. This method bases on a linear model and assumes normal distributions and equal variances, but is fairly robust against violations of these assumptions (e.g. in a comparison of medians with other distributions) as stated in [12] p.405. In addition, a multiple comparison was performed in MATLAB with the method *multcompare*. That method reveals which coding technique differs from another on the basis of an overall level of significance, which is not possible with common two-sample t-tests (for further details see [12] p.425).

In Figure 33 the averaged results for all experimental conditions (2 source positions and 2 systems) are shown. From the simulation studies it is expected that the ranking of the Ambisonic coding techniques for the horizontal source position (panel (a) and (b)) is linked to their according value of r_E. This is indeed proved by the results for both systems,



(a) $M_{2D} = 7$ and $M_{3D} = 3$ ($\theta = 11°$, $\delta = 0°$)

(b) $M_{2D} = 3$ and $M_{3D} = 1$ ($\theta = 11°$, $\delta = 0°$)

(c) $M_{2D} = 7$ and $M_{3D} = 3$ ($\theta = 0°$, $\delta = 37°$)

(d) $M_{2D} = 3$ and $M_{3D} = 1$ ($\theta = 0°$, $\delta = 37°$)

Figure 33: Averaged experimental results for two source positions, horizontal ($\delta = 0°$) and elevated ($\delta = 37°$), and two playback systems, high mixed-order (29LS array) and low mixed-order (11LS array).

the high-order system (left) as well as the low-order system (right), as 3D encoded signals are rated as having a large source expansion, whereas 2D and mixed-order encoded signals are rated as small. Note that even both systems reveal similar assigned values their results are not comparable with each other since both systems have been tested separately. The anchor is thereby placed in between both observations. The range in the individual treatments is relatively constant, considering the interquartile range as well as the range of whiskers, where only the pure 3D case and the anchor reveal a higher individual spread of the results. The variability of the calculated medians between samples for each treatment is indicated by notches displayed in the boxes. A significant difference between the medians of two coding techniques is indicated when there is no overlap between their notches on the basis of a 5% significance level ($p - value < 0.05$). Based on this analysis, there is a significant difference between the 3D and the three coding techniques 2D, mixed and truncated mixed coding, where the last three mentioned ones do not have a significant difference to each other.

Considering the results for the elevated listening condition (panel (c) and (d)), again similar results are obtained for both systems. The essential observation here is that the two mixed cases are not rated worse than the pure 3D coding. Contrarily to that, they are rated with a smaller *apparent source width* than the latter mentioned coding. A significant difference occurs hereby between 3D and the truncated mixed coding in case of the high-order system, which is not present in the lower order system. The multiple comparison reveals though that all 3 coding techniques are significantly different in the high-order system and that the difference between the two mixed-order systems vanishes for the low-order system. The results take a high individual spread in some of the cases. There had been three subjects rating the three coding techniques 3D, mixed and truncated mixed in exactly the opposite trend, meaning an increasing *apparent source width* in the mentioned order. This explains the outliers in the truncated mixed coding in both systems.

Both factors, coding technique and subject, are influencing factors in the given conditions and reveal an interaction based on a $p - value < 0.05$. The detailed ANOVA tables are listed in the Appendix.

The first null hypothesis H_{0_1} can be rejected on the basis of a 95% confidence level for each of the two specific systems, proving H_{A_1} at the same time. Definitely, the second alternative hypothesis H_{A_2} can be rejected, but simultaneously the second null hypothesis H_{0_2} has also to be rejected, since the multiple comparison revealed a better performance of the mixed-order systems (for the specific evaluation angle of $\delta = 37°$).

6.1.3 Discussion

Summarising the results, it can be stated that there is a match between the objective and the subjective evaluation. The principle goals concerning localisation issues of a mixed-order Ambisonic playback system as defined in section 4 are achieved. It is thereby proved that a mixed-order system improves essentially the *sound source focus* in the horizontal plane and keeps a performance that is not worse than the one of a pure 3D system in case of elevated sources.

Even though, the importance of the chosen attribute *apparent source width* regarding to *sound source focus* has been validated in investigations of a mixed-order playback system, it is one out of several possible perceptual attributes (as mentioned in the beginning of this chapter). The awareness of other potential cues were given before the experiments and have been verified by comments from the test-subjects. Each person was shortly interviewed throughout the experimental procedure about other perceivable cues, where the most mentioned ones are as expected unbalanced frequency content (coloration), localisation changes and varying loudness. In addition, some of the subjects experienced front-back reversals, which were most likely to occur for the low-order encoded stimuli and are also likely to be intensified by the short pulsed signal as mentioned in section 2.1. This goes in hand with the exact head position of a test-subject, where perception is very sensitive to head motion and to even slight position changes. Subjects have been instructed to keep an upright head position, but small changes are unavoidable since the subjects had also to interact with the graphical interface. The option of a fixed head position as described in [1] was not selected since a free movable head represents a more realistic listening condition when use is made of a sound reproduction system. Also, different distance perception was mentioned, where between a source in front of the listener and a source that envelopes the listener were distinguished. This refers again to different coding orders present in the experiment. It is hard to uncouple single attributes because they are closely related to each other. Even though, each person was instructed just to focus on the single attribute *apparent source width*, it is not excludable that other cues influenced the subjective evaluation process. The most important relations between different cues in regard to the executed experiment are described in the following.

The described variations in loudness are closely related to differences in the frequency content. It is well-known that the spectral content affects the perceived loudness (for example [16], p.116) and therefore makes this effect plausible. Frequency changes were also perceivable as either pitch changes or filtering/boosting of certain frequency bands. It is thereby not excludable that the changing frequency content influenced the subjects in that way, as low frequency sounds tended to be associated with a big source, whereas high frequency sounds are likely linked to smaller sources. The "sound color" or timbre of the pure 3D system has a more pronounced low frequency content, whereas the two mixed systems and the pure 2D system have similar frequency content that can be described with a brighter timbre. The here made observations coincide well with the objective analysis of the frequency spectrum in section 5.4.

For the elevated source position the localisation of the source tips downwards when the coding is changed from pure 3D, to mixed and finally to truncated mixed, exactly in this order. This effect is observable for the high- as well as for the low-order system. An increased elevation error $\delta_{E_{err}}$ for the mixed-order systems compared to the pure 3D system has already been pointed out in section 5.3. Two possible explanations for the here described effect arising from the mixed-order algorithm itself are provided in the following: Considering again the directivity plots for the different Ambiconic systems in Figure 20 a low mixed-order system has a mainlobe that points towards the front even for sources with an elevation angle of $\delta = 45°$. The direction of the mainlobe in a high mixed-order system is correct, but there is a pronounced sidelobe, that is pointing downwards, visible until the same elevation angle. The effect is larger for a truncated mixed system compared to a non-truncated mixed system. So the effect of a downwards tipping source could be due to higher energy contribution from the horizontal loudspeaker ring or negative elevated speakers compared to a pure 3D reproduction system.

A second plausible explanation can be due to the change in frequency content. As explained in section 2.1 it is the spectral information of the signal that provides the cue for determining elevation. The different filter properties of the system could arise the impression of a down shifting source. It is also possible that a combination of both effects is the explanation for the observed localisation change. It had been highlighted in the previous section that the *apparent source width* of these three coding techniques have been rated as decreasing in the same order. It is possible that the tipping of the source influenced also the perception of the *sound source focus*.

6.2 Experiment B

In the first experiment (Experiment A) an improvement of the 3D system by the two mixed-order approaches could be verified for an artificial and controlled listening condition. The behaviour of these 3 systems in a direct comparison for a complex and more realistic listening situation was tested in a second experiment. The intention of Experiment B is to reveal a tendency, whether the mixed-order approach (for a high- and a low-order system) also provides audible improvement for such a condition which is the desired goal of this implementation. Simultaneously, attributes of special importance should be outlined and support future studies in focussing on individual aspects (attributes) of a complex scene. Due to limited time this experiment should only be understood as a pilot test.

6.2.1 Methods

Procedure and stimuli

A virtual concert scenario was implemented by making use of the LoRA Toolbox ([7]). This toolbox was adapted in order to allow for mixed-order Ambisonic encoding. The simulation bases on room impulse responses (RIRs) for specified source-receiver positions in an acoustical three-dimensional room model (calculated by ODEON), conveying temporal, spectral and directional information. The RIRs were split up into three parts, direct sound, early reflections and late reflections, which were all processed individually by encoding them with Ambisonics: Direct sound and early reflections were encoded with higher-order Ambisonics using the 3D or mixed-order coding techniques. The late reflections are represented by envelopes of the energy and of the vectorial intensity. They were encoded at 1st order (pure 3D) Ambisonics and in addition multiplied by Gaussian noise that is uncorrelated between all loudspeakers in order to recreate the natural high density and diffuseness of late reflections. All three parts were finally added and resulted into a multi-channel room impulse response (mRIR), containing the impulse response for each loudspeaker of the system, i.e. 29 channels in case of the Spacelab used in this experiment. A pop-song with seven instruments - vocals (male), piano, organ, guitar, bass, drums and sound effects (cello or bells) - was chosen and simulated for a room model of the big concert hall in the Royal Danish Academy of Music (DKDM). Accordingly, seven individual source positions on stage and a single receiver position at the ground floor of this hall (empty) for a sitting listener (in a distance of 8m) were defined, resulting into seven source-receiver positions (RIRs). The anechoic mono recordings (24bit, 44.1kHz) of the individual instruments were convolved with their corresponding mRIR and the resulting files were superimposed in a single playback matrix. This process was repeated for all of the three coding techniques, for the high-order ($M_{2D} = 7$, $M_{3D} = 3$, 29LS array) and low-order system ($M_{2D} = 3$, $M_{3D} = 1$, 11LS array) from Experiment A. All files were power-normalised in order to present them at the same comfortable listening level.

The complex scenario introduces a multidimensional field of attributes. The difficulty of finding appropriate attributes is the matching between sensation and verbal description. Methods of finding adequate attributes for multichannel reproduced sounds are for example described in [2] and evaluation procedures for such attributes are presented in [3]. They were taken into account for the design of the present study. Seven attributes were selected according to the chosen listening scenario. The attributes were *room, envelopment, distance, spatial distinctiveness, naturalness and clarity.* In addition, the total *preference* was asked. Explanations of the individual attributes are given in the Appendix, which is the original hand out that was given to each subject. In a proceeding conversation with each subject the attributes were clarified. The attributes were further grouped in order to help the subjects to focus on specific parts of the scene. The first two attributes, *room* and *envelopment*, intend the investigation of the space (the concert hall) that was simulated around the subject. The next two attributes, *distance* and *spatial distinctiveness*, were linked to the perception of the simulated stage in front of the subject, i.e. how far the musicians appeared to be away from the listener and how spatially focussed they could be perceived, respectively. With *naturalness* the total scenario was evaluated, whether it was a realistic or an artificial scene. Concerning the frequency spectrum, *clarity* described whether the sounds were brilliant (high clarity) or muffled (low clarity). On the handout the subjects could optionally write down other attributes or comment on the suggested ones with their own words. The subjects could switch between the three coding techniques, labeled as A (3D), B (mixed-order) and C (truncated mixed-order), on an interface similar to the one used in Experiment A (see Figure 31). They were not told about the different coding techniques themselves. Each attribute was investigated on a new screen in the same order as mentioned here. The task was to rate each sound (A,B, and C) according to an attribute on a scale from *1* to *5* (with an integer stepsize), where *1=bad, 2=poor, 3=fair, 4=good, 5=excellent*[7]as recommended in ITU Recommendation P.800 (08/96) (in concern with the "Subjective determination of transmission quality").

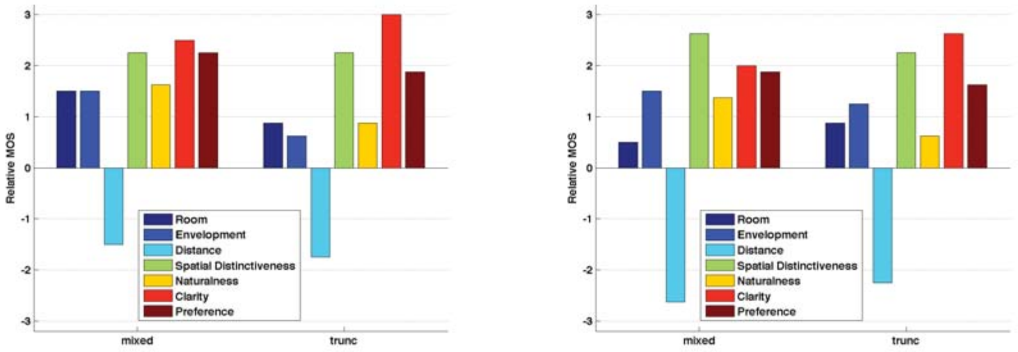
The averaged scores for each attribute resulted into the Mean Opinion Score (MOS). The high and lower Ambisonic systems were evaluated separately. The entire procedure including instructions lasted around one hour. No repetitions have been performed.

Subjects

In total 8 normal hearing subjects (6 male, 2 female) in the age of between 24 and 37 years have been tested. Three of the subjects also participated in Experiment A and all subjects were experienced with psychoacoustic measurements. Their averaged results are presented in the following.

6.2.2 Results and discussion

For the presentation of the results, the 3D system is considered as reference system and the deviation of the two mixed-order systems from this reference is shown in Figure 34. The absolute MOS values are listed in the Appendix.



(a) $M_{2D} = 7$ and $M_{3D} = 3$

(b) $M_{2D} = 3$ and $M_{3D} = 1$

Figure 34: Averaged relative MOS values for 7 attributes in a complex listening scenario for two mixed-order systems (a) high-order and (b) low-order referenced to a pure 3D coding of the specified order M_{3D}.

As can be seen from the figure, both mixed-order approaches get higher scores compared to the pure 3D coding throughout all attributes. This is true for the high order (panel (a)) as well as for the low order system (panel (b)), where similar tendencies are revealed in both cases. A major difference is prominent for the two attributes *spatial distinctiveness* and *clarity* that were assigned with highest scores in both mixed-order approaches. The listeners stated that the space "opens up", shows "more details" and provides "a

[7] In case of the attribute *distance* the scale is adapted, so that *1=very close* and *5=very far*.

more brilliant sound". Contrarily, the 3D system was described as "less focussed" and "muffled". A slightly higher spatial resolution was assigned to the non-truncated system, which is degraded by the second approach as the application of truncation implies. This is due to its reduced directional selectivity in elevation compared to the non-truncated coding technique. The truncated system introduces a higher clarity though, which some subjects found as even "too brilliant" or "too bright". The sensation of the *room* seems to be coupled with the feeling of *envelopment*. An enveloping sound supplied also a better imagination of the room. It was commented that the envelopment was as well linked to the perceived distance to the musicians. It is obvious from the results that the mixed-order systems bring the stage closer to the listener and that this supports the sensation of "being part of the scenario" (more envelopment). At the same time, a closer stage resulted into the feeling of a smaller room as comments revealed. The total scene was evaluated as more *natural* compared to a pure 3D representation. This is even emphasized for the non-truncated coding, which seems to be a decisive factor for the total preference: Both mixed-order approaches are assigned with higher preference ($\sim +2$) compared to the conventional 3D system, where the non-truncated coding is mostly favoured. The most prominent difference in the observations from a high to a low mixed-order system, that is referenced to a 1st order Ambisonic 3D system, is the pronounced perceptional change of distance and spatial distinctiveness. Simultaneously, the difference in clarity is slightly reduced.

Even though these results do not allow for a statistical correct validation due to the pilot character of this experiment and too few subjects, they reveal distinct tendencies. It can be summarised that the mixed-order coding is preferred in contrast to a conventional periphonic system. The preference seems thereby mostly influenced by the improved spatial resolution and a higher clarity. It is surprising that almost identical pronounced effects can be even obtained for a low-order system. Individual statements were that both systems gave similar impressions, but that the difference to the pure 3D coding was stronger audible in case of the low-order system.

A better resolution can at the same time suppress the importance of other attributes and maybe biases the sensation of naturalness. The question is left how detailed a complex scenario in a virtual environment has to be in order to represent a realistic scene. A real-life experience of such a concert might provide less spectral and directional information. Also the perceived distance to the stage might be even too close compared to the specified listening position in case of the mixed-order systems. In the future, for further investigations of a mixed-order system and in the ”fine tuning” of its design, the aspect of authenticity has to be considered additionally. A statistical sophisticating experiment that investigates these issues needs to be performed in future studies.

7 Further and future work

7.1 Transition frequency and energy normalisation

Considering again the energy normalisation from eq. (44) used for the subjective evaluation in Experiment A, the normalisation is not optimal due to the fact that it bases on high frequency considerations. The normalisation should be only applied to higher frequencies where the cut-off frequency can be determined as follows.

Making use of eq. (34) by considering a fixed sweet spot area in dimensions of a head radius of $r = 0.1$m defines maximal frequency limits f_{lim} depending on the order M for that an errorless reproduction (error $< 4\%$) is given according to [7]:

$$f_{lim} = \frac{Mc}{2\pi r}. \tag{46}$$

The according frequency limits for each order M, disregarding if the system is a periphonic or a surround system, is presented in Table 1.

M	1	2	3	4	5	6	7	8
f_{lim} [Hz]	546	1092	1638	2184	2730	3275	3821	4367

Table 1: Frequency limits f_{lim} depending on the order M

These frequency limits give at the same time an indication of the transition frequency f_{trans} above which the energy normalisation should be applied. Below this frequency the contributed energy is unmodified as described in [7] by using a *basic* decoder (eq. (20)). In case of a pure 2D or 3D coding the transition frequency can be read from the table according to the specified order of the system. In case of a mixed-order system, the question arises which order should be chosen, M_{2D} or M_{3D}, to determine the transition frequency. There is no direct simple answer to that question. A first solution could be by evaluating the transition frequency for an averaged value of both orders. A more sophisticated method is by considering both orders individually. The frequency limit gets thereby dependent on the elevation angle since the mixed-order system follows a transition from a 2D ($\delta = 0°$) to a 3D system ($\delta = 90°$). The question is at which elevation angle the according order is to be chosen. The transition area in between is dependent on a point where the influence of the horizontal spherical components can be neglected. As has been outlined in section 4, this is directly linked to the influence of the Legendre functions. A possible intuitive solution could therefore be given with a logical decision element with a threshold t_f, determining whether M_{2D} or M_{3D} should be chosen for calculating the transition frequency. Computing the Legendre functions (with $m = n$) at a specified elevation angle δ of the reproduced source normalised to the maximum value of the Legendre function $max(P_{mm}(\delta)) = P_{mm}(\delta = 0)$ of highest order (determined by M_{3D}) indicates the present influence. The frequency limit is then given by

$$f_{lim} = \begin{cases} \frac{M_{2D}c}{2\pi r} \text{ if } \frac{P_{mm}(\delta)}{P_{mm}(0)} \geqslant t_f \\ \frac{M_{3D}c}{2\pi r} \text{ if } \frac{P_{mm}(\delta)}{P_{mm}(0)} < t_f \end{cases} \tag{47}$$

The task is then to find a proper threshold t_f. The problem with this method is that the knowledge about the source's elevation angle δ must be provided, which is trivial in simulation studies. As soon as real recording material is involved, supplied by an appropriate microphone array, this information is lacking since only information about the Ambisonic channels B^{σ}_{mn} is provided. The advantage of Ambisonics of having a decoupled encoding and decoding would also be destroyed by such an approach. Therefore, there is the need of finding another solution that does not require this additional information. Furthermore, a smooth shift in the transition frequency might be desirable. A possible design for these issues would look like as follows:

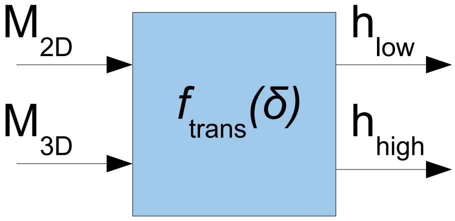


Figure 35: Illustration of determining low and highpass filters for a mixed-order system.

The lowpass and highpass filter h_{low} and h_{high} with transition frequency f_{trans} are applied to separately calculated loudspeaker gains as performed in [7] and then superimposed. Note the time dependence of the signals on the samples n.

$$\begin{aligned} \vec{s}_{ls_{low}}(n) &= D\vec{B}(n) \\ \vec{s}_{ls_{high}}(n) &= D_{norm}\vec{B}(n) \\ \vec{s}_{ls}(n) &= h_{low}(n) * \vec{s}_{ls_{low}}(n) + h_{high}(n) * \vec{s}_{ls_{high}}(n). \end{aligned} \tag{48}$$

The matrix D_{norm} indicates an energy normalised decoding matrix. Energy normalisation is described in [5] p.179 for conventional regular surround and periphonic systems by making use of predictable energy values as given with eq. (37). The normalisation can then directly be applied to the decoding matrix, resulting into

$$D_{norm} = \frac{1}{\sqrt{W_E}} D = \sqrt{\frac{L}{K}} D. \tag{49}$$

There is no single constant energy value in a mixed-order system that can be used for the energy normalisation at $f > f_{trans}$ as the graphical illustrations in chapter 5 have shown. Mixed-order systems are highly dependent on the elevation angle, but lie somewhere in between the predicted estimates of the 2D and 3D systems. Furthermore, this dependency varies for the specified combinations of orders.

Since predictable normalisation constants are not available for a mixed-order system another solution is sought. In [5] p.187 energy normalisation in the treatment of semi-regular loudspeaker configurations[8] is proposed by using just information that is contained in the decoding matrix:

$$D_{norm} = \frac{1}{\sqrt{W_E}} D = \frac{1}{\sqrt{trace(DD^T)}} D. \tag{50}$$

Applying this equation to the 30LS array leads to a shift of the energy curves to a point around the 0dB-level by maintaining their shape, since again just a single constant value is applied to the decoding matrix.

One could therefore imagine to split the decoding matrix into the spherical harmonics ($Y^{(3D)}$) and their added horizontal components ($Y^{(2D)}$) as introduced by the mixed-order system:

$$D = \begin{pmatrix} Y^{(3D)}(\theta_{ls_1}, \delta_{ls_1}) & Y^{(2D)}(\theta_{ls_1}, \delta_{ls_1}) \\ \vdots & \vdots \\ Y^{(3D)} & Y^{(2D)} \\ \vdots & \vdots \\ Y^{(3D)}(\theta_{ls_L}, \delta_{ls_L}) & Y^{(2D)}(\theta_{ls_L}, \delta_{ls_L}) \end{pmatrix}, D^{(3D)} = \begin{pmatrix} \vdots \\ \vdots \\ Y^{(3D)} \\ \vdots \\ \vdots \end{pmatrix}, D^{(2D)} = \begin{pmatrix} \vdots \\ \vdots \\ Y^{(2D)} \\ \vdots \\ \vdots \end{pmatrix} \tag{51}$$

[8] Semi-regularity is defined as fulfilling orthogonality, but not necessarily orthonormality properties of the spherical harmonics.

Doing so allows for a separated normalisation of the two individual groups of components. In order to achieve a constant energy contribution for all elevation angles, decoding matrix $D^{(3D)}$ is normalised by the energy contribution from the 2D-components and the other way around, $D^{(2D)}$ is normalised by the energy contribution from the 3D-components

$$\begin{aligned} D^{(3D)}_{norm} &= \frac{1}{\sqrt{trace(D^{(2D)}D^{(2D)T})}} D^{(3D)} \\ D^{(2D)}_{norm} &= \frac{1}{\sqrt{trace(D^{(3D)}D^{(3D)T})}} D^{(2D)}. \end{aligned} \tag{52}$$

Since this also leads to a total shift of the now flattened energy contribution, the resulting energy contribution of the reassembled decoding matrix $\widetilde{D}$ must also be considered and finally results into the desired normalised decoding matrix D_{norm}.

$$\begin{aligned} \widetilde{D} &= \begin{pmatrix} D^{(3D)}_{norm} & D^{(2D)}_{norm} \end{pmatrix} \\ D_{norm} &= \frac{1}{\sqrt{trace(\widetilde{D}\widetilde{D}^T)}} \widetilde{D} \end{aligned} \tag{53}$$

Note that the effect of applying individual normalisation terms is a weighting of the individual components (as it is done in optimisation techniques described in the next section) and therefore slightly affects the energy vector's magnitude r_E as well. The here proposed method resulted into homogeneous energy contributions in case of the 30LS (±0.5dB) and the 92LS (±1dB) and the effect on r_E is neglectable. The transition frequency can then be found by considering the energy contribution of the 2D-components as a decision element. The effective influence of such an implementation needs to be further investigated for example in terms of the resulting frequency spectrum as done in section 5.4.

7.2 Optimisation decoding methods

Additionally to the in this thesis used *basic* decoder, other decoding principles such as *max* r_E or *in-phase* decoding are available and are compared in [5] pp.184. Their aim is to optimise Ambisonic directional patterns, where always the trade-off between mainlobe width and sidelobes is given. While the latter one removes sidelobes completely but simultaneously enlarges the mainlobe, the first one concentrates energy contributions in direction of the sound source and thereby reduces sidelobes without significantly enlarging the mainlobe. The *max* r_E decoder is of special interest since its application is recommended for medium and high frequencies as stated in [14]. At low frequencies the sidelobes are useful for a proper reproduction of wave propagation properties. Such a decoder is applied by modifying the decoding matrix with order-dependent gains g_m

$$D_{(g_m)} = D\Gamma_{(g_m)}, \tag{54}$$

where $\Gamma_{(g_m)}$ is a diagonal matrix. A full derivation of the appropriate gains g_m is given in [4] pp.18-21, where the energy vector's magnitude

$$r_E = \frac{2\sum\limits_{m=1}^{M} g_m g_{m-1}}{g_0^2 + 2\sum\limits_{m=1}^{M} g_m^2}, \tag{55}$$

is maximised with the boundary condition $\partial r_E/\partial g_m = 0$. This optimisation can be done for conventional 3D and 2D systems and a similar optimisation is desired in terms of a mixed-order system where the horizontal and periphonic order M_{2D} and M_{3D} are taken separately into account.

7.3 Future experiments

In future subjective evaluations it has to be proved that the here suggested or other methods result into a source direction independent frequency content of a mixed-order system. The problem of a changed source localisation for elevated sources that is present in both mixed-order approaches compared to a conventional 3D coding had been outlined in the discussion of Experiment A (section 6.1). It has to be tested whether an equalised frequency content is enough to overcome this issue or if maybe an elevation-transformation needs to be implemented for a mixed-order system.

Including these calibration procedures a statistical sophisticating experiment can be performed for complex listening scenarios (as for example chosen for Experiment B (section 6.2)) by investigating single attributes in comparison tests based on a AB or MUSHRA listening-test design. This is necessary in order to obtain a complete verification of the improvement of a conventional 3D coding by a mixed-order coding technique.

8 Conclusions

In the course of this thesis several aspects of a mixed-order playback system have been outlined. Most important, it has been quantitatively shown that the chosen coding strategy for a basic decoder leads to a higher spatial horizontal resolution of 3D loudspeaker setups that approaches the properties of a conventional 2D system dependent on the number of loudspeakers present in the horizontal plane. This is true for high-order as well as for low-order mixed systems as proved with two exemplary loudspeaker setups.

The algorithm uses the spherical harmonic functions as a basis and extends them with additional horizontal components, which is a straight forward modification of the conventional 3D encoding and decoding principle. The maximal mixed-order of a loudspeaker setup, i.e. the combination of the horizontal order $M_{2D_{max}}$ and the periphonic order $M_{3D_{max}}$, is obtained by considerations of the orthonormality properties of the spherical harmonic functions as it can be performed with the orthonormality error matrix. These considerations revealed a violation of the orthonormality definition and justified a truncated mixed-order coding, that takes the maximal defined periphonic order of the setup into account. This guarantees also a reduced error for the extended components up to an order $M_{2D_{max}}$.

It has been verified with subjective evaluations that the directional focus in the horizontal plane is effectively improved without loosing performance apart from this plane. This has been proved in a simple listening condition (Experiment A) in terms of the single attribute source width. A second subjective pilot test for a complex listening scenario (Experiment B) revealed most prominent changes for the attributes spatial distinctiveness and clarity when comparing the mixed-order approaches with a conventional 3D system. The preference for the mixed-order coding technique, where the non-truncated concept is favoured, underlines the advantage of these systems also in multi-dimensional analyses. Further investigations of single attributes in these multi-dimensional fields (complex scenarios), especially in terms of naturalness, have to be investigated in future studies.

The chosen mixed-order realisation provides an inherent smoothing, given by the Legendre functions, in the transition of horizontal to elevated sources, where a mutation of the system's characteristics from a 2D to a 3D system is present. This introduces artifacts to the system, such as coloration effects and changes in source location, due to the elevation angle dependent energy contribution. The spectral content and source locations get consequently directional dependent as well.

Common methods in order to obtain an energy normalisation are mentioned, but need further investigations, since their treatment is not as simple as in the case of conventional periphonic and surround Ambisonic systems. Optimised decoding techniques, especially *max* r_E, could further improve the system's performance. Some suggestions about the transition frequency between basic decoding and level normalised or energy optimised decoding, have been given, but are as well left for future studies.

9 Appendix

Figure 36: Photo of the Spacelab at the facilities of DTU.

The results for the non-ideal and asymmetric 29LS system are shown in Figures 37 to 38. Similar observations as to the 30LS array can be made, but errors increase dramatically, especially for $\delta_{E_{err}}$ (up to $-50°$) below an elevation angle of about $-40°$ as expected due to the missing loudspeaker at the bottom of the array.

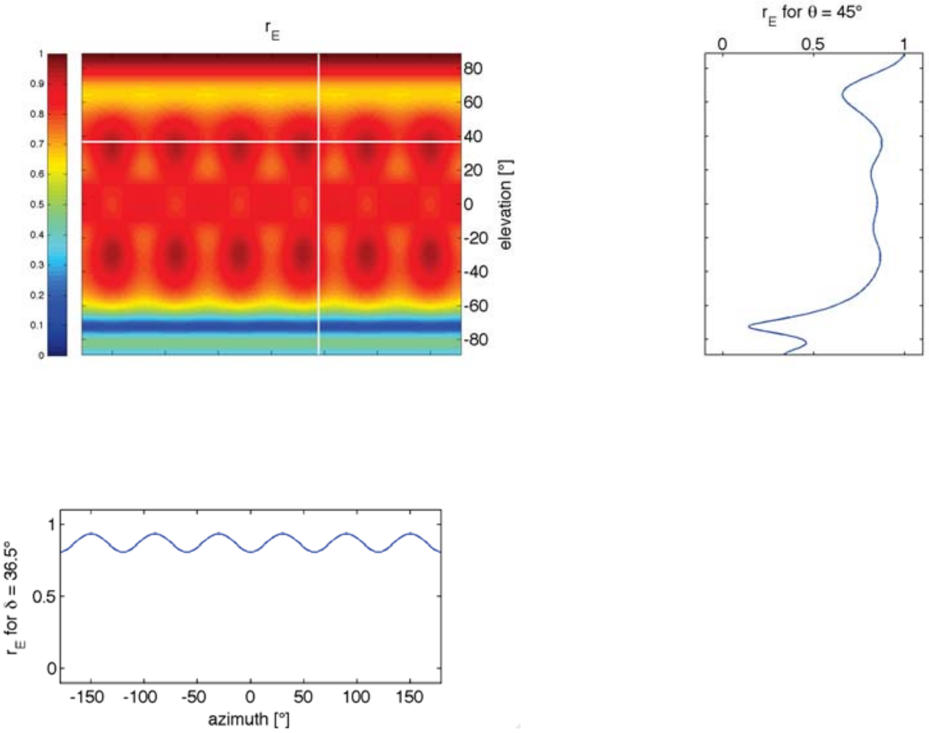


Figure 37: Energy vector magnitude for $M = 4$ for the 29 LS array.

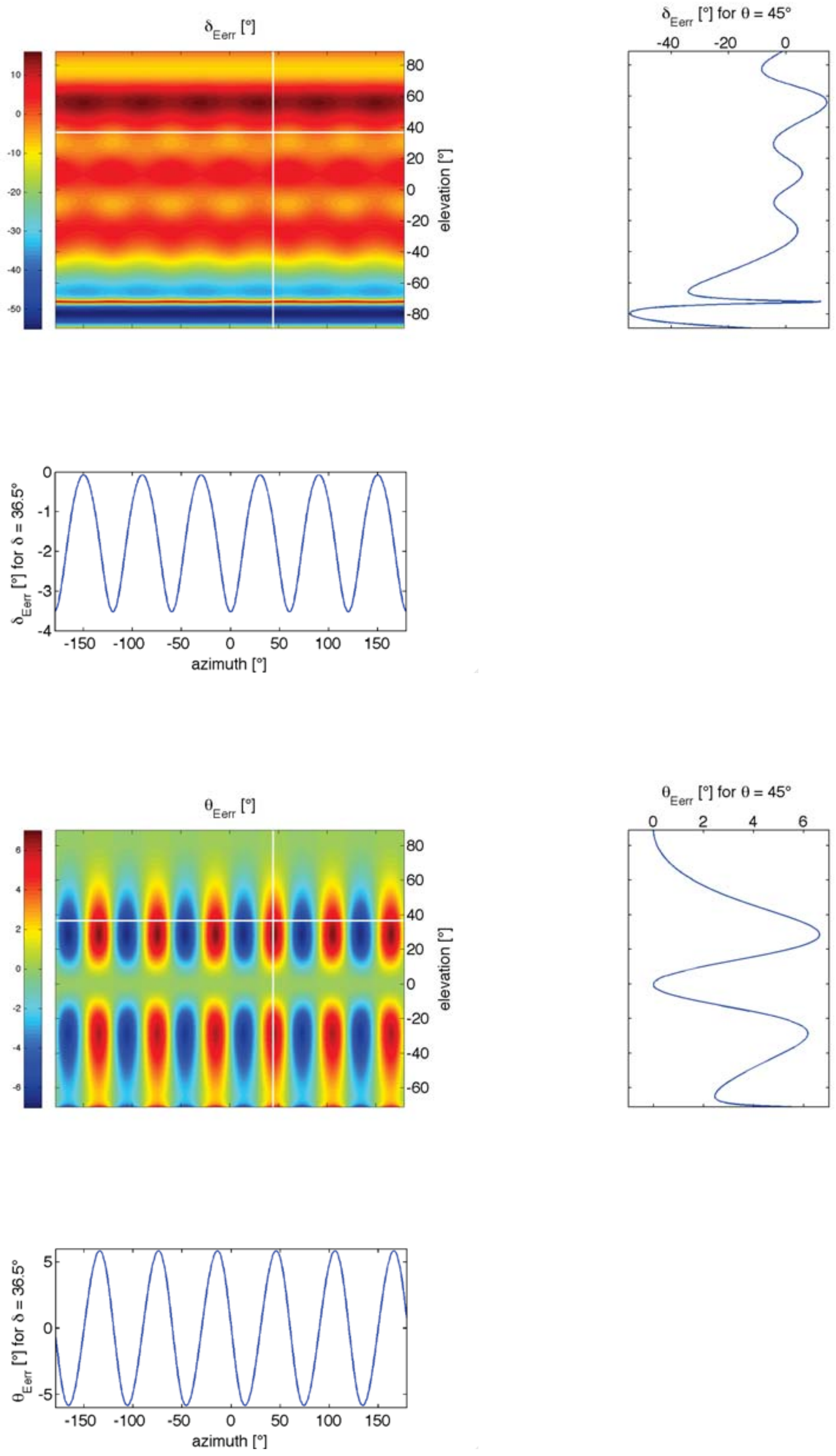


Figure 38: Elevation and azimuth error of the energy vector for $M = 4$ for the 29 LS array.

Investigations of the velocity vector are shown in Figure 39. Only plots in respect to δ are shown, since there are no mentionable changes in respect to the azimuth angle. Considering the magnitude r_V it is unity, but has a peak of 17,6 at $\delta = -67°$. At the same elevation angle also $\delta_{E_{err}}$ takes extremely large values. The reason is that the loudspeaker gains cancel each other (indicating destructive interference), so that the absolute value of W_V has a global minimum at this angle, resulting into $r_V >> 1$ as mentioned in section 3.4. This can be interpreted as a failure of the system for the specified source direction. It is hard to predict exactly the perception of localisation for such a value. Note also that the total energy W_E of the system, here presented in dB, is low at this angle but does not have a corresponding minimum.

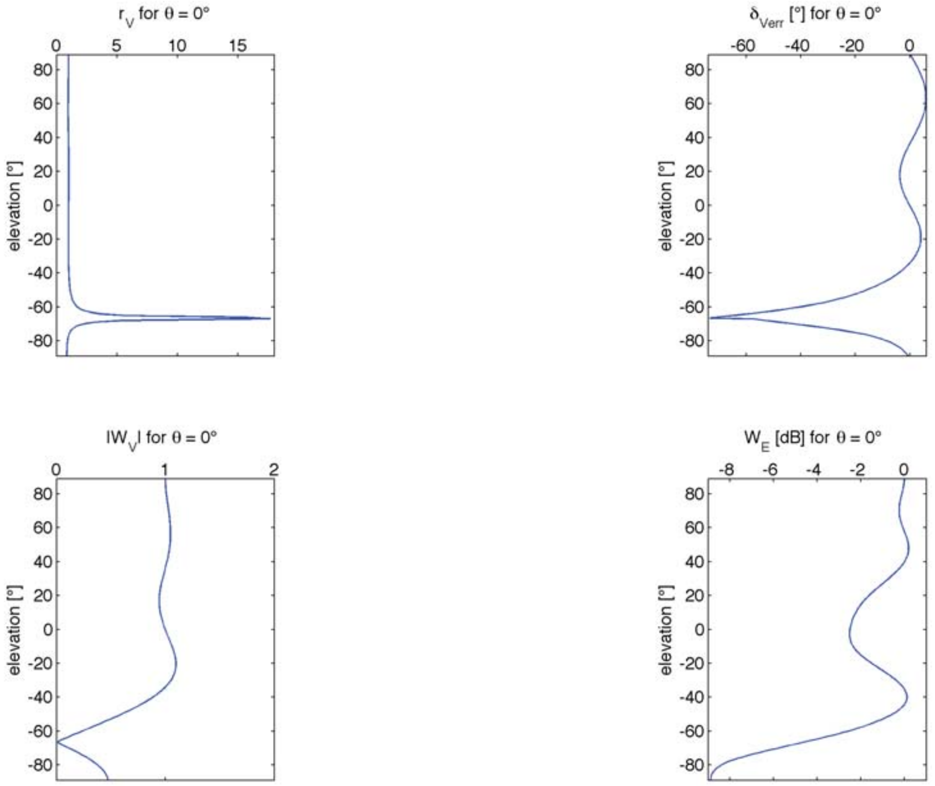
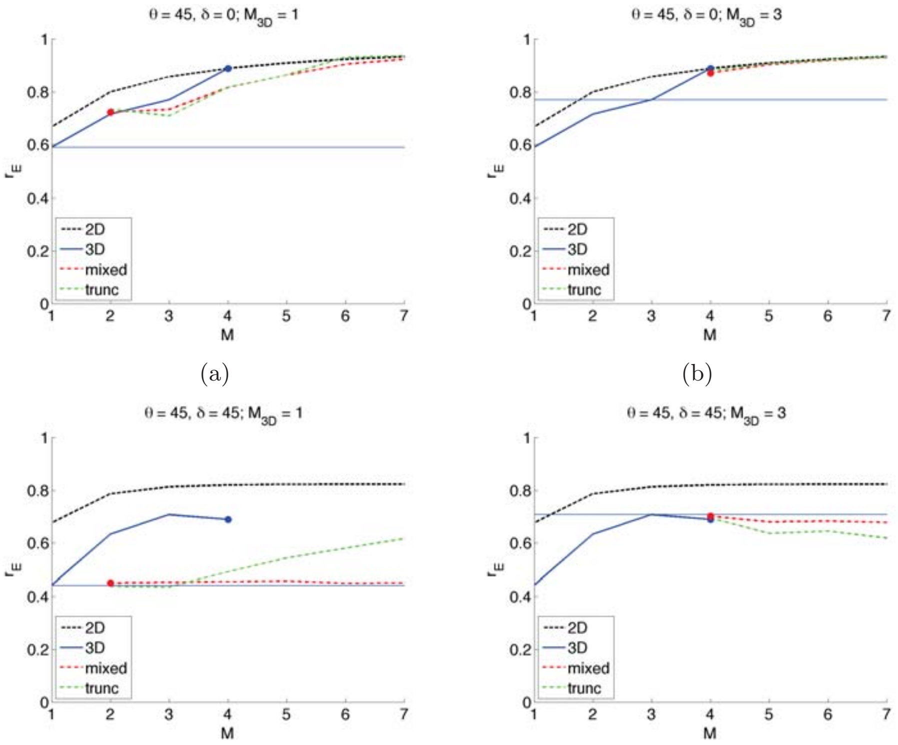


Figure 39: Investigation of the velocity vector for $M = 4$ for the 29 LS array.



(a)

(b)

(c)

(d)

Figure 40: Comparison of the different coding techniques in terms of the energy vector magnitude r_E in dependence of order M for the 29LS array for two specified source positions. The order M refers to the order of a conventional 2D (sources with $\delta \neq 0°$ are plotted by making use of eq. (29)) or 3D representation. The mixed-order implementation has the specified order M_{3D} and $M = M_{2D}$ is the varying parameter. The mixed-order case is to compare with the straight line as this indicates the improvement to a pure 3D coding of this order.

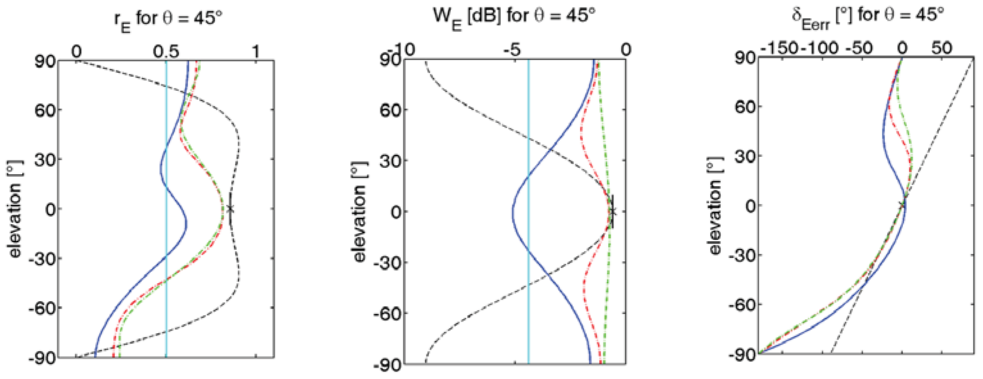
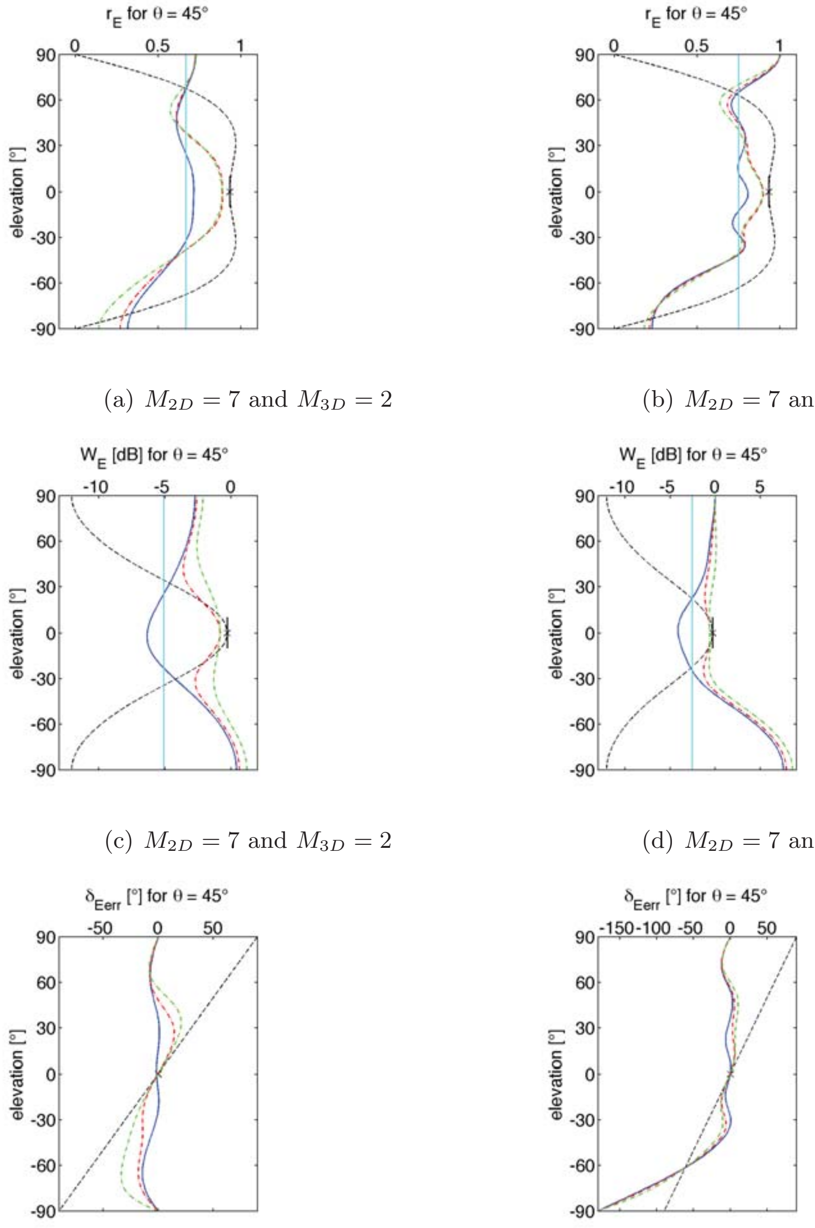


Figure 41: Investigation of the energy vector $\vec{E}$ (r_E, W_E and $\delta_{E_{err}}$) for the 11LS array in case of the different coding techniques for $M_{2D} = 3$ and $M_{3D} = 1$. 2D case: black cross ($\delta = 0$) and dashed-line ($\delta \neq 0°$ acc. to eq. (29)); 3D case: blue solid line; mixed case: red dash-dotted line; truncated mixed case: green dash-dotted line. Estimates of r_E and W_E in case of a regular 3D (cyan solid line) and 2D (short black solid line) loudspeaker setup according to eq. (38) and (37), respectively, are shown additionally.



(a) $M_{2D} = 7$ and $M_{3D} = 2$

(b) $M_{2D} = 7$ and $M_{3D} = 3$

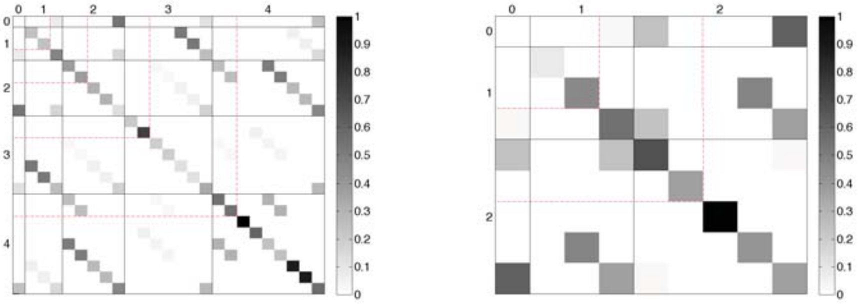
(c) $M_{2D} = 7$ and $M_{3D} = 2$

(d) $M_{2D} = 7$ and $M_{3D} = 3$

(e) $M_{2D} = 7$ and $M_{3D} = 2$

(f) $M_{2D} = 7$ and $M_{3D} = 3$

Figure 42: Investigation of the energy vector $\vec{E}$ (r_E, W_E and $\delta_{E_{err}}$) for the 29LS array in case of the different coding techniques for a constant horizontal order $M_{2D} = 7$ and two different periphonic orders $M_{3D} = 2$ and 3. 2D case: black cross ($\delta = 0$) and dashed-line ($\delta \neq 0°$ acc. to eq. (29)); 3D case: blue solid line; mixed case: red dash-dotted line; truncated mixed case: green dash-dotted line. Estimates of r_E and W_E in case of a regular 3D (cyan solid line) and 2D (short black solid line) loudspeaker setup according to eq. (38) and (37), respectively, are shown additionally.



(a) 29LS array

(b) 11LS array

Figure 43: Orthonormality matrix for the 29LS and 11LS array.

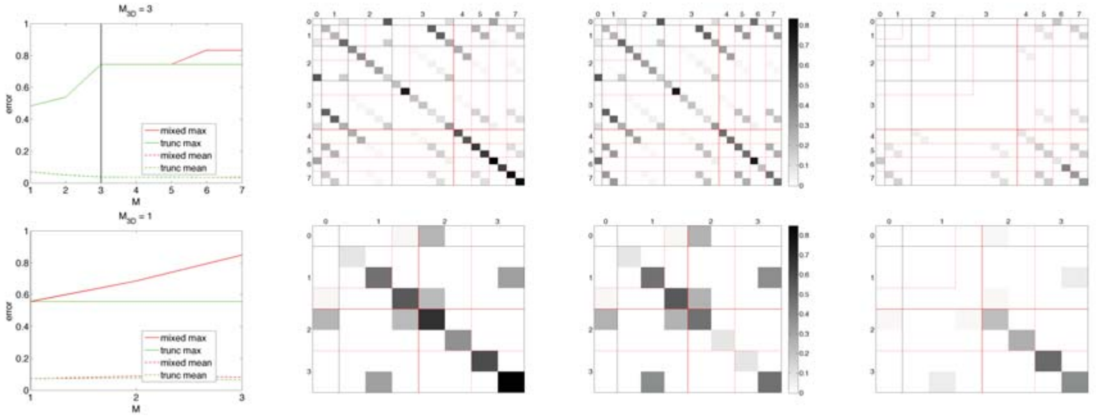
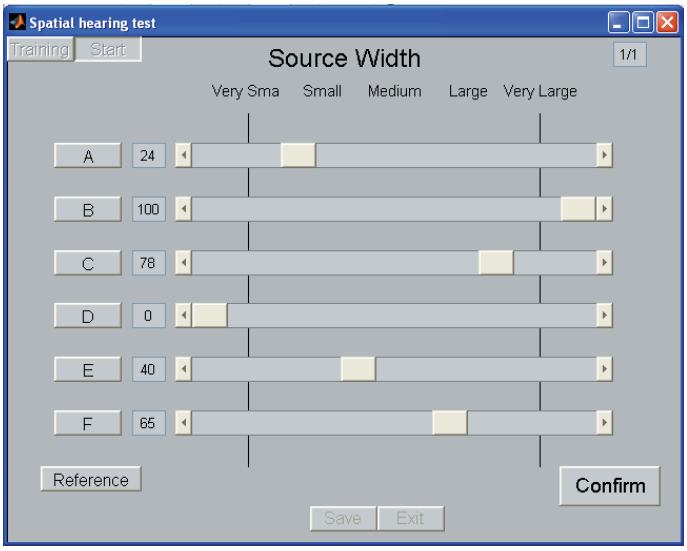


Figure 44: Orthonormality considerations for the 29LS and 11LS array. From left to right: Errors in respect to M_{2D}; Orthonormality matrix for $M_{2D} = 7$, $M_{3D} = 3$: Mixed-order case, truncated mixed-order case and absolute difference between both.

Listening Experiment

In the following experiment you are going to evaluate the attribute ‘*source width*’ of several sound presentations.

You rate each sound on a scale from 0 to 100, where the bottom of the scale refers to a ‘*very small*’ source and the top of the scale to a ‘*very large*’ source. The interface for the experimental procedure is shown in the figure below.



A 24

B 100

C 78

D 0

E 40

F 65

You can listen to a ‘*Sound*’ by pressing one of the buttons indicated with capital letters. One of these sounds will be a ‘*Reference*’ sound (= very small), which can in addition be played with an extra button indicated as such. You can listen to the sounds as often as you want and in random order. Feel encouraged to use the entire range of the scale and identify the reference by rating one of the stimuli as 0. When you have made your decision, ‘*Confirm*’ your evaluation.

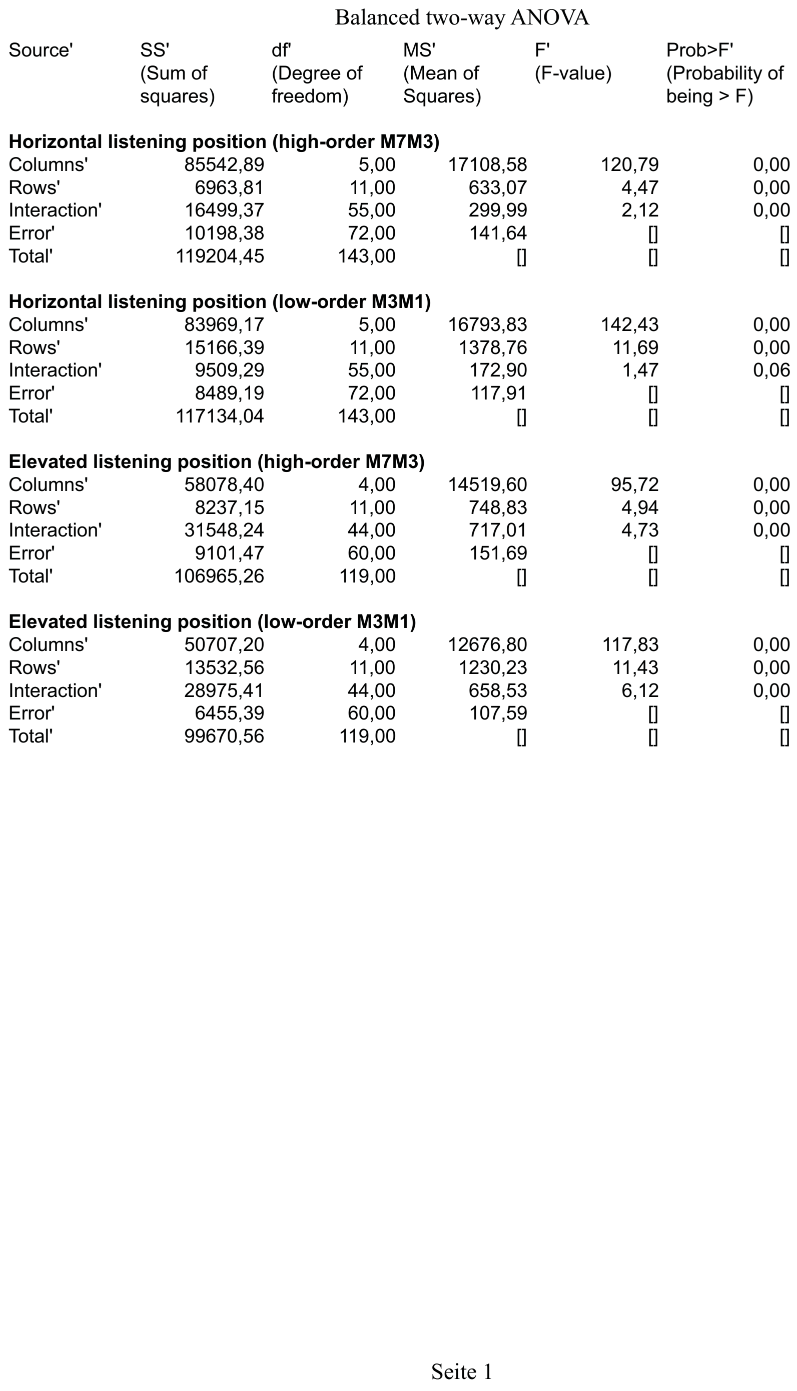
There will be a short preceding training session in order to make you familiar with the interface, the task and the stimuli. The total procedure lasts around 1 hour and you are free to take a break at any point of the experiment.

Please make sure that you keep your head in an upright position by watching to the front when evaluating a sound!

Thank you a lot for your participation and enjoy the experiment.

Johannes Käsbach

Figure 45: Instructions for Experiment A.

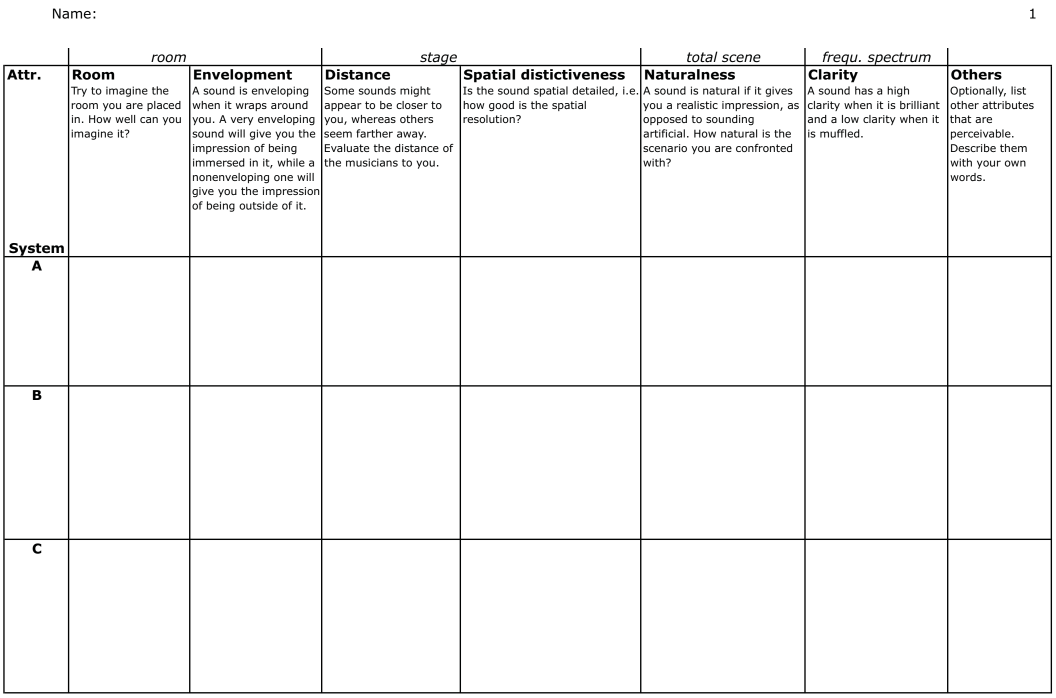


Balanced two-way ANOVA

Source'	SS' (Sum of squares)	df' (Degree of freedom)	MS' (Mean of Squares)	F' (F-value)	Prob>F' (Probability of being > F)
Horizontal listening position (high-order M7M3)					
Columns'	85542,89	5,00	17108,58	120,79	0,00
Rows'	6963,81	11,00	633,07	4,47	0,00
Interaction'	16499,37	55,00	299,99	2,12	0,00
Error'	10198,38	72,00	141,64	[]	[]
Total'	119204,45	143,00	[]	[]	[]
Horizontal listening position (low-order M3M1)					
Columns'	83969,17	5,00	16793,83	142,43	0,00
Rows'	15166,39	11,00	1378,76	11,69	0,00
Interaction'	9509,29	55,00	172,90	1,47	0,06
Error'	8489,19	72,00	117,91	[]	[]
Total'	117134,04	143,00	[]	[]	[]
Elevated listening position (high-order M7M3)					
Columns'	58078,40	4,00	14519,60	95,72	0,00
Rows'	8237,15	11,00	748,83	4,94	0,00
Interaction'	31548,24	44,00	717,01	4,73	0,00
Error'	9101,47	60,00	151,69	[]	[]
Total'	106965,26	119,00	[]	[]	[]
Elevated listening position (low-order M3M1)					
Columns'	50707,20	4,00	12676,80	117,83	0,00
Rows'	13532,56	11,00	1230,23	11,43	0,00
Interaction'	28975,41	44,00	658,53	6,12	0,00
Error'	6455,39	60,00	107,59	[]	[]
Total'	99670,56	119,00	[]	[]	[]

Seite 1

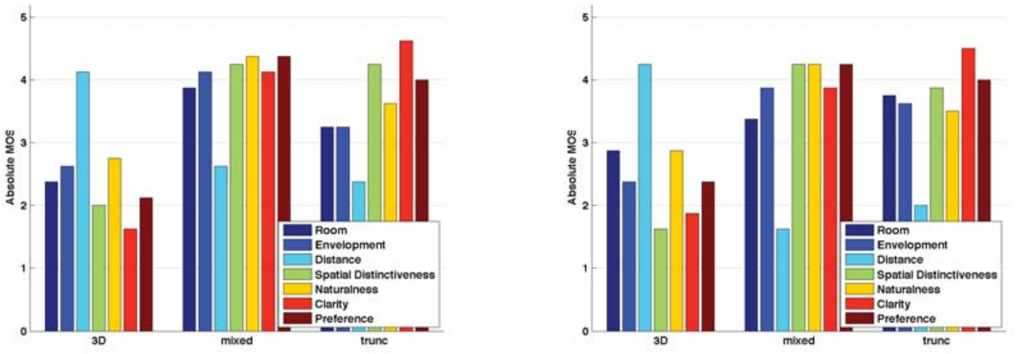
Figure 46: ANOVA Tables for the training and the experimental conditions.



Name: 1

	room		*stage*		*total scene*	*frequ. spectrum*	
Attr. **System**	**Room** Try to imagine the room you are placed in. How well can you imagine it?	**Envelopment** A sound is enveloping when it wraps around you. A very enveloping sound will give you the impression of being immersed in it, while a nonenveloping one will give you the impression of being outside of it.	**Distance** Some sounds might appear to be closer to you, whereas others seem farther away. Evaluate the distance of the musicians to you.	**Spatial distictiveness** Is the sound spatial detailed, i.e. how good is the spatial resolution?	**Naturalness** A sound is natural if it gives you a realistic impression, as opposed to sounding artificial. How natural is the scenario you are confronted with?	**Clarity** A sound has a high clarity when it is brilliant and a low clarity when it is muffled.	**Others** Optionally, list other attributes that are perceivable. Describe them with your own words.
A							
B							
C							

Figure 47: Instructions for Experiment B.



(a) $M_{2D} = 7$ and $M_{3D} = 3$

(b) $M_{2D} = 3$ and $M_{3D} = 1$

Figure 48: Averaged absolute MOS values for 7 attributes in a complex listening scenario for two mixed-order systems (a) high-order and (b) low-order referenced to a pure 3D coding of the specified order M_{3D} (Experiment B).

List of Figures

1 Head-related spherical coordinate system [1]. 12

2 Localisation blur in the horizontal plane for pulsed white noise. 13

3 Localization blur in the median plane for continuous speech [1]. 14

4 Narrowband signals. 15

5 ITD thresholds as a function of signal duration. 16

6 Spherical coordinate system with the three elementary rotation degrees [6]. 17

7 3D spherical harmonics for $m = 0$ to $m = 3$. 21

8 2D circular harmonics for $m = 0$ to $m = 3$. 27

9 Illustration of the loudspeaker arrays. 29

10 Directivity plots for $M = 1$ to $M = 5$. 30

11 Equivalent panning laws for 2D and 3D in polar and linear representation for $M = 1$ to $M = 5$ (normalised to their maximum value). 31

12 Monochromatic soundfield plots for 3D and 2D reproduction systems. . . . 33

13 Energy vector magnitude r_E for $M = 4$ for the 92 LS array. 35

14 Energy vector magnitude for $M = 4$ for the 30 LS array and total energy of the system. 36

15 Elevation error of the energy vector for $M = 4$ for the 30 LS array. 37

16 Azimuth error of the energy vector for $M = 4$ for the 30 LS array. 38

17 Orthonormality matrix for the 92LS and 30LS array. 39

18 Discretisation of continuous associated Legendre functions by the 30LS array. 40

19 Truncated spherical harmonics in comparison with the non-truncated ones. 44

20 Directivity plots for a low- and a high mixed-order system (non-truncated and truncated) in comparison with a pure 3D system illustrated in the xz plane. 47

21 Directivity plots for the mixed-order coding techniques in comparison with pure 2D and 3D coding illustrated in the xy plane. 48

22 Comparison of the different coding techniques in terms of the energy vector magnitude r_E in dependence of order M for the 92LS array for two specified source positions. 50

23 Investigation of the energy vector $\vec{E}$ for the 92LS array in case of the different coding techniques. 51

24 Comparison of the different coding techniques in terms of the energy vector magnitude r_E in dependence of order M for the 30LS array for two specified source positions. 52

25 Investigation of the energy vector $\vec{E}$ for the 30LS array in case of the different coding techniques. 53

26 Orthonormality considerations for the 30LS array in a mixed-order application. 55

27 Comparison of monochromatic soundfield plots for the different coding techniques applied to the 92LS and 30LS array. 55

28 Simulated power spectrum for BRIRs of a B&K dummy head applying different coding techniques to the 29LS (Spacelab) and 11LS array for two source positions. 58

29 Possible source expansions from a reference position. 60

30 Graphical User Interface (GUI) for the experimental procedure. 61

31 Photo of Experimental setup (Experiment A) 63

32 Illustration of the loudspeaker base angle φ_0 [17]. 65

33 Averaged experimental results for Experiment A. 66

34 Averaged relative MOS values (Experiment B). 72

35 Illustration of determining low and highpass filters for a mixed-order system. 76

36 Photo of the Spacelab at the facilities of DTU. 82

37 Energy vector magnitude for $M = 4$ for the 29 LS array. 82

38 Elevation and azimuth error of the energy vector for $M = 4$ for the 29 LS array. 83

39 Investigation of the velocity vector for $M = 4$ for the 29 LS array. 84

40 Comparison of the different coding techniques in terms of the energy vector magnitude r_E in dependence of order M for the 29LS array for two specified source positions. 85

41 Investigation of the energy vector $\vec{E}$ for the 11LS array in case of the different coding techniques. 85

42 Investigation of the energy vector $\vec{E}$ for the 29LS array in case of the different coding techniques. 86

43 Orthonormality matrix for the 29LS and 11LS array. 87

44 Orthonormality considerations for the 29LS and 11LS array. 87

45 Instructions for Experiment A. 88
46 ANOVA Tables for the training and the experimental conditions. 89
47 Instructions for Experiment B. 90
48 Averaged absolute MOS values (Experiment B). 90

List of Tables

1 Frequency limits f_{lim} depending on the order M 75

List of Abbreviations

ANOVA	Analysis of variance
B&K	Brühl & Kjaer
BRIR	Binaural room impulse response
DKDM	Royal Danish Academy of Music
DTU	Technical University of Denmark
HOA	Higher Order Ambisonics
HRTF	Head related transfer functions
ILD	Interaural level difference
IPD	Interaural phase difference
ITD	Interaural time difference
JND	Just noticeable difference
LoRA	Loudspeaker- based Room Auralization
MOS	Mean Opinion Score
MUSHRA	MUlti Stimulus test with Hidden Reference and Anchor
mRIR	Multi-channel room impulse response
RIR	Room impulse response
SPL	Sound pressure level [dB]
WFS	Wave Field Synthesis
VAE	Virtual Acoustic Environments
VBAP	Vector Base Amplitude Panning
VBIP	Vector Base Intensity Panning
VBP	Vector Based Panning

List of Symbols

γ	Angle between source and loudspeaker direction
δ	Elevation angle [°]
$\delta_{E_{err}}, \delta_{V_{err}}$	Elevation error [°]
δ_{pq}	Kronecker symbol
θ, φ	Azimuth angle [°]

$\theta_{E_{err}}, \theta_{V_{err}}$	Azimuth error [°]
σ	Decision element for trigonometric functions
φ_0	Loudspeaker base angle [°]
ω	Angular frequency [rad]
∇^2	Laplace operator
c	Speed of sound [m/s]
f	Frequency [Hz]
f_{lim}	Frequency limit [Hz]
f_{trans}	Transition frequency [Hz]
g	Gain
g_m	Order-dependent gain
h_{low}, h_{high}	Low- and highpass filter
j	Imaginary unit
$j_m(kr)$	Spherical Bessel functions
k	Wavenumber [1/m]
m	Degree
m_T	Truncated degree
n	Order, samples
p	Sound pressure [Pa]
r	Radius [m]
r_E	Magnitude of energy vector
r_V	Magnitude of velocity vector
s_{ls}	Ambisonic loudspeaker gain
$\vec{s}_{ls_{norm}}$	Power-normalised loudspeaker gains
s_{src}	Source signal
$\vec{s}_{VBIP}$	Gains for VBIP
t_f	Threshold
$\vec{u}$	Unity vector
x, y, z	Cartesian coordinates
B_{mn}^{σ}	Ambisonic channels
C	Re-encoding matrix

D	Decoding matrix
D	Energy-normalised decoding matrix
$\vec{E}$	Energy vector
$G(\gamma)$	Equivalent panning functions
H_0	Null hypothesis
H_A	Alternative hypothesis
I_K	Identity matrix
$J_m(kr)$	Cylindrical Bessel functions
K	Total number of Ambisonic components
L	Number of loudspeakers
M	Ambisonic order
M_{2D}	Horizontal Ambisonic order
M_{3D}	Periphonic Ambisonic order
$N2D/N3D$	Full normalisation 2D/3D normalisation
N_{mn}	Schmidt semi-normalisation factor
$P_{mn}(\sin\delta)$	Associated Legendre functions
$P_m(\cos\gamma)$	Unassociated Legendre functions
U	Orthonormality matrix
$\vec{V}$	Velocity vector
W_E	Total energy, sum of squared loudspeaker gains
W_V	Sum of loudspeaker gains
$Y_{mm}^{\sigma(2D)}(\theta,\delta)$	Circular harmonic functions
$Y_{mn}^{\sigma(3D)}(\theta,\delta)$	Spherical harmonic functions

References

[1] Blauert, Jens. ”Spatial Hearing: The psychophysics of human sound localization”. Revised edition, MIT Press, 1997.

[2] Choisel S., Wickelmaier F. ”Extraction of Auditory Features and Elicitation of Attributes for the Assessment of Multichannel Reproduced Sound”. Journal of the Audio Engineering Society, Vol. 54, No. 9, September 2006.

[3] Choisel S., Wickelmaier F. ”Evaluation of multichannel reproduced sound: Scaling auditory attributes underlying listener preference”. Journal of the Acoustical Society of America 121(1), January 2007.

[4] Daniel J., Rault J.B. and Polack J.D.. ”Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions”. Presented at the 105th convention of the Audio Engineering Society, San Francisco, California, September 1998.

[5] Daniel, Jérôme. ”Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia”. Ph.D. thesis, Université Paris 6, 2000.

[6] Daniel J., Nicol R. and Moreau S.. ”Further investigations of High Order Ambisonics and Wavefield Synthesis for holophonic sound imaging”. Presented at the 114th convention of the Audio Engineering Society, Amsterdam, Netherlands, March 2003.

[7] Favrot, Sylvain . ”A loudspeaker-based room auralization system for auditory research”. Ph.D. thesis, CAHR, Technical University of Denmark (DTU), 2010.

[8] Gerzon, Michael A. . ”General Metatheory of Auditory Localisation”. Presented at the 92nd Convention of the Audio Engineering Society, Vienna, Austria, March 1992.

[9] le Goff, Nicolas. ”Processing interaural differences in lateralization and binaural signal detection”. Technische Universiteit Eindhoven, 2009.

[10] Hollerweger, Florian . ”Periphonic Sound Spatialization in Multi-User Virtual Environments”. Master thesis, Institute of Electronic Music and Acoustics (IEM), Center for Research in Electronic Art Technology (CREATE), corrected version, March 14th 2006.

[11] Jacobsen, Finn (DTU) and Juhl, Peter (SDU). ”Radiation of sound”. Acoustic Technology, Technical University of Denmark (DTU) and Institute of Sensors Signals and Electrotechnic, University of Southern Denmark (SDU). May 2008.

[12] Johnson, Richard. ”Miller and Freund’s Probabiity and Statistics for Engineers”. Seventh Edition. Pearson Education, 2005.

[13] Moore, Brian. ”An Introduction to the Psychology of Hearing”. Academic Press, 1997.

[14] Moreau S., Daniel J. and Bertet S.. ”3D Sound Field Recording with Higher Order Ambisonics – Objective Measurements and Validation of Spherical Microphone”. Presented at the 120th convention of the Audio Engineering Society, Paris, France, May 2006.

[15] Pernaux J., Boussard P. (1) and Jot J. (2). ”Virtual Sound Source Positioning and Mixing in 5.1 Implementation on the Real-Time System Genesis”. (1) Steria/Digilog S.A, Aix-en-Provence France and (2) IRCAM, Paris France, 1998.

[16] Plack, Christopher J. ”The Sense of Hearing”. Lawrence Erlbaum Associates, USA, 2005.

[17] Pullki V. and Lokki T. ”Creating Auditory Displays with Multiple Loudspeakers Using VBAP: A Case Study with DIVA Project”. Presented at ICAD’98, the International Community for Auditory Display, University of Glasgow, 1998.

[18] Recommendation ITU-R BS.1534-1. ”Method for the subjective assessment of intermediate quality level of coding systems”. The ITU Radiocommunication Assembly, 2001-2003.

[19] Solvang, Audun. ”Spectral Impairment for Two-Dimensional Higher Order Ambisonics”. Audio Engineering Society, Vol. 56, No.4, April 2008.

[20] Travis, Chris . ”A new mixed-order scheme for Ambisonic signals”. Presented at the Ambisonic Symposium 2009, Graz, Austria, June 2009.

[21] Tukey, J. W.. ”Exploratory Data Analysis”. MA: Addison-Wesley, 1977.

[22] Weller, Tobias. ”Application of a circular hard-sphere microphone array for high-order Ambisonics auralization”. Master thesis, CAHR, Technical University of Denmark (DTU), October 2010.

[23] Williams, Earl G.. ”Fourier Acoustics - Sound Radiation and Nearfield Acoustical Holography”. Naval Research Laboratory Washington, D.C.. Academic Press. 1999.

[24] Zielinski S., Hardisty P. and Hummersone C., Rumsey F..”Potential Biases in MUSHRA Listening Tests”. Presented at the 123rd convention of the Audio Engineering society, New York, USA, October 2007.

[25] Zwicker E. and Fastl H.. ”Psychoacoustics : facts and models”. Springer, 2nd Updated Edition, 1999.

www.elektro.dtu.dk

Department of Electrical Engineering

Hearing Systems
Technical University of Denmark
Ørsteds Plads
Building 348
DK-2800 Kgs. Lyngby
Denmark
Tel: (+45) 45 25 38 00
Fax: (+45) 45 93 16 34
Email: info@elektro.dtu.dk