Acta Acustica 2020, **4**, 19

https://doi.org/10.1051/aacus/2020014

Available online at:
https://acta-acustica.edpsciences.org

TECHNICAL & APPLIED ARTICLE OPEN ACCESS

SCaLAr – A surrounding spherical cap loudspeaker array for flexible generation and evaluation of virtual acoustic environments

Florian Pausch[1,*], Gottfried Behler[2], and Janina Fels[1]

[1] Teaching and Research Area of Medical Acoustics, Institute of Technical Acoustics, RWTH Aachen University, Kopernikusstraße 5, 52074 Aachen, Germany
[2] Chair and Institute of Technical Acoustics, RWTH Aachen University, Kopernikusstraße 5, 52074 Aachen, Germany

Received 13 May 2020, Accepted 11 August 2020

Abstract – *Introduction*: Surrounding spherical loudspeaker arrays facilitate the application of various spatial audio reproduction methods and can be used for a broad range of acoustic measurements and perceptual evaluations.
Methods: Installed in an anechoic chamber, the design and implementation of such an array of 68 coaxial loudspeakers, sampling a spherical cap with a radius of 1.35 m on an equal-area grid, is presented. A network-based audio backbone enables low-latency signal transmission with low-noise amplifiers providing a high signal-to-noise ratio. To address batch-to-batch variations, the loudspeaker transfer functions were equalised by individually designed 512-taps finite impulse response filters. Time delays and corresponding level adjustments further helped to minimise radial mounting imperfections.
Results: The equalised loudspeaker transfer functions measured under ideal conditions and when mounted, their directivity patterns, and in-situ background noise levels satisfy key criteria towards applicability. Advantages and shortcomings of the selected decoders for panning-based techniques, as well as the influence of loudspeaker positioning errors, are analysed in terms of simulated performance metrics. An evaluation of the achievable channel separation allows deriving recommendations of feasible subset layouts for loudspeaker-based binaural reproduction.
Conclusion: The combination of electroacoustic properties, simulated sound field synthesis performance and measured channel separation classifies the system as suitable for its target applications.

Keywords: Spherical loudspeaker array, Array processing, Spatial audio reproduction, Virtual acoustic environment

1 Introduction

Research on loudspeaker-based spatial audio reproduction methods requires carefully designed surrounding spherical loudspeaker arrays to enable practicality of spatialisation methods, aiming at plausible or even authentic perception of simulated virtual acoustic environments (VAEs) [1–3]. Commonly used methods either rely on binaural technology for loudspeaker-based binaural playback in combination with acoustic crosstalk cancellation (CTC) filters [4–7], or on panning methods such as higher-order Ambisonics (HOA) [8–13], vector base amplitude panning (VBAP) and multiple-direction amplitude panning (MDAP) [13–16]. Wave field synthesis shall be mentioned as a further sound field synthesis approach [17–19], although the use of dense linear arrays would be preferred. Since an anechoic chamber is not always available or convenient for permanent array installations, conventional rooms, sometimes acoustically optimised, are regarded as acceptable for the targeted application [20–22]. Although additional reflections in such rooms can be helpful to conceal known perceptual shortcomings of the respective reproduction method [13, 19, 23], particularly objective evaluations require that the array is preferably installed in an anechoic room (e.g., [24, 25]). An unobtrusive mounting construction further helps to decrease sound field distortions that could lead to biased reproduction error metrics and perceptual test results.

The flexible reproduction possibilities facilitate research in a variety of areas. Typical applications include objective evaluations of spatialisation methods [26] and spherical microphone arrays [27, 28], measurements of individual head-related transfer functions (HRTFs), as well as research on hearing aids [29–31]. An optimal reproduction environment also allows to compare perceptual metrics after systematic modifications of system parameters or rendering components. Of particular interest is, for example, how the source localisation differs between virtual sound

*Corresponding author: florian.pausch@akustik.rwth-aachen.de

sources (VSSs) based on individualised HRTFs [32, 33] and real sound sources, usually represented by discrete loudspeakers [34]. In order to continuously approach ecological validity, objective evaluations of simulation methods for VAEs with room acoustics [35] should be supplemented by perceptual evaluations to substantiate their validity [36, 37]. Combined with physiological measures, non-intrusive, controlled experimental environments further represent a tool for assessing environmental noise effects [38]. However, applications are not restricted to the auditory domain but can be extended to multimodal experiments using head-mounted displays for the exploration of auditory perception and cognition [39, 40].

A number of international research groups have already designed and implemented surrounding loudspeaker arrays (e.g., [16, 20–22, 24, 25, 30]). As there is often no room for comprehensive implementation details, this report is intended to provide a compact but more accessible template for similar future projects. We present the design of a surrounding spherical cap loudspeaker array, including a description of the reproduction environment, the array mounting construction, the electroacoustic features of the loudspeakers with custom-made cabinets and the signal network implementation. Based on commonly used performance metrics, the sampling layout is analysed for its suitability using various spatial audio reproduction methods, accounting for physical imperfections.

2 Methods

2.1 Reproduction environment

The loudspeaker array was mounted in the anechoic chamber of the Institute of Technical Acoustics, RWTH Aachen University, with the dimensions 9.2 m × 6.2 m × 5 m (length × width × height), resulting in a room volume of about 285 m^3, see Figure 1. All room surfaces are covered with 0.7 m long glass fibre wedges, determining a lower cut-off frequency of approximately 200 Hz [41]. The wedges and the ceiling construction as well as a net made of steel cables 0.2 m above the floor wedges reduce the effectively usable room dimensions to 7.8 m × 4.9 m × 2.8 m (length × width × height).

2.2 Sampling layout

Since the floor net construction in the anechoic chamber cannot be locally detached, loudspeaker positions at zenith angles close to the south pole are unfeasible, limiting the design to spherical cap layouts when the radius should still be adequately large. Primarily for reasons of sampling efficiency and because of the requirement for a ring-based layout, an equal-area sampling [42] with 75 points and a spherical harmonic (SH) order of N = 7 for a nominal array radius of 1.35 m was generated in MATLAB (MathWorks, Natick, Massachusetts, United States) [43]. The removal of physically impractical loudspeakers at the south pole and on the ring above led to a total number of 68 loudspeakers, populating zenith angles up to about 134°. Serving as a visual anchor, one loudspeaker in the horizontal plane at 0° azimuth represents the default viewing direction of the listener, see Figures 1 and 2.

Figure 1. Physical implementation of the surrounding spherical cap 68-loudspeaker array.

2.3 Motion tracking

Binaural reproduction over loudspeakers in particular requires that the auralisation system knows the listener's current real-world position and orientation, relative to the real-world loudspeaker positions [44, 45]. For behavioural evaluations, the motion tracking data collected from participants when solving experimental tasks may provide crucial information about possible applied listening strategies, thus enabling complementary data analysis [46, 47]. We therefore integrated an optical motion tracking system (OptiTrack, NaturalPoint Inc., Corvallis, Oregon, USA), consisting of four infrared cameras (Flex 13), that transmits the data stream via tracking hub (OptiHub 2) to a desktop computer (Intel® Core i7-8700). Synchronised data logging is possible using the dedicated software (Motive) together with the NatNet software development kit and MATLAB [43]. With an imager resolution of 1280 × 1024 pixels (1.3 MP resolution, 56° × 46° field of view), a maximal native frame rate of 120 Hz and a specified latency of 8.3 ms, the tracking system can resolve objects up to 9 m away at a 3-dimensional accuracy of ±0.2 mm [48].

2.4 Physical implementation

2.4.1 Mounting construction

In order to keep the overall weight and the influence on the reproduced sound field low, the frame construction consists of carbon fibre and aluminium poles with 10 mm

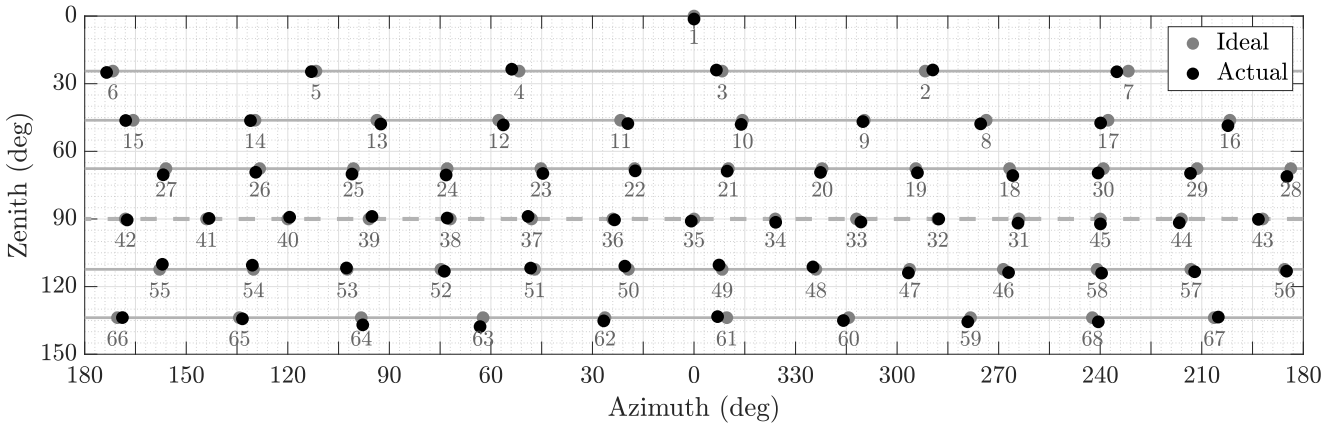


Figure 2. Ideal and actual sampling layouts displayed as grey and black dots, respectively, with loudspeaker channel numbers. Angle definitions are based on a spherical head-related coordinate system. Azimuth angles increase counter-clockwise and are represented by $\varphi \in [0°, 360°)$, and zenith angles are represented by $\vartheta \in [0°, 180°]$. The solid and dashed grey horizontal lines show the loudspeaker rings and the horizontal plane at the listener's nominal ear height, respectively, with loudspeaker 35 indicating the default viewing direction.

diameter, see Figure 1. The loudspeakers in the upper hemisphere were suspended at their centre of gravity on brackets that allow height adjustment, tilting and rotation, and on vertical poles, which in turn were attached to horizontal ceiling poles. To enable correct loudspeaker positioning, these movable horizontal ceiling poles were arranged as spokes between the two concentric ceiling rings with radii of 1.5 m and 0.5 m. For the loudspeakers in the lower hemisphere, including the horizontal plane, we constructed three pentadecagonal rings with decreasing incircle radii of 1.44 m, 1.33 m and 1.07 m that were vertically connected by relocatable aluminium poles. This lower hemispherical ring construction was suspended by 15 vertical aluminium poles from the outer ceiling ring and additionally stabilised with rods to avoid lateral movements. Attaching the loudspeakers to the horizontal carbon rods with sliding and rotating clamping elements allowed the radius and angular orientation to be individually adjusted in the area of the respective centre of gravity.

The door element to enter the array, holding two loudspeakers, demanded careful design as it directly influences the internal stability of the whole mounting construction when opened and general practical usability. A hinge construction, defined end positions of the pole terminations, and additional braces ensure minimal overall mechanical impact and reproducible loudspeaker positions after each use.

2.4.2 Actual loudspeaker positions

The positioning and alignment of the loudspeakers was mainly carried out using laser measurement tools. In a first step, the centre of the north pole loudspeaker was projected onto the ground with a self-levelling cross-line laser (GLL 3-80, Bosch Professional, Gerlingen-Schillerhöhe, Germany). The cross-line laser was then rotatably attached to a compass rose in this ground centre to read the azimuth angles and align the respective loudspeakers accordingly. Additionally, a rod was mounted on the north pole loudspeaker with a bayonet lock, at the end of which a swivelling distance laser (ADM30, FLEX-Elektrowerkzeuge GmbH, Steinheim/Murr, Germany) with a vertical compass rose enabled to measure the respective elevation angles and distances of the loudspeakers. We used an electronic spirit level and the cross line laser to verify the inclinations and heights of the individual loudspeakers per ring, respectively.

For a correct configuration of the auralisation software (e.g., [46, 49, 50]) and to simulate the sound field synthesis performance given loudspeaker positioning errors, we measured the actual loudspeaker positions with the optical tracking system. For the definition of the tracking system's real-world coordinate system, the centre of the array was determined using a light aluminium replica of the original calibration square (CS-100, OptiTrack, NaturalPoint Inc., Corvallis, Oregon, USA). This calibration square was attached to a ball-and-socket joint on the vertical rod of the north pole loudspeaker and aligned in the horizontal plane and with respect to the centre of the floor using the cross line laser. Since reliable and accurate tracking in outside-in systems is limited to a certain detection volume, which develops around the centre of the array depending on camera positions and orientations, a tetrahedral rigid body consisting of four reflective markers was mounted on a 70 cm long carbon fibre rod. The tip of this rod was used to measure the loudspeaker membrane centres by translating the rigid body's pivot point accordingly.

2.4.3 Participant chair

A 600 mm × 600 mm × 30 mm (length × width × height) aluminium grillage, which is supported on the solid ground below the floor wedges by four threaded rods and braced with the floor netting by diagonally and laterally

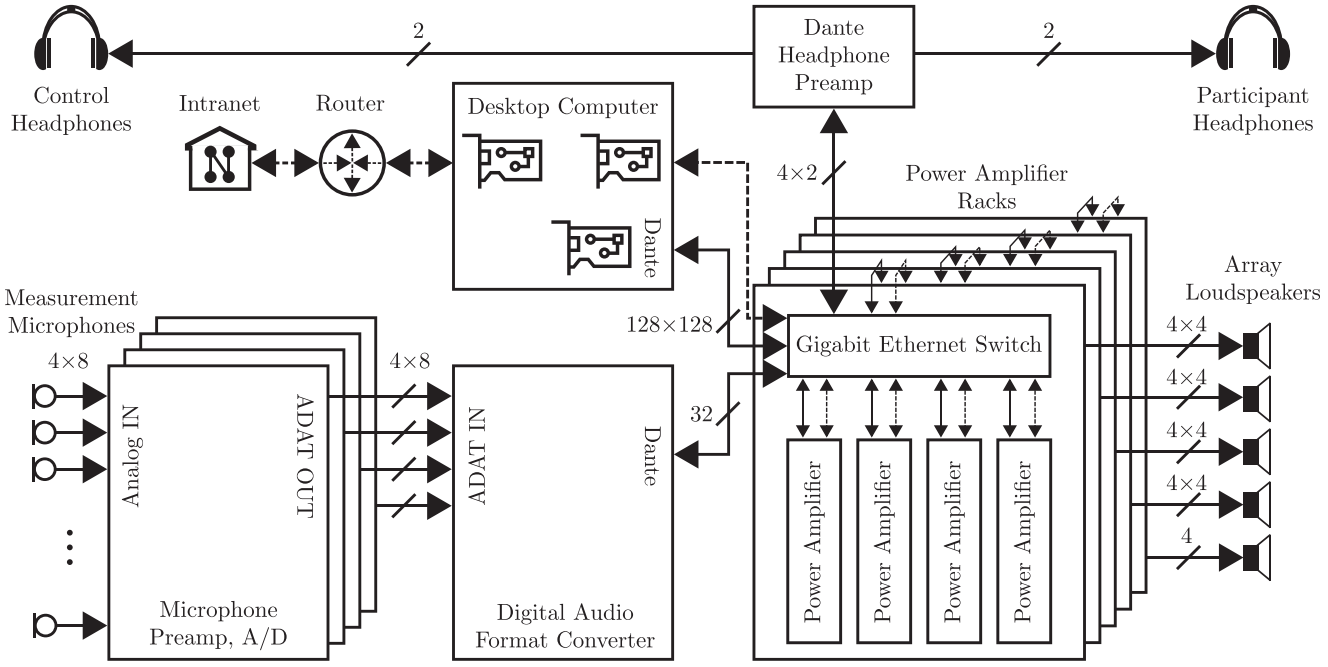


Figure 3. Functional block diagram of the implemented signal network. Audio and control/network signal paths are represented by solid and dashed lines, respectively.

crossed steel cables, allows to attach a socket for the participant chair (IS 1926, Dauphin HumanDesign Group GmbH & Co. KG, Offenhausen, Germany). Optionally, a table covered with absorbers can be mounted to accommodate wireless user input devices (e.g., mouse, keyboard, tablet). To align the participant's interaural axis with respect to the loudspeaker array's centre, the chair can be moved back and forth and adjusted in height. Depending on the experimental design (e.g., a virtual all-around search task), a rotation around the listener's longitudinal axis is made possible or prevented by an adjustment screw. If necessary, the participant's head movements can be restricted using an unobtrusive height-adjustable head rest.

2.5 Signal network

To transmit the audio and control signals a network solution was chosen, interlinking the majority of hardware devices via 16-port Gigabit Ethernet switches (DGS-1016D, Netgear, San Jose, California, USA) and cables, see Figure 3.

2.5.1 Audio and control signals

Audio signals are managed using a PCIe card (RED-NET PCIeR, High Wycombe, UK) with dedicated 128 × 128 Dante® interface and the audio network routing and configuration software Dante Controller (Audinate Pty Ltd, Surry Hills, New South Wales, Australia).

The use of a microphone preamp, A/D converter and ADAT connection protocol (RME Octamic XTC, Audio AG, Haimhausen, Germany) in combination with a digital audio format audio converter (REDNET 3, Focusrite, High Wycombe, UK) enables to capture up to 32 analogue signals for various measurement applications (e.g., measurement microphones, artificial heads, microphone arrays).

Audio playback relies on 17 4-channel power amplifiers with passive cooling, on-board digital signal processor, low-noise Pascal amplifier modules (112 dB(A) output signal-to-noise ratio) and Dante® interface (PPA 1000-4-PC DSP, Four Audio, Herzogenrath, Germany). The amplifier modules were built into custom-made aluminium racks with air vents that hold up to four devices and one network switch. The rack-mounted network switches are connected in series while we used a star-shaped concept for the connection of the amplifiers per rack. Individual device and channel configurations are accomplished via an independent control network with a dedicated PCIe network card. A 4-channel Dante® headphone amplifier (KLANG:QUELLE, KLANG:technologies GmbH, Aachen, Germany) enables playback over open dynamic headphones (HD 650, Sennheiser, Wedemark, Germany). For safety reasons, we installed two kill switches within reach of the experimenter and participant, which immediately interrupt the power supply to all power amplifiers.

2.5.2 Supervision

Visual supervision and verbal communication with participants taking part in perceptual experiments is possible via network camera with night vision (RLC-410, Reolink, Hong Kong, China) and an independent custom-made

talkback system, respectively. A pair of control headphones can be used for channel-based verification of playback signals and to listen to the same material presented to the participant or a binaural downmix [51].

2.6 Electroacoustics

All measurements in the remaining parts of this section, except those related to the in-situ finite impulse response (FIR) filter verification measurements (Sect. 2.6.2), the background noise level (BNL) (Sect. 2.6.4), and CTC performance evaluation (Sect. 2.7.4), were carried out in the hemi-anechoic chamber of the Institute of Technical Acoustics, RWTH Aachen University, with dimensions 11 m × 5.97 m × 4.5 m (length × width × height), exhibiting a lower frequency limit of about 100 Hz [41]. We used the same signal network as described in Section 2.5. Further measurement details are provided in the corresponding subsections.

2.6.1 Loudspeaker cabinet and crossover design

To favour confined spatial sampling and minimise unwanted phase effects of low- and high-frequency drivers located at different positions, we used 4-inch coaxial 2-way loudspeakers (Seas L12RE/XFC H1602-04, Moss, Norway), which were built into 4 L octagonal prismatic cabinets (sidewalls 6 mm, front and backside 12 mm thick birch multiplex), representing a bass reflex system with a cylindrical rear panel port of 100 mm length and 35 mm diameter, see Figure 4. Both cabinet and crossover design were developed based on the acoustic measurement results described below using a simulation software [52].

To measure the individual driver's impulse responses, an exponential sweep with a length of 2^{17} samples at a sampling frequency of 44.1 kHz between 20 Hz and 20 kHz was generated in MATLAB [43] and played back over an example loudspeaker. Positioned on the floor at a distance of 1.4 m, the loudspeaker was tilted towards the measurement microphone (Brüel & Kjær Type 4189 and 2669, Nærum, Denmark), which was used together with a conditioning amplifier (Brüel & Kjær Type 2610, Nærum, Denmark). Finally, the resulting impulse responses were windowed in time domain (Hann window, 1.5 ms fade-in after 0.5 ms, 5 ms fade-out after 25 ms).

2.6.2 Spectral equalisation

For the design of the FIR filters, we repeated the measurements described in Section 2.6.1 for all loudspeakers. Individual 512-taps FIR filters with highpass filter (4-th order Butterworth, 80-Hz cut-off frequency) and 1/12-octave band smoothing were created in the software filter design module (System Designer, Four Audio, Herzogenrath, Germany) and uploaded to the corresponding digital signal processor of the power amplifiers. Adding the power amplifiers' system latency of 2.54 ms, the filter target latency of 8 ms and a maximum 1-ms Dante® network latency results in an overall output latency of 11.54 ms, without taking into account the acoustic delay due to the loudspeaker distance and additional delays for radial corrections, cf. Section 3.1.

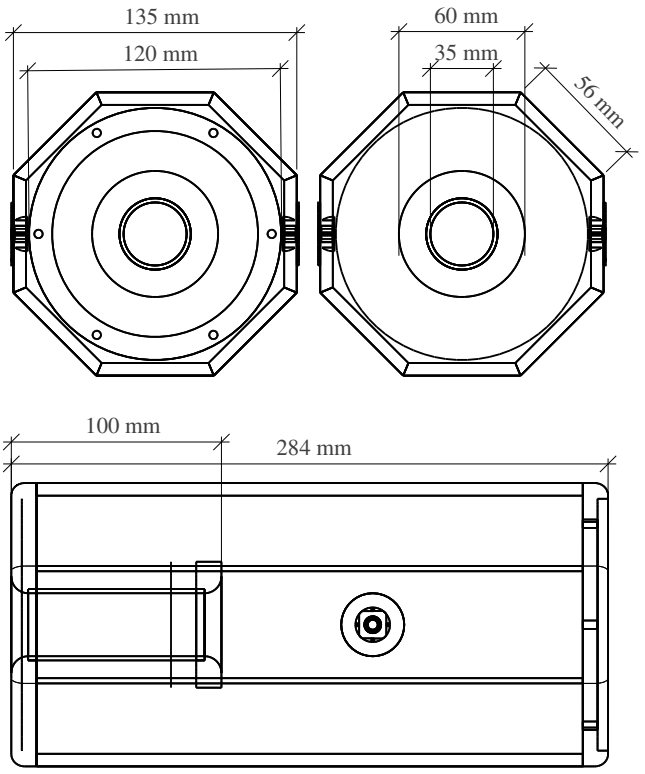


Figure 4. Front view (top left), rear view (top right) and side view (bottom) of the octagonal prismatic cabinet, holding the 4-inch coaxial 2-way loudspeaker, with cylindrical rear panel bass reflex port. The side view additionally shows the fastening screw thread for the brackets in the centre of gravity.

To investigate the acoustic influence of the physical setup, i.e., the loudspeakers and mounting construction, the measurements described in Section 2.6.1 were repeated for an example loudspeaker when installed as part of the array with active FIR filter. The same time window settings were applied to ensure comparability.

2.6.3 Directivity pattern

For this measurement, the example loudspeaker was mounted on a 1 m long pole, which was connected to a custom-made turntable. The turntable was tilted so that the loudspeaker's on-axis direction pointed towards the microphone, positioned on the floor at a distance of 5.4 m. This distance was chosen in order to keep the required inclination of the turntable in a reasonable range and to minimise unfavourable reflections due to the measurement setup as far as possible. A remote-controlled stepping motor (PD4-N, Nanotec, Feldkirchen/Munich, Germany) rotated the turntable in steps of 1° to obtain the loudspeaker's directivity pattern in the horizontal plane, assuming a rotationally symmetric behaviour. The corresponding impulse responses were spectrally smoothed using 1/6-octave band filters, time-windowed (Hann window, 0.5 ms fade-in after 0.9 ms, 5 ms fade-out after 35 ms) and cropped [43].

2.6.4 Background noise level

A low-noise measurement microphone (40HL, GRAS Sound & Vibration A/S, Holte, Denmark) was placed in the array centre and used in combination with a conditioning amplifier (Brüel & Kjær Type 2690-A, Nærum, Denmark) and an otherwise unchanged hardware setup to measure the BNL when all loudspeakers were connected to the power amplifiers. As a baseline, we first measured the BNL with power amplifiers switched off. In a second measurement, all power amplifiers were switched on to investigate the combined noise contribution of the power amplifiers and the loudspeakers during standby operation. Both measurements were performed for a duration of 60 s.

2.7 Spatial audio reproduction

2.7.1 Performance metrics

Previous work introduced the sum of squared loudspeaker gains, $\hat{E}$, as an energy measure to estimate direction-dependent loudness of panning-based spatial audio reproduction methods [53]. The magnitudes of velocity [54] and energy vectors [53], $\hat{\sigma}_{\mathrm{V}}$ and $\hat{\sigma}_{\mathrm{E}}$, respectively, are considered as additional quality indicators for the angular mapping performance [12, 55]. There is perceptual evidence that these metrics correlate with detecting the direction of plane wave incidence at low frequencies (below 700 Hz), while enabling binaural evaluation of the maximal spatial energy concentration for source localisation at higher frequencies and perceived VSS width in the intended sound field synthesis area [13, 23, 53, 54].

The selected metrics were considered to provide an overall performance evaluation of the various decoder strategies described below. Each decoder was calculated for the ideal and actual sampling layout. The resulting decoder-dependent loudspeaker gains were used to estimate $\hat{E}$, $\hat{\sigma}_{\mathrm{V}}$ and $\hat{\sigma}_{\mathrm{E}}$ for directions of incidence on an equiangular grid with a resolution of $1° \times 1°$ in azimuth and elevation, covering the entire sphere [56]. The choice of these directions of incidence should also indicate problems in decoders that are not designed for reproduction outside the spatial range covered by the loudspeaker array. To further assess the perceptual consequences of reproduction errors, the directional deviation between the respective vector direction and the target grid direction, $\hat{\epsilon}_{\mathrm{V}}$ and $\hat{\epsilon}_{\mathrm{E}}$, as well as the perceived source width α [23] were evaluated [56]. In addition to across-decoder comparisons, this strategy allows to investigate the influence of actual loudspeaker positioning errors on the simulated performance of the selected panning-based reproduction techniques and to analyse their general suitability.

2.7.2 Vector base amplitude panning

Since VBAP relies on convex hull triangulation, the available sampling layout provides an unproblematic basis for the selection of valid loudspeakers, pairs or triplets, as long as VSSs are synthesised from directions lying on the spherical cap. For directions towards the south pole, where this approach is likely to suffer from numerical instabilities and unsatisfactory perceptual results, the insertion of one or more imaginary loudspeakers [57] comes in handy by allowing to properly discard unfeasible VSS directions or apply downmixing to nearby loudspeakers [13]. To control for constant source width and reduced colouration, particularly in the case of moving VSSs, the use of auxiliary spreading sources, as done for MDAP, was recommended in previous work [15, 23].

As baseline panning variants, we evaluated the performance of both the VBAP and MDAP decoders. For VBAP, triangulations in the lowest zenith angle range were permitted, since the maximum recommended by-triangle loudspeaker opening angle of 90° is maintained in the given sampling layout [16]. As an example MDAP configuration, we used a number of 12 spreading sources, i.e., virtual auxiliary sources that are equally distributed on a concentric ring around the corresponding target direction of the VSS. This concentric ring was defined at a spreading angle of 10°, i.e., the angle relative to the target direction of the VSS [15]. The final loudspeaker gains based on the entire set of sources per target direction were calculated using the proposed energy-based variant of VBAP, i.e., vector base intensity panning [15, 56, 58].

2.7.3 Higher-order Ambisonics

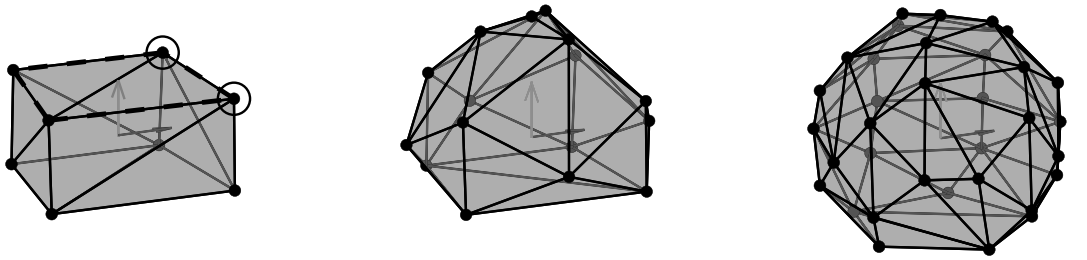
In contrast to vector-based panning methods with discrete loudspeaker weights, HOA allows the sound field to be excited using superposed orthogonal basis functions, represented by SHs, enabling the use of a continuous virtual panning function for spatial reproduction of previously encoded source signals [10, 13]. The limitation of the maximum SH order entails an upper frequency limit for peripheral playback, above which spatial aliasing occurs [59]. With the available number of loudspeakers, determining a maximum and full-sphere equivalent SH order of $\mathrm{N_{eq}} = 7$ [57], and an assumed valid reproduction in a sweet sphere of 15 cm diameter, an upper frequency limit of approximately 5.13 kHz can be estimated [60]. Partial spherical coverage and small array radii pose additional challenges regarding the ambisonic decoder design [57, 61]. For general comparison purposes (see also, [55, 58]), the decoder strategies given below were calculated [56]:

- sampling ambisonic decoder (SAD) [62],
- mode-matching ambisonic decoder (MMAD) [63] with improved regularisation [55],
- energy-preserving ambisonic decoder (EPAD) [64],
- improved all-round ambisonic decoder (AllRAD+) [55, 57].

We applied a virtual t-design of the order $2\mathrm{N} + 1 = 15$ for implementing the AllRAD+ decoder [57, 65].

2.7.4 Acoustic crosstalk cancellation

The original transaural stereo approach required a set of two loudspeakers to convey a binaural audio signal by applying static acoustic CTC filters [4, 5]. The simultaneous



(a) 2/4/8-loudspeaker layouts, the first two being marked by circled black dots and dashed black connecting lines, respectively.

(b) 16-loudspeaker layout.

(c) 32-loudspeaker layout.

Figure 5. Loudspeaker subset layouts to investigate the CTC performance in terms of optimal and achievable channel separation. The horizontal and vertical arrows represent the listener's default orientation by the according view and up vectors.

use of multiple loudspeakers is possible via an L-CTC approach, with L representing the number of involved loudspeakers. Global minimum-phase regularisation allows to remove anti-causal artefacts and generate stable CTC filters [7, 66]. In combination with the optical tracking system, cf. Section 2.3, rotational and translational sweet spot extension enables real-time reproduction of dynamic VAEs [50] with optional plausible room acoustic simulations [35]. Since HRTFs are typically referenced to the centre of the interaural axis with absent head [2] and any change in the listener's head position and orientation should also be referenced to this centre, the pivot point of the head-mounted rigid tracking body must be corrected by applying an individual translational offset [43]. However, it should be noted that individual anthropometric differences to generic HRTFs [67], which are likely to reduce the localisation performance [68], are not considered by this correction and require the use of individualised [32, 33] or the measurement of individual datasets (e.g., [69–72]).

With sufficient processing capacity, it is theoretically possible to use all 68 array loudspeakers for binaural playback. However, we were interested in finding suitable loudspeaker arrangements and estimating the minimum number of loudspeakers required for sufficient channel separation [73]. Therefore, we calculated the optimal and achievable channel separation for a variety of different loudspeaker subset layouts with an increasing number of loudspeakers, cf. Figure 5, and the full layout. The first two simulated arrangements followed the recommendation for the use of elevated loudspeakers [74], with all arrangements aiming for a reasonably uniform spatial distribution [45].

To estimate the optimal and achievable channel separation using the presented experimental setup, the binaural impulse responses of each loudspeaker were measured sequentially from an artificial head with detailed ear and simplified torso geometry [75] in the array centre. As excitation signal, an exponential sweep with a length of 2^{16} samples was used with otherwise unchanged measurement hardware and software settings, as described in Sections 2.5 and 2.6, and activated FIR filters. The raw binaural impulse responses were windowed (Hann window, 1 ms fade-in after 1 ms, 1 ms fade-out after 8 ms). To account for the transducer characteristics of the loudspeakers and artificial head microphones (MK 2H, Schoeps GmbH, Karlsruhe, Germany) and preamp (CMC 6, Schoeps GmbH, Karlsruhe, Germany), the on-axis measurements of both devices were inverted, implemented as minimum-phase filters and convolved with the windowed binaural impulse responses, resulting in the final playback HRTFs, which were cropped to a length of 256 samples. Based on these HRTFs, we calculated CTC filters using an L-CTC approach and a regularisation factor of 1e–3 [66] for the corresponding layouts shown in Figure 5 and the full layout. Since we were interested in seeing how the physical setup affects the channel separation, the CTC filters were applied on the playback HRTFs and the equalised but unwindowed spatial transfer functions, the latter exhibiting a cropped length of 11,025 samples. For these two scenarios, the optimal and achievable channel separation was calculated for the left ear only [7] due to comparable across-ear performance.

3 Results

3.1 Physical implementation

The ideal and actual loudspeaker positions are plotted in Figure 2. Figure 6 shows the resulting deviations, split into radial, azimuth, zenith and great circle central angle error components, ϵ_R, ϵ_φ, ϵ_ϑ and ϵ_γ, respectively. Note that the radial error was calculated after applying digital by-channel delays in steps of two samples on the digital signal processors of the power amplifiers. This allowed discrete corrections of 16 mm at a sampling rate of 44.1 kHz and a speed of sound of 343 m/s to obtain a virtual array radius of 1.44 m, with channel gains adapted accordingly. If more precise radial corrections are required the individual loudspeaker signals can be delayed exactly using, for example, fractional delays [76, 77] or by setting the FIR filter's group delay.

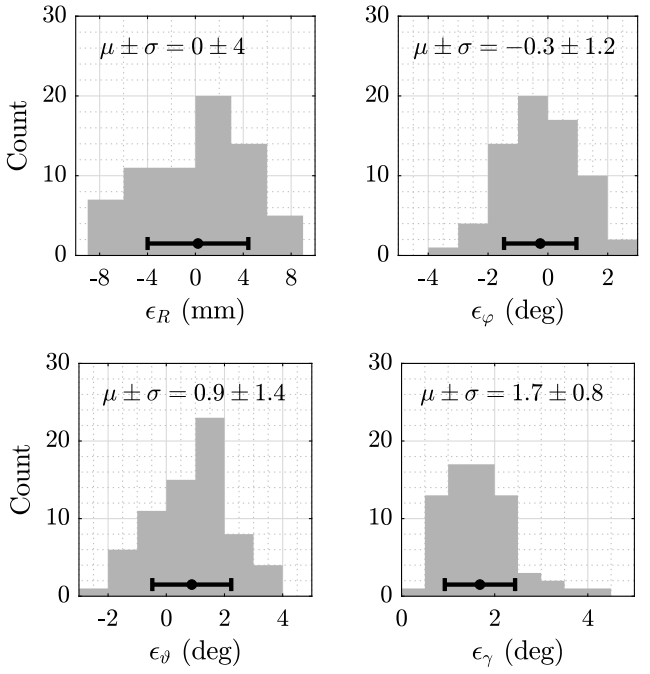


Figure 6. Actual loudspeaker positioning errors. Histograms display the shape and variance of radial, azimuth, zenith and great circle central angle errors, ϵ_R, ϵ_φ, ϵ_ϑ and ϵ_γ, respectively. Dots and error bars represent mean μ and standard deviation σ, respectively. The loudspeakers were delayed to match a nominal virtual array radius of 1.44 m.

3.2 Electroacoustics

3.2.1 Loudspeaker cabinet and crossover design

The measured on-axis sound pressure transfer functions of the two drivers in the final cabinet are shown in Figure 7a. Based on these results, an optimisation routine adapted the electronic components of the crossover circuit in a way that the combination of loudspeaker response and filter matches the target responses of a 4-th order Linkwitz–Riley crossover, see Figure 7b. A crossover frequency of 2 kHz was chosen to take account of the limited power handling capacity of the dome tweeter and the very small volume behind it, both of which advise against operation at lower frequencies. To implement the optimised crossover, depicted in Figure 7c, capacitors with high-quality foils (PE, 100 V) and coils with air core (0.1 mH) and ferrite core (1.2 mH) were used. The electronic components were mounted on a round circuit board and screwed to the plastic holder on the back of the magnet. Banana plugs facilitate the connection via 0.75 mm^2 cables.

Figure 7b displays the resulting transfer functions with adopted energy distribution between the drivers. The roll-off reaches the desired target of 24 dB/octave. However, in the stop band region, the similar magnitude patterns of the two drivers in the direct vicinity of the crossover frequency lead to some remaining deviations that cannot be addressed by the crossover network.

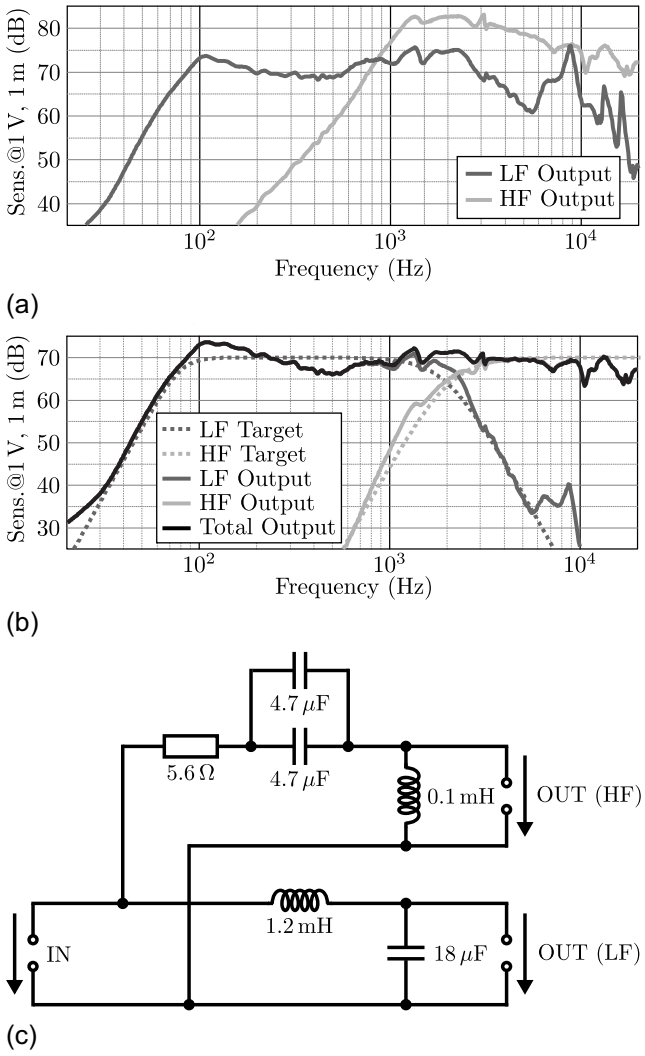


Figure 7. Filtering and energy adjustment of the loudspeaker input signal using a two-way passive crossover network. (a) On-axis sound pressure level transfer functions of the example loudspeaker's low-frequency (LF) and high-frequency (HF) drivers without passive crossover network. (b) Crossover target curves, on-axis sound pressure level transfer functions of the example loudspeaker's LF and HF drivers, and the total output, each filtered with passive crossover. (c) Electrical circuit diagram of the passive crossover network with outputs for LF and HF drivers.

3.2.2 Spectral equalisation

The top panel of Figure 8a displays the results of the on-axis loudspeaker frequency response measurements without FIR filters. The loudspeakers exhibit a mean sensitivity of 69 ± 2 dBV/m ($\mu \pm \sigma$, range: 61–74 dBV/m) between 60 Hz and 20 kHz and a distinct energy increase around the port resonance frequency of about 92 Hz. Further deviations from a linear frequency response can be observed at the crossover frequency of 2 kHz, for reasons mentioned in Section 3.2.1, and around 11 kHz, owing to the coaxial driver geometry and associated phase irregular-

ities. With active individual FIR filters, the batch-to-batch variations could be minimised, resulting in a passband sensitivity (−3 dB re mean sensitivity) of 71 ± 1 dBV/m ($\mu \pm \sigma$, range: 68–72 dBV/m) between 92 Hz and 20 kHz. The transfer function issues related to crossover design and driver geometry effects were largely removed. What remains after FIR filtering are level fluctuations within ±3 dB between 2.4 kHz and 3.7 kHz. This effect can be explained by component-related variations and the high sound pressure behind the textile dome, which has a reduced stability in this frequency range.

Figure 8b shows the on-axis sound pressure level transfer function of an example loudspeaker (no. 8) when mounted in the array. Notable deviations from the response measured under optimal conditions start above frequencies of about 300 Hz. Once the wavelength is within the range of the loudspeaker dimensions and smaller, cf. Figure 4, the influence of other array loudspeakers becomes prominent, resulting in distinct peaks and notches, for example, at about 1 kHz and 1.3 kHz due to interference effects. The pattern at 11 kHz is still visible although smeared. Note that the apparent energy drop above 15.6 kHz is most likely related to the directivity pattern of the microphone, which was pointing to the north pole loudspeaker during the measurement.

3.2.3 Directivity pattern

Figure 9 shows the example loudspeaker's directivity, normalised to its on-axis transfer function. All values were mapped to discrete levels of 3 dB and truncated if they were outside the displayed dynamic range. The example array loudspeaker shows a symmetric homogeneous broadband horizontal directivity pattern (±3 dB) within ±10° and exhibits deviations up to +6 dB between frequencies of 10.7 kHz and 11.7 kHz within ±20°. The previously described on-axis dip around 11 kHz can be explained by ring-shaped diffraction at edges around the dome tweeter. This effect leads to destructive on-axis and constructive off-axis interference, lowering and increasing the sound pressure level, respectively, and is clearly reflected by the two white maxima at 10°–35° off axis. However, a complete correction of this dip is not recommended as the radiated sound power in the affected frequency range would increase too much, resulting in unpleasant and excessive high-frequency pronunciation. Instead, it is advisable to accept this irregularity, as it will hardly be audible using broadband signals.

3.2.4 Background noise level

Figure 10 displays the BNL as unweighted equivalent-continuous sound pressure levels $L_{Z,eq}$ in octave bands with power amplifiers switched on and off, measured in the centre of the loudspeaker array. An increase of inherent noise levels when activating the power amplifiers can be observed in particular for octave bands with centre frequencies above 500 Hz, overall still falling below the NR15 curve, while exceeding the NR10 curve in octave bands above

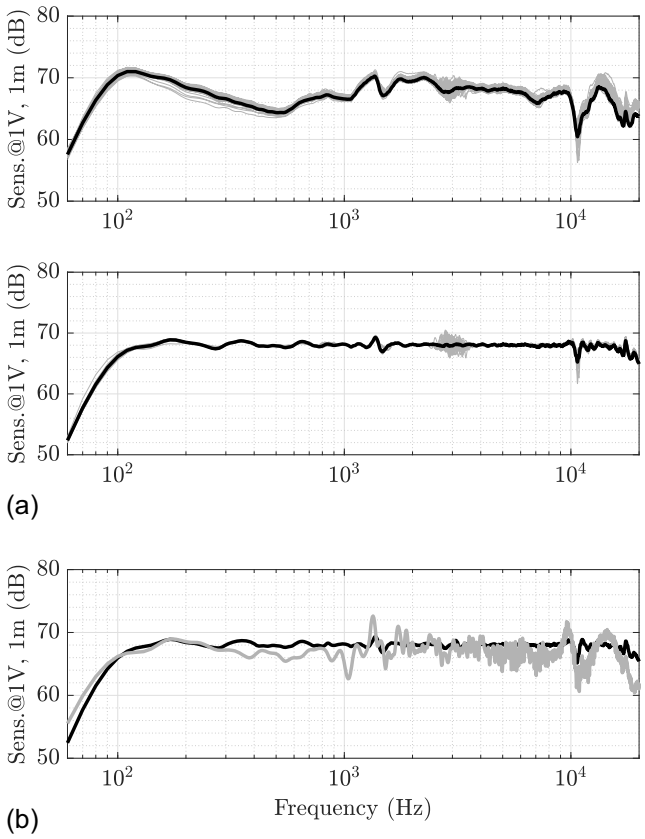


Figure 8. On-axis loudspeaker responses with and without spectral equalisation, measured under optimal conditions and after installation in the array. (a) Top/bottom: On-axis loudspeaker sound pressure level transfer functions without/with FIR filters, measured in the hemi-anechoic chamber. Grey and black lines display individual and mean magnitude spectra, respectively. (b) Comparison of an example loudspeaker's on-axis sound pressure level transfer function, measured in the centre of the loudspeaker array (grey line) and in the hemi-anechoic chamber (black line), with activated FIR filters.

2 kHz [78]. The total sound pressure level changes from about 26 dB to 28 dB.

3.3 Spatial audio reproduction

The results of the performance metrics using panning-based spatial audio reproduction methods are shown in Figure 11 for VBAP, MDAP and the selected HOA decoder strategies (N = 7), which were calculated based on the ideal and actual loudspeaker layouts. A statistical analysis was not carried out, as even the smallest differences in means/medians would lead to significant but perceptually negligible results due to the large sample size and the resulting power of the applied tests.

3.3.1 Vector base amplitude panning

As expected due to the normalisation of loudspeaker weights, both VBAP and MDAP decoders exhibit

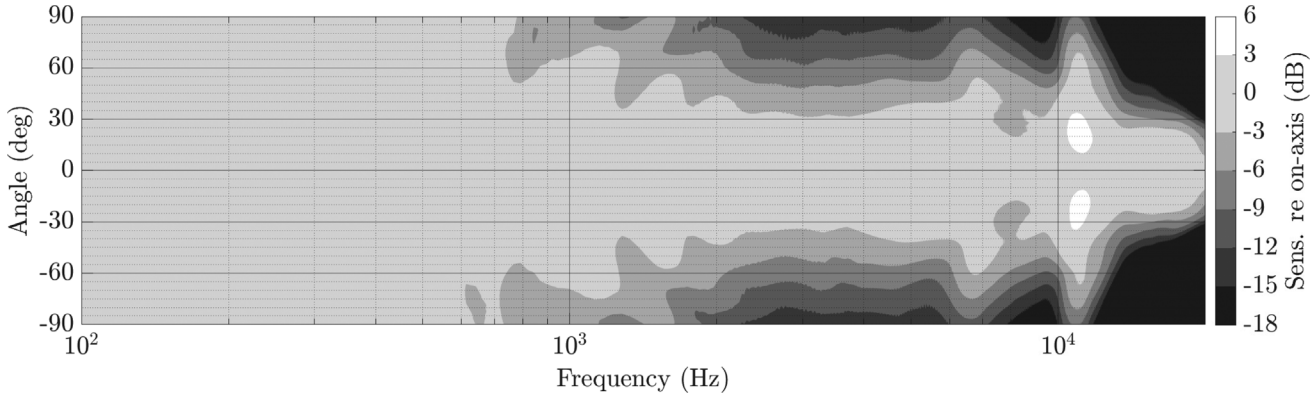


Figure 9. Horizontal directivity pattern of one example array loudspeaker, normalised to its on-axis frequency response. The magnitude spectra were smoothed using filters with constant relative bandwidth of one-sixth octave.

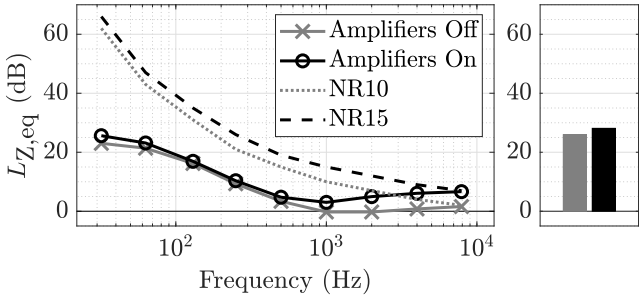


Figure 10. Background noise levels represented by unweighted equivalent-continuous sound pressure levels in octave bands with NR10/NR15 curves (left panel) and corresponding total levels (right panel), measured in the centre of the loudspeaker array. Solid black and grey lines and the corresponding bars show the measurement results with power amplifiers switched on and off, respectively.

direction- independent overall energy. Not surprisingly, $\hat{\sigma}_\mathrm{V}$ and $\hat{\sigma}_\mathrm{E}$ show very values when using VBAP since the decoder relies on a loudspeaker subset selection and thus shifts the velocity as well as the energy vectors in the desired VSS direction. Due to the underlying principle, the MDAP decoder results in a higher variance and generally decreased values in $\hat{\sigma}_\mathrm{V}$. The same tendency can be observed in $\hat{\sigma}_\mathrm{E}$, although not as pronounced for the variance.

In terms of $\hat{\epsilon}_\mathrm{V}$, VBAP outperforms MDAP, which is at the expense of an increased directional energy error $\hat{\epsilon}_\mathrm{E}$. This shortcoming is specifically addressed by MDAP, resulting in minimal and direction-independent errors $\hat{\epsilon}_\mathrm{E}$, which in turn increases the directional velocity error $\hat{\epsilon}_\mathrm{V}$. Due to the high values in $\hat{\sigma}_\mathrm{E}$, a perceived VSS width α of about 15° on average is predicted for the VBAP decoder, the lowest value of all evaluated decoders. The example MDAP configuration entails an increased median value of 16.5°, which is close to the predicted perceived VSS width of the evaluated ambisonic decoders. Compared to the ideal layout, the actual loudspeaker positioning errors have no effect worth mentioning on the performance metrics using VBAP and MDAP decoders.

3.3.2 Higher-order Ambisonics

Apart from considerable differences across HOA decoder strategies but very similar results across ideal and actual loudspeaker layouts, the smallest overall mean energy can be observed for the EPAD decoder followed by MMAD and, at a greater distance, the SAD and AllRAD+ implementations. The energy variances are comparable with slightly higher variations using the SAD decoder. For SAD and MMAD, however, it should be noted that directions below the lowest loudspeaker ring suffer from missing energy, while both EPAD and AllRAD+ manage to widely maintain the energy even for these directions at comparable levels as for directions lying on the spherical cap.

The directional performance in terms of $\hat{\sigma}_\mathrm{V}$ and $\hat{\sigma}_\mathrm{E}$ is also competitive between all HOA decoders, with slightly decreased velocity vector magnitudes and increased energy vector magnitude variance using AllRAD+ and EPAD, respectively. The directional errors in terms of $\hat{\epsilon}_\mathrm{V}$ show similar across-decoder behaviour, with AllRAD+ performing best with low variance. All decoders have similar mean energetic directional errors $\hat{\epsilon}_\mathrm{E}$ with increased and comparable variance ranges between EPAD and AllRAD+. The median values of perceived VSS width α lie around 18° for all decoders with the largest variance for EPAD. The directional error for decoders based on the actual loudspeaker positions is noticeable by (slightly) increased median errors in $\hat{\epsilon}_\mathrm{V}$ and $\hat{\epsilon}_\mathrm{E}$ (SAD, MMAD, EPAD) and a change in variance, the latter surprisingly not always for the worse.

3.3.3 Acoustic crosstalk cancellation

The optimal and achievable channel separation results for the evaluated subset layouts, cf. Figure 5, and the full layout are shown in Figure 12. Table 1 presents the corresponding mean channel separation values, broadband and evaluated for low and high frequency ranges. For each doubling of the number of loudspeakers, the optimal and achievable channel separation increases on average by

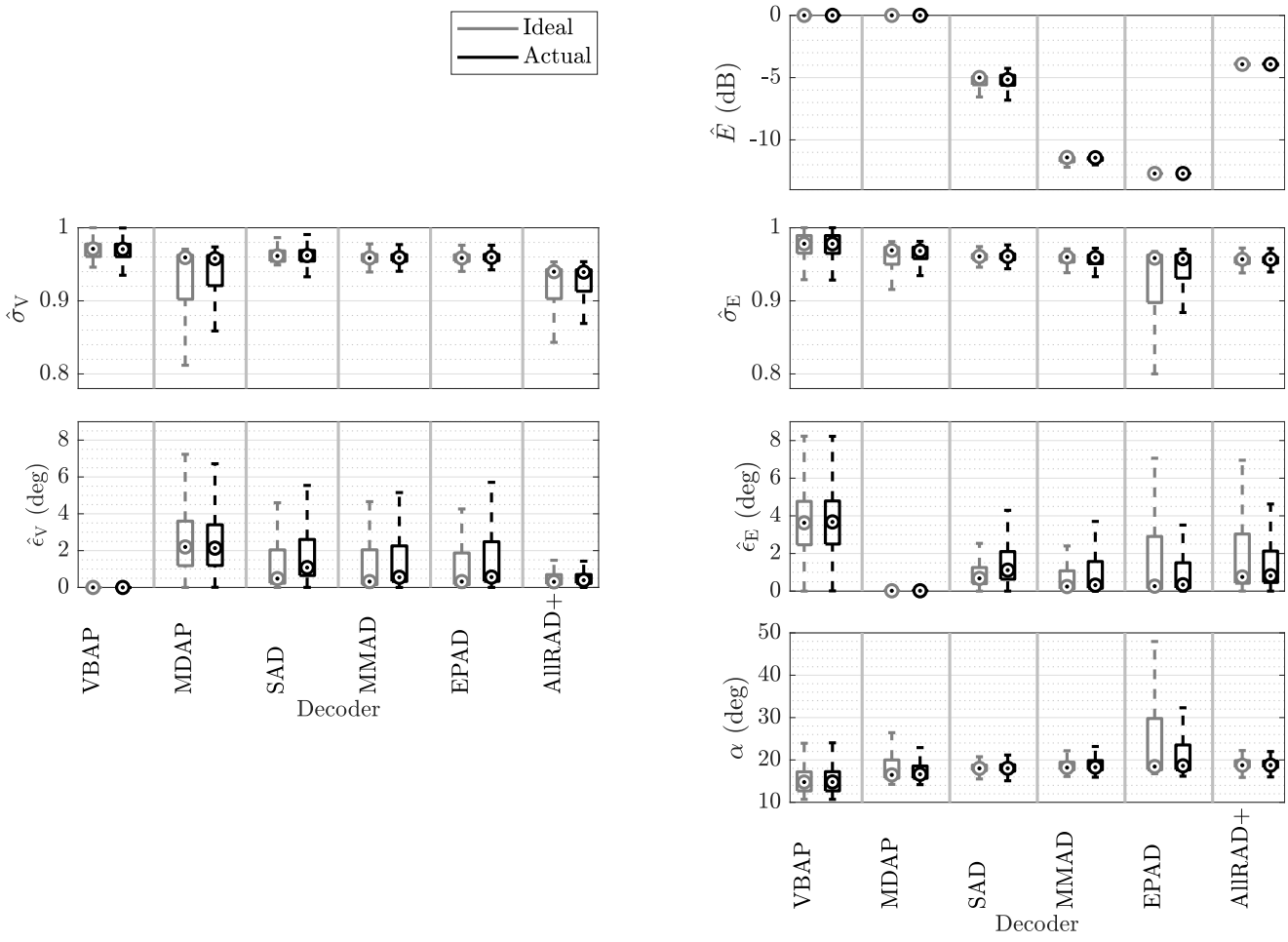


Figure 11. Performance metrics for different panning-based decoder strategies. All ambisonic decoders were calculated for an order of N = 7. The results are based on the ideal and actual sampling layouts, represented by grey and black line colours, respectively. Box plots display medians and interquartile ranges with whiskers covering 1.5 times the interquartile range without showing outliers.

6.8 dB and 2.8 dB, respectively. The optimal channel separation ranges from 40.9–74.7 dB, while decreasing considerably by 34.3 dB on average to 18.2–32.3 dB in the unmatched scenario. Apart from a decrease, a generally reduced variance of achievable channel separation values is observable. The reason for the substantial drop in channel separation can be traced back to the acoustic influence of the experimental setup construction. Sæbø [79] and Kohnen et al. [80] demonstrated decreased filter performance in case of additional reflections. The latter authors presented a CTC filter approach that aimed at compensating for reflections up to the second order and thus increasing channel separation in reflective but geometrically simple acoustic environments. Although positive effects were found in the simulated scenario, the approach could not bring about substantial improvements when applied to measured spatial transfer paths. According to the authors, this is due to the increased acoustic complexity, which cannot be fully described by the geometric acoustic model used in the simulations. Transferred to the present loudspeaker array,

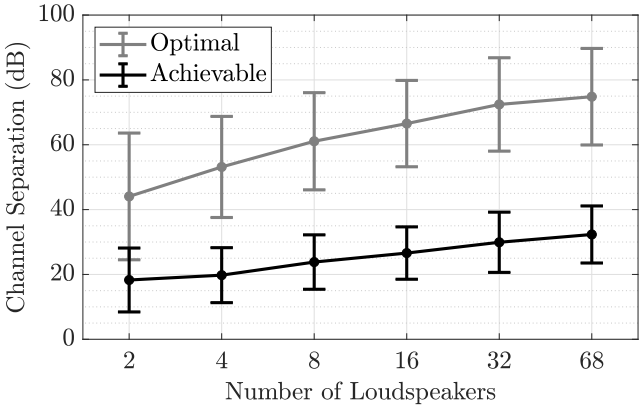


Figure 12. Optimal and achievable channel separation for the evaluated subset layouts, presented in Figure 5, and the full layout, using an L-CTC approach with a regularisation factor of 1e−3. Dots and error bars indicate broadband means and standard deviations, respectively, which were calculated for a frequency range of 90 Hz–20 kHz.

Table 1. Mean optimal and achievable channel separation with standard deviation ($\mu \pm \sigma$) for the evaluated subset layouts, presented in Figure 5, and the full layout. The broadband values were calculated for a frequency range of 90 Hz–20 kHz.

Layout	Optimal			Achievable		
	Broadband	300 Hz–2 kHz	4–16 kHz	Broadband	300 Hz–2 kHz	4–16 kHz
	$\mu \pm \sigma$ (dB)	$\mu \pm \sigma$ (dB)	$\mu \pm \sigma$ (dB)	$\mu \pm \sigma$ (dB)	$\mu \pm \sigma$ (dB)	$\mu \pm \sigma$ (dB)
2-loudspeaker	40.9 ± 19.9	52.5 ± 6.4	41 ± 19.7	18.2 ± 9.8	16.8 ± 5.1	16.3 ± 10.2
4-loudspeaker	49.2 ± 15.7	57.7 ± 7.2	52.5 ± 12.4	19.8 ± 8.5	19.4 ± 4.7	19.3 ± 6.8
8-loudspeaker	57.8 ± 15.2	63 ± 7.1	62.1 ± 10.7	23.8 ± 8.4	21.2 ± 5	23 ± 6.2
16-loudspeaker	63.1 ± 13.1	64 ± 7.7	68.1 ± 6.6	26.6 ± 8.1	24.3 ± 4.6	26.4 ± 5
32-loudspeaker	70.7 ± 14.6	67.4 ± 10.6	75.2 ± 6.4	29.9 ± 9.3	27.1 ± 5.2	31.4 ± 6.6
68-loudspeaker	74.7 ± 15.5	67 ± 10	79.8 ± 6.4	32.3 ± 8.8	26.8 ± 4.8	34.1 ± 5.8

more complex acoustic effects such as scattering or diffraction are also present and further reduce the effectiveness of the CTC filters.

It is noteworthy that in setups with sixteen or more loudspeakers, both optimal and achievable channel separation increase faster in the high frequency range than that in the low frequency range. The high frequency range thus accounts for the major share of the broadband performance, which can be explained by the concept of L-CTC. The resulting CTC system matrix represents an underdetermined system for $L > 2$ and is optimised for minimal energy. Increasing the number of loudspeakers favours a better energy distribution. This usually results in flatter CTC filter curves that require less gain-limiting regularisation, thus preserving the level of detail in the high-frequency structure of the playback HRTFs.

4 Discussion

4.1 Electroacoustics

Due to their largely flat frequency responses under optimal conditions and the covered frequency range, the array loudspeakers are useful for modelling reference point sources in localisation experiments [34] and the theoretically required local sampling of a surrounding sphere for ambisonic reproduction. However, the acoustic influence of other array loudspeakers and the mounting construction cannot be avoided and affects the optimal FIR-filtered loudspeaker responses, although the array was mounted in a highly optimised environment. As a practical side effect, the homogeneous directivity pattern around the on-axis direction subordinates the individual loudspeaker orientations compared to the importance of their spatial positions. This also favours negligible changes in spectral magnitude responses for off-centre listening in case of a moving listener. Sufficient shielding from external noise in combination with the low inherent noise levels and high output signal-noise-ratio render the system a suitable environment for sensitive measurement applications and listening experiments. Since the noise floor lies below the NR15 curve, the experimental setup additionally fulfils the requirements for perceptual assessment of audio systems as per ITU-R BS.1116-3 [81] regarding maximum permissible BNL.

4.2 Spatial audio reproduction

4.2.1 Vector base amplitude panning

Although originally designed for artistic applications without claiming physical correctness [16], the results of the VBAP and MDAP decoders motivate perception-driven reproduction. For real-time auralisation with hybrid room acoustic simulation approaches, vector-based panning techniques represent a flexible implementation approach for direct sound and early reflections [82]. However, one should not ignore the perceptual deficiencies when using VBAP [23] and instead consider MDAP with application-dependent spreading parameter settings. In such a dense loudspeaker arrangement, it would also be an option to use discrete loudspeaker reproduction for the simulation of late reflections [83], as their exact directions of incidence may be perceptually less important.

4.2.2 Higher-order Ambisonics

In general, the observed differences in the directional decoder energy must be taken into account when calibrating the playback level for studies that apply different decoder strategies. If only selective directions of sound incidence are used the calibration should also be tailored to this subset. However, the use of sampling and mode-matching decoding will lead to noticeable loudness variations in the uncovered spatial range [55]. For auralisations that require an all-round incidence of sound waves, decoder strategies such as EPAD or AllRAD+ should thus be preferred to the others, as they are specially designed for irregular layouts. Using the latter strategy, accurate directional performance in terms of minimised loudness variation, correct direction of plane wave incidence and widely direction-independent spatial extent of VSS can be expected. The actual positioning errors of the loudspeakers seem to be within an acceptable range, since an effect on the decoder performance is only most clearly visible for the SAD.

4.2.3 Acoustic crosstalk cancellation

The differences between optimal and achievable channel separation once again point out the necessity of an

optimised reproduction environment, a compact loudspeaker design and an unobtrusive mounting construction [79, 80], which is particularly important for hardware-intense setups, even when installed in anechoic chambers.

To check the applicability of the system, the achievable channel separation results need to be compared to the minimum required. Parodi and Rubak [73] used different stimulus types in a 2-loudspeaker CTC setup, virtually reproduced over headphones, with varying loudspeaker span angles. Based on the results of perceptual tests, they suggested average minimum values around 20 dB for speech, broadband and narrowband noise with centre frequencies below 1 kHz when the listener is located at the nominal centre position. They also reported decreased thresholds of 15 dB for narrowband noise with centre frequencies at 1 and 2 kHz. The most sensitive thresholds of 25 dB were found for centre frequencies above, which was explained by the sensitivity to manipulations of interaural level differences.

According to the current mean broadband results, the proposed minimum channel separation [73] is largely achieved when using sub-layouts with eight and more loudspeakers, representing configurations that also allow complete rotational freedom of the listener in dynamic systems [7, 45]. If necessary, a further improvement of channel separation towards higher frequencies can be achieved by covering the loudspeaker cabinets and the mounting construction with absorbent foam. Reducing the regularisation during CTC filter calculation also leads to an improvement in frequency regions which are characterized by notches in the playback HRTFs. It should be noted, however, that too low a regularisation factor may push the loudspeakers to their physical limits, leading to increased non-linearities. In addition, the narrowband peaks are likely to cause filter ringing and colouration. Although only weak correlations with localisation performance in the saggital plane were observed, channel separation could be a useful predictor of localisation performance in the horizontal plane in matched and mismatched CTC systems [7, 68].

4.3 Limitations

The evaluation of panning-based reproduction methods based on the analysis of the decoder-dependent loudspeaker gains, including loudspeaker positioning errors, only allowed to roughly estimate the performance of the implemented system without accounting for the acoustic influence of the physical setup. A more comprehensive evaluation would require physically sampling the sweet sphere with a microphone array, facilitating a direct comparison of the reproduced sound field with the synthesis target. Alternatively, such an evaluation can be carried out using a sound field reconstruction method based on plane wave decomposition [84] or point source expansion [28]. Furthermore, a measurement-based derivation of energy and velocity vectors [85, 86] potentially provides a more reliable prediction of the overall in-situ performance across decoders.

5 Conclusion

We presented the design and implementation of a surrounding 68-channel spherical cap loudspeaker array. Commonly used spatial audio reproduction methods were assessed based on various performance metrics to test their suitability and practicality. For amplitude panning approaches, recommendations regarding decoder selection were derived on the basis of simulation results. We also suggested suitable subset layouts for loudspeaker-based binaural reproduction, allowing sufficient channel separation, based on in situ measurement results. Collectively, the results indicate that the implemented system provides a good basis for objective and perceptual evaluations of VAEs created by means of the presented spatial audio reproduction methods.

Acknowledgments

The authors would like to thank the team of the mechanical workshop, Uwe Schlömer, Marc Eiker, and Thomas Schaefer, for the design implementation, as well as Rolf Kaldenbach for the electronic implementation, assistance with loudspeaker positioning and general support. We also appreciate the technical assistance and critical proofreading by Marco Berzborn and the help of Jonas Stienen, Michael Kohnen, Manuj Yadav, Mark Müller-Giebeler, Hark Braren, Luiz Otavio Kohler and Robert Henzel. Further thanks go to Karin Charlier for handling financial matters.

Funding

The authors received no specific funding for this work.

Conflict of interest

The authors declared no conflicts of interest.

References

1. Y. Aloimonos: Active Perception, ser. Computer Vision. Lawrence Erlbaum Associates Inc, Hillsdale, NJ, US, 1993.
2. J. Blauert: Spatial Hearing: The Psychophysics of Human Sound Localization. MIT Press, 1997.
3. A. Lindau, S. Weinzierl: Assessing the plausibility of virtual acoustic environments. Acta Acustica United With Acustica 98, 5 (2012) 804–810. https://doi.org/10.3813/AAA.918562.
4. B.B. Bauer: Stereophonic earphones and binaural loudspeakers. The Journal of the Audio Engineering Society 9, 2 (1961) 148–151.
5. B.S. Atal, M.R. Schroeder: Apparent sound source translator. United States Patent 3,236,949, 1966.
6. J. Bauck, D.H. Cooper: Generalized transaural stereo and applications. The Journal of the Audio Engineering Society 44, 9 (1996) 683–705.

7. B. Sanches Masiero: Individualized binaural technology: Measurement, equalization and perceptual evaluation, Ph.D. dissertation. Institute of Technical Acoustics, RWTH Aachen University, Germany, 2012.
8. P. Fellgett: Ambisonics. Part one: General system description. Studio Sound 17, 8 (1975) 20–22.
9. M.A. Gerzon: Ambisonics. Part two: Studio techniques. Studio Sound 17, 8 (1975) 24–26.
10. M.A. Gerzon: Ambisonics in multichannel broadcasting and video. The Journal of the Audio Engineering Society 33, 11 (1985) 859–871.
11. J. Daniel, J.-B. Rault, J.-D. Polack: Ambisonics encoding of other audio formats for multiple listening conditions, in Audio Engineering Society Convention 105. 1998.
12. J. Daniel: Représentation de champs acoustiques, applicationa la transmission eta la reproduction de scenes sonores complexes dans un contexte multimédia, Ph.D. dissertation. Université de Paris, France, 2000.
13. F. Zotter, M. Frank: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality. Springer International Publishing, 2019. https://doi.org/10.1007/978-3030-17207-7.
14. K. Wendt: Das Richtungshören bei der Überlagerung zweier Schallfelder bei Intensitäts-und Laufzeitstereophonie, Ph.D. dissertation. Rheinisch-Westfälische Technische Hochschule Aachen, 1963.
15. V. Pulkki: Uniform spreading of amplitude panned virtual sources, in Applications of Signal Processing to Audio and Acoustics, 1999 IEEE Workshop on. 1999, pp. 187–190. https://doi.org/10.1109/ASPAA.1999.810881.
16. V. Pulkki, Spatial sound generation and perception by amplitude panning techniques, Ph.D. dissertation. Helsinki University of Technology Laboratory of Acoustics and Audio Signal Processing, Finland, 2001, pp. 1456–6303.
17. A.J. Berkhout: A holographic approach to acoustic control. The Journal of the Audio Engineering Society 36, 12 (1988) 977–995.
18. A.J. Berkhout, D. de Vries, P. Vogel: Acoustic control by wave field synthesis. The Journal of the Acoustical Society of America 93, 5 (1993) 2764–2778. https://doi.org/10.1121/1.405852.
19. S. Spors, H. Wierstorf, A. Raake, F. Melchior, M. Frank, F. Zotter: Spatial sound with loudspeakers and its perception: A review of the current state, Proceedings of the IEEE 101, 9 (2013) 1920–1938. https://doi.org/10.1109/JPROC.20132264784.
20. J.M. Zmoelnig, A. Sontacchi, W. Ritsch: The IEM-Cube, a periphonic re-/production system, in Audio Engineering Society Conference: 24th International Conference: Multichannel Audio, The New Reality. 2003.
21. M. Noisternig, T. Carpentier, O. Warusfel: ESPRO 2.0 – Implementation of a surrounding 350-loudspeaker array for 3D sound field reproduction, in Audio Engineering Society Conference: UK 25th Conference: Spatial Audio in Today's 3D World. 2012.
22. L. Gandemer, G. Parseihian, C. Bourdin, Perception of surrounding sound source trajectories in the horizontal plane: A comparison of VBAP and basic-decoded HOA, Acta Acustica United With Acustica 104, 2 (2018) 338–350. https://doi.org/10.3813/AAA.919176.
23. M. Frank: Phantom sources using multiple loudspeakers in the horizontal plane, Ph.D. dissertation. Institute of Electronic Music and Acoustics, University of Music and Performing Arts Graz, Austria, 2013.
24. F.M. Fazi: Sound field reproduction, Ph.D. dissertation. Institute of Sound and Vibration Research, University of Southampton, England, 2010.
25. A. Ahrens: Characterizing auditory and audio-visual perception in virtual environments, Ph.D. dissertation. Hearing Systems Section, Department of Health Technology, Technical University of Denmark, 2019.
26. M. Otani, H. Shigetani: Reproduction accuracy of higher-order Ambisonics with Max-rE and/or least norm solution in decod- ing. Acoustical Science and Technology 40, 1 (2019) 23–28.
27. A. Parthy, C. Jin, A. van Schaik, Evaluation of a concentric rigid and open spherical microphone array for sound reproduction, in Proceedings of Ambisonics Symposium. 2009.
28. E. Fernandez-Grande: Sound field reconstruction using a spherical microphone array. The Journal of the Acoustical Society of America 139, 3 (2016) 1168–1178. https://doi.org/10.1121/1.4943545.
29. P. Minnaar, S.F. Albeck, C.S. Simonsen, B. Søndersted, S.A.D. Oakley, J. Bennedbæk: Reproducing real-life listening situations in the laboratory for testing hearing aids, in Audio Engineering Society Convention 135. 2013.
30. C. Oreinos, J.M. Buchholz: Evaluation of loudspeaker-based virtual sound environments for testing directional hearing aids. Journal of the American Academy of Audiology 27, 7 (2016) 541–556. https://doi.org/10.3766/jaaa.15094.
31. J. Cubick, T. Dau: Validation of a virtual sound environment system for testing hearing aids. Acustica United With Acta Acustica 102 (2016) 547–557. https://doi.org/10.3813/AAA.918972.
32. J.C. Middlebrooks: Virtual localization improved by scaling non-individualized external-ear transfer functions in frequency. The Journal of the Acoustical Society of America 106, 3 (1999) 1493–1510. https://doi.org/10.1121/1.427147.
33. R. Bomhardt: Anthropometric individualization of head-related transfer functions: Analysis and modeling, Ph.D. dissertation. Teaching and Research Area of Medical Acoustics, Institute of Technical Acoustics, RWTH Aachen University, Germany, Berlin, 2017.
34. A.W. Bronkhorst: Localization of real and virtual sound sources. The Journal of the Acoustical Society of America 98, 5 (1995) 2542–2553. https://doi.org/10.1121/1.413219.
35. D. Schröder: Physically based real-time auralization of interactive virtual environments, Ph.D. dissertation. Institute of Technical Acoustics, RWTH Aachen University, Germany, 2011.
36. B.N.J. Postma, B.F.G. Katz: Perceptive and objective evaluation of calibrated room acoustic simulation auralizations. Journal of the Acoustical Society of America 140, 6 (2016) 4326–4337. https://doi.org/10.1121/1.4971422.
37. F. Brinkmann, L. Aspöck, D. Ackermann, S. Lepa, M. Vorländer, S. Weinzierl: A round robin on room acoustical simulation and auralization. Journal of the Acoustical Society of America 145, 4 (2019) 2746–2760. https://doi.org/10.1121/1.5096178.
38. L. Rossi, A. Prato, L. Lesina, A. Schiavi: Effects of low-frequency noise on human cognitive performances in laboratory. Building Acoustics 25, 1 (2018) 17–33. https://doi.org/10.1177/1351010X18756800.
39. S. Rizzi, B. Sullivan: Synthesis of virtual environments for aircraft community noise impact studies, in 11th AIAA/CEAS Aeroacoustics Conference. 2005, p. 2983.
40. I. Muhammad, M. Vorländer, S.J. Schlittmeier: Audio-video virtual reality environments in building acoustics: An exemplary study reproducing performance results and subjective ratings of a laboratory listening experiment. The Journal of the Acoustical Society of America 146, 3 (2019) EL310–EL316. https://doi.org/10.1121/1.5126598.

41. L.L. Beranek, H.P. Sleeper: The design and construction of anechoic sound chambers. The Journal of the Acoustical Society of America 18, 1 (1946) 140–150. https://doi.org/10.1121/1.1916351.
42. P. Leopardi: A partition of the unit sphere into regions of equal area and small diameter. Electronic Transactions on Numerical Analysis 25, 12 (2006) 309–327.
43. M. Berzborn, R. Bomhardt, J. Klein, J.-G. Richter, M. Vorländer: The ITA-Toolbox: An open source MATLAB toolbox for acoustic measurements and signal processing, in 43th Annual German Congress on Acoustics, Kiel (Germany), 6–9 Mar 2017. 2017, pp. 222–225.
44. W.G. Gardner: Head tracked 3-D audio using loudspeakers, in Proceedings of 1997 Workshop on Applications of Signal Processing to Audio and Acoustics, IEEE. 1997, p. 4.
45. T. Lentz, Dynamic crosstalk cancellation for binaural synthesis in virtual reality environments, The Journal of the Audio Engineering Society 54, 4 (2006) 283–294.
46. G. Grimm, J. Luberadzka, T. Herzke, V. Hohmann: Toolbox for acoustic scene creation and rendering (TASCAR): Render methods and research applications, in Proceedings of the Linux Audio Conference. 2015, pp. 9–12.
47. R.A. Viveros Munoz: Speech perception in complex acoustic environments: Evaluating moving maskers using virtual acoustics, Ph.D. dissertation. Teaching and Research Area of Medical Acoustics, Institute of Technical Acoustics, RWTH Aachen University, Germany, 2019. https://doi.org/10.18154/RWTH-2019-07497.
48. OptiTrack, NaturalPoint Inc: Flex 13. https://optitrack.com/products/flex-13/, Accessed on 2020-03-24.
49. M. Geier, S. Spors: Spatial audio with the soundscape renderer, in 27th Tonmeistertagung, Köln, VDT International Convention. 2012.
50. Institute of Technical Acoustics, RWTH Aachen University: Virtual acoustics – A real-time auralization framework for scientific research. http://www.virtualacoustics.org/, Accessed on 2020-04-21.
51. M. Zaunschirm, C. Schörkhuber, R. Höldrich: Binaural rendering of Ambisonic signals by headrelated impulse response time alignment and a diffuseness constraint. The Journal of the Acoustical Society of America 143, 6 (2018) 3616–3627. https://doi.org/10.1121/1.5040489.
52. Institute of Technical Acoustics, RWTH Aachen University: Bassyst 2.1. 2020. http://www.bassyst.de/, Accessed on 2020-03-25.
53. M.A. Gerzon: General metatheory of auditory localisation, in Audio Engineering Society Convention 92. 1992.
54. Y. Makita: On the directional localization of sound in the stereophonic sound field. EBU Review 73, 2 (1962) 1536–1539.
55. F. Zotter, M. Frank, H. Pomberger: Comparison of energy-preserving and all-round ambisonic decoders, in Fortschritte der Akustik, AIADAGA, (Meran). 2013.
56. A. Politis: Microphone array processing for parametric spatial audio techniques, Ph.D. dissertation. Aalto University, Finland, 2016.
57. F. Zotter, M. Frank: All-round Ambisonic panning and decoding. The Journal of the Audio Engineering Society 60, 10 (2012) 807–820.
58. J.-M. Jot, V. Larcher, J.-M. Pernaux: A comparative study of 3-D audio encoding and rendering techniques, in Audio Engineering Society Conference: 16th International Conference: Spatial Sound Reproduction. 1999.
59. F. Zotter, Sampling strategies for acoustic holography/holophony on the sphere. NAG-DAGA, Rotterdam, 2009, pp. 1–4.
60. S. Bertet, J. Daniel, S. Moreau: 3D sound field recording with higher order ambisonics – Objective measurements and validation of spherical microphone, in Audio Engineering Society Convention 120. 2006.
61. J. Daniel: Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonic format, in Audio Engineering Society Conference: 23rd International Conference: Signal Processing in Audio Recording and Reproduction. 2003.
62. D.G. Malham, A. Myatt: 3-D sound spatialization using ambisonic techniques. Computer Music Journal 19, 4 (1995) 58–70.
63. M.A. Poletti: A unified theory of horizontal holographic sound systems. The Journal of the Audio Engineering Society 48, 12 (2000) 1155–1182.
64. F. Zotter, H. Pomberger, M. Noisternig: Energy preserving ambisonic decoding. Acta Acustica United With Acustica 98, 1 (2012) 37–47. https://doi.org/10.3813/AAA.918490.
65. R.H. Hardin, N.J.A. Sloane: McLaren's improved snub cube and other new spherical designs in three dimensions. Discrete & Computational Geometry 15, 4 (1996) 429–441. https://doi.org/10.1007/BF02711518.
66. B. Masiero, M. Vorländer: A framework for the calculation of dynamic crosstalk cancellation filters. IEEE/ACM Transactions on Audio, Speech, and Language Processing 22, 9 (2014) 1345–1354.
67. J. Fels, M. Vorländer: Anthropometric parameters influencing head-related transfer functions. Acta Acustica United With Acustica 95, 2 (2009) 331–342. https://doi.org/10.3813/AAA.918156.
68. P. Majdak, B. Masiero, J. Fels: Sound localization in individualized and non-individualized crosstalk cancellation systems. The Journal of the Acoustical Society of America 133, 4 (2013) 2055–2068. https://doi.org/10.1121/1.4792355.
69. Acoustics Research Institute, Austrian Academy of Sciences: ARI HRTF database. 2020. https://www.kfs.oeaw.ac.at/index.php?view=article&id=608&lang=en, Accessed on 2020-07-20.
70. K. Watanabe, Y. Iwaya, Y. Suzuki, S. Takane, Sato: Dataset of head-related transfer functions measured with a circular loudspeaker array. Acoustical Science and Technology 35, 3 (2014) 159–165. https://doi.org/10.1250/ast.35.159.
71. Institut de Recherche et Coordination Acoustique/Musique: Listen HRTF database. 2020. http://recherche.ircam.fr/equipes/salles/listen/, Accessed on 2020-07-20
72. R. Bomhardt, M. de la Fuente Klein, J. Fels: A high-resolution head-related transfer function and three-dimensional ear model database. Proceedings of Meetings on Acoustics 29, 1 (2016) 050 002. https://doi.org/10.1121/2.0000467.
73. Y.L. Parodi, P. Rubak: A subjective evaluation of the minimum channel separation for reproducing binaural signals over loudspeakers. The Journal of the Audio Engineering Society 59, 7/8 (2011) 487–497.
74. Y.L. Parodi, P. Rubak: Objective evaluation of the sweet spot size in spatial sound reproduction using elevated loudspeakers. The Journal of the Acoustical Society of America 128, 3 (2010) 1045–1055. https://doi.org/10.1121/1.3467763.
75. A. Schmitz: Ein neues digitales Kunstkopfmeßsystem. Acta Acustica United With Acustica 81, 4 (1995) 416–420.
76. V. Valimaki, T.I. Laakso: Principles of fractional delay filters, in 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.00CH37100), Vol. 6. 2000, pp. 3870–3873.
77. U. Zölzer: DAFX: Digital Audio Effects. John Wiley & Sons, 2011.

78. ISO 1996-1: Acoustics – Description, measurement and assessment of environmental noise – Part 1: Basic quantities and assessment procedures. International Organization for Standardization, Geneva, Switzerland, Norm ISO 1996-1:2016, 2016.
79. A. Sæbø: Influence of reflections on crosstalk cancelled playback of binaural sound, Ph.D. dissertation. Faculty of Information Technology, Electrical Engineering, Norwegian University of Science, and Technology, Norway, 2001.
80. M. Kohnen, J. Stienen, L. Aspöck, M. Vorländer: Performance evaluation of a dynamic crosstalk cancellation system with compensation of early reflections, in Audio Engineering Society Conference: 2016 AES International Conference on Sound Field Control, 2016, pp. 1–8.
81. ITU-R BS.1116-3: Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems. International Telecommunication Union, Geneva, Switzerland, Recommendation ITU-R BS.1116-3 (02/2015), 2015.
82. S. Pelzer, L. Aspöck, D. Schröder, M. Vorländer, Interactive real-time simulation and auralization for modifiable rooms, Building Acoustics 211 (2014) 65–73. https://doi.org/10.1260/1351-010X.21.1.65.
83. B. Seeber, S. Kerber, E. Hafter: A system to simulate and reproduce audio-visual environments for spatial hearing research. Hearing Research 260, 1–2 (2010) 1–10. https://doi.org/10.1016/j.heares.2009.11.004.
84. B. Rafaely: Plane-wave decomposition of the sound field on a sphere by spherical convolution. The Journal of the Acoustical Society of America 116, 4 (2004) 2149–2157. https://doi.org/10.1121/1.1792643.
85. R.C. Heyser: Instantaneous intensity, in Audio Engineering Society Convention 81. 1986.
86. H. Hacıhabiboğlu: Theoretical analysis of open spherical microphone arrays for acoustic intensity measurements. IEEE/ACM Transactions on Audio, Speech, and Language Processing 22, 2 (2014) 465–476.

Cite this article as: Pausch F, Behler G & Fels J. 2020. SCaLAr – A surrounding spherical cap loudspeaker array for flexible generation and evaluation of virtual acoustic environments. Acta Acustica, **4**, 19.