Transducer Distribution on Spherical Arrays for Ambisonics Recording and Playback

Daniel Pinardi
Dept. of Engineering and Architecture
University of Parma
Parma, Italy
daniel.pinardi@unipr.it

Angelo Farina
Dept. of Engineering and Architecture
University of Parma
Parma, Italy
angelo.farina@unipr.it

Marco Binelli
Dept. of Engineering and Architecture
University of Parma
Parma, Italy
marco.binelli@unipr.it

***Abstract*—Microphone and loudspeaker arrays are nowadays more and more employed in several applications, such as automotive industry, entertainment, immersive teleconferencing, or remote assistance. The position of the transducers over the surface of the array has a great influence on the beamforming, and so on the spatial performance. In this paper, a recurring geometrical problem is discussed: choosing the optimal locations of transducers for spherical arrays, either microphones or loudspeakers.**

None of the existing systems is currently relying on a spherical design, or t-design, for the distribution of the sampling points over the sphere. It will be shown that such mathematically optimized geometry is an optimal solution for the design of spherical arrays. They are the only known geometries ensuring a lossless transformation back and forth between the two most common spatial audio format: Ambisonics, which makes use of spherical harmonics, and Spatial PCM Sampling, which relies on unidirectional, high directivity virtual microphones.

Keywords—Ambisonics, loudspeaker array, microphone array, spherical harmonics, spatial PCM sampling, spherical design

I. INTRODUCTION

The use of spherical arrays of microphones and loudspeakers has a long history in the field of "spatial audio", dating back to the seventies, when the Soundfield microphone was invented [1] and the first spherical loudspeaker rigs were built, based on the Ambisonics theory [2]. The original Ambisonics approach made use of only four channels for encoding the acoustic spatial information into four signals, corresponding to virtual microphones having directivity patterns described by the Spherical Harmonics (SH) of order 0 and 1 [3]. For improving the spatial resolution of this approach, named First Order Ambisonics (FOA), was later extended to High Order Ambisonics (HOA), including more channels corresponding to virtual microphones having directivity patterns described by SH of order 2, 3, and so on. Due to an initial lack of standardization, many Ambisonics formats were developed with different channels ordering and normalizations. The current standard is named "AmbiX" and defines the Ambisonics Channel Numbering (ACN) and gain-scaling rules (SN3D) for any order [4]. Due to practical constraints, Ambisonics is usually limited to 7th order (64 channels).

For capturing HOA signals, a massive microphone array is required. Several prototypes [5]–[9] have been built, with a number of capsules ranging between 8 and 252. In 2009, the first commercial, studio-quality spherical microphone array was launched: the Eigenmike32™ (EM32 in the following), based on the pioneering work of Elko [10] and featuring 32 capsules arranged over the surface of a rigid sphere having a diameter of 84 mm. More recently, other microphone arrays appeared on the market, such as the Zylia, featuring 19 Micro Electro-Mechanical Systems (MEMS) microphones, the new Eigenmike64 equipped with 64 microphones, the Octamic, the Brahma-8, and the Voyage Audio, featuring eight capsules arranged accordingly to the optimized geometry proposed by Benjamin in [11].

Regarding the reproduction, several loudspeaker rigs were built, albeit only a few of them were spherical. At least the following three must be cited: the 4-loudspeakers rig initially employed by Gerzon and Fellgett [2], which performed quite poorly. The 8-loudspeakers rig (cube) developed by one of the authors in the nineties [12], which also performed sub-optimally with FOA and the 40-loudspeakers spherical rig built at the Institute of Sound and Vibration Research (ISVR), which demonstrated good performance [13].

In most cases, the transducers' location for spherical arrays was guided more by intuition than a rigorous mathematical theory. Of course, a "sampling theorem" must hold also in space, hence the number of transducers had to be in some way proportional to the Ambisonics order and the usage of regular or quasi-regular polyhedrons for large numbers of transducers appeared to be a sensible approach, minimizing the unavoidable "sampling error". It was soon understood that the four vertices of a tetrahedron were a good choice for FOA. Widely accepted choices for HOA were the 12 faces of a dodecahedron for the 2nd order (9 channels) and the 20 faces of an icosahedron for the 3rd order (16 channels). Furthermore, EM32 and other 32-channel arrays used a combination of the previous two: vertices and centers of faces of a dodecahedron, equivalent to centers of faces of a truncated icosahedron.

In this paper, the mathematical basis for the choice of the sampling points on a sphere with the aim of reconstructing correctly the SH signals up to a given order are investigated. It will be shown that, except for the first and second orders, the above assumptions were wrong, and that the choice of the sampling points should follow the "t-design" geometries. T-design is a mathematical theory developed in the seventies [14] to solve the problem of numerical integration on multiple dimensions. It ensures sampling a function over a spherical surface in a way that SH components are not affected by any sampling error. Furthermore, the triple connection between Ambisonics, SPS, and t-design is extensively treated. Despite many papers already dealt with t-design distributions for the measurement of microphone arrays or for performing transformation and conversion of Ambisonics signals [15]–[17], the usage of t-design geometries for converting to SPS domain is still not rigorously formalized.

979-8-3503-1104-4/23/

The paper is arranged as follows: Section II provides the theoretical backgrounds, Section III describes the usage of t-design for lossless reconstruction of SH up to a given order, and in Section IV t-designs are compared to existing microphone arrays. Section V provides instructions for the practical usage of T-designs, while Section VI summarizes the conclusions.

II. Definitions and Theory

This Section provides definitions and theoretical basis for Ambisonics, SPS, and t-design. Several multichannel audio formats are also defined (A-format, B-format, P-format, and T-format), and the conversion method from one to another is explained.

A. Ambisonics

Ambisonics is a method for representing acoustical spatial information in one point in space. Sound arrives in an observing point from many directions, and each contribution corresponds to a traveling wavefront. These wavefronts are not necessarily plane waves: they can exhibit significant curvature, depending on the distance from the source, the type of sound source (point source, line source, surface source, etc.) and its extension. Several of these wavefronts are not radiated directly by the sound sources but have been reflected, scattered, or diffracted from objects or boundaries. This complex mixture is represented with a reduced number of signals, conceptually obtained by placing at the observing point several coincident virtual microphones, characterized by directivity patterns corresponding to SH. These are basis functions with orthonormal properties for the Fourier transform on a sphere [18]. The SH of n^{th} order and m^{th} degree is defined by (1) [3]:

$$Y_n^m(\vartheta,\varphi) \equiv \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}} P_n^m(\cos\vartheta) e^{im\varphi} \qquad (1)$$

where ϑ and φ denote respectively azimuth and elevation angles (ISO 2631 compliant reference system), and P_n^m are the associated Legendre functions. In [19], the authors made available an explicit formulation of the SH up to 7th order. The SH of order 0 is a sphere, corresponding to an omnidirectional microphone, which is effectively a sound pressure sensor. At 1st order, the 2nd, 3rd, and 4th SH are "figure-of-8" microphones, usually existing in the form of "pressure gradient" or "particle velocity" microphones. If the wavefront is a perfectly progressive plane wave, the sound pressure is directly proportional to the particle velocity. When instead the wavefront has some curvature, the value of particle velocity diverges from the value of sound pressure, with a frequency-dependent gain and phase variation, namely "proximity effect". Existing microphone arrays are rather affected by proximity effect and other similar problems occurring at higher orders. This is one of the reasons causing a mismatch between synthetically produced Ambisonics streams, which are based on the theoretical SH formulas (1) and recordings performed with real microphone arrays. The latter are "correct" only at a given distance from the sound sources. This distance can be infinite if the theoretical model of plane wave diffraction is employed, or it can be the finite distance at which the microphone array was calibrated inside an anechoic room or by means of numerical simulations [20].

Regarding microphone arrays, the basic "Ambisonics encoding" problem comes out: raw signals captured by the capsules (usually known as "A-format") must be converted into Ambisonics format, corresponding to the mathematical definition of the SH signals (usually known as "B-format"). This process is called "encoding" and in the specific case of a microphone array it is also known as "A2B conversion". This conversion procedure can be approached in two different ways:

- Linear processing. A matrix of filters which synthesize the required number of virtual microphones from the capsule signals.
- Parametric processing. The capsule signals are employed for performing a parametric spatial analysis of the sound field, extracting the single source signals and their locations, and then the theoretical SH formulas are used for recreating the B-format signals.

The presented work will make use of the linear processing, which was dominant at least until 2018, when a software tool capable of performing parametric A2B conversion was released [21]. When considering the linear processing, there are three possible approaches for deriving the A2B filter matrix:

- Theoretical approach. It makes use of the analytical solution of the equations describing the interaction between sound waves and the microphone array [22]–[25].
- Experimental approach. The microphone array is measured in an anechoic chamber by many directions and the A2B filter matrix is numerically calculated [26], [27].
- Numerical approach. The diffraction of sound waves against the surface of the array is solved with Finite Elements Method (FEM), Boundary Elements Method (BEM), or geometrical acoustics simulations [28], [29].

The theoretical approach is constrained to a few geometries (sphere, cylinder, plane), for which the analytical solution of wave equation is available, whilst the other two approaches can be applied to irregular geometries too. However, the choice of the geometry must consider some acoustic constraints (i.e., a large array works better at low frequencies and worse at high frequencies and vice versa) and some irregular geometries can be tailored towards "mixed-order Ambisonics", improving spatial resolutions in specific regions of the spherical horizon [30], [31].

The Ambisonics signals can be reproduced, using either a loudspeaker array or binaurally over headphones. The two cases are indeed not so different, as usually also for headphone reproduction a virtual loudspeaker array is employed, so that the SH to Binaural processing is conceptually split into a SH to virtual loudspeakers rendering, namely Speaker Feed (SF), followed by the convolution with the Head Related Transfer Functions (HRTF) between the virtual loudspeakers and a binaural dummy head [32]. Therefore, the reproduction of Ambisonics signals over an array of (real or virtual) loudspeakers surrounding the listener is now considered. Also in this case, there are two possible approaches:

- Linear processing. A matrix of filters is employed to process the B-format signals creating the SF.
- Parametric processing. The B-format signals are employed for performing a parametric spatial analysis of the sound scene, extracting the single source signals and their locations and a "diffuse field", and then panning these signals on the loudspeakers, i.e., by using VBAP [33].

One of the convenient features of Ambisonics is the possibility to deal with irregular loudspeaker arrays, usually

caused by geometrical constraints of the room and loudspeakers support structure. Advanced Ambisonics decoders can apply delay and gain compensations in case the loudspeakers are neither at the same distance from the listening point nor uniformly distributed over the sphere [11], [34], [35]. These advanced decoders can even perform some "distance compensation", taking care of the "proximity effect" occurring when the loudspeaker rig is small compared to the wavelength. However, large spherical loudspeaker arrays remain the reference when maximum performance from every direction is required and they are the natural choice for computing virtual loudspeaker signals for binaural rendering. Albeit modern Ambisonics decoders can deal with arbitrary locations of loudspeakers around the listener, in this paper the potential benefits of a regular arrangement of loudspeakers on a sphere according to the optimized t-design distributions are explored. For this assessment a traditional linear decoding was used.

B. Spatial PCM Sampling

Spatial PCM Sampling (SPS) [36], was developed as the spatial equivalent to Pulse-Code Modulation (PCM) sampling of a time-domain waveform (sequence of impulses), exactly as Ambisonics is the spatial equivalent of the Fourier representation (superimposition of sinusoids). In practice, it is an alternative to SH, making use of basis functions emulating unidirectional microphones approximating a "spatial Dirac's Delta function", covering uniformly the surface of a sphere, for capturing the complete spatial information [37]. Raw signals coming from the capsules (A-format) are converted into SPS (P-format) by applying a matrix of Finite Impulse Response (FIR) filters, which synthesize unidirectional virtual microphones having high order cardioid directivity.

SPS signals can be directly reproduced, without any decoder, by employing a loudspeaker rig, either real or virtual. I.e., Mach1 [38], an eight-channel SPS pointing at the vertices of a cube, was developed for spatial audio rendering in virtual reality applications as an alternative to 1st order Ambisonics.

C. T-design and spherical sampling

Spherical t-designs have been introduced by Delsarte et al. [14]. They proposed to approximate a unit sphere in R^n with a finite set of points, namely a spherical t-design, so that an integral over this sphere of a polynomial of degree t or less is equal to the average value of the polynomial evaluated in the set of chosen points. In other words, spherical t-designs provide equal weight quadrature rules on a sphere [39]. Hence, in R^3 a set of N points on a unit sphere S^2 is a spherical t-design if, for any polynomial $P_t(\varphi, \theta)$ of degree at most t, it is satisfied:

$$\int_S P_t(\varphi, \theta) \cdot dS = \frac{1}{N} \cdot \sum_{i=1}^{N} P_t(\varphi_i, \theta_i) \qquad (2)$$

Therefore, given a certain value of N, the aim is to choose the positions of N points that maximize t. This means that such an optimal choice of the positions of these N points ensures capturing the spatial information up to a maximally high spatial frequency. Such t-design application is groundbreaking for microphone arrays design.

Spherical t-designs are generally not unique, although some very rare cases were found to be rigid [40]. The rigid designs are unique for given t and N up to an orthogonal transformation. Another important definition is a tight spherical t-design, meaning it has a minimum cardinality |X|=N. The cardinality of a t-design on S^2 is limited by the following inequality [13]:

$$\begin{cases} |X| \geq (n+1)^2 \\ |X| \geq (n+1)(n+2) \end{cases} \qquad (3)$$

A tight t-design occurs when the equality in (3) holds. Tight designs are also rigid designs [41], meaning they are unique and have the best combinatorial properties at the same time. For the case of R^3, only the following rigid tight designs were identified (corresponding to Platonic solids):

1-design that consists of 2 antipodal points.
2-design that consists of 4 points (a tetrahedron).
3-design that consists of 6 points (an octahedron).
5-design that consists of 12 points (a dodecahedron).

Larger t-designs exceeding the Platonic solids are computed numerically instead of being analytically explicit. More on existence and construction of spherical t-designs can be found in [39], [42]–[45]. The techniques used to compute t-designs are beyond the scope of this paper. Many of the sets were computed and made publicly available in [46], for which the authors express their gratitude to Hardin and Sloane. They found the solutions for N ranging between 1 and 100, showing that, despite a general trend of N increasing with t, some optimal solutions exist, resulting in a large value of t with a relatively small number of sampling points N on the unit sphere. The most relevant t-designs in terms of acoustical applications are the 24-point 7-design (Maclaren's improved snub-cube), the 36-point 8-design and the 60-point 10-design. It is interesting to note that these remarkable geometries do not correspond to known Archimedean solids, being instead the result of numerical optimization, e.g., the Archimedean snub cube, which has 24 vertices, is only a 3-design. Neither do they belong to rigid t-designs, meaning the disposition of points for each of them is not unique. Furthermore, it appears that all possible t-design geometries up to (N=240, t=21) have been already found, and some have been found reaching (N=100200, t=1000) [39]. Therefore, t-design geometries with a suitable number of sampling points can be found for any practical purpose. These geometries are of particular interest for SPS format, as they allow maximizing the extension of the spatial spectrum towards higher orders for a given number of sampling points.

While a spherical t-design distribution is a solution to the equal weights problem, there are other ways of sampling points for the best approximation of functions on a sphere. For instance, *equal-angle sampling* consists in sampling points at uniformly spaced angular positions along θ and φ with uniform angular spacing. As it can be found in [47], this kind of sampling requires a number of points given by:

$$N = 4 \cdot (o+1)^2 \qquad (4)$$

where *o* is the Ambisonics order. The *Gaussian sampling* method is based on the Gauss-Legendre quadrature rule and provides equidistant sampling along the azimuth. According to this method the points result more densely distributed closer to the poles [47]. The number of points for a given order *o* is in this case:

$$N = 2 \cdot (o+1)^2 \qquad (5)$$

Unlike with t-designs, there is no equal weighing of the sampling points, but the error due to modal aliasing for lower number of points can be analyzed in closed form [48]. These two methods, along with *Lebedev grids* or *equirectangular*

grids, do not provide a uniform distribution of sampled points on the surface of a sphere. The *purely uniform sampling* can be fully achieved only with the five Platonic solids, the largest set being the icosahedron, which is the 20-point 5-design (the dodecahedron is also a 5-design with N=12, providing O=2). Sets with larger number of points can only be *quasi-uniform*, an example here being the truncated icosahedron with 32 faces, N=32 points in the center of the faces, used in the EM32 [10]. Besides t-designs, quasi-uniform sampling methods include the *minimum energy Fliege-Maier sampling* and *equal-area triangulation*.

In [48], the truncated icosahedron with a 32-point 7-design and a set obtained via the Gaussian sampling method are compared. However, no substantial difference was pointed out in terms of aliasing. The theoretical minimum number of points required to decompose the sound field up to a given order is shown in Table I for three sampling methods, emphasizing the apparent advantage of t-designs, which, however, is achieved only when the optimal t-designs are chosen. The drawback of t-designs is that these optimal sets are not available for any number of desired points.

TABLE I. MINIMUM NUMBER OF POINTS REQUIRED WITH DIFFERENT SAMPLING METHODS FOR A GIVEN ORDER OF SH

SH order O	Equal-angle $N = 4(o+1)^2$	Gaussian $N = 2(o+1)^2$	t-design N not defined explicitly
1	16	8	4
2	36	18	12
3	64	32	24
4	100	50	36

D. Spherical sampling theorem

In time domain, the sampling theorem ensures that PCM sampling and reconstruction create a perfect copy of the original waveform. This happens if two conditions are met:

- The sampling frequency must be at least twice the maximum frequency contained in the original signal.
- Each sample represents the gain applied to a band limited Sync function.

The second requirement is often neglected, and each sample is instead represented as a Dirac's Delta function (a perfect pulse). A similar case occurs in frequency domain: the spectrum must be divided into bands by means of a filter bank and then all the band-passed signals must be summed together resulting in a perfect reconstruction of the original signal. Each band-pass filter must be defined so that the filtered signals are perfectly reconstructed after summation. Hence, the same must hold also in spatial domain: SPS can be fully reconstructive only if spatial band limited filters (polar patterns) are used for the virtual microphones employed for spatial analysis and reconstruction. In the original SPS formulation, these polar patterns were assumed to be high order cardioids, without any side or rear lobes, defined by the following formula as a function of the angle between the aiming direction of the virtual microphone and the Direction-of-Arrival (DoA) of the sound wave:

$$\boldsymbol{A}(\vartheta, \varphi) = [0.5 + 0.5 \cdot \cos(\vartheta) \cdot \cos(\varphi)]^o \qquad (6)$$

where ϑ and φ denote respectively azimuth and elevation angles, and O is the cardioid order, to be chosen to sample uniformly the unit sphere, depending on the number N of SPS signals. In the case of N=32, using the EM32 directions, it was O=4. Fig. 1 (left) shows the cross-section of the polar pattern of a 4th order cardioid, which in practice corresponds to the virtual microphone feeding each loudspeaker when an "in-phase" decoding scheme is employed.

These 4th order cardioids work well for performing SPS analysis through color maps of the spatial distribution of sound energy [49], [50], as the absence of side and rear lobes avoids false spots. However, these polar patterns are not perfectly reconstructive. Instead, a t-design arrangement of SPS signals becomes fully reconstructive if the virtual microphones have polar patterns corresponding to those employed in the Sampling Ambisonics Decoder (SAD) of order O [51], defined by:

$$Q(\vartheta) = \frac{1}{O+1} \cdot \left[1 + \sum_{i=1}^{O} \cos(i \cdot \vartheta) \right] \qquad (7)$$

In [47], it is suggested that the maximum order O being sampled for a given t-design geometry is related to the value of t by:

$$O = \mathrm{int}\left(\frac{t}{2} \right) \qquad (8)$$

Hence, a perfect reconstruction of the 4th order SH can be achieved with a 36-point 8-design (N=36, t=8), as reported in Section III. Employing a larger value of t for an order O is beneficial in terms of spatial accuracy. Fig. 1 (right) shows the polar pattern corresponding to (7) for the 4th order SAD: it exhibits some side and rear lobes but has a much sharper main lobe than the corresponding 4th order cardioid.

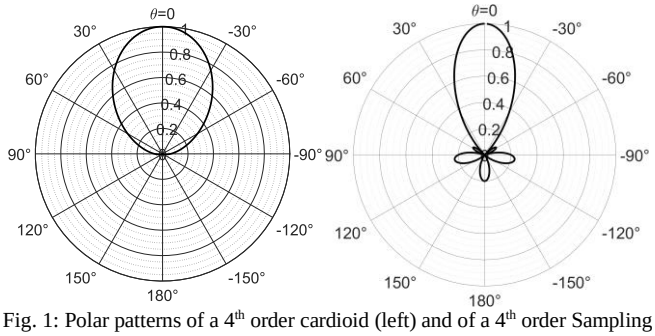

Fig. 1: Polar patterns of a 4th order cardioid (left) and of a 4th order Sampling Ambisonics Decoder (right).

When sound waves impinging on the observing point are captured with an SPS recording system employing a t-design geometry and the virtual microphones have these SAD polar patterns, the resulting multichannel P-format can be renamed as T-format. Hence, we can say that T-format is a particular case of P-format, occurring when the choice of the sampling directions follows a t-design geometry, and the polar patterns are SAD of order O given by (8).

III. OPTIMAL T-DESIGN GEOMETRIES

Based on the available t-design distributions [46] presented in Section I and accordingly to (8), the minimum arrangements of SPS components of Table II are necessary for recording or playing back an Ambisonics B-format signal of order O. One can note the most desirable configurations are the 24-point 7-design and 36-point 8-design: they enable the higher-order decomposition (3rd and 4th order, respectively) in relation to the number of points. Note that the 32 channels of EM32 are not enough for capturing SH of 4th order and the 19 channels of the Zylia are not sufficient even for 3rd order.

Anyway, their geometrical arrangements do not correspond to a t-design (Fig. 5 and Fig. 7).

TABLE II. MINIMUM GEOMETRY FOR A GIVEN AMBISONICS ORDER O

O	t	N (channels)	Geometry
1	2	4	Tetrahedron
2	5	12	Icosahedron
3	7	24	7-design
4	8	36	8-design
5	10	60	10-design
6	12	84	12-design
7	14	108	14-design

The double conversion from Ambisonics to SPS, and from SPS to Ambisonics (B-format to T-format to B-format) can be employed to assess if the SPS geometry entails a spatial information loss (hence it is a P-format instead of a perfectly reconstructing T-format), resulting in an imperfect reconstruction of the original signals. This is assessed by relying on the orthonormality error matrix, defined as:

$$\boldsymbol{D} = 20\log(|\boldsymbol{Y}^t\boldsymbol{Y}|) \qquad (9)$$

where $\boldsymbol{Y}$ are the sampled spherical harmonics and ^{t} denotes the transpose. In the following figures, the orthonormality error between two sampled spherical harmonics is represented by a square.

It was opted to start testing t-designs with a suboptimal configuration, that is 20-point 5-design for 3rd order Ambisonics signals. The conversion to T-format is performed by sampling the set of SH with 20 points over the sphere, then the original 16 signals were reconstructed by encoding back the SPS signals to Ambisonics. Fig. 2 (left) shows the result (color scale in dB). One can note that despite placing 20 transducers in a t-design arrangement (vertices of a dodecahedron), the 3rd order signals are not lossless reconstructed. The result is instead perfect up to the 2nd order, as theoretically predicted (O=2). However, the optimal configuration for the 2nd order reconstruction is the 12-point 5-design (icosahedron), which yields the same result as the 20-point one. The procedure is then repeated for the 24-point 7-design, which exhibits a perfect reconstruction up to the 3rd order, as shown in Fig. 2 (right).

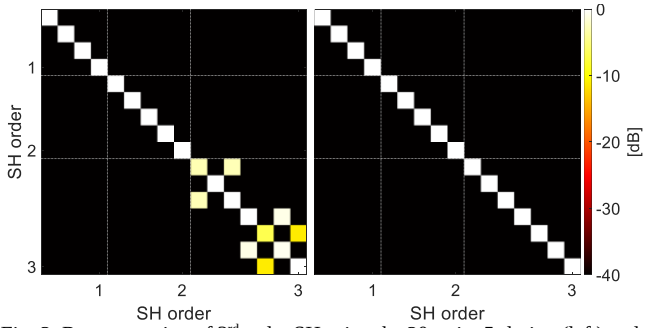

Fig. 2: Reconstruction of 3rd order SH using the 20-point 5-design (left) and the 24-point 7-design (right).

Further t-design geometries for HOA were processed. Fig. 3 presents the result for the 32-point 7-design: it is not enough for the full 4th order Ambisonics but works perfectly for the 3rd order. Therefore, it should be the ideal arrangement for a "T-Eigenmike32", since it ensures perfect reconstruction of each SH without any leakage, which instead occurs with the original EM32's geometry.

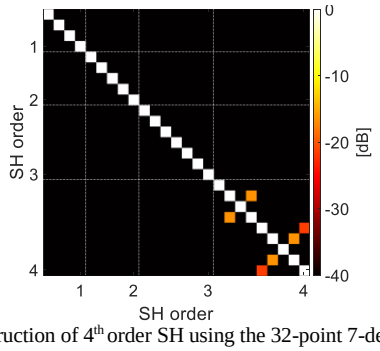

Fig. 3: Reconstruction of 4th order SH using the 32-point 7-design.

Additional optimal t-design geometries tested are the 36-point 8-design and the 60-point 10-design, which can be employed for encoding Ambisonics streams of 4th and 5th order, respectively. The reconstruction of the original Ambisonics streams resulted perfect also in these cases, as shown in Fig. 4.

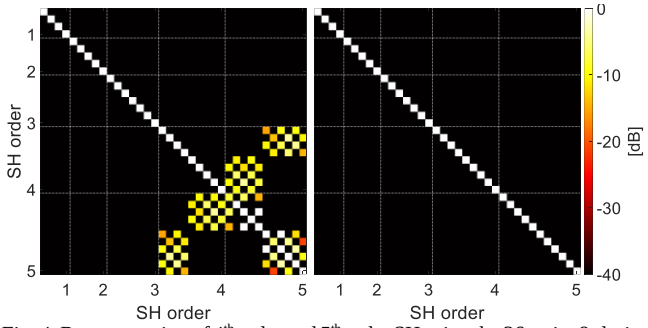

Fig. 4: Reconstruction of 4th order and 5th order SH using the 36-point 8-design (left) and the 60-point 10-design (right).

IV. COMPARISON OF EXISTING MICROPHONE ARRAYS

Several microphone arrays with known capsule locations are now theoretically evaluated, repeating the procedure of sampling SH with a set of SPS virtual microphones, then using the resulting P-format signals to reconstruct the original SH signals. In this way, the sampling/reconstruction error is quantified. In the following, results are shown for Zylia (19 mics), Eigenmike32™ (32 mics), Bruel&Kjaer type WA-1565-W-020 (36 mics), ISVR (40 mics), and Bruel&Kjaer type WA-1565-W-021 (50 mics).

Fig. 5 shows the 19 locations of Zylia microphones in comparison with a spherical 19-point 4-design, whereas Fig. 6 shows the reconstruction errors for the SH up to 3rd order for both geometries. The 19 sampling points of Zylia microphone array are not uniformly scattered over the surface of the sphere, resulting in large errors even at 1st order. Employing instead the same number of capsules with a t-design arrangement (N=19, t=4), the result is perfect up to the 2nd order, as expected, and presents errors only at the 3rd order.

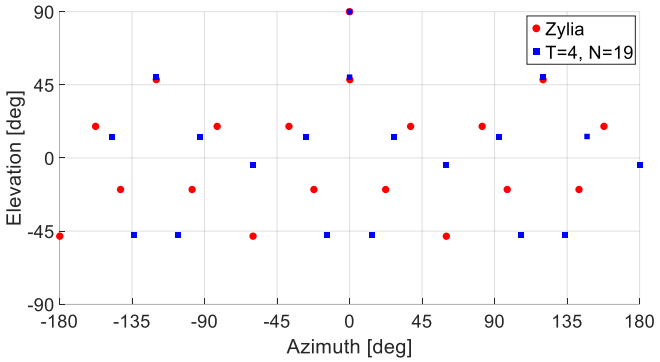

Fig. 5: Equirectangular chart of Zylia's microphone locations (red) vs. 19-point 4-design (blue).

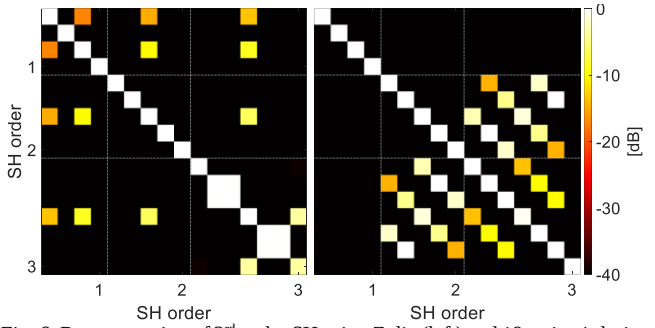

Fig. 6: Reconstruction of 3rd-order SH using Zylia (left) and 19-point 4-design (right).

Then, the case of EM32 is evaluated. Fig. 7 shows that the 32 capsules of EM32 are not positioned in accordance with the optimal 32-point 7-design. The latter is not enough for the 4th order Ambisonics but provides perfect reconstruction up to 3rd order (Fig. 8, right). Note that EM32's arrangement is obtained intersecting two regular polyhedrons, an icosahedron, and a dodecahedron (resulting in a truncated icosahedron – a soccer ball). Except for 8 points (vertices of a cube), the correspondence of the two arrangements is not precise, causing significant errors at 3rd order, and some small errors are present already in the 2nd order (Fig. 8, left).

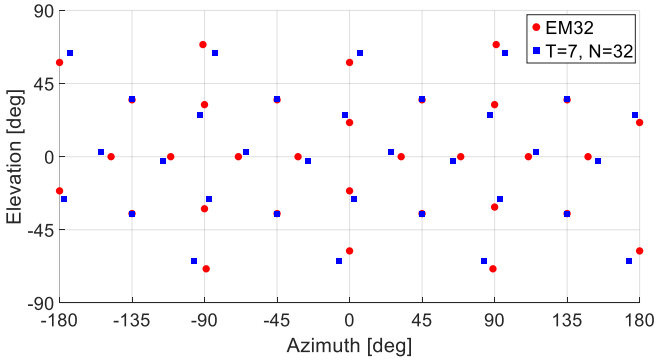

Fig. 7: Equirectangular chart of EM32 microphone locations (red) vs. 32-point 7-design (blue).

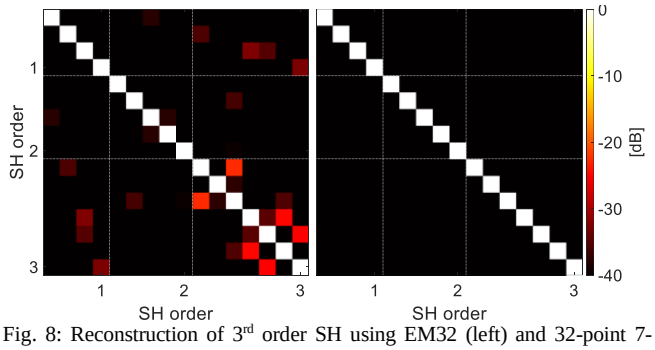

Fig. 8: Reconstruction of 3rd order SH using EM32 (left) and 32-point 7-design (right).

It can be concluded that, despite using more than 24 microphones, which is the minimum number required in a t-design geometry for perfect reconstruction of SH up to 3rd order, the EM32 gives some sampling errors even at 2nd order. However, as it was already shown by previous research [26], [27] when an experimental approach, based on measurements, is preferred with respect to the theoretical approach, based on analytical formulas, the EM32 can provide robust 3rd order signals in a wide frequency range.

Then, Bruel&Kjaer array with 36 microphones (B&K36) and 195 mm in diameter is analyzed. A 36-point 8-design exists, and it allows for a perfect reconstruction of SH up to 4th order. In Fig. 9, B&K36 microphone locations are plotted against the corresponding optimal t-design, and the result of reconstructing SH signals up to 4th order is presented in Fig. 10. Significant errors at all orders can be observed for B&K36 in comparison to the t-design.

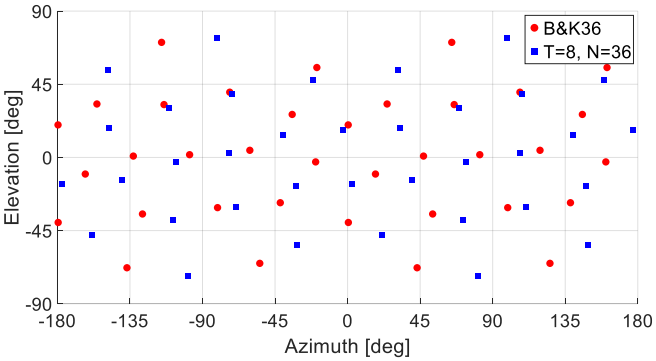

Fig. 9: Equirectangular chart of B&K36's microphone locations (red) vs. 36-point 8-design (blue).

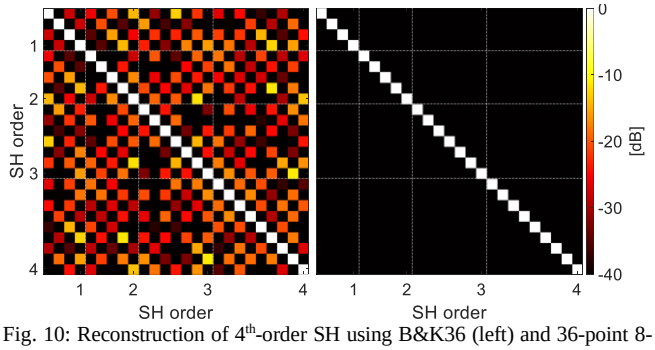

Fig. 10: Reconstruction of 4th-order SH using B&K36 (left) and 36-point 8-design (right).

The next microphone array is by ISVR, with 40 capsules (ISVR40). In Fig. 11, the ISVR40's arrangement is compared to the 40-point 8-design, and the results for 4th order reconstruction are shown in Fig. 12, with significant errors at all orders for ISVR40.

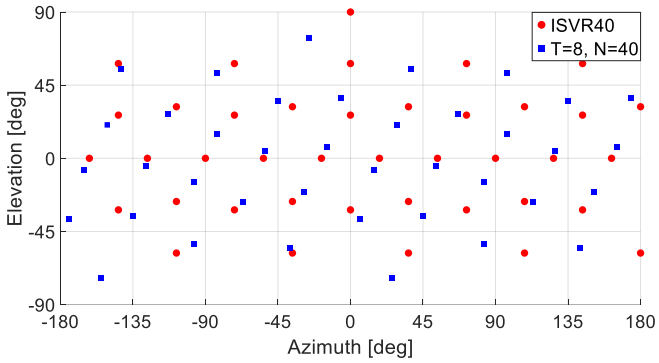

Fig. 11: Equirectangular chart of ISVR40's microphone locations (red) vs. 40-point 8-design (blue).

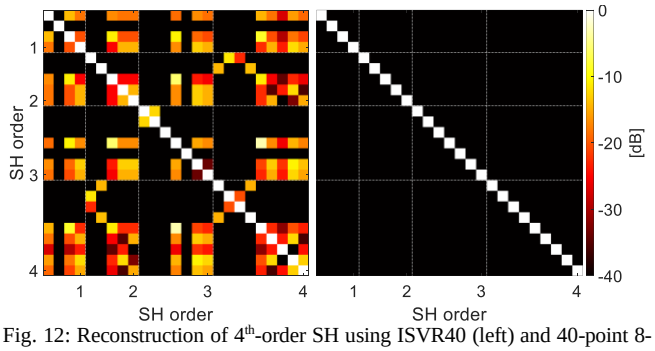

Fig. 12: Reconstruction of 4th-order SH using ISVR40 (left) and 40-point 8-design (right).

Eventually, the 50-capsule arrangement by Bruel&Kjaer (B&K50), having a diameter of 195 mm. In Fig. 13, B&K50 microphone locations in comparison to the 50-point 9-design are plotted, and Fig. 14 provides the results for 4th order reconstruction, still with significant errors for B&K50 except of the 1st order.

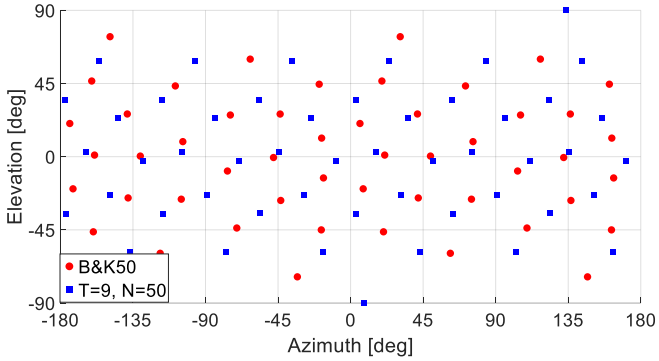

Fig. 13: Equirectangular chart of B&K50's microphone locations (red) vs.50-point 9-design (blue).

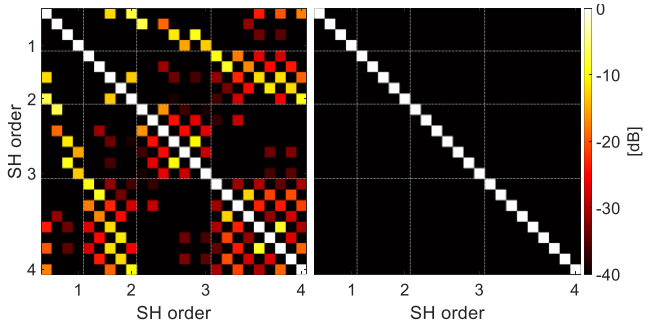

Fig. 14: Reconstruction of 4th order SH using B&K50 (left) and 50-point 9-design (right).

V. Practical Usage of T-Designs

The main practical result of this paper is the formalization of a new T-format multichannel audio stream, which can be seen as a complete alternative to the traditional B-format (Ambisonics). T-format is slightly less efficient than B-format, as it requires a larger number of channels for transporting the same spatial information, as shown in Table III.

TABLE III. Number of channels as a function of Ambisonics order O

Order O	N. channels (B-format)	N. channels (T-format)
1	4	4
2	9	12
3	16	24
4	25	36
5	36	60
6	49	84
7	64	108

However, T-format shares the same advantages already found for P-format, making it easier to perform some operations such as spatial equalization and selective suppression of disturbing noise sources. In addition, it is potentially more robust to channel-independent nonlinear processing tools, such as parametric reverberation, noise reduction, multi-band compression, de-clicking, etc. Furthermore, the T-format signals, albeit representing a scene-based description of the sound field, can also be considered independent "sound object" signals. Hence, they can be transmitted, processed, and rendered employing those systems that are not friendly for scene-based signals (Dolby Atmos or DTS-X) or in subsets of MPEG-H only supporting sound objects. The capability of converting B-format to T-format and back can be seen as an open, bidirectional bridge between scene-based spatial audio and object-based spatial audio.

A. Tools for T-format processing

This Section illustrates the procedure for converting back and forth between B-format and T-format, by employing the SPARTA open-source software [52] consisting in a suite of Virtual Studio Technology (VST) plugins. The conversion from B-format to T-format is performed with an Ambisonics Decoder (AmbiDEC), while the conversion from T-format to B-format is performed with an Ambisonics Encoder (AmbiENC). Both include presets for t-design geometries having 4, 12, 24, 36, 48, and 60 channels. The choice must be made according to Table II. Fig. 15 shows how to configure AmbiDEC: SAD decoder, max-rE not active. Fig. 16 shows how to configure AmbiENC for converting T-format to B-format (3rd order in the example). The choice between Amplitude Preserving (AP) and Energy Preserving (EP) has the effect of modifying the gain of all T-format signals but has no effect on the spatial information. However, it is recommended to use EP as it requires less gain adjustment when reconstructing the original B-format signal from T-format.

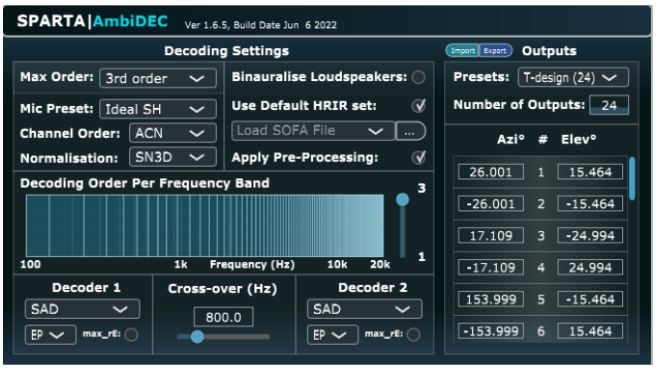

Fig. 15: B-format to T-format conversion using AmbiDEC VST plugin.

Fig. 16: T-format to B-format conversion using AmbiENC VST plugin.

These plugins also allow to perform a "spatial up-mixing" of Ambisonics signals. This is the case when a HOA soundtrack is created, e.g., starting from a FOA recording (4 channels). The traditional approach would rely on the hierarchic property of Ambisonics: the four channels of the FOA recording are fed into the first four channels of the HOA mix, over which other discrete sources are added employing a HOA encoder. However, this approach is suboptimal. Instead, the FOA recording can be converted to T-format with a FOA decoding, and then converted back to HOA with a HOA encoding. This procedure creates a spatial up-mixing by interpolating the spatial information, resulting in a HOA soundtrack, which preserves all the spatial information captured by the FOA microphone array. Even better results can be obtained if the T-format conversion is performed employing a parametric Ambisonics decoder, such as the Compass Decoder (Fig. 17) also available in [52]. This method sharpens the original spatial resolution, resulting in a HOA soundtrack having much better spatial sharpness than the original FOA soundtrack.

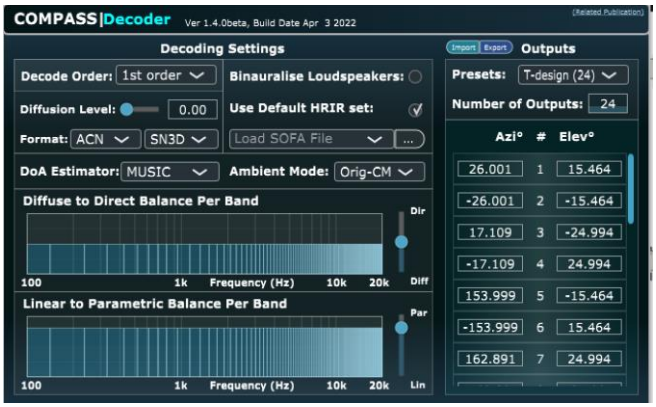

Fig. 17: Parametric conversion from B-format to T-format (spatial up-mixing) using COMPASS Decoder VST plugin.

VI. Conclusions

The relation between Ambisonics, SPS, and t-design is formalized, and a new T-format multichannel audio stream is introduced in this paper. A demonstration is given that t-design is an efficient method to define the position of transducers for an array, whether those being capsules for a microphone array or loudspeakers for a playback system. This is achieved through lossless sound field decomposition into SH up to a given order by a spherical t-design sampling, which is equivalent to perfect reconstruction of the Ambisonics format.

Several t-designs are analyzed, and optimal arrangements are suggested for spherical microphone arrays capable of encoding Ambisonics 3rd, 4th, and 5th order: 24-point 7-design, 36-point 8-design, and 60-point 10-design respectively. Several existing spherical microphone arrays (Zylia, Eigenmike32, B&K36, ISVR40, and B&K50) are studied in comparison with optimal spherical t-designs having the same number of capsules: 19-point 4-design, 32-point 7-design, 36-point 8-design, 40-point 8-design, and 50-point 9-design. None of the existing arrays proves to be perfectly reconstructive, whereas error-free reconstruction was achieved for each of the upper mentioned t-design arrangement keeping unchanged the number of capsules.

Eventually, an efficient method for back-and-forth conversion between Ambisonics and T-format is provided, relying on existing open-source software. This allows us to efficiently convert from scene-based spatial audio to object-based spatial audio and vice versa, and to perform a spatial up-mixing of low order to higher order Ambisonics soundtrack.

References

[1] M. A. Gerzon, "The Design of Precisely Coincident Microphone Arrays for Stereo and Surround Sound," in *Audio Engineering Society Convention 50*, Mar. 1975. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=2466

[2] M. J. Gerzon, "Periphony: With-Height Sound Reproduction," *Journal of The Audio Engineering Society*, vol. 21, pp. 2–10, 1973.

[3] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. San Diego: Academic Press, 1999.

[4] C. Nachbar, F. Zotter, E. Deleflie, and A. Sontacchi, "Ambix—A Suggested Ambisonics Format," Lexington: 3rd Ambisonics Symposium, Jun. 2011.

[5] S. Bertet, J. Daniel, and S. Moreau, "3D Sound Field Recording with Higher Order Ambisonics - Objective Measurements and Validation of Spherical Microphone," in *Audio Engineering Society Convention 120*, May 2006. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=13661

[6] S. Sakamoto, S. Hongo, T. Okamoto, Y. Iwaya, and Y. Suzuki, "Sound-space recording and binaural presentation system based on a 252-channel microphone array," *Acoust Sci Technol*, vol. 36, no. 6, pp. 516–526, 2015, doi: 10.1250/ast.36.516.

[7] D. Pinardi, "A Human Head Shaped Array of Microphones and Cameras for Automotive Applications," in *2021 Immersive and 3D Audio: From Architecture to Automotive, I3DA 2021*, 2021. doi: 10.1109/I3DA48870.2021.9610879.

[8] D. Pinardi *et al.*, "An Innovative Architecture of Full-Digital Microphone Arrays Over A^2B Network for Consumer Electronics," *IEEE Trans Consum Electron*, vol. 68, no. 3, pp. 200–208, Aug. 2022, doi: 10.1109/TCE.2022.3187453.

[9] D. Pinardi, A. Toscani, M. Binelli, L. Saccenti, A. Farina, and L. Cattani, "Full-Digital Microphone Meta-Arrays for Consumer Electronics," *IEEE Transactions on Consumer Electronics*, p. 1, 2023, doi: 10.1109/TCE.2023.3267836.

[10] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *IEEE International Conference on Acoustics Speech and Signal Processing*, IEEE, May 2002, pp. II-1781-II–1784. doi: 10.1109/ICASSP.2002.5744968.

[11] A. Heller, E. Benjamin, and R. Lee, “Design of Ambisonic Decoders for Irregular Arrays of Loudspeakers by Non-Linear Optimization,” vol. 2, May 2010.

[12] A. Farina and E. Ugolotti, “Subjective comparison between Stereo Dipole and 3D Ambisonics surround systems for automotive applications,” in *Proc. of 16th AES International Conference*, Rovaniemi, Apr. 1999.

[13] F. Fazi, P. Nelson, J. Christensen, and J. Seo, “Surround system based on three-dimensional sound field reconstruction,” in *Audio Engineering Society - 125th Audio Engineering Society Convention 2008*, Apr. 2008.

[14] P. Delsarte, J. M. Goethals, and J. J. Seidel, “Spherical codes and designs,” *Geom Dedic*, vol. 6, no. 3, pp. 363–388, 1977, doi: 10.1007/BF03187604.

[15] G. Battista, P. Chiariotti, and P. Castellini, “Spherical Harmonics Decomposition in inverse acoustic methods involving spherical arrays,” *J Sound Vib*, vol. 433, pp. 425–460, 2018, doi: https://doi.org/10.1016/j.jsv.2018.05.001.

[16] D. Pinardi, “Spherical t-Design for Characterizing the Spatial Response of Microphone Arrays,” in *2021 Immersive and 3D Audio: From Architecture to Automotive, I3DA 2021*, 2021. doi: 10.1109/I3DA48870.2021.9610850.

[17] D. Pinardi and A. Farina, “Metrics for Evaluating the Spatial Accuracy of Microphone Arrays,” in *2021 Immersive and 3D Audio: From Architecture to Automotive, I3DA 2021*, 2021. doi: 10.1109/I3DA48870.2021.9610887.

[18] N. M. Ferres, *An elementary treatise on spherical harmonics and subjects connected with them*. Cornell University Library, 1877.

[19] Angelo Farina, “Explicit Ambix formulas for High Order Ambisonics,” Aug. 2017. http://pcfarina.eng.unipr.it/Aurora/HOA_explicit_formulas.htm

[20] D. Pinardi, “Spherical Wave Diffraction for Microphone Arrays Operating in Near Field,” in *2023 Immersive and 3D Audio: From Architecture to Automotive, I3DA 2023*, Bologna, Sep. 2023.

[21] Rode, “Soundfield by Rode.” https://rode.com/it/software/soundfield-by-rode (accessed May 31, 2023).

[22] L. McCormack, S. Delikaris-Manias, A. Farina, D. Pinardi, and V. Pulkki, “Real-time conversion of sensor array signals into spherical harmonic signals with applications to spatially localised sub-band sound-field analysis,” in *144th Audio Engineering Society Convention*, Milan, May 2018, pp. 294–303. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=19456

[23] L. McCormack *et al.*, “Applications of spatially localized active-intensity vectors for sound-field visualization,” *AES: Journal of the Audio Engineering Society*, vol. 67, no. 11, 2019, doi: 10.17743/JAES.2019.0041.

[24] IEM, “IEM Plug-in Suite.” https://plugins.iem.at/download/ (accessed May 31, 2023).

[25] R. O. Duda and W. L. Martens, “Range dependence of the response of a spherical head model,” *J Acoust Soc Am*, vol. 104, no. 5, pp. 3048–3058, Oct. 1998, doi: 10.1121/1.423886.

[26] A. Farina, A. Capra, L. Chiesi, and L. Scopece, “A Spherical Microphone Array for Synthesizing Virtual Directive Microphones in Live Broadcasting and in Post Production,” in *40th International AES Conference: Spatial Audio: Sense the Sound of Space*, Oct. 2010. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=15577

[27] A. Farina, S. Campanini, L. Chiesi, A. Amendola, and L. Ebri, “Spatial sound recording with dense microphone arrays,” in *AES 55th International Conference*, Helsinki, Finland, 2014, pp. 1–8. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=17362

[28] D. Pinardi, A. Farina, and J.-S. Park, “Low Frequency Simulations for Ambisonics Auralization of a Car Sound System,” in *2021 Immersive and 3D Audio: From Architecture to Automotive, I3DA 2021*, 2021. doi: 10.1109/I3DA48870.2021.9610959.

[29] D. Pinardi, K. Riabova, M. Binelli, A. Farina, and J.-S. Park, “Geometrical Acoustics Simulations for Ambisonics Auralization of a Car Sound System at High Frequency,” in *2021 Immersive and 3D Audio: From Architecture to Automotive, I3DA 2021*, 2021. doi: 10.1109/I3DA48870.2021.9610977.

[30] J. Chang and M. Marschall, “Periphony-Lattice Mixed-Order Ambisonic Scheme for Spherical Microphone Arrays,” *IEEE/ACM Trans Audio Speech Lang Process*, vol. 26, pp. 924–936, 2018.

[31] S. Favrot, M. Marschall, J. Kasbach, J. Buchholz, and T. Weller, “Mixed-Order Ambisonics Recording and Playback for Improving Horizontal Directionality,” May 2011.

[32] M. Binelli, D. Pinardi, T. Nili, and A. Farina, “Individualized HRTF for playing VR videos with Ambisonics spatial audio on HMDs,” in *Proceedings of the AES International Conference*, 2018.

[33] V. Pulkki, “Virtual Sound Source Positioning Using Vector Base Amplitude Panning,” *Journal of The Audio Engineering Society*, vol. 45, pp. 456–466, 1997.

[34] P. W. M. Tsang and K. W. K. Cheung, “Development of a re-configurable ambisonic decoder for irregular loudspeaker configuration,” *Circuits, Devices & Systems, IET*, vol. 3, pp. 197–203, May 2009, doi: 10.1049/iet-cds.2009.0007.

[35] D. Arteaga, “An Ambisonics Decoder for Irregular 3-D Loudspeaker Arrays,” in *134th AES Convention*, Rome, Italy: Audio Engineering Society, May 2013.

[36] A. Farina, A. Amendola, L. Chiesi, A. Capra, and S. Campanini, “Spatial PCM Sampling: A New Method For Sound Recording And Playback,” in *52nd international conference of Audio Engineering Society*, Guildford, UK, 2013, pp. 1–12.

[37] A. Farina, A. Amendola, A. Capra, and C. Varani, “Spatial Analysis of Room Impulse Responses Captured with a 32-Capsule Microphone Array,” pp. 13–16, May 2011.

[38] MACH1, “VVBP Standards for Spatial Audio and Agnostic Format Conversions.” https://mach1-public.s3.amazonaws.com/research/Mach1SpatialSystem-WhitePaper_180523.pdf (accessed May 31, 2023).

[39] M. Gräf and D. Potts, “On the computation of spherical designs by a new optimization approach based on fast spherical Fourier transforms,” *Numer Math (Heidelb)*, vol. 119, no. 4, pp. 699–724, 2011, doi: 10.1007/s00211-011-0399-7.

[40] E. Bannai and E. Bannai, “A survey on spherical designs and algebraic combinatorics on spheres,” *European Journal of Combinatorics*, vol. 30, no. 6, pp. 1392–1425, 2009, doi: https://doi.org/10.1016/j.ejc.2008.11.007.

[41] E. Bannai, “Rigid spherical t-designs and a theorem of Y. Hong,” *J. Fac. Sci. Univ. Tokyo*, vol. 34, pp. 485–489, 1987.

[42] E. Bannai, T. Okuda, and M. Tagami, “Spherical designs of harmonic index t,” *J Approx Theory*, vol. 195, pp. 1–18, Jul. 2015, doi: 10.1016/j.jat.2014.06.010.

[43] E. Bannai and R. M. Damerell, “Tight spherical designs, I,” *Journal of the Mathematical Society of Japan*, vol. 31, no. 1, Jan. 1979, doi: 10.2969/jmsj/03110199.

[44] B. Bajnok, “Construction of Designs on the 2-Sphere,” *European Journal of Combinatorics*, vol. 12, no. 5, pp. 377–382, 1991, doi: https://doi.org/10.1016/S0195-6698(13)80013-3.

[45] X. Chen, A. Frommer, and B. Lang, “Computational existence proofs for spherical t-designs,” *Numer Math (Heidelb)*, vol. 117, no. 2, pp. 289–305, Feb. 2011, doi: 10.1007/s00211-010-0332-5.

[46] R. H. Hardin and N. J. A. Sloane, “McLaren’s improved snub cube and other new spherical designs in three dimensions,” *Discrete Comput Geom*, vol. 15, no. 4, pp. 429–441, Apr. 1996, doi: 10.1007/BF02711518.

[47] B. Rafaely, *Fundamentals of Spherical Array Processing*. in Springer Topics in Signal Processing. Springer, Cham: Springer International Publishing, 2019. doi: 10.1007/978-3-319-99561-8.

[48] H. Teutsch, “Wavefield Decomposition Using Microphone Arrays and Its Application to Acoustic Scene Analysis,” Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), 2006.

[49] A. Farina, D. Pinardi, M. Binelli, M. Ebri, and L. Ebri, “Virtual reality for subjective assessment of sound quality in cars,” in *144th Audio Engineering Society Convention 2018*, 2018.

[50] D. Pinardi, L. Ebri, C. Belicchi, A. Farina, and M. Binelli, “Direction Specific Analysis of Psychoacoustics Parameters inside Car Cockpit: A Novel Tool for NVH and Sound Quality,” *SAE Technical Papers*, no. 2020, 2020, doi: 10.4271/2020-01-1547.

[51] A. Politis, “Compact higher-order Ambisonics library,” *2015*. http://research.spa.aalto.fi/projects/ambi-lib/ambi.html (accessed Jun. 01, 2023).

[52] L. McCormack and A. Politis, “SPARTA suite.” https://leomccormack.github.io/sparta-site/docs/plugins/sparta-suite/ (accessed Jun. 01, 2023).