First-order ambisonic coding with quaternion-based interpolation of PCA rotation matrices

Pierre Mahé, Stephane Ragot, Sylvain Marchand

FIRST-ORDER AMBISONIC CODING WITH QUATERNION-BASED INTERPOLATION OF PCA ROTATION MATRICES

Pierre Mahé[1,2] **Stéphane Ragot**[2] **Sylvain Marchand**[1]

[1] L3i, Université de La Rochelle, France
[2] Orange Labs, Lannion, France
pierre.mahe@orange.com, stephane.ragot@orange.com, sylvain.marchand@univ-lr.fr

ABSTRACT

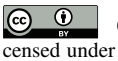
We present a new first-order ambisonic (FOA) coding method extending existing speech/audio codecs such as EVS or Opus. The proposed method is based on Principal Component Analysis (PCA) and multi-mono coding with adaptive bit allocation. The PCA operating in time-domain is interpreted as adaptive beamforming. To guarantee signal continuity between frames, beamforming matrices are interpolated in quaternion domain. The performance of the proposed method is compared with naive multi-mono coding with fixed bit allocation. Results show significant quality improvements at bit rates from 52.8 kbit/s (4×13.2) to 97.6 kbit/s (4×24.4) using the EVS codec.

1. INTRODUCTION

The current codecs used in telephony are mostly limited to mono. With the emergence of devices supporting spatial audio capture and playback, including multi-microphone smartphones, there is a need to extend traditional codecs to enable immersive communication conveying a spatial audio scene. There are different spatial audio coding approaches. For multichannel (or channel-based) audio, one can for instance use channel pairing with a stereo codec, parametric coding with downmixing/upmixing or residual coding. The most recent spatial audio codecs [1–4] handle various input types (channel, object, scene-based audio) and playback setups. Due to its flexibility, ambisonics is potentially an interesting internal coding representation to handle the multiplicity of input/output formats.

To code ambisonics, a naive approach is to extend an existing mono codec to spatial audio by coding each ambisonic component by separate codec instances; this approach is hereafter referred to as *multi-mono coding*. Informal listening tests showed that the naive multi-mono approach may create various spatial artifacts at low bit rates. These artifacts can be classified in three categories: diffuse blur, spatial centering, phantom source. They stem from the correlated nature of ambisonic components. A way to improve this is to use a fixed channel matrixing followed by multi-mono or multi-stereo coding as implemented in the ambisonic extension of Opus [5] or e-AAC+ [6]; this allows to better preserve the correlation structure after coding. Another approach is to analyze the audio scene to extract sources and spatial information. To code first-order ambisonic signal, the DirAC method [7] estimates the dominant source direction and diffuseness parameters in each time/frequency tile and re-creates the spatial image. This method has been extended to High-Order Ambisonics (HOA) in the so-called HO-DirAC [7] where the sound field is divided into angular sectors, for each angular sector, one source is extracted. More recently, Compass [8] was proposed: the number of sources is estimated and a beamforming matrix derived by Principal Component Analysis (PCA) is used to extract sources. In MPEG-H 3D Audio [1] a similar sound scene analysis is performed, e.g. by Singular Value Decomposition (SVD), to code ambisonics, and predominant and ambiance channels are extracted. The ambiance is transmitted as an FOA dowmix. When using PCA or SVD in time domain, transformed components may change dramatically between consecutive frames causing channel permutation and signal discontinuities [9]. The MPEG-H 3D Audio codec already employs overlap-add and channel re-alignent. In [9, 10], improvements were proposed, in particular performing SVD in frequency domain to ensure smooth transitions across frames.

In this work, we investigate how spatial audio can be coded by extending codecs currently used in telephony. We focus on FOA coding because this is a starting point before considering higher orders and FOA coding is required in some codec designs (e.g FOA truncation for ambiance [1] or HOA input in [11]). We reuse the Enhanced Voice Services (EVS) codec [12] which supports only mono input and output signals, however the proposed method can be applied to other codecs such as Opus. We wanted to avoid assumptions on the sound field (e.g. presence of predominant sources, number of sources). The aim was to use PCA to decorrelate FOA components prior to multimono coding; the PCA matrix can be seen as a matrix of beamformer weights. The continuity of the components across frames is guaranteed by an interpolation of 4D rotation ma-



doi:10.25836/sasp.2019.19

trices in quaternion domain.

This paper is organized as follows. Section 2 gives an overview of ambisonics and beamforming. Section 3 provides some background on quaternions. Section 4 describes the proposed coding method. Section 5 presents subjective test results comparing the proposed method with naive multi-mono coding.

2. AMBISONICS

Ambisonics is based on a decomposition of the sound field in a basis of spherical harmonics. Initially limited to first order [13], the formalism was extended to high orders [14]. We refer to [15] for fundamentals of ambisonics. To perfectly reconstruct the sound field, an infinite order is required. In practice, the sound field representation is truncated to a finite order N. For an given order the number of ambisonic components is $n = (N+1)^2$.

In this work, we focus on first-order ambisonic (FOA) where $N = 1$ with $n = 4$ components (W, X, Y, Z). A plane wave with pressure $p(t)$ at azimuth θ and elevation ϕ (with the mathematical convention) is encoded to the following B-format representation:

$$\mathbf{y}(t) = \begin{bmatrix} w(t) \\ x(t) \\ y(t) \\ z(t) \end{bmatrix}^T = \begin{bmatrix} 1 \\ \cos\,\theta\,\cos\,\phi \\ \sin\,\theta\,\cos\,\phi \\ \sin\,\phi \end{bmatrix}^T p(t) \qquad (1)$$

To renderer ambisonics on loudspeakers, many decoding methods have been proposed – see for instance [16,17]. We only consider here the simplest method which recombines ambisonic components by a weight matrix $\mathbf{V} = [v_1, ..., v_n]$ to compute the signal feeds for loudspeakers located at known positions. This decoding is a conversion of the ambisonic representation to the loudspeaker domain. Similarly, it is possible to transform the ambisonic representation to another sound field representation using a transformation matrix $\mathbf{V}$ of size $n \times n$:

$$\mathbf{A} = \mathbf{V}\mathbf{Y} \qquad (2)$$

where $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_n]$ is the input matrix of n components, $\mathbf{A} = [\mathbf{a}_1, \cdots, \mathbf{a}_n]$ is the output matrix of the equivalent sound field representation. It is possible go from one sound field representation to another, provided that the matrix $\mathbf{V}$ is unitary:

$$\mathbf{V^T V} = \mathbf{I} \qquad (3)$$

where $\mathbf{I}$ is the identity matrix. If the matrix $\mathbf{V}$ does not satisfy this condition, some spatial deformation will occur.

It is possible to convert the representation $\mathbf{A}$ to the ambisonic representation $\mathbf{Y}$ by inverting the matrix $\mathbf{V}$. This principle of conversion from B-format to A-format is used in [6] or in the equivalent spatial domain (ESD) [18]. This conversion by a unitary matrix $\mathbf{V}$ allows interpreting the Principal Component Analysis (PCA) described hereafter, in terms of ambisonic transformation to an equivalent spatial representation.

To render ambisonics over headphones, a binaural rendering of ambisonic signals can be used. The simplest approach is to decode the signals over virtual loudspeakers, convolve the resulting feed signals by Head-Related Impulse Responses (HRIRs) and combine the results for each ear. A more optimized method using generic HRIRs in B-format domain is used in [14].

3. QUATERNIONS

Quaternions were introduced in 1843 by Hamilton [19] to generalize complex numbers. They have many applications in mathematics [20], computer graphics [21] and physics (aerospace, robotics, etc.) [22]. A quaternion q is defined as $q = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$, where a, b, c, d are real and $\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk} = -1$. Quaternions are often used as a parametrization of 3D rotations, especially for rotation interpolation. We recall that the set of 3D rotations can be mapped to the unit-norm quaternions under a one-to-two mapping [20–22], i.e. each 3D rotation matrix maps to two antipodal unit-norm quaternions: q and $-q$. Spherical linear interpolation (slerp) consists in the following principle [22]:

$$slerp(q_1, q_2, \gamma) = q_1(q_1^{-1}q_2)^\gamma \qquad (4)$$

where q_1 and q_2 are respectively the starting and ending quaternions and $0 \leq \gamma \leq 1$ is the interpolation factor. This is equivalent to [22]:

$$slerp(q_1, q_2, \gamma) = \frac{\sin((1-\gamma)\Omega)}{\sin(\Omega)} q_1 + \frac{\sin(\gamma\Omega)}{\sin(\Omega)} q_2 \qquad (5)$$

where $\Omega = \arccos(q_1 \cdot q_2)$ is the angle between q_1 and q_2 and $q_1 \cdot q_2$ is the dot product of q_1 and q_2. This boils down to interpolating along the grand circle (or geodesics) on a unit sphere in 4D with a constant angular speed as a function of γ. To ensure that the quaternion trajectory follows the shortest path on the sphere [21], the relative angle between successive unit-norm quaternions needs to be checked to choose between $\pm q_2$.

In this work, we used double quaternions to represent 4D rotation matrices. The product $\mathbf{R} = \mathbf{Q}^\star.\mathbf{P} = \mathbf{P}.\mathbf{Q}^\star$ of an anti-quaternion matrix $\mathbf{Q}^\star$ and a quaternion matrix $\mathbf{P}$, where

$$\mathbf{Q}^\star = \begin{pmatrix} a & b & c & d \\ -b & a & -d & c \\ -c & d & a & -b \\ -d & -c & b & a \end{pmatrix} \qquad (6)$$

and

$$\mathbf{P} = \begin{pmatrix} w & -x & -y & -z \\ x & w & -z & y \\ y & z & w & -x \\ z & -y & x & w \end{pmatrix} \qquad (7)$$

associated to $q = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$ and $p = w + x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$, is a 4D rotation [20]. Conversely, given a 4D rotation $\mathbf{R}$ one may compute two quaternions q and p (up to sign) using factorization methods described in [20, 23]. The interpolation of 4D rotation matrices can be done by interpolating separately the associated pairs of quaternions, for instance using slerp interpolation. However, it is important to

doi:10.25836/sasp.2019.19

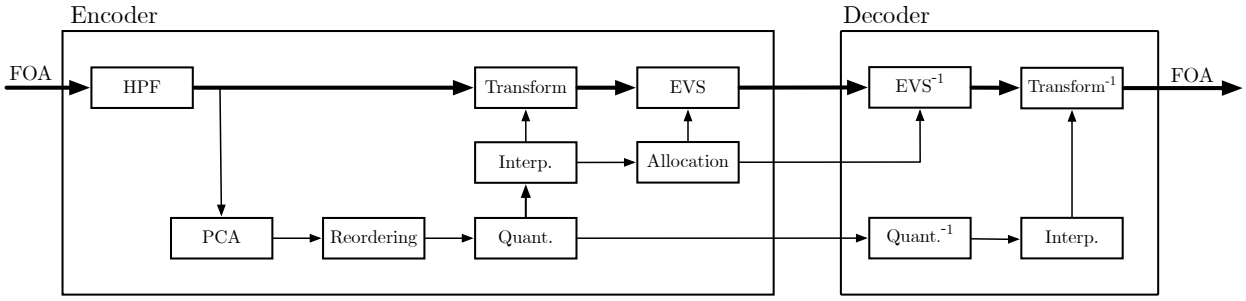


Figure 1. *Overview of proposed coding method.*

keep the sign consistent between double quaternions when constraining the shortest path.

4. PROPOSED CODING METHOD

The proposed coding method relies on a pre-processing of ambisonic components to decorrelate them prior to multi-mono coding. We illustrate this method using the EVS codec as a mono core codec. The input signal is a first-order ambisonic signal, with $n = 4$ ambisonic components labeled with an index $i = 1, ..., n$. The ambisonic channel ordering has no impact, therefore it is not specified here. The coding method operates on successive 20 ms frames, which is the EVS frame length. The overall codec architecture is shown in Figure 1. In the following we describe separately the pre-processing part based on PCA and the multi-mono coding part. We refer to [24] for more details on the codec description.

4.1 FOA pre-processing based on PCA

4.1.1 Beamforming matrix estimation

In each frame, the covariance matrix $\mathbf{C_{YY}}$ is estimated in time domain:

$$\mathbf{C_{YY}} = \mathbf{Y}^T\mathbf{Y} \tag{8}$$

where $\mathbf{Y} = [\mathbf{y}_1, ..., \mathbf{y}_n]$ is the matrix of $n = 4$ ambisonic components. The covariance matrix $\mathbf{C_{YY}}$ is factorized by eigenvalue decomposition as:

$$\mathbf{C_{YY}} = \mathbf{V}\Lambda\mathbf{V}^T \tag{9}$$

The matrix $\mathbf{V} = [\mathbf{v}_1, ..., \mathbf{v}_n]$ is a 4D rotation matrix if the transformation matrix is orthogonal and if $\det(\mathbf{V}) = +1$. We ensure that the eigenvector matrix defines a rotation matrix by inverting the sign of $\mathbf{v}_n$ if $\det(\mathbf{V}) = -1$. This matrix $\mathbf{V}$ is the transformation matrix to convert the components to another spatial domain equivalent to the original ambisonic B-format. In the following, the rotation matrix in the current frame of index t will be denoted $\mathbf{V}_t$.

To avoid bias from low frequencies in PCA, the input components are pre-filtered by a 20 Hz high-pass IIR filter from the EVS codec [25].

4.1.2 Re-alignment of beams

From frame to frame, the eigen decomposition can switch the order of eigenvectors or invert their sign. These changes may modify significantly the weighting coefficients of the beamforming matrix. Therefore the beam directions might change and these modifications might create discontinuities in signals, which can degrade audio quality after multi-mono coding. To improve signal continuity between successive frames, a signed permutation was applied to the eigenvector matrix in the current frame $\mathbf{V}_t$ to maximize similarity to the eigenvector matrix $\mathbf{V}_{t-1}$. The signed permutation is obtained in two steps:

First, a permutation is found to match the eigenvectors of frames t and $t-1$. This problem is treated as an assignment problem where the goal is to find the closest beam, in terms of direction. As in [9], the Hungarian algorithm was used with the following optimization criterion:

$$\mathbf{J}_t = tr(|\mathbf{V}_t.\mathbf{V}_{t-1}^T|) \tag{10}$$

Where $tr(|.|)$ is the trace of the matrix $|\mathbf{V}_t.\mathbf{V}_{t-1}^T|$ whose coefficients are the absolute values. It is noted that only the beams direction it is considered in this step.

Second, to avoid sign inversion of component between consecutive frames, the autocorrelation was computed:

$$\mathbf{\Gamma}_t = \tilde{\mathbf{V}}_t.\mathbf{V}_{t-1}^T \tag{11}$$

A negative diagonal value in $\mathbf{\Gamma}_t$ indicates a sign inversion between two frames. The sign of the respective columns of $\tilde{\mathbf{V}}_t$ was inverted to compensate for this change of direction.

4.1.3 Quantization of beamforming matrix

In [26], 2D and 3D rotation matrices were converted by angle parameters. A similar idea was used to quantize the beamforming matrix in each frame. A rotation matrix of size $n \times n$ can be parametrized by $n(n-1)/2$ generalized Euler angles [27]. The 4D rotation matrix $\mathbf{V}_t$ is converted to 6 generalized Euler angles, with 3 angles in $[-\pi, \pi)$ and 3 angles in $[-2\pi, 2\pi)$. These angles are coded by scalar quantization with a budget of respectively 8 and 9 bits for angles defined over a support of length π and 2π. The overall budget for 6 angles is 51 bits per frame.

4.1.4 Interpolation of beamforming matrices

To improve continuity and guarantee smooth transition between beams across consecutive frames, the 4D rotation matrix of the current frame and previous frames are interpolated by subframes. The rotation matrices are converted to pairs of quaternions (q, p) and the interpolation is done

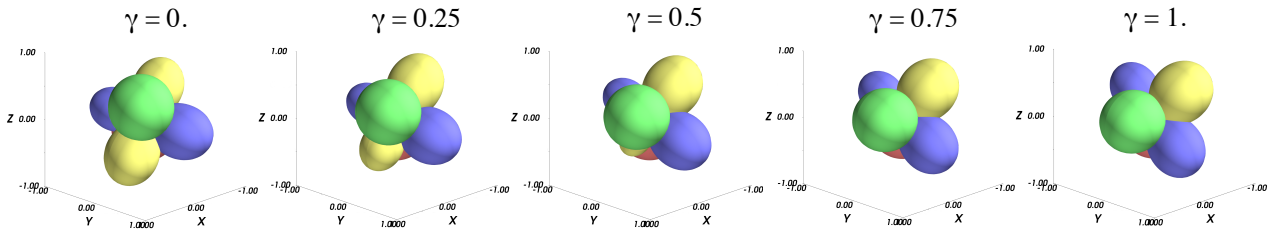


Figure 2. *Beamforming interpolation.*

in this double quaternion domain, in aim to interpolation of rotation with constant angular velocity. Each frame is divided into K subframes and for each subframe of index $1 \leq k \leq K$ in the current frame, the corresponding quaternion pairs (q_{t-1}, q_t) and (p_{t-1}, p_t) are interpolated used spherical linear interpolation (slerp). In the proposed coding method, the frame of $L = 640$ samples (20 ms at the 32 kHz sampling rate) is divided into K sub-frames. We used $K = 128$ which gives a subframe length of $L/K = 10$ samples (0.3125 ms) and the interpolation factor in Eq. 5 was set to $\gamma = k/K$. The interpolated pairs of quaternions were converted back to a 4D matrix.

The beamforming matrix interpolation is illustrated in Figure 2, for interpolation factors $\gamma = 0, 0.25, 0.5, 0.75$ and 1.

4.1.5 PCA matrixing

The pre-processed FOA signal is transformed into 4 principal components by applying the interpolated 4D rotation matrix (beamforming matrix) in each sub-frame.

4.2 Multi-mono coding with adaptive bit rate allocation

In naive multi-mono coding the bit rate is the same for each component. It was observed experimentally that signal energy after PCA matrixing may vary significantly between components and it was found that an adaptive bit allocation is necessary to optimize quality. The EVS codec quality [28] does not increase according to rate-distortion theoretic predictions when increasing bit rate. In this work the audio quality for each bit rate was modeled by energy-weighted average MOS (Mean Opinion Score) values. We used a greedy bit allocation algorithm which aims at maximizing the following score:

$$S(b_1, \cdots, b_n) = \sum_{i=1}^{n} Q(b_i).E_i^{\beta} \qquad (12)$$

where b_i and E_i are respectively the bit allocation and the energy of the i^{th} component in the current frame and $Q(b_i)$ is a MOS score for a bit rate corresponding to b_i bits. These values $Q(b_i)$ may be take from the EVS characterization report [28] the values used in this work are defined in [24]. This optimization is subject to the constraint $b_1 + \cdots + b_n \leq B$ where B is the budget allocated for multi-mono coding. Note that if another core codec than EVS is used, the values $Q(b_i)$ can be adjusted accordingly; for instance, a quality evaluation of Opus can be found in [29]. The bit allocation to individual audio channels was restricted to all EVS bit rates $\geq$ 9.6 kbit/s to ensure a super-wideband coded bandwidth. Details for bitstream structure and a bit rate allocation example can be found in [24].

In each 20 ms frame, the selected bit allocation is transmitted to the decoder and used for multi-mono EVS coding.

4.3 Decoding

The decoding part consists in multi-mono decoding based on the received bit allocation and PCA post-processing (which is the inverse of the pre-processing) in each frame.

5. EXPERIMENTAL RESULTS

5.1 Test setup

We conducted a subjective test according to the MUSHRA methodology [30] to compare the performance of naive multi-mono coding and the proposed coding method. For each item, subjects were asked to evaluate the quality of conditions with a grading scale ranging of 0 to 100 (Bad to Excellent). The test conditions included three specific items: the hidden reference (FOA) and two anchors. MUSHRA tests for mono signals typically use a low anchor (3.5kHz low-pass filtered original) and a medium anchor (7kHz low-pass filtered original). For MUSHRA tests with stereo, it is suggested to use a "reduced stereo image" as anchors [30]. There is no clear recommendation for spatial alterations for MUSHRA tests with ambisonics. In this work we used the following spatial reduction:

$$FOA = \begin{pmatrix} W \\ \alpha X \\ \alpha Y \\ \alpha Z \end{pmatrix}, \qquad \alpha \in [0, 1] \qquad (13)$$

with $\alpha = 0.65$ and $\alpha = 0.8$ for the low and medium anchors, respectively. All FOA items were binauralized with the Resonance Audio renderer [31]. All test conditions are summarized in Table 1. The test items consisted of 10 challenging ambisonic items: 4 voice items, 4 music items and 2 ambient scenes. The synthetic items were generated in Orange Labs, the recorded items were captured and mixed by Orange Labs or done jointly with partners, see [24] for more details. All subjects conducted the listening test with the same professional audio hardware in a dedicated listening room at Orange Labs. In total 11 listeners participated

doi:10.25836/sasp.2019.19

Table 1. *List of MUSHRA conditions.*

Short name	Description
HREF	FOA hidden reference
LOW_ANCHOR	3.5 kHz LP-filtered and spatially-reduced FOA ($\alpha = 0.65$)
MED_ANCHOR	7 kHz LP-filtered and spatially-reduced FOA ($\alpha = 0.8$)
MULTI52	FOA coded by multimono EVS at 4×13.2 kbit/s
MULTI65	FOA coded by multimono EVS at 4×16.4 kbit/s
MULTI97	FOA coded by multimono EVS at 4×24.4 kbit/s
PCA52	FOA coded by proposed method at 52.8 kbit/s
PCA65	FOA coded by proposed method at 65.6 kbit/s
PCA97	FOA coded by proposed method at 97.6 kbit/s

in the test; all of them are expert or experienced listeners without hearing impairments. Each item was coded at three bit rates for multi-mono coding: 52.8, 65.6, 97.6 kbit/s which corresponds to a fixed allocation of 13.2, 16.4 and 24.4 kbit/s per channel (respectively). For the proposed coding method, as explained in Section 4, the bit rate was dynamically distributed between channels; however the target (maximum) bitrate was set to the same bit rate as multi-mono coding for a fair comparison.

5.2 Subjective test results

The MUSHRA test results, including the mean and 95% confidence intervals, are presented in Figure 3. They show that significant quality improvement over multi-mono coding. For instance, the proposed coding method at 52.8 kbit/s is is equivalent to multi-mono coding at 65.6 kbit/s.

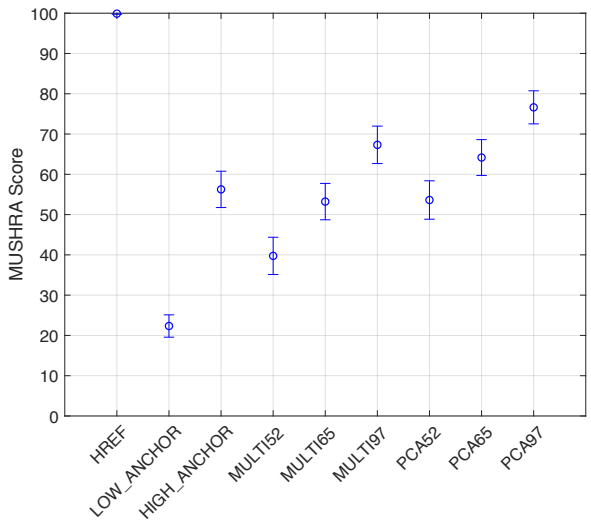


Figure 3. *MUSHRA test results.*

Spatial artifacts were noted at every bit rate for multi-mono coded items. They can be classified in three categories: diffuse blur, spatial centering, phantom source. With the proposed coding method, these artifacts are mostly removed because the correlation structure is less impacted by coding. This explanation was supported by the feedback from some subjects, after they conducted the subjective test.

6. CONCLUSION

This article presented a spatial extension of an existing codec (EVS), with a pre-processing decorrelating ambisonic components prior to multi-mono coding. The proposed method operate in time domain to avoid extra delay and allow maximum compatibility with existing codecs which are used as a black box. In each frame, a beamforming basis is found by PCA; the PCA matrices are interpolated in quaternion domain to guarantee smooth transitions between beamforming coefficients. Subjective test results showed significant improvements over naive multi-mono EVS coding for bit rates from 4×13.2 to 4×24.4 kbit/s, which may be explained by the combination of the use of PCA matrixing and adaptive bit allocation.

7. ACKNOWLEDGMENTS

The authors thank all participants in the subjective test. They also thank Jérôme Daniel for discussions on spatial reduction for MUSHRA anchors items.

8. REFERENCES

[1] J. Herre, J. Hilpert, A. Kuntz, and J. Plogsties, "MPEG-H audio—the new standard for universal spatial/3D audio coding," *Journal of the Audio Engineering Society*, vol. 62, no. 12, pp. 821–830, 2015.

[2] ETSI TS 103 190 V1.1.1, "Digital Audio Compression (AC-4) Standard," April 2014.

[3] ATSC Standard, Doc. A/52:2018, "Digital Audio Compression (AC-3, E-AC-3)," January 2018.

[4] ETSI TS 103 491 V1.1.1, "DTS-UHD Audio Format; Delivery of Channels, Objects and Ambisonic Sound Fields," April 2017.

[5] J. Skoglund, "Ambisonics in an Ogg Opus Container." IETF RFC 8486, Oct. 2018.

[6] 3GPP TS 26.918, "Virtual Reality (VR) media services over 3GPP, clause 6.1.6," 2018.

[7] V. Pulkki, A. Politis, M.-V. Laitinen, J. Vilkamo, and J. Ahonen, "First-order directional audio coding (DirAC)," in *Parametric Time-Frequency Domain Spatial Audio*, ch. 5, John Wiley & Sons, 2018.

[8] A. Politis, S. Tervo, and V. Pulkki, "Compass: Coding and multidirectional parameterization of ambisonic sound scenes," in *Proc. ICASSP*, pp. 6802–6806, 2018.

[9] S. Zamani, T. Nanjundaswamy, and K. Rose, "Frequency domain singular value decomposition for efficient spatial audio coding," in *Proc. WASPAA*, pp. 126–130, 2017.

[10] S. Zamani and K. Rose, "Spatial Audio Coding with Backward-Adaptive Singular Value Decomposition," in *145th AES Convention*, 2018.

[11] D. McGrath *et al.*, "Immersive Audio Coding for Virtual Reality Using a Metadata-assisted Extension of the 3GPP EVS Codec," in *Proc. ICASSP*, May 2019.

[12] S. Bruhn *et al.*, "Standardization of the new 3gpp evs codec," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5703–5707, IEEE, 2015.

[13] M. A. Gerzon, "Periphony: With-height sound reproduction," *Journal of the Audio Engineering Society*, vol. 21, no. 1, pp. 2–10, 1973.

[14] J. Daniel, *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*. PhD thesis, Université Paris 6, 2000. `http://gyronymo.free.fr`.

[15] B. Rafaely, *Fundamentals of spherical array processing*. Springer, 2015.

[16] A. J. Heller, R. Lee, and E. M. Benjamin, "Is My Decoder Ambisonic?," in *125th AES Convention*, 2008.

[17] F. Zotter and M. Frank, "All-round ambisonic panning and decoding," *Journal of the Audio Engineering Society*, vol. 60, no. 10, pp. 807—820, 2012.

[18] 3GPP TS 26.260, "Objective test methodologies for the evaluation of immersive audio systems," 2018.

[19] W. R. Hamilton, *On a new Species of Imaginary Quantities connected with a theory of Quaternions*, vol. 2. 1844.

[20] P. De Casteljau, *Les quaternions*. Dunod, 1987.

[21] K. Shoemake, "Animating rotation with quaternion curves," *ACM SIGGRAPH Computer Graphics*, vol. 19, no. 3, pp. 245—-254, 1985.

[22] A. Hanson, *Visualizing Quaternions*. Morgan Kaufmann Publishers, 2006.

[23] A. Perez-Gracia and F. Thomas, "On Cayley's factorization of 4D rotations and applications," *Advances in Applied Clifford Algebras*, vol. 27, no. 1, pp. 523–538, 2017.

[24] P. Mahé, S. Ragot, and S. Marchand, "First-Order Ambisonic Coding with PCA Matrixing and Quaternion-Based Interpolation," in *Proc. DAFx*, 2019.

[25] 3GPP TS 26.445, "Codec for Enhanced Voice Services (EVS); Detailed algorithmic description," 2019.

[26] M. Briand, *Études d'algorithmes d'extraction des informations de spatialisation sonore : application aux formats multicanaux*. PhD thesis, INPG Grenoble, 2007.

[27] D. K. Hoffman, R. C. Raffenetti, and K. Ruedenberg, "Generalization of Euler Angles to N-Dimensional Orthogonal Matrices," *Journal of Mathematical Physics*, vol. 13, no. 4, pp. 528–533, 1972.

[28] 3GPP TS 26.952, "Codec for Enhanced Voice Services (EVS); Performance Characterization," 2019.

[29] A. Rämö and H. Toukomaa, "Voice quality characterization of IETF Opus codec," in *Proc. Twelfth Annual Conference of the International Speech Communication Association*, 2011.

[30] ITU-R Rec. BS.1534–3, "Method for the subjective assessment of intermediate quality level of coding systems," 2015.

[31] "Resonance audio : Rich, immersive, audio." `https://resonance-audio.github.io/resonance-audio`.

 doi:10.25836/sasp.2019.19