Coding Higher Order Ambisonics in 3GPP IVAS – Scaling Parametric Audio Coding to Higher Bitrates

Christoph Hold*†[1], Dominik Weckbecker*[1], Guillaume Fuchs*, Markus Multrus*,
Rishabh Tyagi‡, Stefanie Brown‡, Juan Torres‡, Stefan Bruhn‡
* *Fraunhofer IIS, Erlangen, Germany, {firstname.lastname}@iis.fraunhofer.de*
† *Aalto University, {firstname.lastname}@alumni.aalto.fi*
‡ *Dolby Laboratories, {firstname.lastname}@dolby.com*

***Abstract*—Immersive Voice and Audio Services (IVAS) is the immersive audio communications codec for 5G networks recently standardized by 3GPP (3rd Generation Partnership Project). It supports multiple immersive audio formats, including higher-order Ambisonics (HOA). The latter is known to be particularly challenging to compress given the bitrate and complexity limits in a mobile-communications scenario. Directional Audio Coding (DirAC) has been shown to be an effective parameterization of first-order Ambisonics and has been adopted for low-to-medium bitrates in IVAS. Higher-order parametrization can scale the codec to higher quality, overcoming known challenges with the parameterization from first-order Ambisonics. These challenges are addressed by combining the strengths of the parametric higher-order Directional Audio Coding (HO-DirAC) and Spatial Reconstruction (SPAR). Scaling to higher bitrates is achieved by a hybrid parameterization scheme and a sector-based higher-order DirAC architecture specifically adapted to IVAS. The latter estimates two directions of arrival (DoAs) but also a single global diffuseness, facilitating a deep integration into the IVAS codec and bitrate switching. Here we detail the Ambisonics coding in IVAS and its extensions at high bitrates. We show that the proposed system improves the perceptual quality for challenging audio scenes while keeping the complexity within the required limits.**

***Index Terms*—Directional audio coding, parametric spatial audio, IVAS, Ambisonics, mobile communications**

I. Introduction

Ambisonics expresses a sound scene around a given listener position in terms of spherical harmonics (SH) expansion coefficients of the sound pressure on the sphere [1]–[6]. The spatial resolution of the sound scene is linked to the Ambisonics order N_{sph}, which is linked to the number of basis functions given as $L = (N_{\mathrm{sph}} + 1)^2$. Streaming the L SH coefficients conveniently corresponds to streaming L audio channels. This suggests that multi-channel audio codecs may be employed to transport the Ambisonics audio scene. However, a straightforward coding of the individual discrete audio channels is expensive in terms of bitrate, as the number of channels scales quadratically with the order. Furthermore, lossy compression on the individual channels, unaware of the properties of the Ambisonic signals, may alter the inter-channel features and therefore the directional properties of the scene. Instead, parametric techniques such as DirAC [7]–[9] and SPAR [10] encode the crucial directional features more compactly and reproduce them more accurately.

Directional Audio Coding (DirAC) is a perceptually motivated parametrization of the sound field into a mixture of a directional and a diffuse component. It utilizes a direction of arrival (DOA) and a diffuseness parameter that are linked to the inter-aural coherence [8]. These two spatial cues are encoded by the DirAC parameters $\mathbf{\Theta}_D$ (DoA angle) and Ψ (diffuseness), which are estimated at the encoder and transmitted in the bitstream. Traditionally, DirAC estimates a single DoA angle and diffuseness for each frequency band from the first-order component channels of the input signal (FO-DirAC) [7], [8]. Recently, Fuchs et al. [9] have shown that higher-order Ambisonics coding based on FO-DirAC can be a viable solution to meet the requirements of low bitrate real-time communications. The DirAC encoder in [9] only operates on first-order signals. The decoder uses the parameters to expand into higher-order Ambisonic components. While very effective at low to medium bitrates, it was also found that increasing the bitrate could not increase the perceptual quality to the expected amount, with premature quality saturation for complex sound scenes, suggesting the potential for system optimizations. At the same time, the limited FO-DirAC parametrization has been shown to be susceptible to audible artefacts in complex scenes where the assumption of a single DoA is insufficient. To this end, this paper explains a set of mechanisms in IVAS HOA coding, which aim to improve quality at higher bitrates than presented in [9].

A complementary residual/parametric method for Ambisonics coding is Spatial Reconstruction (SPAR). SPAR [10], [11] encodes a compact representation of the Ambisonics signal by estimating the covariance between channels over a set of psychoacoustically spaced frequency bands. Bandwise correlation between channels is exploited to produce low-energy, highly compressible channel residuals, a subset of which is transmitted. The decoder precisely recreates these channels from the transmitted residuals and prediction metadata, with the only sources of error being metadata quantization and core audio coder losses. The remaining channels, for which residuals are not available, are approximated using transient-aware decorrelators. DirAC and SPAR approaches complement each other in the sense that, while both approaches seek to preserve the integrity of the Ambisonics spatial scene,

[1]These authors contributed equally.

DirAC focuses on estimating and preserving directional audio components, while SPAR seeks to reconstruct the individual Ambisonics channels directly.

The 3GPP IVAS codec employs an advanced combination of DirAC and SPAR that achieves good quality at low bitrates from $13.2\,\mathrm{kbps}$ on, scaling to high bitrates up to $512\,\mathrm{kbps}$ [11]. Bitrate scaling is achieved by varying the number of transport audio channels and parameter accuracy. This contribution explains the implemented DirAC extension that enables the additional analysis of the input signal up to the second order using a newly formulated two sector-based HO-DirAC approach. We further present specific tunings of SPAR for operation at high bitrates. These changes help IVAS overcome the quality limitations of FO-DirAC. It builds on earlier works by Politis et al. [12] and Hold et al. [13], [14] but is strongly adapted to meet the demands of IVAS.

In this contribution, we will provide more implementation details of HOA coding in IVAS and show that (i) the perceptual quality of difficult scenes is improved as compared to the system where the bitrate is increased without the extensions, (ii) the complexity is kept within the limits defined by 3GPP [15], and (iii) IVAS outperforms a conventional multi-channel HOA codec. Furthermore, we describe in detail the implementation of the sector-based HO-DirAC method in combination with SPAR in the framework of the IVAS codec, extending the description of the lower bitrate system in Ref. [11]. Finally, we analyze the perceptual quality and computational complexity of the IVAS codec at the two highest supported bitrates.

II. System design

The described system consumes HOA input and decodes to HOA output, albeit IVAS supports many other formats. The bitrates for Ambisonics coding up to third order range from 13.2 to $512\,\mathrm{kbps}$, where the presented system is active at the higher bitrates between 384 and $512\,\mathrm{kbps}$ [11]. Besides the discrete coding of audio channels, IVAS supports parametric spatial audio coding techniques, whose parameters can be transmitted across time and frequency.

Parametric spatial audio techniques such as sector-based HO-DirAC have been proposed for HOA compression [14]. The IVAS codec features a special adaptation of this method. Specifically, the HO-DirAC system adopted in IVAS (i) is implemented on top of the existing hybrid SPAR-DirAC system, (ii) integrates deeply with the rest of the codec, (iii) meets the complexity and delay requirements, and (iv) allows for bitrate switching between the FO- and HO-DirAC modes.

Driven by these criteria, the encoder design is shown in Fig. 1. The design shares the structure detailed in [11], but includes additional processing of the higher-order component channels. Specifically, the higher-order Ambisonics input signal x^l_{in} is separated into two sector signals:

$$x^{l'}_s = \sum_l w^{l,l'}_s \, x^l_{\text{in}}, \tag{1}$$

where $x^{l'}_s$ are the SH coefficient signals of the sectors with $s = 1, 2$, x^l_{in} those of the HOA2 input signal, and $w^{l,l'}_s$ the

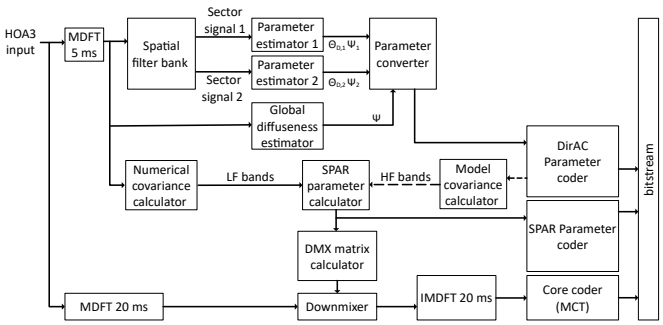

Fig. 1. Encoder block diagram

filter coefficients of the sector s. The time and frequency indices are omitted from the notation. The index l runs over the spherical harmonics functions up to second order, creating four first-order equivalent signals indexed by l'. The HOA signals are constrained by two complementary cardioids pointing in opposite directions, where preservation of the sound scene pressure amplitude is ensured by the sector design [13].

The resulting sector signals correspond to first-order SHD signals that have been directionally constrained by the corresponding cardioid patterns. Each of these sector signals is then fed into one FO-DirAC parameter estimator, yielding one DoA $\theta_{D,s}$ and one diffuseness Ψ_s for each sector. From the sector diffuseness parameters, the sector diffuseness ratio is computed as

$$a_1 = \frac{1 - \Psi_1}{(1 - \Psi_1) + (1 - \Psi_2)}, \tag{2}$$

which satisfies $a_2 = 1 - a_1$, and is assumed to correspond to $\frac{E_{\text{dir},1}}{E_{\text{dir},1} + E_{\text{dir},2}}$. The energies $E_{\text{dir},1/2}$ are those of the directional components of the sector signal, such as a plane-wave. In contrast to previous HO-DirAC architectures, IVAS also employs a global diffuseness estimator (see Fig. 1), which computes the global diffuseness parameter Ψ from the first-order of the input signal. In the metadata, only the parameter set $\{\theta_{D,1}, \theta_{D,2}, \Psi, a_1\}$ is transmitted. This choice of parameter set lends itself to a tight integration of the HO-DirAC parameter coding into the IVAS framework. Specifically, the parameter quantization and coding is compatible and shared with Ambisonics coding at the lower bitrates [9], [11], [16] as well as with the Metadata-Assisted Spatial Audio (MASA) coding mode [16], [17] in IVAS. Furthermore, this design facilitates switching between the FO- and HO-DirAC bitrates as the processing of the global diffuseness path remains consistent across bitrates.

The other blocks in Fig. 1 are the same as in [11]. The SPAR encoder estimates the bandwise covariance between Ambisonics channels and computes prediction and decorrelation metadata, then creates a 4-channel downmix consisting of W and residuals Y', Z', X'. The downmix channels are coded jointly using the multichannel coding tool (MCT), which dynamically forms channel pairs based on the remaining interchannel correlation [18]. At the decoder, the first-order channels are reconstructed from the residuals and prediction metadata,

whilst remaining channels are recovered parametrically using decorrelators to preserve inter-channel covariance.

The covariance and SPAR parameters can be estimated from the DirAC parameters and vice versa [11]. At $384\,\mathrm{kbps}$, the HF SPAR parameters [10], [11], [16] are calculated from the model-based covariance using only a single DoA. Here, that of the left sector is chosen. Although this is exact for scenes with only one DoA, it is only an approximation to the correct values in more complex scenes. However, listening tests have shown that the error introduced here is perceptually limited for typical scenarios. At $512\,\mathrm{kbps}$, SPAR parameters are transmitted directly for all frequency bands. At both 384 and $512\,\mathrm{kbps}$, DirAC parameters are transmitted for all frequency bands. This is in contrast to the lower bitrate system described in [11]. The channels reconstructed by SPAR and DirAC are indicated in Tab. I.

The corresponding IVAS decoder block diagram is shown in Fig. 2. The SPAR decoder reconstructs first-order Ambisonics channels and additional higher-order channels according to Table I. These are generated in the domain of the Complex Low-Delay Filterbank (CLDFB) [11], [16].

The reconstructed first-order channels are fed into the HO-DirAC-decoder block together with the DirAC metadata from the DirAC metadata decoder (cf. [9], [11]). Conceptually, DirAC splits the decoding into directional (upper) and diffuse (lower) decoding paths, respectively. This enables optimized rendering for both distinct types, where both are mixed according to the diffuseness estimate, i.e., weighted by $1-\Psi$ and Ψ, respectively. Note the lack of decorrelators in the present DirAC decoding here.

A plane-wave is fully described as its (pressure) signal x_p and DoA angle θ_D, which allows expanding the corresponding SHD components directly. The two directional signals to be expanded to the SHD are obtained according to the cardioid beamformers utilized at the encoder, available from the FOA signals x^l_{FOA} (output of the SPAR decoder)

$$x_{p,s} = \sum_l w_{s,l}\, x^l_{\mathrm{FOA}}\,, \tag{3}$$

which are then panned/expanded to produce the directional signal for each sector s up to the third Ambisonics order:

$$x_{s,l} = (1-\Psi)\, a_s\, x_s\, Y_l(\theta_D)\,. \tag{4}$$

$Y_l(\theta_D)$ is the spherical harmonic with the combined index l evaluated at the DoA angle θ_D. Therefore, the global diffuseness Ψ regulates the global balance of the plane-wave reencoding, where the sector diffuseness ratio a_s balances the contribution of each individual sector. Intuitively, the sector whose signal is estimated to be closer to a plane-wave will contribute more to the directional signal component of the output signal.

The diffuse stream rendering is given directly by a scaling of the FOA signal with the global diffuseness estimate. The assumption here is that a diffuse sound signal is less direction-dependent and can be approximated by a FOA signal. According to the model, the energy of higher SHD components

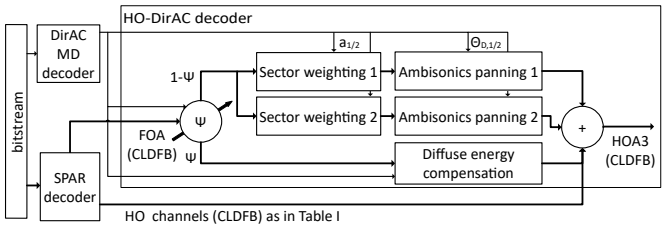

Fig. 2. Decoder block diagram of the IVAS Ambisonics decoder at high bitrates, described in this contribution.

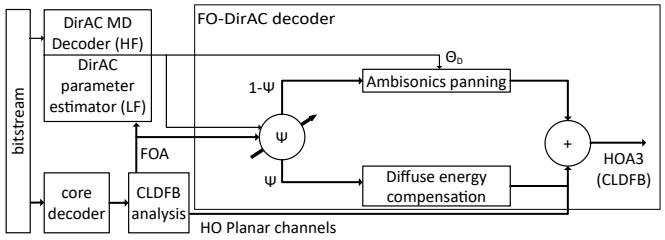

Fig. 3. Decoder block diagram of the FO-DirAC system described in Ref. [9], active at lower bitrates.

is pushed towards the FOA components for a diffuse signal, implemented by a diffuseness dependent amplification of the first-order signal components by the factor [9], [16]

$$1 + g(\Psi) = \sqrt{1 + \Psi\left(\frac{H+1}{L+1} - 1\right)}\,. \tag{5}$$

However, this may alter the ambience in some scenes as uncorrelated contents of individual HOA channels can not be exactly reconstructed. Finally, the diffuseness-compensated first-order and directional second- and third-order channels are combined.

TABLE I
HO-DIRAC (D) AND SPAR (S) SYNTHESIS METHODS FOR AMBISONICS OUTPUT CHANNELS IN ACN ORDER.

Ambisonics Order (l)	384 kbps	512 kbps
0	S	S
1	S S S	S S S
2	S D D D S	S S S S S
3	S D D D D D S	S D D D D D S

III. SUBJECTIVE QUALITY EVALUATION

Parametric spatial audio codecs have been shown to be an effective coding strategy for Ambisonics. In that context, DirAC and HO-DirAC have been found to be a suitable model choice and have shown their perceptual effectiveness [9], [11], [14]. In particular for IVAS, [9] showed the evaluation of DirAC with a first-order estimator, and [11] evaluated the SPAR-DirAC hybrid approach, both for bitrates lower than in the present study focus. Both indicated potential for further quality improvements at higher bitrates. To this extent, the present evaluation highlights the performance gain available between IVAS in the low-bitrate configuration [9] depicted in

Fig. 3 and the presented high-bitrate configuration depicted in Fig. 2.

In order to limit the comparison to systematic differences between the architectures, both systems operate with the same number of coded audio channels, which means more than in the low-bitrate system presented in [9]. In other words, system Fig. 3 with more transport channels can be seen as the straightforward scaling of the low-bitrate system [9] and was hence chosen for comparison in this study.

In order to assess the perceptual quality of the IVAS Ambisonics coding system, we have conducted a MUSHRA [19] listening test. Compared are the architecture presented in the present paper (Fig. 2), the previously presented architecture of [9] (Fig. 3), and Opus [20], [21] as a baseline state-of-the-art multichannel perceptual audio codec for HOA, for two bitrates of interest (384 and 512 kbps). The baseline Opus codec codes all Ambisonics channels, without downmixing. The HOA channel mapping employs linear (signal-independent) mixing and demixing matrices at the encoder and decoder, respectively, aiming to preserve more of the inter-channel SHD features, as opposed to coding each SHD signal independently. Opus was configured with default parameters (AMBIX format, channel-mapping family 3, hard CBR).

The test comprised mostly critical audio scenes, which comprise multiple directional sound sources and additional ambiance as described in Tab. II. Items were decoded in third-order Ambisonics (HOA3) format and then binauralized with the publicly available toolbox used in the IVAS development [22]. The anchor was a mono downmix with 3.5 kHz low-pass filtering. The test was presented to expert listeners over headphones. The results at 384 kbps are plotted in Fig. 4. We find that the presented high-bitrate IVAS system outperforms the single-DOA DirAC system (FO-DirAC) at the same bitrate. The difference between IVAS and channel-based methods (represented in this test by Opus) is even larger. This is due to the large number of audio channels to code and the comparatively low per-channel bitrate, which underlines the usefulness of parametric spatial audio coding techniques in high-channel count scenarios.

Analyzing the content dependency of the codecs, we observe a general trend favoring the proposed system. Within the test items curated to represent particularly challenging scenarios, a small (non-significant) exception may be found for items 6 and 9. We conjecture that this difference is related to the additional core-coder channels, which may benefit items where the higher-order channels contain a significant amount of uncorrelated signals.

For 512 kbps, IVAS still outperforms the FO-DirAC based system. One-sided paired t-tests (DF=99) indicate that, at a 95% significance level, the proposed merged system scores statistically significantly better than both Opus ($p < 0.01$) and the FO-DirAC based system ($p < 0.01$). The results indicate no statistically significant difference between Opus and the FO-DirAC based system ($p = 0.39$). At higher bitrates, we observe a trend of all systems towards the reference. This tendency is expected because the coding of all Ambisonics channels

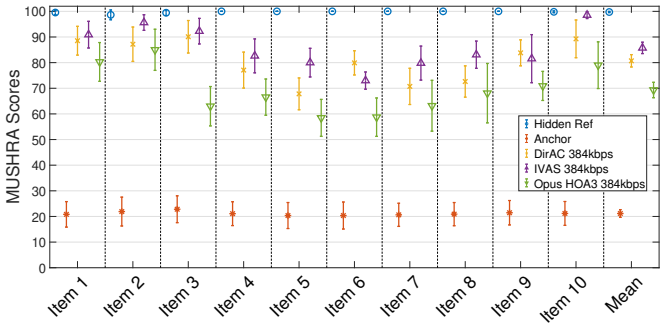

Fig. 4. Results of a MUSHRA listening tests with 11 expert listeners at 384 kbps. Confidence intervals are 95% with a Student's t-distribution.

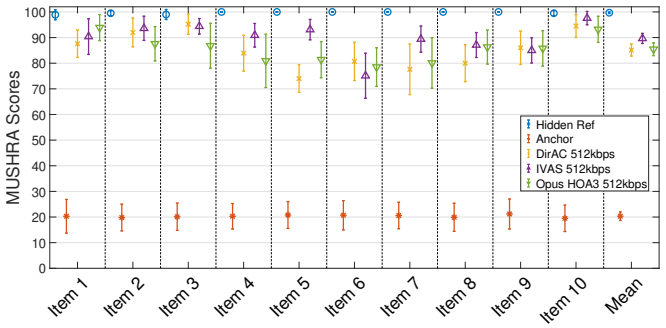

Fig. 5. Results of a MUSHRA listening tests with 10 expert listeners at 512 kbps. Confidence intervals are 95% with a Student's t-distribution

becomes the most accurate method when approaching very high bitrates. Nevertheless, IVAS exhibits very good scaling to its highest supported bitrate.

In previous internal listening tests with a DirAC system without SPAR, we found a quality improvement with the presented two-sector approach as compared to a single DoA and diffuseness approach. Furthermore, it was found in internal tests that the inclusion of higher-order planar SPAR channels in IVAS improves the perceptual quality. Hence, we concluded that both the two-sector approach and the tuning of SPAR contribute to the quality improvement of the combined system deployed in IVAS. This is in line with the general consent that more model parameters and more reconstructed channels can generally improve the perceptual quality, as long as the bitrate can accommodate the additional data.

IV. Complexity Measurements

Analogously to the lower bitrates in [11], we have measured the computational complexity of the codecs in the above perceptual-quality testing with the ITU-T weighted-million-operations-per-second (WMOPS) counting tool [23], [24]. As the Opus code is incompatible with the measurement tool, we have measured multi-mono EVS as a proxy for discrete coding of all Ambisonics channels. The EVS bitrate for each of these channels was 24.4 kbps and 32 kbps to compare to the other codecs at 384 and 512 kbps, respectively. This is the closest approximation of the per-channel bitrates with the

TABLE II
DESCRIPTIONS OF THE AUDIO SCENES IN THE LISTENING TESTS.

number	description
1	birds, background noise, non-overlapping speech
2	recording of a coffee break at a conference with overlapping speech and background noise
3	two overlapping speakers, one low- and one high-pitch
4	complex synthetic scene with animal sounds, musical instruments, and noise
5	complex synthetics scene with thunder, rain, and noise
6	recording with forest sounds and music, strong ambience
7	recording of pouring water
8	recording of a helicopter flying by
9	recording of speech and background sounds
10	panned overlapping speakers, one low and one high-pitch

TABLE III
WORST CASE COMPLEXITY FOR IVAS WITH HO-DIRAC, FO-DIRAC [9], AND MULTI-MONO EVS. ENCODING FROM AND DECODING TO 3RD ORDER AMBISONICS ARE COMBINED.

bitrate [kbps]	**complexity [WMOPS]**		
	IVAS	FO-DirAC	multi EVS
384	742.323[a]	788.507[b]	16 x 104.015 = 1664.24[c]
512	793.623[a]	791.542[b]	16 x 101.875 = 1630.0[d]

[a] 4 Downmix channels. [b] 8 Downmix channels.
[c]EVS at 24.4 kbps. [d]EVS at 32 kbps.

available EVS bitrates. The results are shown in Tab. III. We find that the computational complexity of IVAS is comparable to the FO-DirAC system of [9] and much lower than that of multi-mono EVS. While there is additional complexity for the parametric processing for HO-DirAC and the additional SPAR channels in IVAS (system fig. 2), scaling the low-bitrate IVAS architecture (system fig. 3) utilizes more core-coder channels. IVAS satisfies the complexity requirements of 3GPP for level 3 at both bitrates [15].

V. CONCLUSIONS

In summary, delivering HOA in IVAS may reach excellent perceptual quality. The effectiveness of the high-bitrate HOA codec structure in IVAS has been demonstrated. The combination of the particular HO-DirAC architecture with high-bitrate tunings for SPAR has been shown to improve the perceptual quality. If less bandwidth is available, the codec can still switch back to low-bitrate modes and scale down to a lower-order parametrization. The computational complexity is well within the limits defined by 3GPP and is not significantly higher than that of an equivalent system using FO-DirAC only. Hence, the IVAS codec scales well to higher bitrates, making it attractive not only for conference calls and voice communication but also streaming of high-quality immersive HOA content over 5G mobile networks.

ACKNOWLEDGMENT

The authors acknowledge contributions to the codec by Stefan Bayer.

REFERENCES

[1] F. Zotter and M. Frank, *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*. Springer International Publishing, 2019, vol. 19.
[2] D. G. Malham, "Higher order ambisonic systems for the spatialisation of sound," in *ICMC*, 1999.
[3] M. A. Gerzon, "Periphony: With-height sound reproduction," *J. Audio Eng. Soc*, vol. 21, pp. 2–10, Feb. 1973.
[4] C. Nachbar, F. Zotter, E. Deleflie, and A. Sontacchi, "AMBIX - A Suggested Ambisonics Format," in *Ambisonics Symposium 2011*, Lexington, 2011.
[5] M. A. Poletti, "Three-Dimensional Surround Sound Systems Based on Spherical Harmonics," *J. Audio Eng. Soc.*, vol. 53, no. 11, 2005.
[6] D. Ward and T. Abhayapala, "Reproduction of a plane-wave sound field using an array of loudspeakers," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 6, pp. 697–707, Sep. 2001.
[7] V. Pulkki, "Spatial Sound Reproduction with Directional Audio Coding," *J. Audio Eng. Soc*, vol. 55, no. 6, pp. 503–516, 2007.
[8] V. Pulkki, A. Politis, G. D. Galdo, and A. Kuntz, "Parametric spatial audio reproduction with higher-order B-format microphone input," 2013.
[9] G. Fuchs, F. Ghido, D. Weckbecker, and O. Thiergart, "A first-order DirAC-based parametric Ambisonic coder for immersive communications," in *ICASSP*. Hyderabad, India: IEEE, Apr. 2025, pp. 1–5.
[10] D. McGrath, S. Bruhn, H. Purnhagen, M. Eckert, J. Torres, S. Brown, and D. Darcy, "Immersive Audio Coding for Virtual Reality Using a Metadata-assisted Extension of the 3GPP EVS Codec," in *ICASSP 2019*. Brighton, United Kingdom: IEEE, May 2019, pp. 730–734.
[11] D. Weckbecker, S. Brown, J. Torres, M. Multrus, A. Tamarapu, and G. Fuchs, "Ambisonics Coding in IVAS: A Hybrid SPAR and DirAC System," in *ICASSP 2025*, Hyderabad, Apr. 2025, pp. 1–5.
[12] A. Politis, J. Vilkamo, and V. Pulkki, "Sector-Based Parametric Sound Field Reproduction in the Spherical Harmonic Domain," *IEEE J. Sel. Top. Signal Process.*, vol. 9, no. 5, pp. 852–866, Aug. 2015.
[13] C. Hold, V. Pulkki, A. Politis, and L. McCormack, "Compression of Higher-Order Ambisonic Signals Using Directional Audio Coding," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 32, pp. 651–665, 2024.
[14] C. Hold, L. McCormack, A. Politis, and V. Pulkki, "Perceptually-Motivated Spatial Audio Codec for Higher-Order Ambisonics Compression," in *ICASSP 2024*. Seoul, Korea, Republic of: IEEE, Apr. 2024, pp. 1121–1125.
[15] "IVAS Design Constraints," 3GPP, Tdoc S4-231031, May 2023.
[16] "Codec for Immersive Voice and Audio Services (IVAS); Detailed Algorithmic Description including RTP payload format and SDP parameter definitions," 3GPP, TS 26.253, Jul 2024.
[17] J. Paulus, L. Laaksonen, T. Pihlajakuja, M.-V. Laitinen, J. Vilkamo, and A. Vasilache, "Metadata-Assisted Spatial Audio (MASA) – An Overview," in *2024 IEEE 5th International Symposium on the Internet of Sounds (IS2)*. Erlangen, Germany: IEEE, Sep. 2024, pp. 1–10.
[18] G. Marković, E. Fotopoulou, J. F. Kiene, and C. R. Helmrich, "Efficient MDCT-Based Multi-Channel Coding with Perceptual Whitening and Broadband ILD Compensation," in *ICASSP 2025*. Hyderabad, India: IEEE, Apr. 2025, pp. 1–5.
[19] *Method for the subjective assessment of intermediate quality level of audio systems*, International Telecommunication Union Std. BS.1534-3, 2015.
[20] J. Valin, K. Vos, and T. Terriberry, "Definition of the Opus Audio Codec," RFC Editor, Tech. Rep. RFC6716, Sep. 2012. [Online]. Available: https://www.rfc-editor.org/info/rfc6716
[21] J. Skoglund and M. Graczyk, "Ambisonics in an Ogg Opus Container," RFC Editor, Tech. Rep. RFC8486, Oct. 2018. [Online]. Available: https://www.rfc-editor.org/info/rfc8486
[22] I. P. Collaboration, "Ivas processing scripts," https://forge.3gpp.org/rep/ivas-codec-pc/ivas-processing-scripts, 2025, accessed: 02 07, 2025.
[23] *Software tools for speech and audio coding standardization*, International Telecommunication Union Std. ITU-T G.191, 5 2024.
[24] "ITU-T Software Tool Library (STL)," https://github.com/openitu/STLhttps://github.com/openitu/STL.