COMPRESSING HIGHER ORDER AMBISONICS OF A MULTIZONE SOUNDFIELD

Panji Setiawan and Wenyu Jin

Huawei Technologies Düsseldorf GmbH
European Research Center
Munich, Germany

ABSTRACT

At present, the majority of the studies in the area of multizone soundfield reproduction are focused on the decoding of the soundfield. In this work, we propose an approach to encode the multizone soundfield within the desired reproduction region based on higher order Ambisonics (HOA) formats. The B-format signals for the complex multizone soundfield can be derived based on the coefficients of a formulated planewave expansion. The multizone soundfield B-format signals are then compressed using two state-of-the-art audio codecs as if they were individual mono signals. The results confirm the effectiveness of this HOA-based multizone soundfield encoding. Additionally, a significant reduction on the encoding rate of the desired multizone soundfield with sufficient accuracy can be achieved by quantitatively analyzing the reproduction performance of the system. We use the acoustic energy contrast between the selected bright zone and quiet zone as the performance evaluation measure.

1. INTRODUCTION

Recently, there is an increasing interest in the area of spatial soundfield reproduction systems. Well known approaches include wave field synthesis (WFS) [1, 2], higher order Ambisonics (HOA) [3, 4] and least squares methods [5–8]. The HOA system, in particular, is based on the decompositions of the desired soundfield and the reproduced soundfields due to the secondary sources (i.e., loudspeakers) within the desired listening area into spherical harmonics (3-D) or cylindrical harmonics (2-D) [9, 10]. In this work the latter is considered for the ease of demonstration. The decompositions of the relevant soundfields into spatial basis harmonic functions facilitate the method of mode-matching that leads to an equation system. For a two-dimensional representation and an order of M, $2M+1$ components are required to achieve an accurate representation of the soundfield. However, it usually leads to a high total bit rate that can be a huge burden for both storage and network transmission.

In [11], the authors explore a method for reducing the bit rate needed for transmitting and storing HOA signals. The so-called HOA B-format signals were directly compressed using MPEG-4 High Efficiency Advanced Audio Coding (HE AAC) [12], as if they were individual mono signals. Simulation results suggest that by assigning the higher order components with a low bit rate compression does not lead to a reduced audio quality. The resulting error is significantly low at the center of the reproduction region, but increases as a function of distance from the center. Therefore, the reduction of the total bit rate for HOA formats can be achieved by assigning the higher order components with a low bit rate in practice. This finding may lead to the conclusion that other single channel/mono speech and audio codecs can basically be used as alternatives to HE AAC, such as MPEG-D USAC [13], the ITU-T standardized

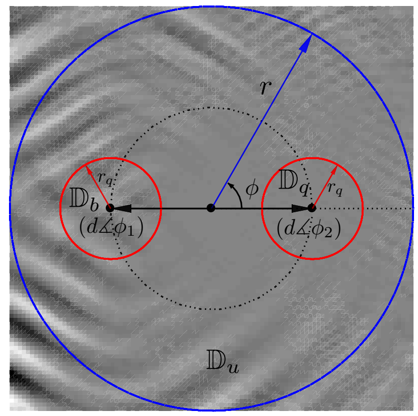


Fig. 1: The multizone soundfield reproduction over the desired reproduction region using Q loudspeakers.

audio codecs [14], the open, royalty-free Opus audio codec [15, 16], and the 3GPP standardized audio codecs (e.g., Enhanced Voice Services (EVS) [17]). Further reduction of the bit rate is feasible by exploiting the spatial redundancy as demonstrated by the recently developed MPEG-H 3D Audio codec [18], which supports the spatial compression of HOA B-format signals of up to order 6 resulting in a maximum total bit rate of 1.2 Mbps.

The aforementioned techniques are designed for the single sweet-spot or zone formats. Multizone soundfield is a more challenging problem in spatial soundfield reproduction and has drawn the researchers' attention recently. It aims at providing an individual sound environment to each of the listeners simultaneously using a set of loudspeakers. Currently, the majority of the studies in this area are focused on the rendering of the multizone soundfield, i.e., the decoding end [19–24]. At present, no study is found that specifically looks at the reduction of encoding rate for multizone soundfield.

In this paper, we propose a method to encode the desired multizone soundfield within the reproduction region based on HOA formats. The HOA B-format signals for the desired multizone soundfield can be derived based on the coefficients for a well-formulated planewave expansion. Two state-of-the-art audio codecs, i.e., 3GPP EVS and MPEG-4 HE AAC, are employed to compress the multizone soundfield HOA B-format signals. The recently standardized EVS codec represents the latest technology deployed for communications and HE AAC has been one of the widely deployed codec technology for streaming and broadcasting. We examine the acoustic contrast between the selected bright and quiet zones of the rendered soundfield at the decoding end and the results confirm the feasibility of encoding the multizone soundfield using the same framework of HOA format encoding.

2. PROBLEM FORMULATION

The study aims to investigate the feasibility of encoding and compressing the desired multizone soundfield based on the HOA formats. In this work, we mainly focus on the 2-D (height invariant) case and disc-shaped zones are considered. The theory is readily extended to 3-D space.

As illustrated in Figure 1, the desired reproduction region $\mathbb{D}$ is the entire control zone of interest with a radius of r, which includes both the acoustic bright zone $\mathbb{D}_b$ and the quiet zone $\mathbb{D}_q$ of radius r_q. The target bright and quiet zones are located at angles ϕ_1 and ϕ_2, respectively. The remaining area in $\mathbb{D}$ is defined as the unattended zone $\mathbb{D}_u$. $k = 2\pi f/c$ is the wavenumber, where f is the frequency and c is the speed of sound propagation. Throughout this paper, we use k instead of f to represent the frequency since we assume c to be constant. We assume no sound sources or scattering objects being present inside the reproduction area. The number of employed loudspeakers is Q.

3. B- AND D-FORMAT FOR MULTIZONE SOUNDFIELD

In [8], the authors formulated a soundfield orthogonal basis expansion to describe any feasible soundfield over the desired reproduction region with a suitable set of planewave functions as input. It reduces the multizone sound reproduction problem to the reconstruction of a set of planewave functions over the desired region. In this section, we follow the outcome of this approach, which allows us to derive the B-format signals for the desired multizone soundfield.

To express the approximation of any feasible 2-D soundfield $S^a(\mathbf{x}, k)$, we can write it as a weighted series of planewaves arriving from various angles [8]:

$$S^a(\mathbf{x}, k) = \sum_{j=1}^{2M+1} P_j(k) F_j(\mathbf{x}, k), \tag{1}$$

where P_j is a set of coefficients for the planewave functions and $\mathbf{x} \equiv (\|\mathbf{x}\|, \angle\theta_x)$ denotes an arbitrary spatial observation point in polar coordinate. Note that $M = kr$ is the truncational length of the desired reproduction region. The planewave function $F_j(\mathbf{x}, k) = e^{ik\mathbf{x}\cdot\hat{\phi}_\mathbf{j}}$, where $\hat{\phi}_\mathbf{j} \equiv (\mathbf{1}, \hat{\phi}_\mathbf{j})$ is the unit vector in the direction of the jth planewave and $\hat{\phi}_j = \frac{2\pi(j-1)}{2M+1}$. More details on the derivations of P_j can be found in [8].

A horizontal 2-D soundfield can be expressed in terms of its cylindrical harmonics decomposition [25]:

$$\begin{aligned} S(\mathbf{x}, k) = B_{00}^{+1} J_0(k\|\mathbf{x}\|) + \sum_{m=1}^{M} J_m(k\|\mathbf{x}\|) B_{mm}^{+1} \sqrt{2}\cos(m\theta_x) \\ + \sum_{m=1}^{M} J_m(k\|\mathbf{x}\|) B_{mm}^{-1} \sqrt{2}\sin(m\theta_x), \end{aligned} \tag{2}$$

where $J_m(\cdot)$ is the m-th order Bessel function and M is the truncational length. The coefficients B_{mm}^{+1} are the so-called B-format signals in Ambisonics. The B-format signals can be found either by encoding each virtual source signal individually [4] or by using a multi-element microphone that extracts the B-format signals by processing the microphone signals [26]. For example, the B-format signals for a planewave function arriving from the angle θ_j are given by: $B_{mm}^{+1}(\theta_j) = i^m e^{-im\theta_j}$, $B_{mm}^{-1}(\theta_j) = i^{-m} e^{im\theta_j}$, where $m = 0, \cdots, M$.

Incorporating the coefficients for the planewave functions P_j, we can derive the B-format signals for the desired multizone soundfield for $m = 0, \cdots, M$:

$$\begin{aligned} B'^{+1}_{mm}(k) &= \sum_{j=1}^{2M+1} P_j(k) i^m e^{-im\theta_j}, \\ B'^{-1}_{mm}(k) &= \sum_{j=1}^{2M+1} P_j(k) i^{-m} e^{im\theta_j}. \end{aligned} \tag{3}$$

To derive the loudspeaker signals for the Q loudspeakers (the D-format) from the B-format, the so-called mode-matching is used [10,27]. Assuming that loudspeakers are far enough from the desired reproduction region, we have the following relationship based on the matrix product

$$\mathbf{B}'(k) = \mathbf{C}\boldsymbol{\ell}(k), \tag{4}$$

where $\boldsymbol{\ell}(k) = [\ell_1(k), \cdots, \ell_Q(k)]^T$ represents the loudspeaker weights, $\mathbf{B}'(k) = [B'^{+1}_{00}(k), B'^{+1}_{11}(k), B'^{-1}_{11}(k), \cdots, B'^{-1}_{MM}(k)]^T$, and $\mathbf{C}$ is a $[(2M+1) \times Q]$ matrix featuring the B-format signals for the planewaves from the directions of the loudspeakers. Therefore, the decoding matrix $\mathbf{D}$ can be defined as

$$\mathbf{D} = pinv(\mathbf{C}) = (\mathbf{C}^H\mathbf{C})^{-1}\mathbf{C}^H, \tag{5}$$

provided that there are sufficient loudspeakers, i.e., $Q \geq 2M + 1$. The decoding operation is given by:

$$\boldsymbol{\ell}(k) = \mathbf{D}\mathbf{B}'(k). \tag{6}$$

The reproduced soundfield $S^a(\mathbf{x}, k)$, due to loudspeakers radiating planewaves can be written as:

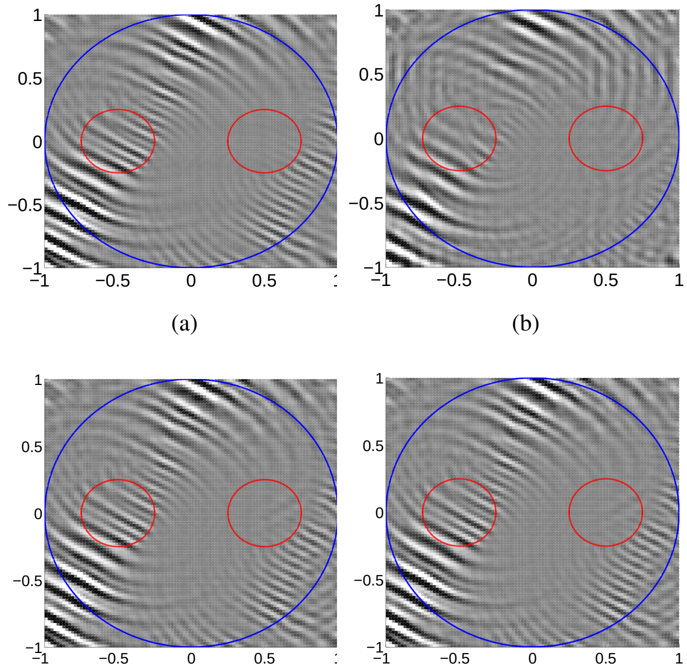
$$S^a(\mathbf{x}, k) = \sum_{q=1}^{Q} \ell_q(k) e^{ikr\cdot\cos(\theta_q - \theta_\mathbf{x})}, \tag{7}$$

where θ_q is the angle of the q-th loudspeaker. Only a short introduction to HOA-based B-format and D-format is given in this section since the complete theory is thoroughly presented in [4, 27].

4. MULTIZONE SOUNDFIELD COMPRESSION

It has been shown in [11] that the suboptimal way of compressing the HOA B-format signals with HE AAC seems to give a remarkably well outcome and that the distortion is comparable whether it is the D- or B-format signals that have been compressed. It also went further by mentioning that large soundfield distortion was observed when lower bit rates were used (in particular at 12 kbps). This is due to the Spectral Band Replication (SBR) technique being active in HE AAC, although perceptually the quality drop is not as big.

In this work, a similar compression method was adopted for the multizone soundfield case. The multizone B-format signals were initially extracted and subsequently compressed. In addition to HE AAC implementation obtained from [28], the recently standardized EVS audio codec was also selected. EVS is an attractive candidate due to several reasons. Firstly, EVS represents the latest single channel audio compression technology as a result of intensive collaboration efforts between 12 leading telecommunications companies and research institutions. It will be interesting to observe its performance for compressing a multizone soundfield. Secondly, EVS has undergone extensive listening tests [29] and can be concluded that it performs really well at low bit rates. It is certainly a better alternative choice to eliminate the SBR distortion effect in HE AAC at low bit



(a) (b) (c) (d)

Fig. 2: Soundfield snapshots (2×2 m) of a single source chirp signal reproduced with HOA order 50. The smaller circle to the left in each subfigure is the bright zone and to the right is the quiet zone. (a) Original uncompressed soundfield, (b) HE AAC compressed soundfield at 6.04 Mbps, (c) EVS compressed soundfield at 6.04 Mbps, (d) EVS compressed soundfield at 2.104 Mbps.

rates. Thirdly, if similar and consistent observations as in [11] are obtained, this will certainly add more weights to the initial thought that any single channel audio codec may actually be employed in compressing the HOA B-format signals in general.

Having encoded a virtual source signal by extracting its multizone soundfield B-format signals following (3), an inverse Fourier transform (IFFT) is performed to obtain the corresponding time domain signals. Typically it is done in a framewise overlap-add manner employing a window and an IFFT length of around 20-30 ms and an overlap factor of 50%. These time domain signals are then compressed using the previously mentioned audio codecs, i.e., EVS and HE AAC. At the receiving end, the decompressed signals are transformed back to the frequency domain by means of Fourier transform (FFT). The loudspeaker weights for the desired multizone soundfield are finally obtained using the distorted frequency domain B-format signals following (6). The soundfield reproduction is then performed using (7).

5. RESULTS AND DISCUSSION

For the ease of demonstration, we investigate the dualzone case in this work. However, the system is readily extended to the case with more than two selected zones. The encoding rate is not proportional to the number of zones in the desired multizone soundfield. This is due to the fact that the sound basis function approach in Section 3 enables us to match the desired multizone soundfield over the entire region using a single set of HOA B-format signals.

The speed of sound c is 343 m/s in our simulations. A total amount of $Q = 101$ loudspeakers, representing the highest HOA order 50, are evenly distributed along a concentric circle. The centers of $\mathbb{D}_b$ and $\mathbb{D}_q$ lie on a circle of radius $d = 0.5$ m within $\mathbb{D}$. The target bright and quiet zones were located at 180° and 0°, respectively, with $r_q = 0.25$ m and $r = 1$ m, as shown in Figure 2. The desired soundfield over $\mathbb{D}_b$ is selected to be a planewave arriving from 60° for the following simulations. The FFT/IFFT and frame lenghts were chosen to be 1024 samples. Both objective and subjective performance analysis was conducted to assess the performance.

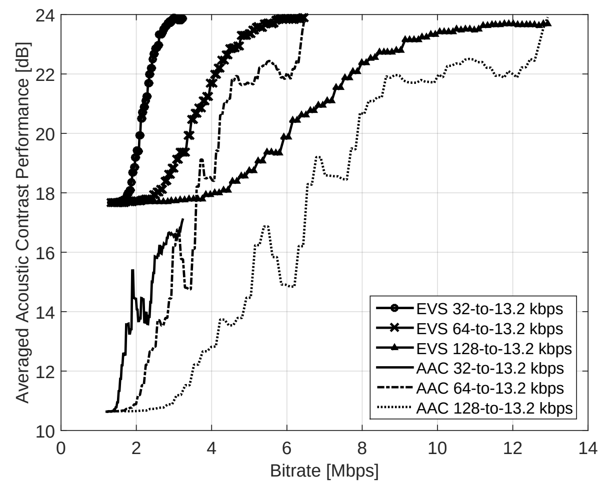


Fig. 3: Multizone soundfield HOA-based compression using EVS and HE AAC. 128-to-13.2 kbps means the curve was generated by firstly individually compressing the B-format signals with 128 kbps and finally compressing them all with 13.2 kbps. The intermediate bit rates were obtained by gradually changing the bit rate from 128 to 13.2 kbps starting from the highest to the lowest order component.

5.1. Snapshot Observation

The most straightforward way of analysing the objective performance is by visually observing the soundfield snapshots. A half second long and fullband 48 kHz sampled chirp signal was used as the virtual source. The values of weighting function $w(\mathbf{x})$ that specifies the importance of reproduction accuracy for $\mathbb{D}_b$, $\mathbb{D}_q$ and $\mathbb{D}_u$ were 1, 10 and 0.05, respectively. Comparing Fig. 2(a), the uncompressed, and 2(b), the compressed soundfield, the distortions are clearly visible on the compressed soundfield in both the bright and quite zones, however, the wave contour indicating the right propagation direction is still feasible in the bright zone, as well as the attenuation in the quiet zone. This implies that the compression has indeed formed a distorted soundfield but it is still somewhat acceptable. The compressed soundfield was obtained using HE AAC with the first 41 lower order HOA components are subject to 128 kbps compression rate each and the remaining higher order components being assigned to 13.2 kbps giving a total bit rate of around 6 Mbps. Less distortions are observed in Figure 2(c), where EVS was operated at the same total bit rate of around 6 Mbps. This certainly highlights the superiority of EVS compared to HE AAC. This is made more evident by observing Figure 2(d), where EVS operating at around 2.1 Mbps shows similar distortions to that at around 6 Mbps due to the reason described later in this Section. This operating bit rate is attractive considering the maximum bit rate of MPEG-H 3D audio codec at 1.2 Mbps tested with up to order 6, whereas here we have order 50.

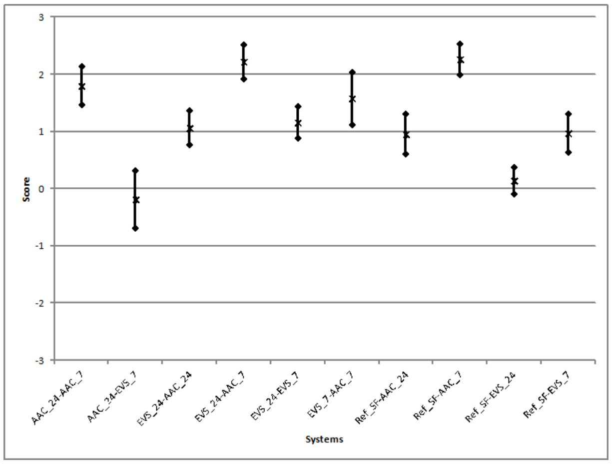


Fig. 4: Mono and Narrowband AB7 listening test.

5.2. Acoustic Contrast Performance

To further evaluate the objective performance of multizone sound rendering system, we use the following measure of the acoustic brightness contrast between the bright zone $\mathbb{D}_b$ and the quiet zone $\mathbb{D}_q$ at the decoding end, which quantifies the sound leakage between the two zones:

$$\zeta(t) = 10 \cdot \log_{10} \frac{\frac{1}{\|\mathbb{D}_b\|} \int_{\mathbb{D}_b} |S(\mathbf{x}, t)|^2 d\mathbf{x}}{\frac{1}{\|\mathbb{D}_q\|} \int_{\mathbb{D}_q} |S(\mathbf{x}, t)|^2 d\mathbf{x}}, \tag{8}$$

where $\|\mathbb{D}_b\|$ and $\|\mathbb{D}_q\|$ mark the sizes of $\mathbb{D}_b$ and $\mathbb{D}_q$. The error is then calculated frame by frame, and averaged over time. It should be noted that this error measure does not take any perceptual aspects into account. Furthermore, the authors in [30] suggest that an acoustic energy contrast of 20 dB between the bright and quiet zone denotes a perceivable multizone sound implementation.

Figure 3 depicts the performance curves following (8) of both EVS and HE AAC codecs for the multizone soundfield compression. For the following simulation, the chirp signal used in Section 5.1 has been bandlimited to 5 kHz. Each codec has three different curves corresponding to a gradual decrease of bit rates from the maximum equal rate compression of individual B-format signal with 128, 64, and 32 kbps, to the minimum equal rate compression of 13.2 kbps.

Obviously, EVS performs better than HE AAC in all cases. Clear performance advantage of EVS is shown when all B-format signals were equally compressed with 13.2 kbps. Here, EVS can still achieve the performance of slightly lower than 18 dB whereas HE AAC stands at around 10.5 dB. These values represent the lowest achievable performance levels. Moreover, when all B-format signals were equally compressed with 32 kbps, HE AAC performance stands at around 17 dB whereas the somewhat 24 dB reference is still maintained by EVS, which indicates a significant drop of HE AAC performance for the 32-to-13.2 kbps curve. Finally, given a certain performance level requirement, e.g., 20 dB, EVS will always deliver the minimum total bit rate provided that the same bit rate assignment is applied.

Further conclusions can be drawn by relating Figure 3 to the soundfield snapshots in Figure 2. Figure 2(b) of HE AAC scores the performance level of around 15 dB at roughly 6 Mbps, whereas EVS in Figure 2(c) approximately scores 20 dB having the same total bit rate. The same performance level of 20 dB can be achieved by EVS at a total bit rate of around 2.1 Mbps as depicted in Figure 2(d).

5.3. Listening Test Performance

A mono and narrowband AB7 listening test was conducted to assess the subjective performance. AB7 is an A-B comparison listening test having a 7-point response scale (+/-3) where the score 0 is labeled as ”A is the same as B” and the scores 1, 2, 3, denote ”A is slightly better, better, and much better than B”, respectively. The negative scores indicate the reverse comparison of B to A. A total of 8 listeners, 6 males and 2 females, were taking part to the test. Five listeners are expert listeners in audio coding and/or reproduction and three are naïve listeners. Four audio materials consisting of 2 speech and 2 music signals were extracted from SQAM database and downsampled to 8 kHz. The final test items were extracted from the reproduced soundfield at the center of bright zone. These mono narrowband items were played back as stereo using a pair of stereo headsets during test. The values of weighting function $w(\mathbf{x})$ assigned to $\mathbb{D}_b$, $\mathbb{D}_q$ and $\mathbb{D}_u$ were adjusted to 1, 2 and 0.01, respectively.

Five different conditions were extracted consisting of the reference reproduced soundfield without any compression (REF_SF) and the reproduced soundfield having the B-format signals individually compressed with EVS and AAC codecs at 7.2 kbps (i.e., EVS_7 and AAC_7, respectively) and at 24.4 kbps (i.e., EVS_24 and AAC_24, respectively). We have used the EVS bitrates as a reference since we are able to operate the AAC at any specified bitrate. The chosen bitrates correspond to the lowest and highest possible narrowband bitrates as specified by EVS without the DTX (Discontinuous Transmission) feature. Please note that the bitrate of 7.2 kbps is not part of the recommended AAC operating bitrates.

Figure 4 shows the listening test results analyzed using 95 % confidence interval. Each of the five conditions is compared with the remaining conditions yielding ten different comparison pairs. It shows that two pairs, i.e., REF_SF-EVS_24 and AAC_24-EVS_7, are having the bar crossing the score 0 which implies that the conditions are perceptually similar. Four pairs are having the bar crossing the score 2 which means that these pairs are the ones showing a clearly perceptual difference. These four pairs are consistently having AAC_7 as one of the conditions, thus we conclude that the degradation caused by AAC_7 can be significantly perceived by the listeners and rated worse. This result confirms our expectation due to the reasoning presented earlier regarding the recommended operating bitrate. The other four pairs are having the bar crossing the score 1 which means that there is a perceptually slight significant difference between these conditions. The highest perceptual narrowband quality can actually be obtained by the EVS_24 compression method followed by AAC_24 then EVS_7. The lowest EVS operating bitrate of 7.2 kbps is therefore very appealing because it allows for a slight degradation while gaining a considerable amount of total bitrate saving. A total bitrate of 727.2 kbps is therefore needed to allow for this narrowband dualzone reproduction.

6. CONCLUSION

We presented in this paper a HOA-based multizone soundfield encoding and compression employing two state-of-the-art audio codecs, i.e., EVS and HE AAC. It has been shown that compressing a large number of B-format signals is highly feasible as demonstrated by compressing of order 50 B-format signals into a total bit rate of merely 2.1 Mbps using EVS. The superiority of EVS has also been demonstrated compared to HE AAC in handling lower bit rates compression. This proposed compression method for a multizone soundfield can also take advantage from any additional pre-processing which may reduce the number of B-format signals prior to compression.

7. REFERENCES

[1] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *The Journal of the Acoustical Society of America*, vol. 93, no. 5, pp. 2764–2778, 1993.

[2] D. de Vries, "Sound reinforcement by wavefield synthesis: Adaptation of the synthesis operator to the loudspeaker directivity characteristics," *J. Audio Eng. Soc*, vol. 44, no. 12, pp. 1120–1131, 1996.

[3] M. A. Gerzon, "Periphony: With-height sound reproduction," *J. Audio Eng. Soc*, vol. 21, no. 1, pp. 2–10, 1973.

[4] J. Daniel, S. Moreau, and R. Nicol, "Further investigations of high-order ambisonics and wavefield synthesis for holophonic sound imaging," in *Audio Engineering Society Convention 114*, Mar 2003.

[5] O. Kirkeby and P. A. Nelson, "Reproduction of plane wave sound fields," *The Journal of the Acoustical Society of America*, vol. 94, no. 5, pp. 2992–3000, 1993.

[6] W. Jin, W. B. Kleijn, and D. Virette, "Multizone soundfield reproduction using orthogonal basis expansion," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, May 2013, pp. 311–315.

[7] W. Jin and W. B. Kleijn, "Multizone soundfield reproduction in reverberant rooms using compressed sensing techniques," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, May 2014, pp. 4728–4732.

[8] W. Jin and W. B. Kleijn, "Theory and design of multizone soundfield reproduction using sparse methods," *Audio, Speech, and Language Processing, IEEE/ACM Transactions on*, vol. 23, no. 12, pp. 2343–2355, Dec 2015.

[9] S. Spors and J. Ahrens, "Comparison of higher-order ambisonics and wave field synthesis with respect to spatial aliasing artifacts," in *19th International Congress on Acoustics*, 2007, pp. 2–7.

[10] M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *J. Audio Eng. Soc*, vol. 53, no. 11, pp. 1004–1025, 2005.

[11] I. Burnett, E. Hellerud, A. Solvang, and U. P. Svensson, "Encoding higher order ambisonics with AAC," in *Audio Engineering Society Convention 124*, May 2008.

[12] M. Wolters, K. Kjorling, D. Homm, and H. Purnhagen, "A closer look into MPEG-4 High Efficiency AAC," in *Audio Engineering Society Convention 115*, Oct 2003.

[13] M. Neuendorf, P. Gournay, M. Multrus, J. Lecomte, B. Bessette, R. Geiger, S. Bayer, G. Fuchs, J. Hilpert, N. Rettelbach, R. Salami, G. Schuller, R. Lefebvre, and B. Grill, "Unified speech and audio coding scheme for high quality at low bitrates," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, April 2009, pp. 1–4.

[14] R. V. Cox, S. F. de Campos Neto, C. Lamblin, and M. H. Sherif, "ITU-T coders for wideband, superwideband, and fullband speech communication," *Communications Magazine, IEEE*, vol. 47, pp. 106–109, Oct 2009.

[15] K. Vos, K. V. Sørensen, S. S. Jensen, and J.-M. Valin, "Voice coding with Opus," in *Audio Engineering Society Convention 135*, Oct 2013.

[16] J.-M. Valin, G. Maxwell, T. B. Terriberry, and K. Vos, "High-quality, low-delay music coding in the Opus codec," in *Audio Engineering Society Convention 135*, Oct 2013.

[17] 3GPP Specification: TS 26.445, "Codec for enhanced voice services (EVS); detailed algorithmic description v12.3.0," Jun 2015.

[18] J. Herre, J. Hilpert, A. Kuntz, and J. Plogsties, "MPEG-H 3D audio–the new standard for coding of immersive spatial audio," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 9, no. 5, pp. 770–779, Aug 2015.

[19] M. Poletti, "An investigation of 2-D multizone surround sound systems," in *Audio Engineering Society Convention 125*, Oct 2008.

[20] Y. J. Wu and T. D. Abhayapala, "Multizone 2D soundfield reproduction via spatial band stop filters," in *Applications of Signal Processing to Audio and Acoustics, 2009. WASPAA '09. IEEE Workshop on*, Oct 2009, pp. 309–312.

[21] T. D. Abhayapala and Y. J. Wu, "Spatial soundfield reproduction with zones of quiet," in *Audio Engineering Society Convention 127*, Oct 2009.

[22] Y. J. Wu and T. D. Abhayapala, "Spatial multizone soundfield reproduction: Theory and design," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 6, pp. 1711–1720, Aug 2011.

[23] P. Coleman, P. J. B. Jackson, M. Olik, M. Møller, M. Olsen, and J. Abildgaard Pedersen, "Acoustic contrast, planarity and robustness of sound zone methods using a circular loudspeaker array," *The Journal of the Acoustical Society of America*, vol. 135, no. 4, pp. 1929–1940, 2014.

[24] T. Okamoto, "Generation of multiple sound zones by spatial filtering in wavenumber domain using a linear array of loudspeakers," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, May 2014, pp. 4733–4737.

[25] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*, Elsevier, 1999.

[26] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, May 2002, vol. 2, pp. II–1781–II–1784.

[27] M. A. Poletti, "A unified theory of horizontal holographic sound systems," *J. Audio Eng. Soc*, vol. 48, no. 12, pp. 1155–1182, 2000.

[28] *Nero AAC codec V1.5.1.0*, 2009, `http://www.nero.com`.

[29] ETSI TR 126 952, "Universal mobile telecommunications system (UMTS); LTE; codec for Enhanced Voice Services (EVS); performance characterization (3GPP TR 26.952 version 12.2.0 release 12) v12.2.0," July 2015.

[30] J.-H. Chang, C.-H. Lee, J.-Y. Park, and Y.-H. Kim, "A realization of sound focused personal audio system using acoustic contrast control," *The Journal of the Acoustical Society of America*, vol. 125, no. 4, pp. 2091–2097, 2009.