Evaluation of Compressed Higher-Order Ambisonics at Off-Center Listening Positions

Matthias Frank
Institute of Electronic Music and Acoustics
University of Music and Performing Arts
Graz, Austria
frank@iem.at

Sonja Lutz
University of Music and Performing Arts
Graz, Austria
sonja-claudia.lutz@student.kug.ac.at

***Abstract*—This paper evaluates the quality of OPUS-compressed 1st-, 3rd-, and 5th-order Ambisonics at off-center listening positions in a studio with a hemispherical loudspeaker arrangement. The audio signals were compressed with 16, 32, and 64 kb/s per channel and compared to an uncompressed 5th-order reference in a listening experiment. Audio scenes comprised frontally panned speech and music with and without surrounding reverberation. The experiment reveals the requirements of Ambisonics order and channel bandwidth to achieve excellent quality in dependence of the audio material and listening position for streaming of spatial audio across different venues.**

***Index Terms*—Ambisonics, OPUS, compression, off-center**

I. Introduction

Networked Music Performances (NMP) with distributed musicians and audiences [1] are an important component of the Internet of Sounds (IoS) [2]. Current research indicates the benefit of spatial audio for NMP with increased perceptual ratings for localization, immersion, social presence, realism, and connection with other musicians in comparison to stereo [3], [4] and thus its application is emerging [5], also for interactive listening experiences in museums and cultural heritage sites [6]. In addition to spatial audio rendering, Turchet categorized performance and functional requirements for such applications that include low latencies and network traffic predictions [7]. Moreover, it seems beneficial to adapt the rendering to the available computational capabilities and listening situation, e.g. headphones or loudspeakers. End-to-End latency must be kept below 30 ms to achieve conditions that are similar to those in traditional in-presence musical performances [8], [9]. Besides the network itself, one major contributor to the latency is the audio codec and in particular its frame size. Specifically designed for interactive speech and audio transmission, the OPUS codec [10] provides small frame sizes down to 2.5 ms. Moreover, it supports up to 255 audio channels, which makes it suitable for transmission of spatial audio, especially for Higher-Order Ambisonics.

Ambisonics is a spatial audio technology [11]–[13] that represents the soundfield by a sum of basis functions. The Ambisonics order N defines the spatial resolution and the number of basis functions $(N+1)^2$ and channels. The area around the center of an Ambisonic playback system, where the sound pressure of a desired sound field can be recreated accurately, is rather small and it increases with the order and decreases with frequency. Using 1st-order Ambisonics, this area tightly includes the head of a single listener for frequencies up to 700 Hz [14]. With Higher-Order Ambisonics (HOA), which has been researched with increasing interest around the 2000s [15], providing physically accurate sound to four listeners simultaneously would require at least an order of 14 (196 loudspeakers for a full sphere) for an upper frequency limit of 1.4 kHz. Nevertheless, practical experience and perceptual evaluation indicated the *perceptual sweet area* to be much larger [16]–[19].

One of the advantages of Ambisonics is the separation of the recording/production process and the playback. Thus, Ambisonics can be played back on arbitrary surrounding loudspeaker arrangements [20] and headphones [21], [22] including efficient incorporation of head rotations. This is especially interesting for streaming and transmitting between different venues [23], which do not provide standard loudspeaker arrangements [24]. Moreover, the size of the sweet area can be adjusted by applying weighting to the Ambisonics signal that attenuates sidelobes, such as max-r_{E} [15]. The number of transmission channels in HOA is not depending on the number of actual sound objects, which helps predicting network traffic. Moreover, the order for decoding can easily be adapted to the available computational resources at the receiver. As the number of transmission channels in HOA is typically high, compression could be helpful.

The studies by Narbutt [25]–[27] investigated localization accuracy and listening quality on headphones for 1st- to 3rd-order Ambisonics using the OPUS codec 1.2 with bandwidths of 8, 16, and 32 kb/s per channel. While the quality mainly benefited from higher bandwidth, localization improved with higher orders. Rudzki [28] extended research towards higher orders (up to 5th order), channel bandwidths (up to 64 kb/s), and loudspeaker playback in an acoustically-treated listening room. The results indicate that strong compression also impairs localization and that at least 3rd order is necessary for complex audio scenes [29]. At the central listening position, localization with 5th order and 64 kb/s was comparable to the uncompressed 5th-order reference. Moreover, 32 kb/s was sufficient for speech [30] and there was barely any difference between channel mapping families 2 and 3 in OPUS 1.3.

Experiments in a studio environment were done in [31] to compare bandwidth requirements for stereo and surround (5.1 and 7.1.4) playback with OPUS 1.3. While 64 kb/s were necessary for stereo playback that was indistinguishable from the reference, 48 kb/s was already enough for surround. In the same setup, there were no audible artefacts for 3rd-order Ambisonics at 48 kb/s [32].
Recent publications presented more sophisticated codecs that try to make use of the redundancy in HOA [33]. While the compression rate is high for dry audio scenes with a low number of audio objects, it can be reduced by complex scenes with reverberation [34]. To save bandwidth in such audio scenes, some codecs employ convolutional neural networks [35]. Alternatively, only lower-order signals could be transmitted and upmixed to higher orders at the receiver [36]. Parametric spatial audio compression follows a similar idea [37].
However, this paper sticks to the publicly available OPUS codec that is implemented on a variety of platforms. We extend existing studies by evaluating off-center listening positions to investigate the influence of compression on the perceptual sweet area, as it has been analyzed theoretically in [38]. First, we describe the setup of the listening experiment, its audio scenes and the conditions that combine 1st-, 3rd-, and 5th-order Ambisonics with 16, 32, 64 kb/s bandwidth per channel. Subsequently, the results are presented and discussed in comparison to previous experiments. Finally, the findings are summarized and we propose order/bandwidth combinations for excellent quality in dependence of audio material and size of the target listening area.

II. Setup and Conditions

The experiment compared nine conditions combining three different Ambisonics orders (1, 3, 5) with three different channel bandwidths (16 kb/s, 32 kb/s, 64 kb/s) against the uncompressed 5th-order reference. Compression employed the OPUS codec 1.3 with channel mapping 255 (independent channels) and a frame size of 20 ms as it is available in the digital audio workstation Reaper[1] that was used for rendering. The four different audio scenes comprised speech (first sentence of EBU's male English speech reference recording [39]) and music with percussive elements (a 7 seconds long excerpt of the song "What's Trumps" by the band Rhythmussportgruppe [40] which is part of the DEGA stimulus database) both with and without reverberation. In all audio scenes, the direct sound was panned to 0° azimuth and elevation, i.e. the position of the front loudspeaker C in Figure 1. Surrounding reverberation was created by a 64×64 feedback delay network in the *FdnReverb*[2] plug-in with a reverberation time of 2.0 seconds below 1 kHz that was gradually reduced to 1 second at 10 kHz. The room size parameter was set to 20 and the 64 output channels were encoded to 64 directions that were evenly distributed on a sphere.
The experiment was done at the IEM CUBE, a $10.3\,\mathrm{m} \times 12\,\mathrm{m} \times 4.8\,\mathrm{m}$ studio, cf. Figure 1, with a reverberation time of 0.5 s.

[1]https://www.reaper.fm/

[2]part of the open-source plug-in suite from https://plugins.iem.at/

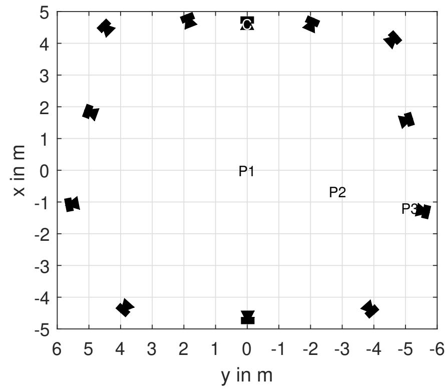


Fig. 1. Horizontal loudspeaker arrangement (C indicates frontal center loudspeaker) and listening positions P1, P2, and P3 in the experiment.

A hemispherical arrangement of 25 d&B loudspeakers was used for playback. It comprises a ring of 12 loudspeakers (12S-D) at ear height, two elevated rings of 8 and 4 loudspeakers (8S), and a voice-of-god loudspeaker (8S) above the central listening position. Ambisonic decoding used AllRAD [20] including appropriate max-r_{E} weighting [15] for each playback order and no delay compensation of the different loudspeaker distances to the center in order to achieve best results in terms of localization and coloration at off-center listening positions [17], [41].
The experiment was performed at three different listening positions, cf. Figure 1: P1 at the center, P2 halfway to a lateral loudspeaker, and P3 half a meter away from that loudspeaker. A multi-stimulus comparison procedure was used to compare the nine conditions for each audio scene and at each listening position to the reference. In order to avoid a larger number of conditions, there was no hidden reference. Also, an explicit lower anchor was omitted, as the 1st-order conditions with 16 kb/s channel bandwidth were expected to perform at best poorly. Listeners could seamlessly switch between them during looped playback to rate the perceived quality. The rating was based on the MUSHRA recommendations [42] with a scale from 0 to 100 and additional attributes from *bad* (0 to 20), over *poor* (20 to 40), *fair* (40 to 60), *good* (60 to 80) to *excellent* (80 to 100), see also Figure 2.
On average, listeners needed 40 minutes (min. 22, max. 51) to complete the 12 = 3 (listening positions) × 4 (audio scenes) trials. The order of the listening positions was chosen randomly, and within each listening position, the order of the audio scenes was also randomized. Before the actual experiment, the listeners were instructed on how to perform the experiment by an oral explanation of the task, as well as a description in the graphical user interface. Moreover, every listener received an individual demonstration of the expected compression and spatial artifacts by listening to the dry music scene in the reference condition, as well as with 16 kb/s and 64 kb/s channel bandwidth in 1st- and 5th-order Ambisonics. During the demonstration, they were encouraged to walk between the three listening positions and focus on

localization and spatial extent of the direct sound, as well as distortion and noise artifacts. While generally facing the frontal center loudspeaker C, listeners were allowed to turn their head during the experiment.

In total, 20 listeners (4 female, 16 male) participated in the experiment with a median age of 25 years (min. 21, max. 43). All of them were students or staff at the Institute of Electronic Music and Acoustics, with a background in audio engineering and about half of them were experienced with listening experiments on spatial audio.

III. Results

Figure 2 shows the quality ratings for different audio scenes and at different listening positions as median values and confidence intervals for each combination of Ambisonics order and bandwidth per channel. The following statistical analysis of each audio scene is based on pairwise Wilcoxon signed-rank tests with Bonferroni-Holm correction. As all conditions with 16 kb/s achieved at best fair quality (and poor for the music scenes), the order-dependent analysis of these conditions is omitted. The same is true for the bandwidth-dependent analysis of the 1st-order conditions.

A. Speech dry

At the central listening position P1 and for the dry speech scene, the significantly best ($p \geq 0.013$) quality was achieved for the condition 5/64 (5th-order Ambisonics and 64 kb/s per channel). Within the 64 kb/s-conditions, the quality significantly increased with the Ambisonics order ($p \leq 0.003$), while at 32 kb/s, the difference between 1st and 3rd/5th order is significant ($p \leq 0.003$) and there was no difference between 3rd and 5th order ($p = 0.38$). Quality significantly increased with bandwidth for 3rd order ($p \leq 0.045$) and 5th order ($p \leq 0.013$). Excellent quality (median value ≥ 80) was achieved by the order/bandwidth combinations 5/64, 3/64, and 5/32. In addition, good quality (median values ≥ 60) was achieved by 3/32.

At position P2 (halfway to the side), 5/64 was rated significantly best ($p \geq 0.013$). At 64 kb/s, the quality significantly increased with the order ($p \leq 0.013$). Similarly to P1, the difference between orders 3 and 5 was not significant ($p = 0.78$), but both were significantly better than 1st order ($p \leq 0.004$). At 3rd order, there was no significant difference between 64 and 32 kb/s ($p = 0.36$), but to 16 kb/s ($p \leq 0.004$). For 5th order, the quality increased with the channel bandwidth ($p \leq 0.02$). Excellent quality was achieved for 5/64 and 3/64, while good quality was achieved by 5/32 and 3/32. At the outmost listening position P3, 5/64 was again significantly best ($p \leq 0.017$). Within the conditions with 64 kb/s, the quality significantly increased with the order ($p \leq 0.017$). However, there were no significant differences between the conditions with 32 kb/s ($p \geq 0.10$). For both 3rd and 5th order, the quality significantly increased with the channel bandwidth ($p \leq 0.02$ and $p \leq 0.003$ respectively). Excellent quality was only achieved for 5/64, while good quality was achieved also for 3/64.

B. Speech reverberant

At P1 for the reverberant speech scene, 5/64 was the best condition ($p \leq 0.02$), however it was not different from 5/32 ($p = 0.134$). Within the conditions with 64 kb/s and 32 kb/s, the quality significantly increased with the order ($p \leq 0.02$ and $p \leq 0.006$, respectively). Within the 5th-order conditions, the condition with 16 kb/s was rated worst ($p \leq 0.04$), however there was no difference between 32 and 64 kb/s ($p = 0.134$). Similarly, there was no difference between 32 and 64 kb/s at 3rd order ($p = 0.09$), but to 16 bk/s ($p \leq 0.013$). Excellent quality was achieved for 5/64, 3/64, and 5/32, while good quality was achieved for 3/32.

At P2, 5/64 was rated best ($p \leq 0.01$), but not better than 3/64 ($p = 0.052$). Within the conditions with 64 kb/s and 32 kb/s, the quality generally increased with the Ambisonics order ($p \leq 0.01$ and $p \leq 0.02$), however there was no difference between 3rd and 5th order ($p = 0.052$ and $p \approx 1$). Excellent quality was achieved for 5/64, 3/64, and 5/32, and good quality was achieved for 3/32.

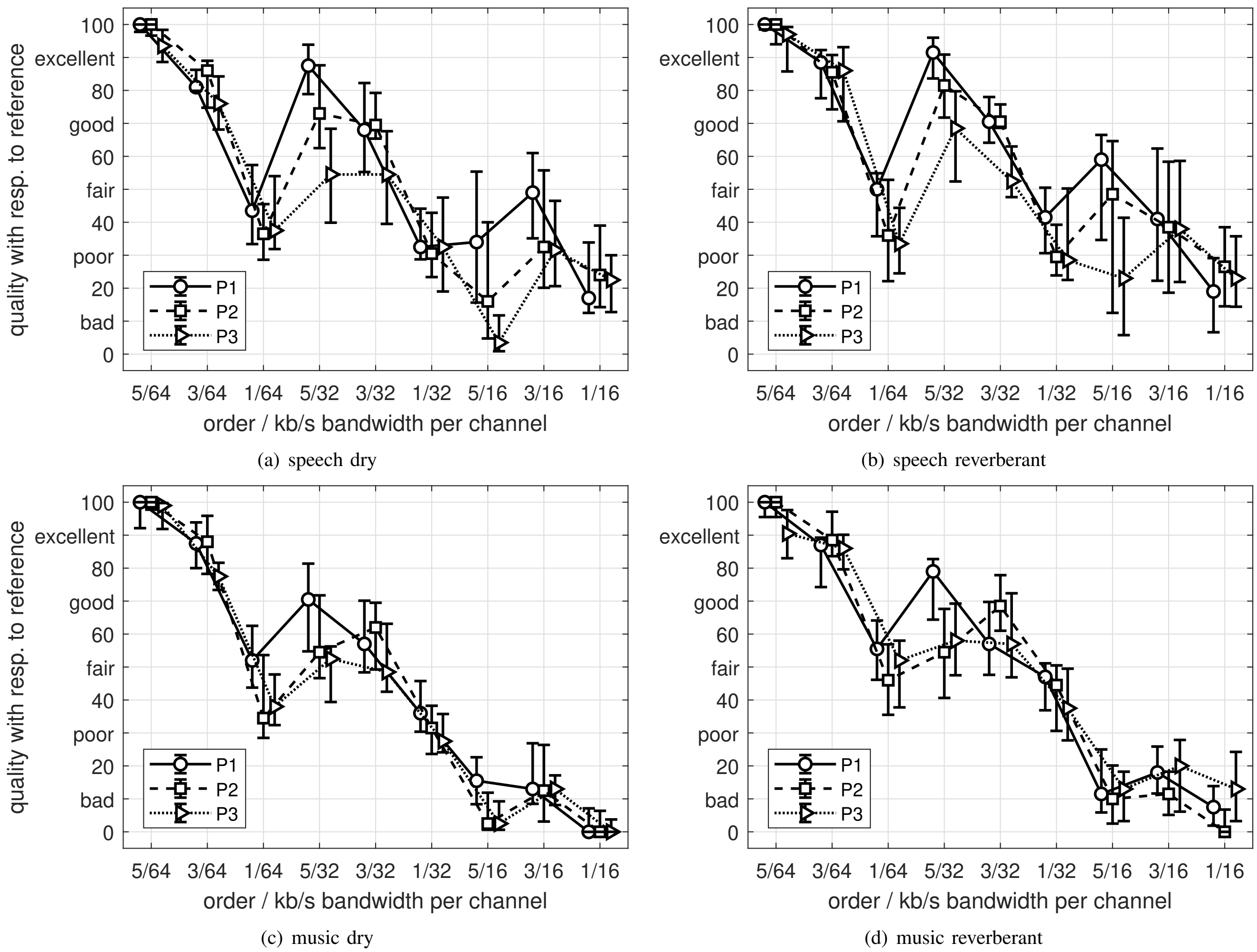
At P3, 5/64 was again rated best ($p \leq 0.03$), but not better than 3/64 ($p = 0.25$). Nevertheless, the decrease of the spatial resolution to 1/64 leads to significantly worse results ($p \leq 0.009$). Within the conditions with 32 kb/s, the quality increased with the order ($p \leq 0.01$). Similarly, within the 5th-order conditions, the quality increased with the channel bandwidth ($p \leq 0.03$). However, for 3rd order, 64 kb/s was the best ($p \leq 0.03$), while 3/32 and 3/16 were not different ($p = 0.27$). Excellent quality was achieved for 5/64 and 3/64. Good quality was achieved for 5/32.

C. Music dry

For the dry music scene at P1, 5/64 was the best condition ($p \leq 0.003$), however not better than 3/64 ($p = 0.11$). At 32 kb/s channel bandwidth, there was also no difference between 3rd and 5th order ($p = 0.25$), but to 1st order ($p \leq 0.03$). For both 3rd and 5th order, there was a significant increase in quality with channel bandwidth ($p \leq 0.03$). Excellent quality was achieved by 5/64 and 3/64, while 5/32 achieved good quality.

At P2, 5/64 was the best condition ($p \leq 0.024$). There was a significant increase with the order for the conditions with 64 kb/s ($p \leq 0.024$). With 32 kb/s, 5th and 3rd order were similar ($p = 0.87$), however the differences to 1st order were significant ($p \leq 0.05$). Within the conditions with 3rd and 5th order, channel bandwidth significantly increased quality ($p \leq 0.03$ and $p \leq 0.02$). Excellent quality was achieved for both 5/64 and 3/64, and good quality for 3/32.

At P3, 5/64 was again the best condition ($p \leq 0.03$). There was a significant increase of the quality within the conditions with 64 kb/s ($p \leq 0.03$). For 32 kb/s, there was no difference between 3rd and 5th order ($p = 0.94$), while both are different from 1st order ($p \leq 0.04$). Within the 3rd- and 5th-order conditions, the quality increased with the channel bandwidth ($p \leq 0.006$ and $p \leq 0.003$). Excellent quality was only achieved for 5/64 and good quality for 3/64.



(a) speech dry

(b) speech reverberant

(c) music dry

(d) music reverberant

Fig. 2. Median vales and 95% confidence intervals of quality ratings for different audio scenes and at different listening positions.

D. Music reverberant

For the reverberant music scene at P1, 5/64 was the best condition ($p \leq 0.025$). The quality increased with the order for the conditions with 64 kb/s ($p \leq 0.025$). However with 32 kb/s, only the difference between 1st and 5th order was significant ($p = 0.02$), 1/32 vs. 3/32 and 3/32 vs. 5/32 were not different ($p \geq 0.26$). Within the 5th-order conditions, the quality increased with the channel bandwidth ($p \leq 0.005$). Within the conditions with 3rd order, 16 kb/s was worst ($p \leq 0.003$), but 32 and 64 kb/s were similar ($p = 0.08$). Excellent quality was achieved for 5/64 and 3/64, while good quality was achieved for 5/32.

At P2, 5/64 was again the best condition ($p \leq 0.003$), however not better than 3/64 ($p \approx 1$). Similarly, at 32 kb/s, only 1st and 5th order were rated differently ($p = 0.013$), but not 1/32 vs. 3/32 and 3/32 vs. 5/32 ($p = 0.18$). Again, the increase of the channel bandwidth yielded an increase in quality for 3rd order ($p \leq 0.001$) and 5th order ($p \leq 0.003$). Excellent quality was achieved for 5/64 and 3/64, with good quality ratings for 3/32.

Finally, at P3, 5/64 was again the best condition ($p \leq 0.003$), with the exception of 3/64 ($p \approx 1$). 3rd and 5th order were also similar at 32 kb/s ($p \approx 1$), while 1st and 5th order were different ($p = 0.04$), and 1st and 3rd were not ($p = 0.076$). The increase of the channel bandwidth yielded an increase in quality for 3rd order ($p \leq 0.001$) and 5th order ($p \leq 0.003$). Excellent quality was achieved by 5/64 and 3/64, while no other condition archived at least good quality.

IV. Discussion

The conditions with 16 kb/s per channel were rated at best fair for the speech scenes and poor or even bad for music. Similarly, all 1st-order conditions achieved at best fair ratings. Thus, orders 3 and 5, as well as bandwidths of 32 to 64 kb/s were necessary to achieve good to excellent quality ratings, similar as in [29], [31], [32]. The combination of 5th order and 64 kb/s was obviously rated to be the best condition, however, as there was no hidden reference rated in the experiment, we could not draw statistically meaningful conclusions whether this combination would be significantly different from the reference or not. For speech, 32 kb/s was sometimes sufficient for excellent quality. The higher bandwidth requirements for music agree with the results in [30]. Moreover, 3rd order was sufficient for the reverberant audio scenes at the central listening position. This agrees with the findings in [43], where the increase from 3rd to 5th order was only beneficial at off-center listening positions.

Tables I and II summarize the conditions that achieved excellent quality (median values $\geq$ 80) for the speech and music scenes. The quality ratings are presented in dependence of the size of the listening area, e.g. P1 for only the central listening position and P1+P2+P3 for all positions. In comparison to the bitrate of around 4 Mb/s for a lossless nearly delay-free codec at 3rd order [44], our excellent-rated conditions require only a half or a quarter. However, for applications with stricter latency requirements, slightly higher bitrates might be necessary to maintain quality: Decreasing the frame size from 20 ms to 5 ms required about 25% greater bitrates for the same quality in [10].

TABLE I
CONDITIONS WITH EXCELLENT QUALITY FOR BOTH SPEECH SCENES IN DEPENDENCE OF THE LISTENING AREA.

Positions	Conditions	min. total bitrate
P1	5/64, 3/64, 5/32	1024 kb/s
P1+P2	5/64, 3/64	1024 kb/s
P1+P2+P3	5/64	2304 kb/s

TABLE II
CONDITIONS WITH EXCELLENT QUALITY FOR BOTH MUSIC SCENES IN DEPENDENCE OF THE LISTENING AREA.

Positions	Conditions	min. total bitrate
P1	5/64, 3/64	1024 kb/s
P1+P2	5/64, 3/64	1024 kb/s
P1+P2+P3	5/64	2304 kb/s

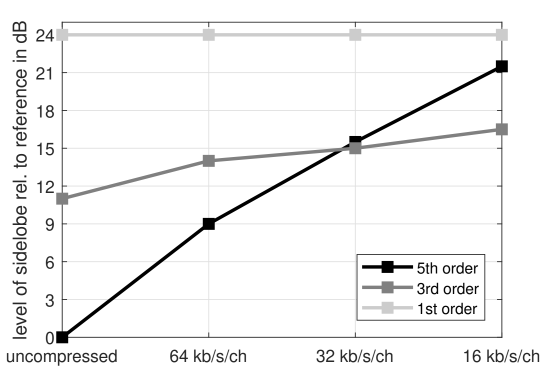


Fig. 3. Level of direct sound in lateral loudspeaker in dependence of order and channel bandwidth relative to 5th-order uncompressed reference.

In the dry audio scenes, listeners reported an increased level of the sidelobe from the lateral loudspeaker close to listening position P3 even for 5th order. Due to the panning direction of the direct sound to the front and the max-r_{E} weighting, the sidelobe in the uncompressed 5th-order reference is only audible when pressing one's ear directly onto the loudspeaker grille. Figure 3 shows that for 5th order, the level of the sidelobe strongly increases for lower channel bandwidths. Interestingly, although starting with much higher levels already for the uncompressed case, the level increases less the lower the order. At 32 kb/s, the benefit of the stronger sidelobe attenuation of 5th order gets lost in comparison to 3rd order. The stronger sensitivity at higher orders can be explained by adding up of codec artefacts, such as non-linear distortions, due to the higher number of channels that are added or subtracted in the Ambisonics decoder. The surrounding reverberation in the reverberant audio scenes could partly mask the increased sidelobes of the direct sound.

Moreover, listeners reported that rotating their head by around 90° sometimes helped them in detecting spatial differences to the reference in the direct sound. This finding agrees with the results in [45], where differences due to order truncation were found to be most prominent at lateral directions.

V. Conclusion

This paper presented a listening experiment to evaluate the quality of OPUS-compressed 1st-, 3rd-, and 5th-order Ambisonics at different listening positions in a studio environment with a hemispherical loudspeaker arrangement of approximately 5 m radius. The audio signals were compressed with 16, 32, and 64 kb/s per channel and the experiment employed four audio scenes with uncompressed 5th-order playback as a reference: speech and music panned to the frontal center loudspeaker both with and without surrounding reverberation.

Excellent quality, which is a prerequisite for professional applications, was achieved by using either 5th- or 3rd-order Ambisonics with 64 kb/s per channel at the two more central listening positions. The same per-channel bandwidth is used in current on-demand and streaming solutions that support Ambisonics up to an order of 4 [46], [47]. While the outmost position required 5th-order Ambisonics, speech got along with only 32 kb/s at the central listening position. The compression of the Ambisonics signals increased the lateral sidelobes of the frontal direct sound, especially at higher orders. The surrounding reverberation could partly mask this artifact.

Future research might look into other channel mapping families of OPUS or more sophisticated compression algorithms and their quality for complex, reverberant audio scenes at off-center listening positions. As the presented experiment evaluated the streaming application of Ambisonics with a frame size of 20 ms, future research could investigate how much the required per-channel bandwidths would increase for smaller frame sizes that are necessary for interactive Network Music Performances. On the other hand, the required number of transmission channels might be lower, as such applications typically use headphone rendering, which is similar to the central listening position evaluated here, where 3rd-order Ambisonics was already sufficient for excellent quality.

Acknowledgment

This work was partly supported by the project Spatial Audio Network Europe (SANE), which is funded by the EU under the CREA2027 programme.

References

[1] C. Rottondi, C. Chafe, C. Allocchio, and A. Sarti, "An Overview on Networked Music Performance Technologies," *IEEE Access*, vol. 4, pp. 8823–8843, 2016.

[2] L. Turchet, M. Lagrange, C. Rottondi, G. Fazekas, N. Peters, J. Østergaard, F. Font, T. Bäckström, and C. Fischione, "The Internet of Sounds: Convergent Trends, Insights, and Future Directions," *IEEE Internet of Things Journal*, vol. 10, no. 13, pp. 11 264–11 292, 2023.

[3] M. Tomasetti and L. Turchet, "Playing With Others Using Headphones: Musicians Prefer Binaural Audio With Head Tracking Over Stereo," *IEEE Transactions on Human-Machine Systems*, vol. 53, no. 3, pp. 501–511, 2023.

[4] P. Cairns, A. Hunt, D. Johnston, J. Cooper, B. Lee, H. Daffern, and G. Kearney, "Evaluation of metaverse music performance with bbc maida vale recording studios," *Journal of the Audio Engineering Society*, vol. 71, pp. 313–325, June 2023.

[5] P. Cairns, T. Rudzki, J. Cooper, A. Hunt, K. Steele, G. A. Martínez, A. Chadwick, H. Daffern, and G. Kearney, "Singer and Audience Evaluations of a Networked Immersive Audio Concert," *Journal of the Audio Engineering Society*, vol. 72, pp. 467–478, August 2024.

[6] C. Rinaldi, F. Franchi, A. Marotta, F. Graziosi, and C. Centofanti, "On the Exploitation of 5G Multi-Access Edge Computing for Spatial Audio in Cultural Heritage Applications," *IEEE Access*, vol. 9, pp. 155 197–155 206, 2021.

[7] L. Turchet, C. Rinaldi, C. Centofanti, L. Vignati, and C. Rottondi, "5G-Enabled Internet of Musical Things Architectures for Remote Immersive Musical Practices," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 4691–4709, 2024.

[8] S. Farner, A. Solvang, A. Sæbo, and P. Svensson, "Ensemble Handclapping Experiments under the Influence of Delay and Various Acoustic Environments," *Journal of the Audio Engineering Society*, vol. 57, pp. 1028–1041, December 2009.

[9] P. F. Driessen, T. E. Darcie, and B. Pillay, "The Effects of Network Delay on Tempo in Musical Performance," *Computer Music Journal*, vol. 35, no. 1, pp. 76–89, 2011.

[10] 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*, October 2013.

[11] D. H. Cooper and T. Shiga, "Discrete-matrix multichannel stereo," *Journal of the Audio Engineering Society*, vol. 20, no. 5, pp. 346–360, 1972.

[12] M. A. Gerzon, "With-height sound reproduction," *Journal of the Audio Engineering Society*, vol. 21, pp. 2–10, 1973.

[13] F. Zotter and M. Frank, *Ambisonics: A practical 3D audio theory for recording, studio production, sound reinforcement, and virtual reality*. Springer Nature, 2019.

[14] 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, September 2001.

[15] J. Daniel, "Représentation de champs acoustiques, application à la transmission et à la reproduction de scénes sonores complexes dans un contexte multimédia," Ph.D. dissertation, Université Paris 6, 2001.

[16] G. Kearney, E. Bates, F. Boland, and D. Furlong, "A comparative study of the performance of spatialization techniques for a distributed audience in a concert hall environment," in *Audio Engineering Society Conference: 31st International Conference: New Directions in High Resolution Audio*, London, 2007.

[17] M. Frank and F. Zotter, "Localization experiments using different 2D Ambisonics decoders," in *25th Tonmeistertagung*, Leipzig, 2008.

[18] P. Stitt, S. Bertet, and M. van Walstijn, "Off-centre localisation performance of ambisonics and HOA for large and small loudspeaker array radii," *Acta Acustica united with Acustica*, vol. 100, no. 5, 2014.

[19] M. Frank and F. Zotter, "Exploring the Perceptual Sweet Area in Ambisonics," in *Audio Engineering Society Convention 142*. Audio Engineering Society, 2017.

[20] 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.

[21] M. Zaunschirm, C. Schörkhuber, and R. Höldrich, "Binaural rendering of ambisonic signals by head-related impulse response time alignment and a diffuseness constraint," *Journal of the Acoustical Society of America*, vol. 143, no. 6, pp. 3616–3627, 2018.

[22] C. Schörkhuber, M. Zaunschirm, , and R. Höldrich, "Binaural rendering of ambisonic signals via magnitude least squares," in *Fortschritte der Akustik, DAGA*, vol. 44, 2018, pp. 339–342.

[23] M. Frank and A. Sontacchi, "Case study on ambisonics for multi-venue and multi-target concerts and broadcasts," *Journal of the Audio Engineering Society*, vol. 65, no. 9, pp. 749–756, 2017.

[24] ITU, "BS.2051-3 Advanced sound system for programme production," 2022.

[25] M. Narbutt, S. O'Leary, A. Allen, J. Skoglund, and A. Hines, "Streaming VR for immersion: Quality aspects of compressed spatial audio," in *2017 23rd International Conference on Virtual System & Multimedia (VSMM)*. IEEE, 2017, pp. 1–6.

[26] M. Narbutt, A. Allen, J. Skoglund, M. Chinen, and A. Hines, "AMBIQUAL-a full reference objective quality metric for ambisonic spatial audio," in *2018 tenth international conference on quality of multimedia experience (QoMEX)*. IEEE, 2018, pp. 1–6.

[27] M. Narbutt, J. Skoglund, A. Allen, M. Chinen, D. Barry, and A. Hines, "Ambiqual: Towards a quality metric for headphone rendered compressed ambisonic spatial audio," *Applied Sciences*, vol. 10, no. 9, p. 3188, 2020.

[28] T. Rudzki, I. Gomez-Lanzaco, J. Stubbs, J. Skoglund, D. T. Murphy, and G. Kearney, "Auditory localization in low-bitrate compressed ambisonic scenes," *Applied Sciences*, vol. 9, no. 13, p. 2618, 2019.

[29] T. Rudzki, I. Gomez-Lanzaco, P. Hening, J. Skoglund, T. McKenzie, J. Stubbs, D. Murphy, and G. Kearney, "Perceptual evaluation of bitrate compressed ambisonic scenes in loudspeaker based reproduction," in *Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio*. Audio Engineering Society, 2019.

[30] T. Rudzki, "Improvements in the Perceived Quality of Streaming and Binaural Rendering of Ambisonics," Ph.D. dissertation, University of York, 2023.

[31] A. Devantier, C. Tejeda-Ocampo, C. Zhongran Wang, W. Saba, and S. Bharitkar, "Bit Rate Requirements for an Audio Codec for Stereo, Surround and Immersive Formats," in *Audio Engineering Society Convention 151*, 2021.

[32] E. Souza-Blanes, C. Tejeda-Ocampo, C. Wang, and S. Bharitkar, "Bitrate Requirements for Opus with First, Second and Third Order Ambisonics reproduced in 5.1 and 7.1.4," in *Audio Engineering Society Convention 152*, 2022.

[33] J. Xu, Y. Niu, X. Wu, and T. Qu, "Higher order ambisonics compression method based on independent component analysis," in *Audio Engineering Society Convention 150*. Audio Engineering Society, 2021.

[34] E. Hellerud, A. Solvang, and U. P. Svensson, "Spatial redundancy in Higher Order Ambisonics and its use for lowdelay lossless compression," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 269–272.

[35] T. Qu, J. Xu, Z. Yuan *et al.*, "Higher order ambisonics compression method based on autoencoder," in *Audio Engineering Society Convention 153*. Audio Engineering Society, 2022.

[36] A. Wabnitz, N. Epain, A. McEwan, and C. Jin, "Upscaling ambisonic sound scenes using compressed sensing techniques," in *2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE, 2011, pp. 1–4.

[37] C. Hold, V. Pulkki, A. Politis, and L. McCormack, "Compression of higher-order ambisonic signals using directional audio coding," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 651–665, 2023.

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

[39] EBU. (2008) EBU SQAM CD: Sound Quality Assessment Material recordings for subjective tests. [Online]. Available: https://tech.ebu.ch/publications/sqamcd

[40] D. Leckschat, C. Epe, N. Dahlheimer, M. Bier, L. Prinz, and H. Schulte, "Tonproduktion zweier Titel des Funk-Ensembles "Rhythmussportgruppe"," Feb. 2020. [Online]. Available: https://doi.org/10.5281/zenodo.3601032

[41] M. Frank, "Phantom sources using multiple loudspeakers in the horizontal plane," Ph.D. dissertation, University of Music and Performing Arts Graz, Austria, 2013.

[42] ITU, *ITU-R BS.1534-1: Method for the subjective assessment of intermediate sound quality (mushra)*, International Telecommunication Union Std., 2006.

[43] M. Frank and F. Zotter, "Spatial impression and directional resolution in the reproduction of reverberation," in *Fortschritte der Akustik, DAGA*, Aachen, 2016.

[44] M. Namazi and K. Rose, "Ultra-Low Delay Lossless Compression of Higher Order Ambisonics," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 791–795.

[45] S. Bertet, J. Daniel, E. Parizet, and O. Warusfel, "Investigation on localisation accuracy for first and higher order Ambisonics reproduced sound sources," *Acta Acustica united with Acustica*, vol. 99, no. 4, pp. 642–657, 2013.

[46] T. Deppisch, N. Meyer-Kahlen, B. Hofer, T. Latka, and T. Zernicki, "HOAST: A higher-order ambisonics streaming platform," in *Audio Engineering Society Convention 148*, 2020.

[47] B. Mróz, P. Odya, P. Danowski, and M. Kabacinski, "A commonly-accessible toolchain for live streaming music events with higher-order ambisonic audio and 4k 360 vision," in *AES International Conference on Spatial and Immersive Audio*, August 2023.