Surround Sound Processed by Opus Codec: A Perceptual Quality Assessment

Franziska Trojahn, Martin Meszaros, Michael Maruschke and Oliver Jokisch

Hochschule für Telekommunikation Leipzig, Germany
trojahn@hft-leipzig.de

Abstract: The article describes the first perceptual quality study of 5.1 surround sound that has been processed by the Opus codec standardised by the Internet Engineering Task Force (IETF). All listening sessions with up to five subjects took place in a slightly sound absorbing laboratory – simulating living room conditions. For the assessment we conducted a Degradation Category Rating (DCR) listening-opinion test according to ITU-T P.800 recommendation with stimuli for six channels at total bitrates between 96 kbit/s and 192 kbit/s as well as hidden references. A group of 27 naive listeners compared a total of 20 sound samples. The differences between uncompressed and degraded sound samples were rated on a five-point degradation category scale resulting in Degradation Mean Opinion Score (DMOS). The overall results show that the average quality correlates with the bitrates. The quality diverges for the individual test stimuli depending on the music characteristics. Under most circumstances, a bitrate of 128 kbit/s is sufficient to achieve acceptable quality.

1 Introduction

Nowadays, a high number of different speech and audio codecs are implemented in several kinds of multimedia applications; including audio / video entertainment, broadcasting and gaming. In recent years the demand for low delay and high quality audio applications, such as remote real-time jamming and cloud gaming, has been increasing. Therefore, current research objectives do not only include close to natural speech or audio quality, but also the requirements of low bitrates and a minimum latency. Furthermore, several streaming and cloud-based applications use 5.1 surround sound or would like to offer multichannel sound in the near future [1, 2].

An audio codec that offers high quality, low delay and the capability of coding multiple channels is the IETF standardised Opus Codec [3]. The Allround-Codec supports several audio bandwidths from Narrowband (3.4 kHz) to Fullband (20 kHz), making it suitable for all audible audio signals including speech and music in high quality. Additionally, it has the capability to switch seemlessly between coding modes. Various studies have surveyed the quality of music coding with the Opus Codec [4, 5].

With the standardisation of Web Real-Time Communication (WebRTC), Opus was implemented by many browser manufacturers into their web browsers. That means the Opus codec is included by default in web browsers such as Google Chrome, Mozilla Firefox, Opera and Microsoft Edge [6]. In our study “Review of the Opus Codec in a WebRTC Scenario for Audio and Speech Communication” [7], we have proven that the Opus codec conforms to its standardised definition in a real world scenario of WebRTC. Among other applications, this makes it possible for Video on Demand (VoD) services, like Netflix or Amazon Prime Video, to offer surround sound using the web browser embedded Opus audio codec. Consequently, the installation of additional plug-ins or software would not be necessary. Opus' source code is freely available and published under a liberal license and thus can be used without paying royalty fees.

Our contribution deals with the perceptual quality assessment of 5.1 surround sound processed by the Opus Codec. The main research question is how the degradation of the surround sound is perceived at different bitrate conditions. For the assessment we performed a DCR listening test as described in the ITU-T P.800 recommendation. The test stimuli were presented in pairs of uncompressed and degraded surround sound audio samples. The degradation was rated by study participants using DMOS on a five point degradation category scale.

At the beginning, we present the significance of the Opus Codec for trending applications. In section 3, we introduce our testing concept and setup. In section 4 we discuss the perceptual assessment results of the Opus codec processing surround sound in detail. Finally, we conclude with an overall summary of the results.

2 Previous studies

Standardised in RFC 6716 by the IETF, Opus was designed as an all-purpose interactive speech and audio codec [3]. Thus, Opus is suitable for multiple use cases like Voice over IP (VoIP), video conferencing, online gaming and audio on demand. Depending on the application, Opus can be utilised for low bitrate speech coding as well as high quality multichannel music coding. In order to realise both, high quality and dynamic characteristics, Opus combines the linear prediction-based SILK codec and the Modified Discrete Cosine Transform (MDCT)-based Constrained Energy Lapped Transform (CELT) codec. For flexible usage, the Opus codec supports the frequency band types Narrowband (NB), Wideband (WB), Super-Wideband (SWB) and Fullband (FB). Consequently, Opus encodes speech and music within a bitrate range from 6 kbit/s to 510 kbit/s. Furthermore, Opus has codec delay times between 5 and 66.5 ms[1] for all relevant sample rates - from 8 kHz up to 48 kHz.

Additionally, Opus is not limited to mono or stereo scenarios: It supports a maximum of 255 audio channels. Opus can not only be used for storing and streaming multichannel music or providing surround sound in movies, but also for interactive multichannel audio scenarios, like remote real-time jamming and cloud gaming. The low delay characteristics of the Opus codec make it suitable for these scenarios despite their challenging low processing time requirements. Due to the interactive type of the named applications, high latency would result in a bad user experience.

With cloud gaming, the player's device, e.g. computer, console or mobile device, is only used for input interaction plus displaying the game. The processing power is moved into remote server farms and a high quality video is streamed to the user. Respectively, the network parameters have a high impact on the Quality of Experience (QoE). A perceptual evaluation of the QoE when using cloud gaming under different network conditions was conducted in JARSCHEL et al. [8]. The results show that a Round Trip Time (RTT) of 160 ms[2] leads to a noticeable degradation of the QoE.

Traditional codecs, like Advanced Audio Coding (AAC) or MP3, were not designed for realtime coding and come with a high algorithmic delay. In several studies the amount of delay introduced by those codecs is investigated [e.g. 9]. In practice, with a network bandwidth much higher than the bitrate of the codec and a processor workload during the encoding process of 30 %, a delay of 63 ms is introduced by AAC and 107 ms by MP3. Given a one way delay (remote server farm to client) of around 80 ms, with AAC most of the delay budget would be spent only by audio encoding, leaving 17 ms for transmission over the network and processing

[1] Opus' algorithmic delay highly depends on the selected framesize whereas a value of 20 ms is default.

[2] Time, the user input is transferred over the network to the data centre which transmits the result back to user.

the data on the client side, whilst with MP3 the delay budget would be exceeded.

Consequently, well-known codecs like AAC or MP3 are not suitable for real-time scenarios although multichannel audio is supported by them. Additional to the impact of latency, the QoE is influenced by the overall listening impression. Several listening tests have assessed and compared the quality of different speech and audio codecs with that of the Opus codec. A summary can be found in HOENE et al. [10] and an assessment including the Enhanced Voice Services (EVS) codec, which was standardised in 2014, in RÄMÖ and TOUKOMAA [11]. The performance of the Opus codec surpassed the performance of all other audio codecs except for EVS in these tests, especially in the wider bands if applicable. We conducted a study comparing the music coding ability of Opus and EVS in which both had a similiar performance [5]. Nevertheless, all mentioned tests included only mono or stereo mode. Furthermore, EVS supports stereo only as coding two mono channels [12].

3 Test design

3.1 Testing concept

The Opus codec has been assessed in different studies as described in section 2, but only in mono or stereo mode. We experimented with 5.1 surround sound samples in order to survey the quality of multichannel coding with Opus at different bitrates. The Opus codec requires a minimal bitrate of 96 kbit/s for six channels. In a preliminary test, we observed that five experienced listeners could not detect quality differences between the bitrates 192 kbit/s and 256 kbit/s for four out of five audio samples. For this reason, and in order not to overstrain the participants, we chose to test the codec with the bitrates 96, 128 and 192 kbit/s in comparison to the uncompressed original music files. The sound material from *2L High Resolution Music* [13] includes four accoustic samples with the bit depth of 24 bit and a sample rate of 96 kHz. The Free Lossless Audio Codec (FLAC) files are provided free of charge for evaluation. The fifth sample was featured on *Dolby Atmos Demonstration Disc* [14] as Dolby True HD (lossless) at 24 bit and 48 kHz for demonstration purposes. Details can be found in Table 1.

For the listening test type we selected the DCR method which is described in ITU-T [15]. Listeners rate the degradation of an audio sample compared to a quality reference on a five point scale. The results are displayed as degradation mean opinion score. We chose the DCR method, because of the following characteristics:

- It is intended to detect small impairments;
- minimizes influence of personal taste in music and loud speaker characteristics and
- resembles the rating of a first impression better than e.g. the MUSHRA test [16], that allows repeated listening of each sample and reference.

ID	Title	Artist	Composer	Source
Unfold	Dolby Atmos Unfold	unknown	unknown	synthetic
Soprano	Recitative from Cantata RV 679	T. Wik & Barokkanerne	A. Vivaldi	period instr.
Piano	Living	J. G. Hoff	J. G. Hoff	piano
String quartet	String Quart. in D, Op. 76, No. 5 Finale, Preston	Engegård Quartet	J. Haydn	string instr.
String ensemble	Frank Bridge Variations - Romance	Trondheim-Solistene	B. Britten	string instr.

Table 1 – Tested surround music samples in overview

3.2 Experimental setup

The DCR listening-opinion test was conducted in a lecture room that fullfilled the requirements according to ITU-T [15]. The Reverberation Time (RT) 60 was approximately 400 ms. The loud speakers were aligned as described in ITU-R [17] and as depicted in Figure 1. A maximum of five subjects participated in the listening test at the same time. The test equipment consisted of a Canton Movie 95 5.1 satellite speaker system and an active subwoofer which were driven by a Pioneer VSX-422 Audio/Video (AV)-receiver. The speakers were calibrated with the AV-receivers' built-in Multi-Channel Acoustic Calibration System. The audio samples were played back with a laptop connected to the receiver via a High Definition Multimedia Interface (HDMI) cable. No further audio processing was carried out by the receiver except for the room calibration adjustments.

The following testing conditions were used:

- five different audio samples (four acoustic pieces including one with singing voice, one computer generated sample);
- 20 audio samples overall (Opus at three different bitrates plus FB reference);
- encoding and decoding with libopus 1.1.3 [18];
- 27 naive listeners, students ranging from 18 to 30;
- listening assessment on five point DMOS scale (see Table 2);
- four practice samples before the assessment and
- hidden reference.

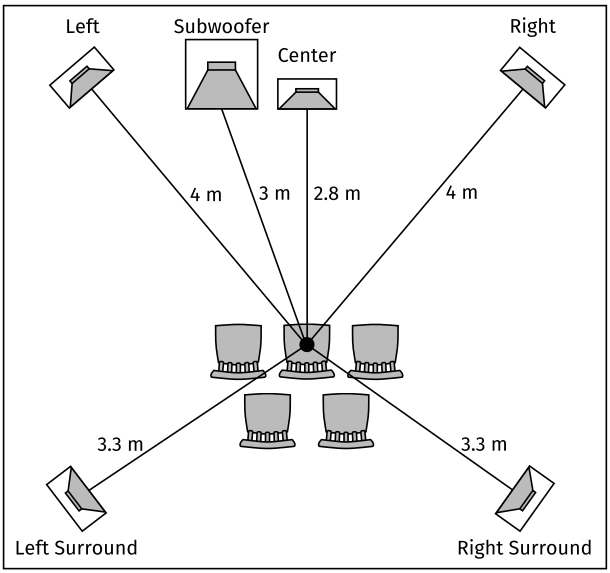

Figure 1 – Room layout and test setup

DMOS Rating	Description
5	Degradation is inaudible.
4	Degradation is audible but not annoying.
3	Degradation is slightly annoying
2	Degradation is annoying.
1	Degradation is very annoying.

Table 2 – DMOS Scale

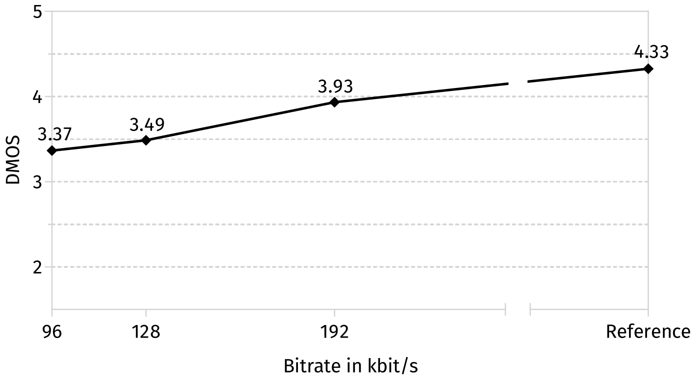

Figure 2 – Perceived overall quality (DMOS) depending on Opus coding bitrate

4 Results and discussion

4.1 Influence of the coding bitrate

Figure 2 shows the results averaged over all music samples for the bitrates 96, 128 and 192 kbit/s. As reference, the mean rating of all uncompressed audio files is shown. The rating of the reference with a DMOS of 4.33 illustrates the challenging test conditions and the critical absolute ratings of our subjects. A significant drop of 0.4 from the reference to the highest, compressed bitrate (192 kbit/s) can be surveyed. As expected, the saturation is not yet reached at 192 kbit/s. The graph proceeds to decrease almost linearly and the DMOS reaches its lowest value 3.37 at 96 kbit/s. This rating is still significantly above a score of 3.0 (degradation is slightly annoying) which we have defined as a minimal criterion for acceptable quality, considering the relatively low score of the reference and that the DCR method is intended to detect minor differences. On average, the degradation ranges from slightly annoying to not annoying but audible. Hence, the audio quality is at least acceptable at all tested bitrates.

4.2 Effect of the music characteristics

Figure 3 shows the subjective quality for each audio sample at all bitrates under test. In order to challenge the codec, we chose a variety of music types (Table 1). The expected rise, smiliar to the overall quality, can be observed in three of the five music pieces, namely *Soprano*, *Piano* and *String quartet*. They grow constantly from the lowest to the highest bitrate but the amount of increase differs. The *Piano* received very good ratings over all bitrates whereas *Soprano* and *String quartet* were very challenging for the codec. The results for the lowest bitrate range from a DMOS of 1.74 (*String quartet*) to 4.07 (*Piano*).

The sample *String ensemble* follows the same trend as the overall assessment, except for the rating at 96 kbit/s, which is remarkably high in comparison with the other samples and bitrates. According to ITU-T Rec. P.800, an effect dependent on the order of the presentation should not be observed in a DCR listening-opinion test. Nevertheless, we assume that such an effect might be the reason for the inconsistency of the assessment, as the P.800 recommendation has not been tested with multichannel music. The sample *String ensemble* with the bitrate 96 kbit/s was presented to the subjects directly after the training samples.

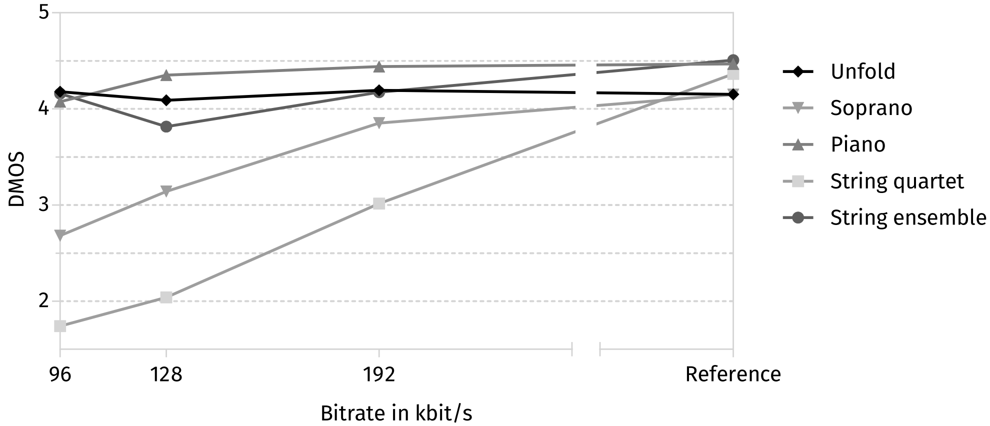

Figure 3 – Perceived quality (DMOS) depending on music characteristics and Opus coding bitrate

For the sample *Unfold*, the scores do not differ significantly for all bitrates (difference below 0.1 DMOS). The subjects were not able to detect the lowest bitrate or the reference. There are two possible explanations for this: On the one hand, the listeners expressed that they found it difficult to concentrate on specific differences between reference and compressed file. Due to the high distinctions between the six channels and the impression of movement generated by the surround sound, *Unfold* seems to be complex for human test participants. On the other hand, some channels do not contain significant information throughout the whole time which is also highlighted by the compression ratios: In average, *Unfold* manifests a compression ratio of 62.8, whereas all other samples result in lower average ratios with a minimum of 50.2. In general, a signal with lower spectral entropy, as given in *Unfold*, requires a lower coding bitrate for achieving a comparable quality to a signal with higher entropy [cf. 19, chapter 2].

5 Conclusion and future research

We tested the Opus multichannel coding ability using the DCR method and 5.1 surround sound. The overall perceived quality correlates with the bitrates whereas the saturation is not yet reached at 192 kbit/s for all tested music samples. In the overall trend, the quality is assessed as acceptable for all bitrates, but diverges for the individual test stimuli. Hence, the minimal bitrate required by the codec for tolerable perceived quality depends strongly on the music characteristics of the audio sample. Under most circumstances, a bitrate of 128 kbit/s is sufficient to achieve a minimal DMOS of 3.0 (*degradation slightly annoying*).

In the future, we would like to conduct research on the delay of multichannel audio coding using Opus codec. Furthermore, we will perform tests in order to determine and compare delay values of established audio codecs.

Acknowledgement

We would like to thank all our listeners who volunteered in this study and Jens Berger from SwissQual AG, Switzerland (a Rhode & Schwarz Company) for useful hints. Further thanks goes to GWT-TUD GmbH for supplying audio equipment.

References

[1] *Video on Demand Streaming Flatrates im Vergleich*. Trialo. Nov. 2016. URL: https://www.trialo.de/vod-streaming/.

[2] *Data use and stream quality on PlayStation Now*. Playstation. URL: https://www.playstation.com/en-gb/get-help/help-library/services/playstation-now/data-usage-and-stream-quality-on-playstation-now/ (visited on 06/01/2017).

[3] VALIN, J.-M., K. VOS and T. TERRIBERRY: *Definition of the Opus Audio Codec*. RFC 6716 (Proposed Standard). Internet Engineering Task Force, Sept. 2012. URL: http://www.ietf.org/rfc/rfc6716.txt.

[4] VALIN, J.-M., G. MAXWELL, T. TERRIBERRY and K. VOS: *High-quality, low-delay music coding in the opus codec*. In: *Audio Engineering Society Convention 135*. Audio Engineering Society. 2013.

[5] MARUSCHKE, M., O. JOKISCH, M. MESZAROS, F. TROJAHN and M. HOFFMANN: *Quality Assessment of Two Fullband Audio Codecs Supporting Real-Time Communication*. In: *Speech and Computer. 18th International Conference, SPECOM 2016*. Springer International Publishing, 2016, pp. 571–579. DOI: 10.1007/978-3-319-43958-7_69.

[6] VALIN, J. and C. BRAN: *WebRTC Audio Codec and Processing Requirements*. RFC 7874. RFC Editor, May 2016.

[7] MARUSCHKE, M., O. JOKISCH, M. MESZAROS and V. IAROSHENKO: *Review of the Opus Codec in a WebRTC Scenario for Audio and Speech Communication*. In: *Speech and Computer*. Springer, 2015, pp. 348–355.

[8] JARSCHEL, M., D. SCHLOSSER, S. SCHEURING and T. HOSSFELD: *An Evaluation of QoE in Cloud Gaming Based on Subjective Tests*. In: *Fifth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*. June 2011, pp. 330–335. DOI: 10.1109/IMIS.2011.92.

[9] LUTZKY, M., G. SCHULLER, M. GAYER, U. KRÄMER and S. WABNIK: *A guideline to audio codec delay*. In: *AES 116th convention, Berlin, Germany*. 2004, pp. 8–11.

[10] HOENE, C., J. VALIN, K. VOS and J. SKOGLUND: *Summary of OPUS listening test results*. Tech. rep. -. Available: http://tools.ietf.org/html/draft-ietf-codec-results-03. May 2013.

[11] RÄMÖ, A. and H. TOUKOMAA: *Subjective Qualitiy Evaluation of the 3GPP EVS Codec*. In: *IEEE International Conference on Acoustics, Speech and Signal Processing; Brisbane; Australia* (Apr. 2015), pp. 5157–5161.

[12] DIETZ, M. et al.: *Overview of the EVS codec architecture*. In: *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Apr. 2015, pp. 5698–5702. DOI: 10.1109/ICASSP.2015.7179063.

[13] *2L High Resolution Music .:. free TEST BENCH*. 2L. URL: http://www.2l.no/hires/index.html (visited on 03/01/2017).

[14] *Dolby Atmos Demonstration Disc*. Aug. 2014.

[15] ITU-T: *Methods for objective and subjective assessment of quality- Methods for subjective determination of transmissen quality*. REC P.800. International Telecommunication Union (Telecommunication Standardization Sector), Aug. 1996. URL: http://www.itu.int/rec/T-REC-P.800-199608-I/en.

[16] ITU-R: *Method for the subjective assessment of intermediate quality levels of coding systems*. REC BS.1534. International Telecommunication Union (Radiocommunication Sector), Oct. 2015. URL: http://www.itu.int/rec/R-REC-BS.1534/en.

[17] ITU-R: *Multichannel stereophonic sound system with and without accompanying picture*. REC BS.775-3. International Telecommunication Union (Radiocommunication Sector), Aug. 2012. URL: http://www.itu.int/rec/R-REC-BS.775-3-201208-I/en.

[18] *Downloads – Opus Codec*. URL: http://www.opus-codec.org/downloads/ (visited on 06/01/2017).

[19] STRUTZ, T.: *Bilddatenkompression: Grundlagen, Codierung, Wavelets, JPEG, MPEG*. 4th ed. Springer-Verlag, 30th Sept. 2009. 408 pp.