Quality Analysis of Audio-Video Transmission in an OFDM-Based Communication System

Monika Zamlynska[1], Grzegorz Debita[1], and Przemyslaw Falkowski-Gilski[2](✉)

[1] Faculty of Management, General Tadeusz Kosciuszko Military University of Land Forces, Czajkowskiego 109, 51-147 Wroclaw, Poland
grzegorz.debita@awl.edu.pl

[2] Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, Narutowicza 11/12, 80-233 Gdansk, Poland
przemyslaw.falkowski@eti.pg.edu.pl

Abstract. Application of a reliable audio-video communication system, brings many advantages. With the spoken word we can exchange ideas, provide descriptive information, as well as aid to another person. With the availability of visual information one can monitor the surrounding, working environment, etc. As the amount of available bandwidth continues to shrink, researchers focus on novel types of transmission. Currently, orthogonal frequency division multiplexing (OFDM) is widely utilized both in wired and wireless transmission. In this paper we investigate the quality of service (QoS) parameters of a simulated data transmission system, dedicated particularly to audio and video content distribution with orthogonal frequency division multiplexing. The audio research part involves a group of four language sets, namely: American English, British English, German, and Polish, processed using the Ogg Vorbis format. Whereas, in the video part we investigate a set of MPEG-4 video sequences coded at standard resolution of 480 × 270 pixels. Tests were performed under varying network and bandwidth conditions, including signal-to-noise ratio (SNR) and bit error rate (BER). Results of this study may aid parties interested in designing additional backup or supplementary services for portable devices and user terminals, including reliable means of contact, surveillance and monitoring for the Industry 4.0 and Internet of things (IoT) concept.

Keywords: IoT · Quality evaluation · Reliability · Video coding

1 Introduction

Designing and maintaining reliable communication services is a challenging task. With the outbreak of both desktop and mobile devices, followed by novel modulation and coding schemes, user expectations regarding the level of quality of a particular service continues to grow [1–4]. Currently, more and more solutions are based on orthogonal frequency division multiplexing (OFDM) [5, 6]. With fluctuating bandwidth conditions

T. Hara and H. Yamaguchi (Eds.): MobiQuitous 2021, LNICST 419, pp. 724–736, 2022.
https://doi.org/10.1007/978-3-030-94822-1_47

in heterogeneous networks, it is important to study how does it affect the quality of transmission.

Nowadays, its utilization in industrial networks seems to be of great interest, particularly in smart grids and the Industry 4.0 concept, including various signals, e.g. audio and video [7–10]. Considering the importance of the analyzed problem, we intended to determine the impact of several factors on the transmission of audio and video content. In our case, we utilize the orthogonal frequency division multiplexing technique. The study involved a set of speech samples as well as video sequences, coded and then processed in our custom-build OFDM telecommunication system, with respect to quality of service (QoS) requirements.

2 Audio Quality Evaluation

2.1 Audio Signal Processing

According to the 3rd Generation Partnership Project (3GPP) [11], when examining digital voice communication services, they can be divided into three groups. Table 1 describes principle voice (speech) services with their requirements, including delay and bit error rate (BER).

Table 1. Principle voice communication services with QoS requirements.

Heading level	Delay [ms]	BER
Conversational voice (real-time streaming)	100	10^{-2}
Non-conversational voice (buffered streaming)	300	10^{-6}
Interactive voice (live streaming)	100	10^{-3}

This particular study is focused on speech (voice) communication. According to Table 1, these services require a delay from less than 100 ms (conversational or interactive voice) to less than 300 ms (non-conversational voice). Whereas, when it comes to error rate, the accepted threshold ranges from 10^{-2} or 10^{-3} up to 10^{-6}, depending on the variant. This of course can affect the quality of transmitted speech samples.

2.2 Speech Signal Samples

The signal samples used during study were sourced from ITU-T P.501 [12], and consisted of sentences spoken by both female and male lectors in different languages. When examining the international profile of the broadcasting and streaming industry, we have selected 4 language sets: American English (AE), British English (EN), German (GE), and Polish (PL).

These samples were originally available in the WAV 16-bit PCM format. Next, each sample was coded using the Ogg Vorbis format [13, 14], the bitrate was set to 32 kbps, whereas the initial sampling frequency was changed to 44.1 kHz. Then all of them were transmitted via our communication system.

2.3 Simulated Communication System

In order to assess the quality of transmission in the modeled telecommunication system, we have utilized the Matlab/Simulink environment. The simulated model enabled to determine the BER for the transmission of audio files. The block diagram is shown in Fig. 1.

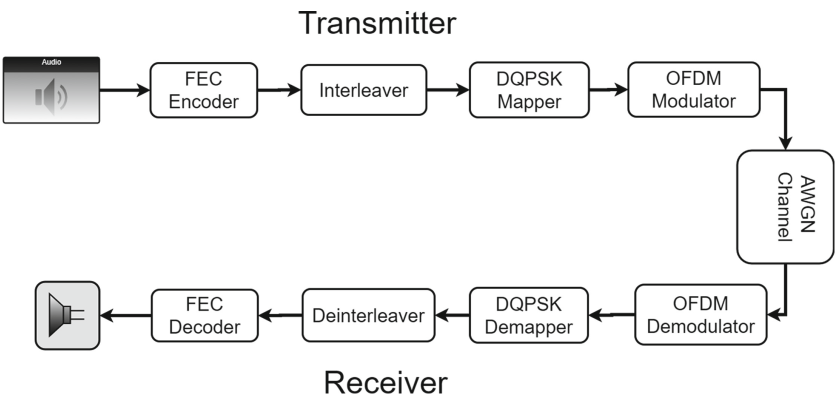

Fig. 1. Block diagram of simulated audio transmission system.

The modeled communication link was based on a well-known broadcasting chain utilized e.g. in digital audio broadcasting (DAB) [15–17]. The audio content used during the simulation was encoded using the Ogg Vorbis codec. It was transmitted over the channel in a single channel (mono) mode. Therefore, it did not require multiplexing during transmission. The coding itself was based on the DAB system.

2.4 Results

Not surprisingly, SNR has a profound impact on the overall error rate. Similar results may be observed regardless of the language set. An overall summary of the dependency of BER on SNR is shown in Fig. 2.

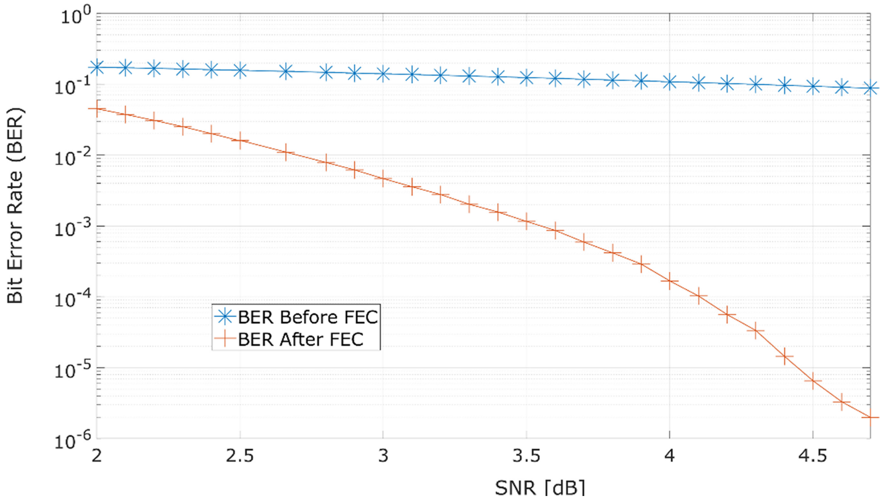

Fig. 2. Dependency of BER on SNR with and without the application of FEC for audio content.

2.5 Discussion

In this experiment, the Ogg Vorbis audio files, with a bitrate of 32 kbps and a sampling frequency of 44.1 kHz, were fed into the simulator. The input was composed of four language sets, namely: American English, British English, German, and Polish. Furthermore, in each case both female and male voices were distinguished.

All files were processed one by one, in a queue. The initial (input) as well as resulting (output) format remained unaltered. Additionally, the content processing chain did not cause change in either bitrate nor sampling frequency.

Overall, the test involved 24 audio files, with signal-to-noise ratio (SNR) ranging from 2.66 to 4.8 dB. The BER was monitored both before forward error correction (FEC) protection coding (so-called channel BER) and after FEC protection coding. Results of this experiment are described in Tables 2, 3, 4, and 5. Results for the American English (AE) language set are described in Table 2. Whereas, results for the following sets of speech samples, that is: British English (EN), German (GE), and Polish (PL), are described in Tables 3, 4, and 5, respectively.

Table 2. Results for transmitted speech samples in American English.

File name	SNR [dB]	BER after FEC	BER before FEC
AEfemale1	2.660	1.1088E−02	1.5166E−01
	3.530	1.0507E−03	1.2295E−01
	4.080	1.1567E−04	1.0594E−01
	4.450	1.0092E−05	9.4984E−02
	4.745	1.1755E−06	8.6652E−02

(*continued*)

Table 2. (*continued*)

File name	SNR [dB]	BER after FEC	BER before FEC
AEfemale2	2.660	1.0806E−02	1.5184E−01
	3.530	1.0295E−03	1.2319E−01
	4.080	1.0437E−04	1.0602E−01
	4.450	1.0999E−05	9.5014E−02
	4.745	1.2548E−06	8.6657E−02
AEmale1	2.660	1.0932E−02	1.5179E−01
	3.530	1.0671E−03	1.2311E−01
	4.080	1.2096E−04	1.0597E−01
	4.450	1.3148E−05	9.4996E−02
	4.745	1.0253E−06	8.6653E−02
AEmale2	2.660	1.1064E−02	1.5163E−01
	3.530	1.0716E−03	1.2297E−01
	4.080	1.1415E−04	1.0593E−01
	4.450	1.0590E−05	9.6452E−01
	4.745	1.0473E−06	8.6232E−02

According to obtained results, FEC has an enormous impact on the final error rate. The initial range of approx. 10^{-1} and 10^{-2} is shifted to a new range from 10^{-2} up to even 10^{-6}. What is worth mentioning, this increase is observable regardless of the audio sample.

Table 3. Results for transmitted speech samples in British English.

File name	SNR [dB]	BER after FEC	BER before FEC
ENfemale1	2.660	1.1316E−02	1.5173E−01
	3.530	1.0991E−03	1.2306E−01
	4.080	1.1305E−04	1.0589E−01
	4.450	1.1924E−05	9.4968E−02
	4.745	1.1621E−06	8.6317E−02
ENfemale2	2.660	1.1182E−02	1.5170E−01
	3.530	1.0625E−03	1.2297E−01
	4.080	1.2698E−04	1.0587E−01

(*continued*)

Table 3. (*continued*)

File name	SNR [dB]	BER after FEC	BER before FEC
	4.450	1.0897E−05	9.5876E−02
	4.745	1.0247E−06	8.6664E−02
ENmale1	2.660	1.0953E−02	1.5168E−01
	3.530	1.0119E−03	1.2299E−01
	4.080	1.1090E−04	1.0591E−01
	4.450	1.0980E−05	9.4998E−02
	4.745	1.3083E−06	8.5947E−02
ENmale2	2.660	1.1126E−02	1.5177E−01
	3.530	1.0424E−03	1.2314E−01
	4.080	1.2626E−04	1.0606E−01
	4.450	1.1053E−05	9.5845E−02
	4.745	1.1258E−06	8.6790E−02

The number of erroneous bits has been reduced by 10 times, in case of the lowest SNR value of 2.660 dB. Whereas, in case of the highest SNR value, that is 4.745 dB, the quality has been raised more than a thousand times.

Table 4. Results for transmitted speech samples in German.

File name	SNR [dB]	BER after FEC	BER before FEC
GEfemale1	2.660	1.1190E−02	1.5169E−01
	3.530	1.0604E−03	1.2308E−01
	4.080	1.0578E−04	1.0602E−01
	4.450	1.0543E−05	9.4999E−02
	4.745	1.0241E−06	8.6644E−02
GEfemale2	2.660	1.1114E−02	1.5177E−01
	3.530	1.0345E−03	1.2307E−01
	4.080	1.0335E−04	1.0594E−01
	4.450	1.1356E−05	9.6466E−02
	4.745	1.2803E−06	8.6679E−02
GEmale1	2.660	1.1301E−02	1.5172E−01
	3.530	1.0926E−03	1.2304E−01

(*continued*)

Table 4. (*continued*)

File name	SNR [dB]	BER after FEC	BER before FEC
	4.080	1.1373E−04	1.0592E−01
	4.450	1.1790E−05	9.5002E−02
	4.745	1.0334E−06	8.6656E−02
GEmale2	2.660	1.1299E−02	1.5181E−01
	3.530	1.0610E−03	1.2317E−01
	4.080	1.1319E−04	1.0613E−01
	4.450	1.0558E−05	9.5014E−02
	4.745	1.1809E−06	8.6596E−02

It can be seen that independently from the type of lector, that is either a female or male individual, obtained BER values are close in range. In case of both after and before FEC, obtained results most often differ only at the second or third decimal position.

Table 5. Results for transmitted speech samples in Polish.

File name	SNR [dB]	BER after FEC	BER before FEC
PLfemale1	2.660	1.0930E−02	1.5182E−01
	3.530	1.0603E−03	1.2309E−01
	4.080	1.0517E−04	1.0592E−01
	4.450	1.1673E−05	9.4971E−02
	4.745	1.0515E−06	8.6651E−02
PLfemale2	2.660	1.1021E−02	1.5179E−01
	3.530	1.0678E−03	1.2312E−01
	4.080	1.1854E−04	1.0560E−01
	4.450	1.0614E−05	9.4982E−02
	4.745	1.4726E−06	8.6631E−02
PLmale1	2.660	1.1245E−02	1.5172E−01
	3.530	1.0859E−03	1.2303E−01
	4.080	1.2338E−04	1.0588E−01
	4.450	1.1215E−05	9.5551E−02
	4.745	1.0268E−06	8.6481E−02

(*continued*)

Table 5. (*continued*)

File name	SNR [dB]	BER after FEC	BER before FEC
PLmale2	2.660	1.1096E−02	1.5177E−01
	3.530	1.0816E−03	1.2309E−01
	4.080	1.2623E−04	1.0597E−01
	4.450	1.1520E−05	9.5605E−02
	4.745	1.1515E−06	8.6650E−02

Not surprisingly, SNR has a profound impact on the overall error rate. Similar results may be observed regardless of the language set.

3 Video Quality Evaluation

3.1 Video Signal Processing

According to the 3rd Generation Partnership Project (3GPP) [11], video services can be divided into two main categories, depending on delay and bit error rate, as described in Table 6.

Table 6. Principle video transmission services with QoS requirements.

Heading level	Delay [ms]	BER
Conversational video (live streaming)	150	10^{-3}
Non-conversational video (buffered streaming)	300	10^{-6}

In this particular study we focused on BER ranging from 10^{-3} to 10^{-6}. In our case delay may be neglected.

3.2 Video Signal Samples

The processed video content, available in the MP4 format, coded with 8-bit resolution, was sourced from [18]. Each sample consisted of 5 sequences:

1. Walking man – static background with a single man walking along the street;
2. Windmill – static angle with numerous fast moving (rotating) windmills;
3. Traffic – static angle with cars passing along the road, at various speeds.
4. Toddler fountain – playful child walking around a fountain, with lots of movement in the background;

5. Toddler montage – similar material with a playful child recorded in slow-motion (much higher framerate);

   available in a number of bitrates (qualities), from standard definition (SD) up to Full-HD (1920 × 1080). For the purpose of this test, we have selected one video file, available in 480 × 270 resolution. We have chosen this resolution, the outcome of dividing 1920 × 1080 pixels by 4 in each axis, due to its scalability with the Full-HD format as well as popularity among many portable devices and user terminals, including media players and consoles [19].

3.3 Simulated Communication System

In order to assess the quality of transmission in the modeled telecommunication system, we have utilized the Matlab/Simulink environment. The simulated model enabled to determine the BER for the transmission of video files. The block diagram is shown in Fig. 3.

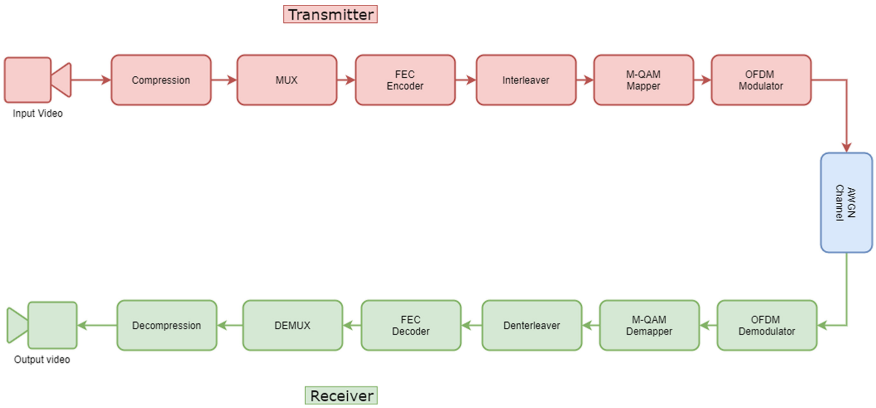

Fig. 3. Block diagram of simulated video transmission system.

The modeled communication link was based on a well-known broadcasting chain utilized e.g. in digital video broadcasting – terrestrial (DVB-T) [6, 20]. The video content was encoded using the MPEG-4 codec [21], undeniably on of the most popular formats for content processing and distribution, including a variety of consumer devices and streaming services [4]. It was transmitted over the channel in the red, green, blue (RGB) mode. The coding and multiplexing of particular streams was based on DVB [15]. Later on, data was interleaved with a random permutation, and then converted into 256-QAM symbols combined with OFDM [5, 22]. OFDM maintains the orthogonality of the sub-carriers, which reduces the risk of interference.

3.4 Results

During the experiment, the MPEG-4 video files, with a resolution of 480 × 270 pixels, were fed into the simulator. All files were processed one by one, in a queue. The original (input) as well as processed (output) format remained unaltered. Overall, the test involved varying SNR ranging from 24 to 29 dB, as well as monitored BER before and after forward error correction (FEC). Results of this analysis are shown in Fig. 4.

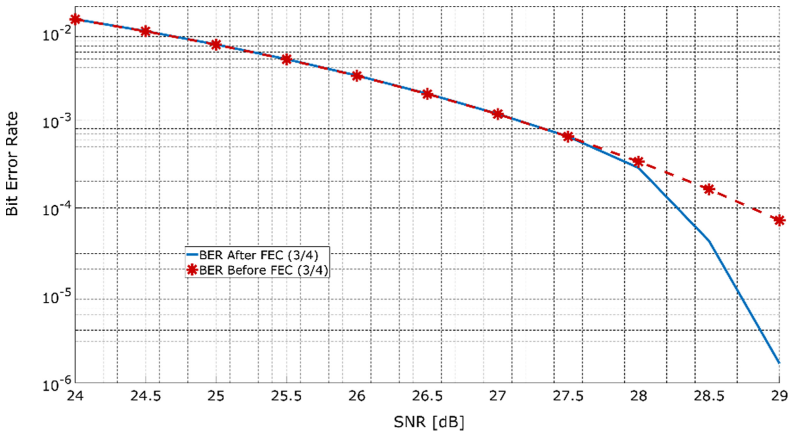

Fig. 4. Dependency of BER on SNR with and without the application of FEC for video content.

According to obtained results, FEC has a noticeable impact on the final error rate only from SNR equal to 27.5 dB and higher. The initial range of approx. 10^{-4} and 10^{-3} is shifted to a new range, in which the number of erroneous bits has been reduced by tens or even hundreds of times. However, the impact of both SNR and BER on quality of video content needs to be determined in a future subjective evaluation study.

3.5 Discussion

According to obtained results for lowest SNR equal to 24 dB (see first row of Fig. 5) was very poor. No details were visible, colors were distorted, whereas numerous elements did not have sharp edges. They looked grainy and/or blurred. Undeniably, stripes and checkered patterns were also noticeable, as a result of compression (coding) algorithms. This effect was clearly visible with a black and/or dark background. The walking man and cars were clearly noticeable, whereas the windmill presented something hard to determine. When it comes to the toddler, it was hard to determine whether it was water falling or just noise.

The second level of SNR equal to 24.4 dB (see second row of Fig. 5) provided similar observations. No significant changes were noticeable.

The third level of SNR equal to 27 dB (see third row of Fig. 5) was definitely better than the previous ones. The quality was still not acceptable, many pixel artefacts did occur. On the other hand, there was a visible decrease in the level of noise. The image was still noticeably divided into smaller segments (squares), but the colors were more precise and closer to reality. In this case, the windmill could be adequately labeled, so were the fountains.

In the fourth level of SNR equal to 28.3 dB (see fourth row of Fig. 5) there was a noticeable upgrade, however color mismatch and artefacts among many pixels were still present.

Fig. 5. Distorted frame of video sequence transmitted at different SNR.

Whereas the fifth level of SNR equal to 28.7 dB (see fifth row of Fig. 5) had only a little less noise, artefacts, and color distortions.

4 Conclusions

As shown, FEC correction coding can raise the quality of audio and video content processed and transmitted via a OFDM communication channel. Depending on the initial SNR value, this increase may be equal to ten times or even more than a thousand times. Furthermore, it has been proved that additional error correction mechanisms can raise the BER in case of all speech samples, regardless of the type of lector (female or male) or even spoken language (American English, British English, German, and Polish). As observed, similar results were obtained for all language sets. A small difference was only noticed on the second or third decimal point. When it comes to video coding, FEC had a noticeable impact on the final error rate only from SNR equal to 27.5 dB and higher.

This small difference in obtained BER, with and without FEC coding, when comparing each and every language set with one other, proved the correctness and accuracy of implementation of the transmission link. Moreover, it may be said that other language sets, more popular in various regions of the world, would perform similarly as the chosen and investigated portion of samples. Additionally, it would be also interesting to evaluate a different set of video content, including more static as well as dynamic sequences. This fact makes future studies, including various additional quality aspects, even more promising. One should keep in mind that, all in all, each system and service is designed to operate and interact with human end users. This implies many technical aspects, especially dependability and reliability [23].

Certainly, it would be interesting to determine the impact of SNR and BER, related with objective QoS, on the subjective judgements of the end user, referred to as quality of experience (QoE). Such an investigation would surely be valuable to both researchers and professionals active in the content creation and distribution link. They may include typical streaming or broadcasting services, as well as specific dedicated auxiliary solutions, e.g. utilizing portable mobile terminals. A good source of inspiration may be found in [24].

Further research should and will therefore focus on evaluating the set of both audio and video processed signal samples in a subjective listening test and video quality evaluation study. It would be interesting to directly link both SNR and BER parameters with a standard mean opinion score (MOS) judgement. Furthermore, it would be stimulating to broaden the range of content as well as research scenarios, followed by a subjective user evaluation. Future studies could be performed according to recommendations as well as best practices when it comes to crowdsourcing, which may be found in [25–28].

References

1. Hossfeld, T., et al.: Best practices for QoE crowdtesting: QoE assessment with crowdsourcing. IEEE Trans. Multim. **16**(2), 541–558 (2014)
2. Boz, E., Finley, B., Oulasvirta, A., Kilkki, K., Manner, J.: Mobile QoE prediction in the field. Pervas. Mobile Comput. **59**, 101039 (2019)
3. Kostek, B.: Music information retrieval – the impact of technology, crowdsourcing, big data, and the cloud in art. J. Acoust. Soc. Am. **146**(4), 2946 (2019)
4. Falkowski-Gilski, P., Uhl, T.: Current trends in consumption of multimedia content using online streaming platforms: a user-centric survey. Comput. Sci. Rev. **37**, 100268 (2020)
5. Cioni, S., Corazza, G.E., Neri, M., Vanelli-Coralli, A.: On the use of OFDM radio interface for satellite digital multimedia broadcasting systems. Int. J. Satell. Commun. Network. **24**(2), 153–167 (2006)
6. Aragón-Zavala, A., Angueira, P., Montalban, J., Vargas-Rosales, C.: Radio propagation in terrestrial broadcasting television systems: a comprehensive survey. IEEE Access **9**, 34789–34817 (2021)
7. Lazaropoulos, A.G., Cottis, P.G.: Transmission characteristics of overhead medium-voltage power-line communication channels. IEEE Trans. Power Deliv. **24**(3), 1164–1173 (2009)
8. Henry, P.S.: Interference characteristics of broadband power line communication systems using aerial medium voltage wires. IEEE Commun. Mag. **43**(4), 92–98 (2005)

9. Debita, G., Falkowski-Gilski, P., Habrych, M., Miedziński, B., Wandzio, J., Jedlikowski, P.: Quality evaluation of voice transmission using BPL communication system in MV mine cable network. Elektron. Elektrotech. **25**(5), 43–46 (2019)
10. Debita, G., et al.: BPL-PLC voice communication system for the oil and mining industry. Energies **13**(18), 4763 (2020)
11. 3GPP Technical Specification 23.203: Technical specification group services and system aspects; Policy and charging control architecture (2011)
12. ITU Recommendation P.501: Test signals for telecommunication systems (2017)
13. Kosaka, A., Yamaguchi, S., Okuhata, H., Onoye, T., Shirakawa, I.: VLSI implementation of Ogg Vorbis decoder for embedded applications. In: Proceedings of 15th Annual IEEE International ASIC/SOC Conference, pp. 20–24. IEEE, Rochester (2002)
14. Kosaka, A., Okuhata, H., Onoye, T., Shirakawa, I.: Design of Ogg Vorbis decoder system for embedded platform. IEICE Trans. Fundam. Electron. Commun. Comput. Sci. **88**(8), 2124–2130 (2005)
15. Harada, H., Prasad, R.: Simulation and Software Radio for Mobile Communications, 1st edn. Artech House, Norwood (2002)
16. Kim, G., Lee, Y.T., Park, S.R., Lee, Y.H.: Design and implementation of a novel emergency wake-up alert system within a conventional T-DMB service network. IEEE Trans. Consum. Electron. **60**(4), 574–579 (2014)
17. Zhang, H., Wang, H., Wang, G., Lu, M.: Design and implementation of the DAB/DMB transmitter identification information decoder. Int. J. Circuit. Syst. Signal Process. **11**, 59–64 (2017)
18. Netflix Open Content – Chimera Database. http://download.opencontent.netflix.com/?prefix=AV1/Chimera/Old/. Accessed 15 July 2021
19. Miranda, G., Macedo D.F., Marquez-Barja, J.M.: A QoE inference method for DASH video using ICMP probing. In: Proceedings of 16th International Conference on Network and Service Management, pp. 1–5. CNSM, Izmir (2020)
20. ETSI Standard EN 300 744: Digital video broadcasting (DVB); framing structure, channel coding and modulation for digital terrestrial television (2009)
21. Thibeault, J.: Streaming video fundamentals. SMPTE Motion Imaging J. **129**(3), 10–15 (2020)
22. Russell, M., Stüber, G.L.: Terrestrial digital video broadcasting for mobile reception using OFDM. Wirel. Pers. Commun. **2**, 45–66 (1995)
23. Zamojski, W., Mazurkiewicz, J., Sugier, J., Walkowiak, T., Kacprzyk, J. (eds.): Theory and Applications of Dependable Computer Systems: Proceedings of the Fifteenth International Conference on Dependability of Computer Systems DepCoS-RELCOMEX, June 29 – July 3, 2020, Brunów, Poland, 1st edn. Springer, Cham (2020)
24. Jeena Jacob, I., Kolandapalayam Shanmugam, S., Piramuthu, S., Falkowski-Gilski, P. (eds.): Data Intelligence and Cognitive Informatics: Proceedings of ICDICI 2020. Springer, Singapore (2021)
25. Korshunov, P., Shuting, C., Touradj, E.: Crowdsourcing Approach for Evaluation of Privacy Filters in Video Surveillance. In: Proceedings of ACM Workshop on Crowdsourcing for Multimedia, pp. 35–40. ACM, Nara (2012)
26. Figuerola Salas, Ó., Adzic, V., Shah, A., Kalva. H.: Assessing internet video quality using crowdsourcing. In: Proceedings of 2nd ACM International Workshop on Crowdsourcing for Multimedia, pp. 23–28. ACM, Barcelona (2013)
27. Anegekuh, L., Sun, L., Ifeachor, E.: A screening methodology for crowdsourcing video QoE evaluation. In: Proceedings of 2014 IEEE Global Communications Conference, pp. 1152–1157. IEEE, Austin (2014)
28. ITU Technical Report PSTR-CROWDS: Subjective evaluation of media quality using a crowdsourcing approach (2018)