CEUR-WS.org/Vol-2654/paper25.pdf

Comparative Analysis of Modern Formats of Lossy Audio Compression

Bohdan Zhurakovskyi [1][0000-0003-3990-5205], Nataliia Tsopa [1][0000-0002-3215-0711], Yevhenii Batrak [1][0000-0003-3027-4507], Roman Odarchenko[2,3][0000-0002-7130-1375] and Tetiana Smirnova[4][0000-0001-6896-0612]

[1]National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
[2]National Aviation University, Kyiv, Ukraine
[3]Yessenov University, Aktau, Kazakhstan
[4]Central Ukrainian National Technical University, Kropivnitskiy, Ukraine
zhurakovskiybyu@tk.kpi.ua

Abstract. In the modern world, computer engineering is developing high-speed – frequency and performance of processes are growing, storage space is increasing, and storage access time is speeding up. However, while the speed of different devices is extremely growing, the speed of data transfer is increasing significantly slower. The peculiarity of the majority types of data is their redundancy. Due to this, nowadays the compression algorithms are widely used to provide efficient storage and transfer of information. The objective of the work is the study and analysis of modern formats of lossy audio information compression. At present, a significant number of formats of audio compression are used. They all have pros and cons. This study is performed to find out which format is the most effective and successfully replaced MP3 – one of the most popular formats.

Keywords: compression, data, digital audio, lossy compression, quantization.

1 Introduction

In the years since the first famous phonautograph audio recording in 1860, the technology of audio recording and reproduction has continuously been developing. The 20th century was the beginning of the period of professional phonograph recorders and engineers, the epoch of audio transfer via radio waves, significant achievements in sound quality and technology, and further development of audio industry and trade in general.

In 1982, the world of audio took its first steps into the new century with the first digital audio format – compact-disk (CD). Built based on the innovative technology of pulse coding (PCM), a CD could store analogue sound waves as digital values by quantizing them to the nearest supported digital value [1].

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). CybHyg-2019: International Workshop on Cyber Hygiene, Kyiv, Ukraine, November 30, 2019.

Pulse Code Modulation (PCM) caused a new era of innovations for digital audio formats. Modern codecs like MP3 and WAV were growing in intensity during the decade. At the beginning of the 2000s, the first wave of lossless audio codecs, which increased the digital quality but led to big file sizes, occurred.

However, these formats were not ready for the capturing of markets by MP3 players at the beginning of the 2000s. iPod Apple and other similar devices introduced the world of digital audio to millions of people, and MP3 became a standard for sound reproduction all over the world. Today MP3 has become out-of-date and is no longer supported. Currently, there is a significant number of audio codecs, which are used in different industries. They all have pros and cons. The objective of this work is the study and analysis of modern formats of lossy audio information compression to define the most effective that could replace MP3 – one of the most popular formats.

The most advanced and standard codecs have been chosen for review in this work. This list includes MP3, AAC and OGG (Vorbis and Opus).

2 Statement of the research problem

2.1 MPEG-1 Audio Layer3

The history of MP3 began back in 1987 when Fraunhofer Institute started to work with audio coding as part of the Digital Audio Broadcasting (DAB) project.

In 1988, Motion Pictures group of experts was created, and in 1992, Audio Layer 3 standard was created based on Fraunhofer developments.

Fraunhofer released the first MP3 player, but it was not very easy to use. Soon the AMP free player appeared, which was designed by university students, and after the development of user interfaces for Windows, Mac-WinAMP and MacAMP music became an admiration that ended as Napster, iTunes and similar Internet services.

The reasons why MPEG-1 Layer-3 compression technology, in particular, became the main one are the following [2]:

1. Open standard. The specification is available (for a separate fee) to everyone interested in the standard implementation. While there is a range of patents that include MPEG Audio coding and decoding, all owners have declared that they licenced patents under reasonable terms for everyone. There is no separate company that owns the standard. There is a public example of code realization, created to avoid text misunderstandings of the standard. The format is well defined. Except for some incomplete realizations, there are no problems with the interaction of hardware and software from different providers.

2. Coders and decoders availability. Initially generated by the demands of professional use for radio broadcasting, hardware (DSP) and software decoders have been available for many years.

3. The popularity of other necessary related technologies.

Main MP3 characteristics.

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). CybHyg-2019: International Workshop on Cyber Hygiene, Kyiv, Ukraine, November 30, 2019.

1. Filterbank used in MP3 belongs to the class of hybrid banks of filters. It is built by combining two different kinds of filterbank: at first multiphase filterbank, then an additional one – modified discrete cosine transform (MDCT).

2. The psychoacoustic model mainly defines the realization quality of this coder. Additional work has been done in this part of the coder since the appearance of the original version.

3. The psychoacoustic model uses a separate filterbank, which combines calculations of energy values (for masking calculation), and the main filterbank. The output of the psychoacoustic model consists of the values of the masking threshold or the allowed noise for each frequency band. In MP3, these bands are approximately equivalent to the critical bands of human hearing. If the quantization noise can be restricted by the values of the lower threshold of masking for each frequency band, then the compression result must not differ from the original signal.

4. Quantized values are coded by Huffman coding. The optimal Huffman table is selected from the range of available variants for the adaptation of coding process with different local statistics. Different Huffman code tables may be selected for different parts of the spectrum to get even better adaptation to signal statistics. Since Huffman coding is the method with variable code length, and the noise formation is necessary for quantization noise to be lower than the masking threshold, the level value (that defines the size of quantization step) and the spectral values (that define the noise level for each band) are applied to the actual quantization.

MP3 has three compression modes:

1) CBR (constant bitrate) – compression with constant bitrate.

2) VBR (variable bitrate) – bitrate may change over time for its use optimization.

3) ABR (average bitrate) – average bitrate, a combination of the later ones.

Main MP3 defects.

Unlike analogue equipment of FM-broadcasting, psychoacoustic compression methods using too low bitrate or incorrect parameters demonstrate the defects, which in most cases differ from the noise or distortions we are all used to.

Loss of bandwidth. If a coder runs out of bits, other words it does not find the path for the coding of sound data unit with the necessary precision (e. g. allowed noise per critical frequency band) in the scope of the allowed bitrate, some frequencies may be equated to zero (removed). The most common manifestation of this is situations in which some high-frequency content was lost. If the loss of bandwidth is not constant but changes from frame to frame (e. g. every 24 ms), the effect becomes less significant than in the case of a constant decrease of bandwidth.

Echoes before the signal. This is a very common and known artefact for psychoacoustic systems of high frequency coding. The principal coding artefact is the noise that spreads for a while before the sound event, which causes masking.

Roughness. There is a discrepancy between the time permission of the coder (at least in a more effective "normal unit" mode), and the demand resulted from the temporary signal structure, especially in case of low bit frequencies and low discretization frequencies. This effect is the most observable for verbal signals and during listening via earphones. The effect is sometimes called a "double word" because one voice may sound like the recording is made twice and these two variants are overlapped.

The last and the most significant defect of MP3 is that it is not developing and is not supported anymore. This process started back in 1995, not long before the first successful practical realization. From this moment, the standard has not been changing, and only the coders have been improving (to be more exact, their psychoacoustic model) by efforts of different companies. As of 2019, the potential development of the format is officially revoked, though it remains the most common format of lossy audio compression.

2.2 Advanced audio coding

The first version of Advanced AudioCoding was standardized in 1994 as a part of the MPEG-2 standard. Based on the development experience of MP3 and other codecs, the primary authors, AT&T, Dolby, Fraunhofer IIS and Sony, have started to create the modern audio codec from scratch.

In comparison to MPEG-2, the MPEG-4 standard defines a set of coding tools, which can be combined to comply with requirements in the best way. Some combinations for different programs are defined in the MPEG-4 standard as profiles. The MPEG-2 AAC codec was expanded into the MPEG-4 standard using adding the tools of psychoacoustic noise suppression (PNS), spectral band replication (SBR) and parameterised stereo (PS) [3].

The main MPEG-4 AAC profile is “AAC profile” (see Fig. 1), which is usually represented as AAC-LC (standard with low complexity). The most famous implementation of this profile is iTunes. The combination of AAC-LC with SBR tool has caused the appearance of high-efficient AAC (HE-AAC). The SBR tool expands the efficiency of coding, especially at low bitrates [4].

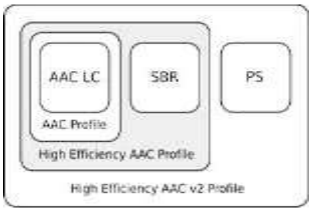


Fig.1. AAC profiles.

For further improvement of coding efficiency, HE-AAC may be combined with the PS (see Fig. 2) tool to create "High-Efficiency AAC v2" profile (HE-AAC v2). By creating a new profile by adding new tools to the main profile, the coder has all the tools necessary to create a bit flow. The same applies to the decoder. HE-AAC v2 decoder may decode AAC-LC, HE-AAC and HE-AAC v2 bit flow.

From the moment, the calculation complexity became insignificant for SBR and PS tools, all broadcasting standards, except for ISDB in Japan, define HE-AAC v2 profile as a standard to provide the listeners with the maximum quality. ISDB standard had been completed before the MPEG-4 standardization work was done, and therefore it is based on the MPEG-2 AAC standard. AAC low delay profile (AAC-LD) and

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). CybHyg-2019: International Workshop on Cyber Hygiene, Kyiv, Ukraine, November 30, 2019.

the enhanced AAC low delay profile (AAC-ELD) were designed for communication apps, as well as for some types of attached links. While the coding delay of AAC-ELD is up to 15 ms, these codecs still provide excellent efficiency per bitrate. For the highest quality and archiving purposes, high-definition AAC profile (HD-AAC) provides scaled lossy audio compression.

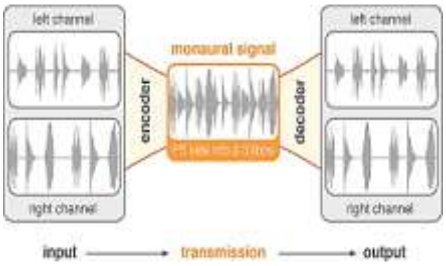


Fig.2. PS tool operation.

Advantages of AAC vs MP3:

- Higher discretization frequency (from 8 kHz to 96 kHz) vs MP3 (from 16 kHz to 48 kHz)
- Up to 48 channels (MP3 supports up to two channels in MPEG-1 mode and up to 5.1 channels in MPEG-2 mode)
- Arbitrary bitrate and variable frame length
- Higher efficiency and more accessible filterbank (non-hybrid MP3 coding, AAC uses clean MDCT)
- Higher efficiency of coding for stationary signals
- Higher accuracy of coding for transitory signals
- Better processing of frequencies higher than 16 kHz
- More flexible supported stereo (PS tool)
- Additional modules (tools) added for compression efficiency increase. Modules may be combined into different coding profiles

2.3 Ogg: Opus

Opus is the most recent codec of lossy audio compression, which was released in 2012 and designed to become the only standard for several allocations. Opus was created, taking into account the needs of the modern world, where the primary needs are high quality, and low delay adapted to network communications and lived musical performances. Its delay may be reduced to 5 ms compared to the majority of other codecs that are unlikely to provide even 100 ms delay.

Opus codec is based on two original independent developments. Xiph.org started to work on the codec named CELT in 2007 to reduce the gap between Vorbis (their audio codec for high bitrates) and Speex (their verbal codec) for programs where both high-quality sound and low delay are preferred. At the same time, Skype was working on SILK program, a new generation codec for their VoIP software. After combining their efforts in 2010, the workgroup on Internet development (IEFT) approved Opus for standardization as RFC 6716 two years later, and on 11 September 2012, the first stable versions were released [5].

What makes Opus especially attractive for use is the fact that it combines the advantages of audio codecs specific to music and broadcasting [6]. That became possible due to parallel managing of two coders based on the modified CELT coder and enhanced SILK (see Fig. 3).

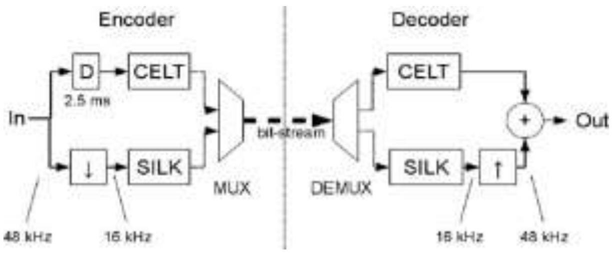


Fig.3. Opus coding and decoding process.

SILK codec, which is based on linear predictive coding (LPC), is primarily intended for speech transfer at the low delay, and that is why it is not well suited for music. CELT codec, on the other hand, is a broadband codec of general intended use, based on the modified discrete cosine transform (MDCT) [7] and coding purposes for both speech and music at higher bitrates.

To get the best of both worlds, Opus combines these two codecs in three different modes:

1. SILK mode is particularly suited for broadcasting from narrowband to the broadband range.

2. CELT mode is the mode for coding of music and other complex broadband information.

3. Opus also provides a hybrid model that uses CELT for high frequencies (> 8 kHz) and SILK for low frequencies (< 8 kHz). This mode may be used for speech coding with higher quality than the mode, which can provide only SILK.

The mode that is used is defined implicitly via selected bitrate (which, by the way, can be changed on-the-go). Future versions of the coder allow the algorithm of music/speech detection to inform about the mode selection automatically.

Advantages of Opus:

- Bitrate from 6 kbit/s to 510 kbit/s
- Discretization frequency from 8 kHz to 48 kHz
- Support of constant (CBR) and variable (VBR) bitrates
- Support for narrowband and broadband sound
- Support of voice and music, and different algorithm optimization for each
- Support of stereo and mono
- Capability to restore audio flow in case of frame loss (PLC)
- Support of up to 255 channels

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). CybHyg-2019: International Workshop on Cyber Hygiene, Kyiv, Ukraine, November 30, 2019.

▪ Realizations with the use of arithmetic with drifting and fixed point

Codec is distributed under the BSD licence and is completely removed from all patent claims. Additionally, approved as an Internet standard. Opus may be used in any project, including commercial without sharing the content of the project.

2.4 Compression formats comparison

In the beginning, diffident metrics were used for analysis. As a result, it was decided to choose for final comparison such characteristics as file size, and audio analysis of quality and spectrum. In addition, the use of codecs for different directions in the music industry was taken into account. For the file size parameter, each song was converted into each of the four formats in three different bitrates. This analysis of the file size metrics indicates how each coder is correlated with the others. The sound quality value was obtained by blind listening based on the results of own listening and the studies of Xiph team. Finally, the spectrum of each file was analysed to define how each codec performs separation of frequency scale bands. For the comparison of each codec, the lossless WAV format was used as a baseline.

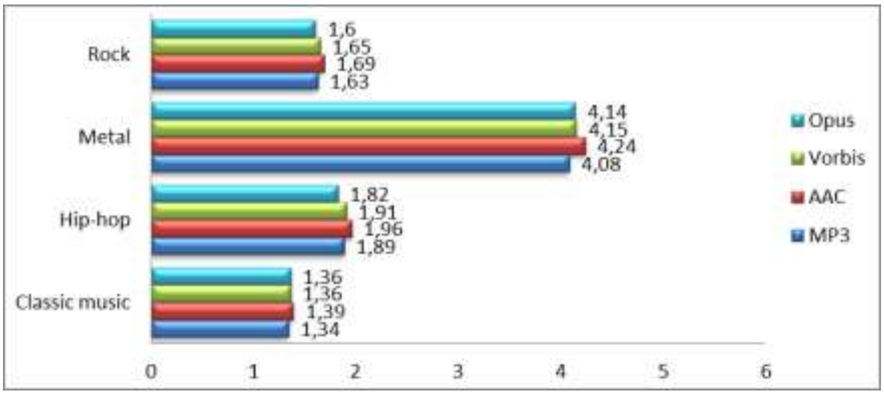


Fig.4. File size of different compression formats for different types of audio information at 64 kbit/s bitrate.

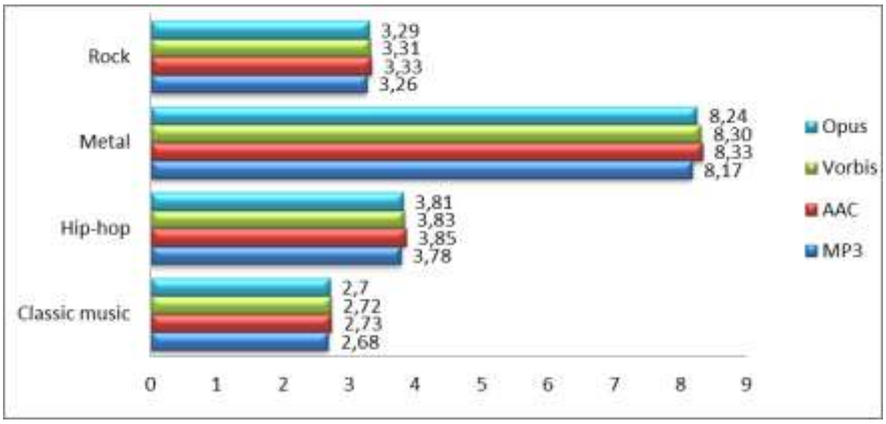


Fig.5. File size of different compression formats for different types of audio information at 128 kbit/s bitrate.

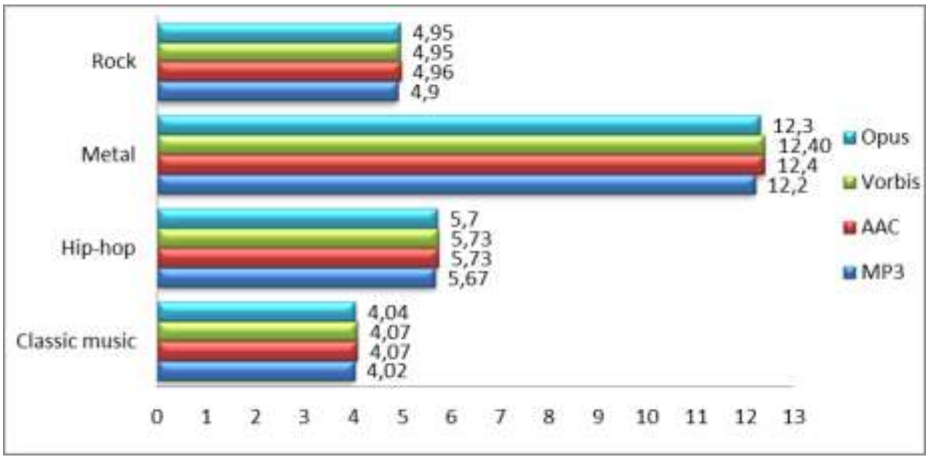


Fig.6. File size of different compression formats for different types of audio information at 192kbit/s bitrate.

These data show that MP3-codec creates a file of the smallest size and AAC creates the most important file that is true for all three bitrates. In general, the file size of different codecs almost equals to each other. Due to this factor, it can be stated that the analysis of file size metrics is not determinative.

The second test was performed by «blind listening». Each song was listened in three different bitrates.. An assessment table was created. After that, the results were summarised and sorted from the best to the worst.

Table 1. Table of sound quality assessment depending on the type of codec for different types of bitrates

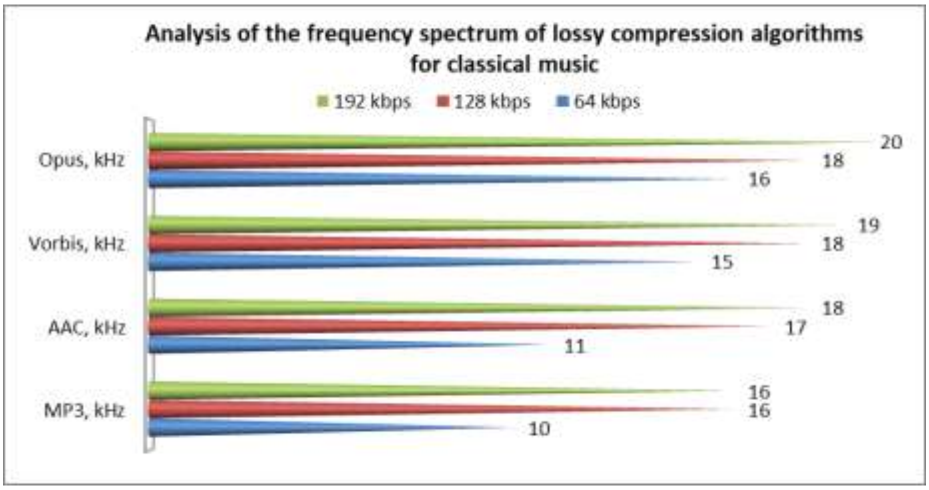
Rating	64 kbit/s	128 kbit/s	192 kbit/s

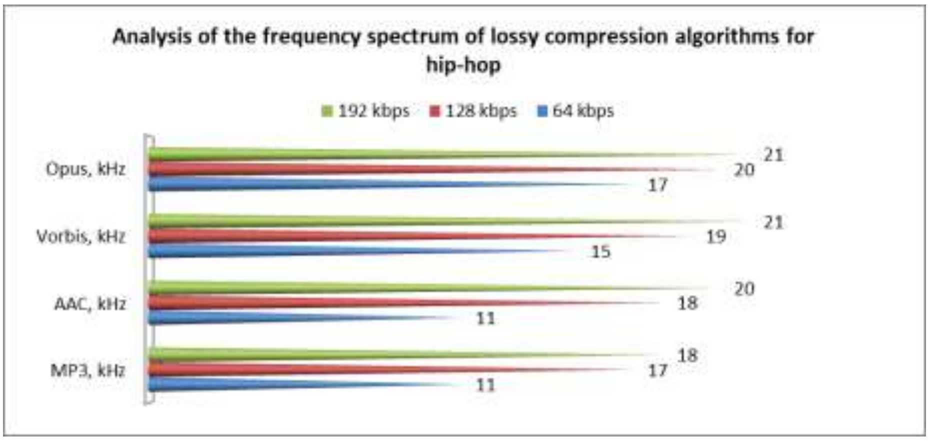
Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). CybHyg-2019: International Workshop on Cyber Hygiene, Kyiv, Ukraine, November 30, 2019.

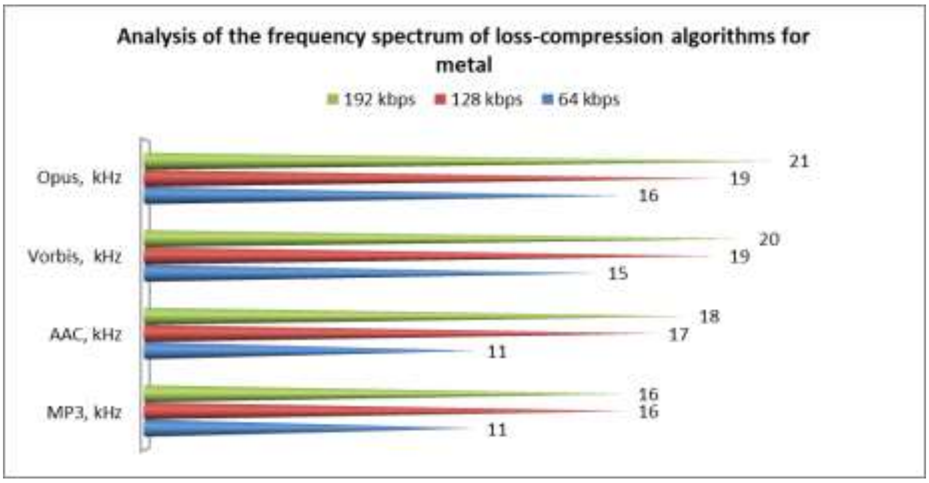
1	Opus	Opus	Opus
2	Vorbis	Vorbis	Vorbis
3	MP3	MP3	AAC
4	AAC	AAC	MP3

The information mentioned above shows that Ogg codecs were the best in all bitrates and the last Ogg Opus codec became the winner. The second was MP3, followed by AAC. The reason why AAC occupied the last place is that it is better adapted to the audio compression with high bitrates and loses out to MP3 at low bitrates.

The analysis of the frequency spectrum of each song shows how well each code compression algorithm works in comparison to WAV and each other. Its shows when frequency cross-section of the codec leads to the loss of high-frequency part of the sound. Then codec with the biggest frequency range is the codec that provides the nearest reproduction value of the full frequency spectrum in comparison to non-compressed WAV with the range of approximately 22 kHz.







Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). CybHyg-2019: International Workshop on Cyber Hygiene, Kyiv, Ukraine, November 30, 2019.

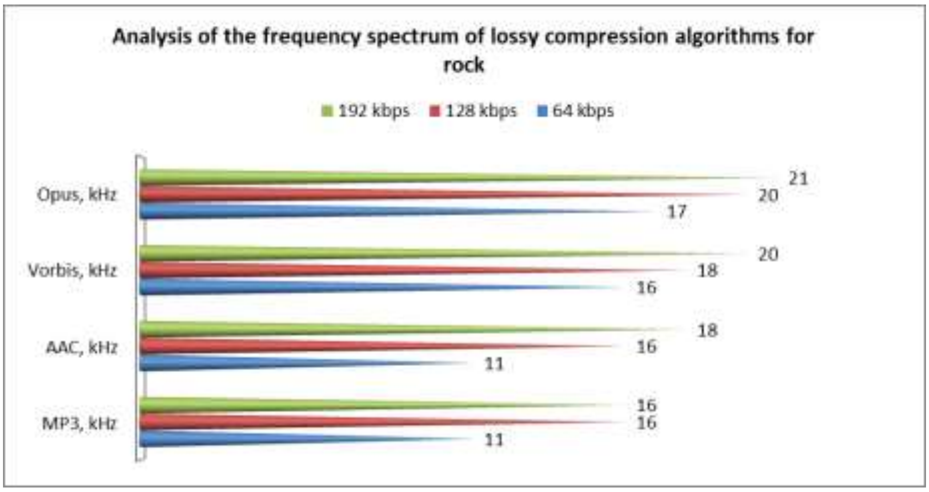


Fig.7. Analysis of the frequency spectrum of lossy compression algorithms for different types of music.

While comparing the codecs by listening, Xiph team received the results [8] that, in general, are the same as the obtained results. With further bitrate increase, the difference in quality between different codecs is aligned [9-13].

3 Conclusions and discussions

MP3 remains a more common lossy data compression format. Its defects are complete obsolescence, absence of further support and low-quality level per bitrate compared to formats that are more modern.

AAC has not become a standard format for audio storage, even though it was designed in 1994 as a direct replacement for MP3. Its use mostly ends with the audio storage in video clips. In general, AAC provides the solution of problems related to MP3 and a fresh approach to coding tools use. In addition, the problem is that AAC is a proprietary closed format, unlike all other Ogg types. It has a lower level of quality than Opus, especially concerning the issue of transitory processes of signal dynamics.

Opus is the most prospective lossy audio compression format. The quality per bitrate is currently the best. In addition, like AAC, this codec is not accessible for storing clean audio materials. It is widely spread in the modern Internet since WebM video data format, which stores sound using Opus, is the part of HTML5 standard. Nowadays, YouTube, as many other services mainly use Opus and AAC at a lower level.

Additionally, in those cases when a minimum bitrate is needed for verbal signals and upper frequencies are not critical (phone and other) other Ogg codec, Speex, shows quite good results. Since Opus, as Vorbis follower, was designed for qualitative compression of complex broadband signals (music and other), Speex was created because of human speech transfer.

The obtained results show that Opus codec has the biggest frequency range in all bitrates for all files. It can be observed that the capability to transfer high-frequency MP3 spectrum is approximately 16 kHz, while Ogg codecs are capable of the range up to 20–21 kHz at a bitrate of 192 bits. This codec has the biggest frequency range that provides the nearest reproduction value of the full frequency spectrum in comparison to non-compressed WAV with the range of approximately 22 kHz. In general, AAC codec has the third range in size, while MP3 codec moves to the bottom of the list.

Generally, while bitrate is increasing the difference between the codecs is decreasing. The principle moment here is the conditional point of "transparency".

This value is the highest for MP3 and is approximately 250 kbit/s. It is less for AAC (220–230 kbit/s) and even less for Opus (190–200 kbit/s).

References

1. Zhurakovskyi B., Srochinskaya G. Analysis of effectiveness recognition of methods to the identification of information. Management systems, navigation and audio communication, № 1(35), pp. 114 - 117. (2014)
2. Finlayson R., More A.: Loss-Tolerant RTP Payload Format for MP3 Audio. Internet Engineering Task Force, 22 p. (2008). Homepage, https://www.rfc-editor.org/info/rfc5219
3. ISO. ISO/IEC 13818-7:1997, Information technology. Generic coding of moving pictures and associated audio information. Part 7: Advanced Audio Coding (AAC) (1997). Homepage, https://www.iso.org/standard/25040.html
4. Logvinov M . Fundamentals of lossy audio coding. Beta testing Opus 1.3 (2018). Homepage, https://habr.com/ru/post/346532/
5. Goodwin, M.: Primary-Ambient Signal Decomposition and Vector-Based Localization for Spatial Audio Coding and Enhancement. In Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP), Honolulu, HI, USA, 15–20 April; pp. 9–12. (2007)
6. Bismor D., Bartnicki K.: Real-time Advanced Speech Enhancement using Low-power, Fixed-point Hardware. 2016 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA) IEEE Xplore Digital Library (2016). Homepage,https://ieeexplore.ieee.org/document/7763588/authors#authors
7. Chen, S.X.; Xiong, N.: Spatial parameters for audio coding: MDCT domain analysis and synthesis. Multimed. Tools Appl. vol. 48, 225–246. (2010)
8. Network Working Group. RTP Payload Format and File Storage Format for Opus Speech and Audio Codec. Opus codec. IETF (2015). Homepage, https://tools.ietf.org/html/rfc7587
9. Three MUSHRA-type tests performed in March 2011 at Google. Homepage,https://www.opus-codec.org/comparison/GoogleTest1.pdf
10. Al-Azzeh J.S., Al Hadidi M., Odarchenko R., Gnatyuk S., Shevchuk Z., Hu Z. Analysis of self-similar traffic models in computer networks, International Review on Modelling and Simulations, № 10(5), pp. 328-336, 2017.
11. Holub S., Khymytsia N., Holub M., Fedushko S. The Intelligent Monitoring of Messages on Social Networks. CEUR Workshop Proceedings. Vol 2616: COAPSN-2020, 2020. p. 308-317. http://ceur-ws.org/Vol-2616/paper26.pdf

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). CybHyg-2019: International Workshop on Cyber Hygiene, Kyiv, Ukraine, November 30, 2019.

12. Hryshchuk R., Molodetska K., Syerov Y. Method of Improving the Information Security of Virtual Communities in Social Networking Services. CEUR Workshop Proceedings. 2019. Vol 2392: COAPSN-2019. p. 23–41.
13. Odarchenko R., Abakumova A., Polihenko O., Gnatyuk S. Traffic offload improved method for 4G/5G mobile network operator, Proceedings of 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET-2018), pp. 1051-1054, 2018.