The Opus Codec

High-quality, low-delay music codec

Sevag Hanssian

MUMT 621, Winter 2021

Feburary 09, 2021

2021-02-09

The Opus Codec

The Opus Codec
High-quality, low-delay music codec

Sevag Hanssian

MUMT 621, Winter 2021

Feburary 09, 2021

Xiph.org

Xiph.org is a collection of open source, multimedia-related projects.[1]

1. Codecs: FLAC, Vorbis, Opus, Speex, Daala, Theora
2. Misc: RNNoise (**R**ecurrent **N**eural **N**[etwork])

The most aggressive effort works to put the foundation standards of Internet audio and video into the public domain, where all Internet standards belong.

1. Closed software and protocols are not evil or worse than open source, but by definition only exist serve the bottom line of a corporation
2. The Internet is built on open development, free exchange of ideas, and intellectual cooperation

[1] *Xiph.org: About*. URL: https://xiph.org/about/.

Xiph.org

Xiph.org is a collection of open source, multimedia-related projects.[1]

1. Codecs: FLAC, Vorbis, Opus, Speex, Daala, Theora
2. Misc: RNNoise (Recurrent Neural N[etwork])

The most aggressive effort works to put the foundation standards of Internet audio and video into the public domain, where all Internet standards belong.

1. Closed software and protocols are not evil or worse than open source, but by definition only exist serve the bottom line of a corporation
2. The Internet is built on open development, free exchange of ideas, and intellectual cooperation

[1] Xiph.org: About. URL: https://xiph.org/about/.

- I.e. they are not in the public's best interest
- Google AMP, facebook, twitter
- Google AMP - mention technical difficulty
- This is really becoming a problem today

Why multimedia needs open standards

1. MPEG – **M**oving **P**ictures **E**xpert **G**roup – "is the name of a family of standards used for coding audio-visual information (e.g., movies, video, music) in a digital compressed format.[2]"
2. "Working group of ISO/IEC (**I**nternational **O**rganization for **S**tandardization, **I**nternational **E**lectrotechnical **C**ommission) in charge of the development of international standards for compression, decompression, processing, and coded representation of moving pictures, audio and their combination."
3. RIAA – **R**ecording **I**ndustry **A**ssociation of **A**merica –[3]

   > *We work to protect artists' creative freedom and promote the unique work that labels do to support them. [...] We work to protect artists and all music creators from the damaging impact of music theft.*

[2] *MPEG – The Moving Picture Experts Group*. URL: https://www.mpegstandards.org/.

[3] *What We Do – RIAA*. URL: https://www.riaa.com/what-we-do/.

Why multimedia needs open standards

1. MPEG – **M**oving **P**ictures **E**xpert **G**roup – "is the name of a family of standards used for coding audio-visual information (e.g., movies, video, music) in a digital compressed format.[2]"
2. "Working group of ISO/IEC (**I**nternational **O**rganization for **S**tandardization, **I**nternational **E**lectrotechnical **C**ommission) in charge of the development of international standards for compression, decompression, processing, and coded representation of moving pictures, audio and their combination."
3. RIAA – **R**ecording **I**ndustry **A**ssociation of **A**merica –[3]
   We work to protect artists' creative freedom and promote the unique work that labels do to support them. [...] We work to protect artists and all music creators from the damaging impact of music theft.

[2] *MPEG – The Moving Picture Experts Group*. URL: https://www.mpegstandards.org/.
[3] *What We Do – RIAA*. URL: https://www.riaa.com/what-we-do/.

MP3 in 1998

Fraunhofer/Thompson (two industrial giants holding MP3-related patents) started demanding royalties in 1998[4]:

> *Since 1997, we have been working with the MP3 source code released by the ISO. [...] Then we got an e-mail [...] "As you may know, both the Fraunhofer Institute and THOMSON have done important work to develop MPEG Layer-3 audio compression (before and after it became part of the MPEG standards). This work has resulted in many inventions and several patents, covering the MPEG Layer-3 standard. Our files do not show that you have a valid license agreement with us. This means that the products infringe the patent rights of Fraunhofer and THOMSON."*

RIAA sued an MP3 player manufacturer, Diamond, in the late 90s[5]

[4] *What's New – Oct. 31, 1998*. URL: https://web.ncf.ca/aa571/wn103198.htm.

[5] Stephen W. Webb. "RIAA v. Diamond Multimedia Systems: The Recording Industry Attempts to Slow the MP3 Revolution, Taking Aim at the Jogger Friendly Diamond Rio". In: *7 RICH. J.L. & TECH. 5*. 2000. URL: https://core.ac.uk/download/pdf/232774502.pdf.

MP3 in 1998
Fraunhofer/Thompson (two industrial giants holding MP3-related patents) started demanding royalties in 1998[4]:
Since 1997, we have been working with the MP3 source code released by the ISO. [...] Then we got an e-mail [...] "As you may know, both the Fraunhofer Institute and THOMSON have done important work to develop MPEG Layer-3 audio compression (before and after it became part of the MPEG standards). This work has resulted in many inventions and several patents, covering the MPEG Layer-3 standard. Our files do not show that you have a valid license agreement with us. This means that the products infringe the patent rights of Fraunhofer and THOMSON."
RIAA sued an MP3 player manufacturer, Diamond, in the late 90s[5]
[4] What's New – Oct. 31, 1998. URL: https://web.ncf.ca/aa571/wn103198.htm.
[5] Stephen W. Webb. "RIAA v. Diamond Multimedia Systems: The Recording Industry Attempts to Slow the MP3 Revolution, Taking Aim at the Jogger Friendly Diamond Rio". In: 7 RICH. J.L. & TECH. 5. 2000. URL: https://core.ac.uk/download/pdf/232774502.pdf.

- story of computall
- two forms of evil here
  - corporate giants who pretend to develop open standards, then apply a royalty
  - music publishers who are terrified of open standards
- birth of vorbis right here

Containers and codecs

1. A *container* is associated to the file extension – it describes which codecs are used for its video/audio contents, followed by the actual encoded video/audio data, and extra data such as subtitles
2. A *codec* defines how to *encode* raw audio/video into data to put in a container (i.e. file), and how to *decode* data from the container back to a form suitable for playback[6]

File extension	Audio codec	Video codec	Container
.webm	Vorbis or Opus	VP8 or VP9	Matroska
.mkv	Any	Any	Matroska
.ogg	Vorbis	n/a	Ogg
.opus	Opus	n/a	Ogg
.mp4	AAC	MPEG-4	MP4

[6] Jean-François Fortin Tam. *Understanding codecs and containers*. URL: http://www.pitivi.org/manual/codecscontainers.html.

Containers and codecs
1. A *container* is associated to the file extension – it describes which codecs are used for its video/audio contents, followed by the actual encoded video/audio data, and extra data such as subtitles
2. A *codec* defines how to *encode* raw audio/video into data to put in a container (i.e. file), and how to *decode* data from the container back to a form suitable for playback[6]
File extension Audio codec Video codec Container
.webm Vorbis or Opus VP8 or VP9 Matroska
.mkv Any Any Matroska
.ogg Vorbis n/a Ogg
.opus Opus n/a Ogg
.mp4 AAC MPEG-4 MP4
[6] Jean-François Fortin Tam. *Understanding codecs and containers*. URL: http://www.pitivi.org/manual/codecscontainers.html.

The Opus Codec

Audio codec designed for the Internet[7]

- Open-source, royalty-free
- Lossy
- Can trade off quality to reduce latency
- Derives from:
  - CELT (**C**onstrained-**Energy** and **L**apped **T**ransform)
  - SILK, Skype speech codec
- Replaces Vorbis (music) and Speex (speech) in a single codec

Opus can operate in three modes:

1. SILK mode for speech – low bitrate narrowband speech
2. CELT mode for music – high bitrate, high quality music
3. Hybrid – SILK <8kHz, CELT >8kHz

[7] Jean-Marc Valin et al. *High-Quality, Low-Delay Music Coding in the Opus Codec*. 2016. arXiv: 1602.04845 [cs.MM]. URL: https://arxiv.org/abs/1602.04845.

The Opus Codec

Audio codec designed for the Internet[7]

- Open-source, royalty-free
- Lossy
- Can trade off quality to reduce latency
- Derives from:
  - CELT (**C**onstrained-**Energy** and **L**apped **T**ransform)
  - SILK, Skype speech codec
- Replaces Vorbis (music) and Speex (speech) in a single codec

Opus can operate in three modes:

1. SILK mode for speech – low bitrate narrowband speech
2. CELT mode for music – high bitrate, high quality music
3. Hybrid – SILK <8kHz, CELT >8kHz

[7] Jean-Marc Valin et al. *High-Quality, Low-Delay Music Coding in the Opus Codec*. 2016. arXiv: 1602.04845 [cs.MM]. URL: https://arxiv.org/abs/1602.04845.

- SILK is not an acronym
- CELT used to be a standalone algorithm, derived from Vorbis, now its the music part of opus
- It can change modes within a stream – making it good for some degradation that occurs as the internet gets choppy
- 8khz? segue into speech v music

Speech and music

Sampling rate (Hz)	Max frequency (Hz)	Name
8000	4000	Narrowband
16000	8000	Wideband
44100	22050	CD
48000	24000	Fullband (DVD)

80% of perceptually important spectrum in *voiced* speech is <4kHz, however speech up to 8kHz is preferred in subjective tests (due to *unvoiced* speech)[8]. Humans can hear from 20Hz-20kHz, and music requires a larger frequency range than speech[9].

[8] Julien Epps and W.H. Holmes. "A new technique for wideband enhancement of coded narrowband speech". In: Feb. 1999, pp. 174–176. ISBN: 0-7803-5651-9. DOI: 10.1109/SCFT.1999.781522.

[9] Brian Moore. "Effects of Sound-Induced Hearing Loss and Hearing Aids on the Perception of Music". In: *Journal of the Audio Engineering Society* 64 (Mar. 2016), pp. 112–123. DOI: 10.17743/jaes.2015.0081.

Speech and music

Sampling rate (Hz)	Max frequency (Hz)	Name
8000	4000	Narrowband
16000	8000	Wideband
44100	22050	CD
48000	24000	Fullband (DVD)

80% of perceptually important spectrum in *voiced* speech is <4kHz, however speech up to 8kHz is preferred in subjective tests (due to *unvoiced* speech)[8]. Humans can hear from 20Hz-20kHz, and music requires a larger frequency range than speech[9].

[8] Julien Epps and W.H. Holmes. "A new technique for wideband enhancement of coded narrowband speech". In: Feb. 1999, pp. 174–176. ISBN: 0-7803-5651-9. DOI: 10.1109/SCFT.1999.781522.

[9] Brian Moore. "Effects of Sound-Induced Hearing Loss and Hearing Aids on the Perception of Music". In: *Journal of the Audio Engineering Society* 64 (Mar. 2016), pp. 112–123. DOI: 10.17743/jaes.2015.0081.

- Refresher on the Nyquist frequency: Audio sampled at X kHz can only contain frequencies up to $\frac{X}{2}$ kHz
- Omitted 12000, 6000 which is mediumband and 24000/12000 which is superwideband
- 32000, 16000, nb cassette's are analog but they can't contain more than 18000hz
- Voiced signals are produced when the vocal cords vibrate during the pronounciation of a phoneme
- Unvoiced signals, by contrast, do not entail the use of the vocal cords
- f versus v, s versus z
- There is 96000, ultrasonic - supporting frequencies up to 48000 hz. not useful for humans

Lossy compression and psychoacoustics

Lossy versus *lossless* encoding is the primary issue in data compression[10]
Lossy compression sacrifices data for space savings

A question of fidelity: can we perceive the lost data? In the case of audio encoding, we need to consider psychoacoustics and perception.

> *The basis of lossy psychoacoustical compression methods is the omission of information from the audio signal so that it does not result in perceived difference.*[11]

https://wiki.xiph.org/Opus_Recommended_Settings says to prefer FLAC (lossless) for archival to avoid generation loss[12]

---

[10] Steven S. Skiena. *The Algorithm Design Manual*. London: Springer, 2008. DOI: 10.1007/978-1-84800-070-4.

[11] Péter Rucz. *Examination of lossy audio compression methods*. 2018. URL: https://last.hit.bme.hu/sites/default/files/documents/audio_labor_en.pdf.

[12] Frank Kurth. "An Audio Codec For Multiple Generations Compression Without Loss Of Perceptual Quality". In: (Aug. 2002).

Lossy compression and psychoacoustics

Lossy versus *lossless* encoding is the primary issue in data compression[10]
Lossy compression sacrifices data for space savings

A question of fidelity: can we perceive the lost data? In the case of audio encoding, we need to consider psychoacoustics and perception.
The basis of lossy psychoacoustical compression methods is the omission of information from the audio signal so that it does not result in perceived difference.[11]

https://wiki.xiph.org/Opus_Recommended_Settings says to prefer FLAC (lossless) for archival to avoid generation loss[12]

[10] Steven S. Skiena. *The Algorithm Design Manual*. London: Springer, 2008. DOI: 10.1007/978-1-84800-070-4.
[11] Péter Rucz. *Examination of lossy audio compression methods*. 2018. URL: https://last.hit.bme.hu/sites/default/files/documents/audio_labor_en.pdf.
[12] Frank Kurth. "An Audio Codec For Multiple Generations Compression Without Loss Of Perceptual Quality". In: (Aug. 2002).

Opus timeline

- 1993 — Xiph starts on music codec Vorbis
- 1998 — Vorbis development accelerated due to Fraunhofer Society MP3 royalties
- 2003 — Xiph starts on speech codec Speex
- 2007 — Xiph starts CELT starts as successor to Vorbis
- 2009 — Skype opens SILK protocol
- 2010 — Beginnings of Opus with hybrid CELT-SILK format
- 2012 — Xiph gets Opus standardized in IETF https://tools.ietf.org/html/rfc6716
- 2017 — Opus updated https://tools.ietf.org/html/rfc8251

Opus timeline
1993 Xiph starts on music codec Vorbis
1998 Vorbis development accelerated due to Fraunhofer Society MP3 royalties
2003 Xiph starts on speech codec Speex
2007 Xiph starts CELT starts as successor to Vorbis
2009 Skype opens SILK protocol
2010 Beginnings of Opus with hybrid CELT-SILK format
2012 Xiph gets Opus standardized in IETF https://tools.ietf.org/html/rfc6716
2017 Opus updated https://tools.ietf.org/html/rfc8251

- skype works with xiph and others to get something SILK-like in IETF - ultimately opus is the one who wins in IETF
- vorbis officially deprecated for opus
- speex is more softly discouraged. but not as strongly

Opus details – speech

The SILK half of Opus:[13]:

- *for speech, linear prediction techniques, such as Code-Excited Linear Prediction (CELP), code low frequencies more efficiently than transform (e.g., MDCT) domain techniques*
- Based on LPC (**L**inear **P**redictive **C**oding). Larynx emits simple signal (white noise or impulse train) through the articulatory system (throat, etc.) with coefficients. Sender sends articulatory coefficients, receiver recreates original sound by driving larynx signal through it[14]
- Computes LPC coefficients for *voiced* and *unvoiced* speech differently, using results of pitch analysis

[13] K. Vos et al. "Voice coding with opus". In: *135th Audio Engineering Society Convention 2013* (Jan. 2013), pp. 722–731. URL: https://jmvalin.ca/papers/aes135_opus_silk.pdf.

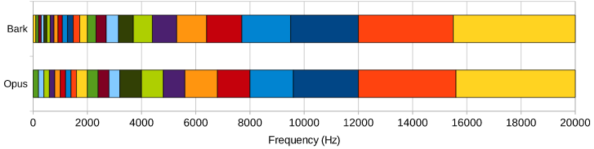
[14] Shahram Shirani. "Speech Compression". In: *ELEC 728, McMaster University Department of Electrical Engineering* (2010). URL: https://www.ece.mcmaster.ca/~shirani/multi10/speech%20compression.pdf.

- LPC is hugely popular in speech
- To make a musical analogy, it's like listening to a piano performance, transcribing it, sending over the score, and getting someone to play it on the other end... The result on the other end will be close to the original performance, but only a representation has been sent over.

Opus details – music

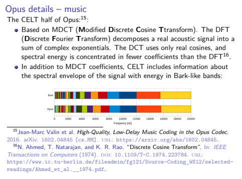
The CELT half of Opus:[15]:

- Based on MDCT (**M**odified **D**iscrete **C**osine **T**transform). The DFT (**D**iscrete **F**ourier **T**ransform) decomposes a real acoustic signal into a sum of complex exponentials. The DCT uses only real cosines, and spectral energy is concentrated in fewer coefficients than the DFT[16].
- In addition to MDCT coefficients, CELT includes information about the spectral envelope of the signal with energy in Bark-like bands:



[15] Jean-Marc Valin et al. *High-Quality, Low-Delay Music Coding in the Opus Codec*. 2016. arXiv: 1602.04845 [cs.MM]. URL: https://arxiv.org/abs/1602.04845.

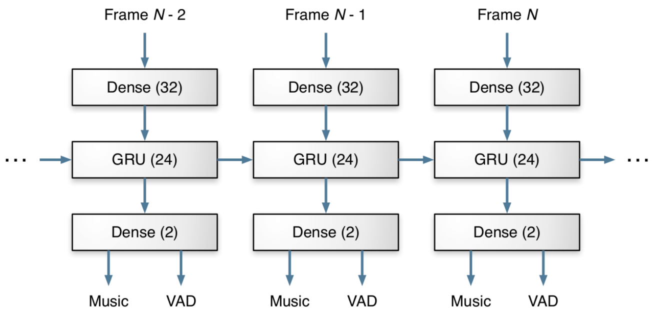
[16] N. Ahmed, T. Natarajan, and K. R. Rao. “Discrete Cosine Transform”. In: *IEEE Transactions on Computers* (1974). DOI: 10.1109/T-C.1974.223784. URL: https://www.ic.tu-berlin.de/fileadmin/fg121/Source-Coding_WS12/selected-readings/Ahmed_et_al.__1974.pdf.



- like lpc is dominant for speech, mdct is dominant for audio codecs
- lots of DCTs, details were too much to get into
- fewer coefficients means that the DCT is very used in compression

Auto-detect music and speech

Opus can automatically detect whether its input is speech or music, and choose the optimal encoding mode accordingly. GRU (**G**ated **R**ecurrent **U**nit) with just 4986 weights (that fit in less than 5 kB) and takes about 0.02% CPU to run in real-time[17]



[17] Jean-Marc Valin. *Opus 1.3 Released*. URL: https://jmvalin.ca/opus/opus-1.3/.

Auto-detect music and speech
Opus can automatically detect whether its input is speech or music, and choose the optimal encoding mode accordingly. GRU (Gated Recurrent Unit) with just 4986 weights (that fit in less than 5 kB) and takes about 0.02% CPU to run in real-time[17]
[17] Jean-Marc Valin. Opus 1.3 Released. URL: https://jmvalin.ca/opus/opus-1.3/.

Ambisonics, spatial audio

> *Opus 1.3 adds support for immersive audio using ambisonics that surrounds the listener in a full-sphere sound field. This is done through two new (soon to be RFC 8486) Ogg mapping families for Opus ambisonics. Unlike other multi-channel surround formats, ambisonics is independent of speaker layout. This allows for flexible speaker configurations and scalable audio to efficiently transmit 3D audio soundtracks.*[18]

https://tools.ietf.org/html/rfc8486

[18] Jean-Marc Valin. *Opus 1.3 Released.* URL: https://jmvalin.ca/opus/opus-1.3/.

Ambisonics, spatial audio
Opus 1.3 adds support for immersive audio using ambisonics that surrounds the listener in a full-sphere sound field. This is done through two new (soon to be RFC 8486) Ogg mapping families for Opus ambisonics. Unlike other multi-channel surround formats, ambisonics is independent of speaker layout. This allows for flexible speaker configurations and scalable audio to efficiently transmit 3D audio soundtracks.[18]
https://tools.ietf.org/html/rfc8486
[18] Jean-Marc Valin. Opus 1.3 Released. URL: https://jmvalin.ca/opus/opus-1.3/.

- jake mentioned
- note that the change is to Ogg, the container, to include multiple ambisonic channels of Opus-encoded audio

Sound samples

https://opus-codec.org/examples/

Sound samples

https://opus-codec.org/examples/