

Subjective quality evaluation of modern video codecs

R. Macoratti

Multimedia Video, laurea degree in Digital Communication, academic year 2014-2015

Table of contents

1. Introduction
2. Testing video sequences
 - a. Selected codecs
 - b. Video sources
3. Evaluation
 - a. Testing environment
 - b. Observers
 - c. Testing procedures
4. Results
5. References

1. Introduction

With the variety of video codecs developed in the time being, choosing one among them to compress general purpose video sequences is a common problem.

There are a plethora of open source specifications and implementations, as well as some proprietary ones. Some are backed by big organizations and others live as relatively small projects.

But the real question is what codec provides the capability of maintaining an acceptable quality compared to an uncompressed source, while still remaining into the limits of a modest physical size.

The aim of the test is not the objective quality or the correctness of the ideas behind the realization of the codec, but instead it is the subjective quality, which an observer can perceive in comparison to a reference video source. In the end, the codec (or codecs) that best resembles the overall quality of the source will score better and be considered the preferred choice.

2. Testing video sequences

2a. Selected codecs

Since the subject of testing were commonly used codecs, a handful of modern and popular video codecs has been chosen (see *Table 1*).

Among the selected ones, there are some created by the Joint Video Team^[1], that is a union team composed by MPEG and VQEG groups, which are MPEG-2, MPEG-4 and MPEG-H, and others created by Google^[2], such as VP8 and VP9, for their extraordinarily compression needs with YouTube. Other open source codecs, such as Theora^[3] and Dirac^[4], are present. Finally it also appears proprietary Apple's ProRes^[5], a format created specifically for archiving video works.

All selected codecs support, and also recommend, YUV colorspace with 4:2:0 chroma subsampling, with the exception of ProRes, which uses the same colorspace with 4:2:2 color subsampling.

For a better evaluation of codecs capabilities, two version have been created: a standard one and a high quality one. As it can be seen ahead, many video codecs perform better with more quality at their disposition.

Created by	Codec	Colorspace
ISO/IEC MPEG ITU-T VCEG	H.262/MPEG-2 Part 2	YUV 4:2:0 Progressive
	H.264/MPEG-4 AVC	YUV 4:2:0 Progressive
	H.265/MPEG-H Part 2	YUV 4:2:0 Progressive
Google	VP8	YUV 4:2:0 Progressive
	VP9	YUV 4:2:0 Progressive
Xiph.Org	OGG Theora	YUV 4:2:0 Progressive
Apple	ProRes 422	YUV 4:2:2 Progressive
BBC	Dirac/Schroedinger	YUV 4:2:0 Progressive

Table 1. Codecs used in test.

2b. Video sources

As uncompressed video sources, testing files provided by the Video Quality Experts Group have been used. These are five ten second scenes taken from Swedish television, 1920 pixels wide and 1080 pixels high, 50 frames per seconds. Every frame is provided individually in the form an uncompressed image. The final testing footage is a juxtaposition of the five smaller sequences.

The open source tool `ffmpeg`^[6] has been used to convert and compress video sequences.

For the standard quality video sequences a medium quality preset has been maintained and framerate has been reduced to 25 frames per second. On the other hand, for the high quality video sequences a higher preset has been set. Framerate, size, colorspace, and chrome subsampling have been left unchanged.

3. Evaluation

2a. Testing environment

The display used in the experiment is 21 inches, 96 dpi, BARCO prototype. It is 43 cm wide and 33 cm high, with a resolution of 1600×1200 pixels and a refresh rate of 60 Hz (see *Image 1*).

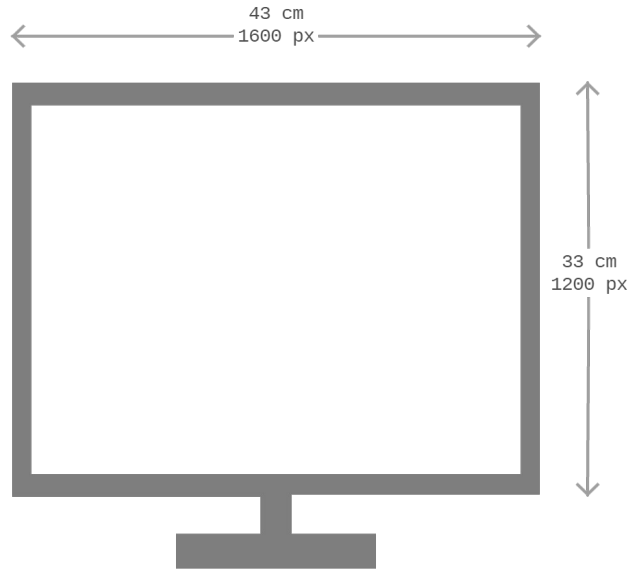


Image 1. Monitor schema.

Observers sat on an adjustable office chair, their eyes middle point vertically and horizontally aligned with the center of the screen.

The distance of the chair from the monitor has been calculated using the $3 \times \text{monitor}_{\text{height}}$ formula, resulting in approximately 1 meter of distance.

2b. Observers

Observers were young boys and girls, all in the 20 – 30 years old segment. No particular experience in the video evaluation was required.

2c. Testing procedures

A simple application has been developed, which sequentially gives the option to display before the uncompressed source sequence and right after the compressed footage for every chosen codec. The

testers had the possibility to view both the source and the compression again, how many times they needed.

In order to provide a good evaluation base and help the observers give a unadulterated opinion, a Likert-type scale for every testing video sequence has been given to them. The selectable scores are shown in *Table 2*.

After each pair of sequences source-compression has been shown the observer, he/she can proceed to evaluate the discrepancy in quality between the two, marking the preferred value of the Likert scale.

Quality	Value
Bad	1
Poor	2
Fair	3
Good	4
Excellent	5

Table 2. Likert scale.

When all the questionnaire data has been acquired, some statistical calculation over the results has been done, in order to extract a rank. The weight of the evaluation of a codec is given by both the quality in comparison to source and the size of the compressed video file. For what concerns quality, a Mean Opinion Score (refer to Formula 1), abbreviated M.O.S., of all scores given by the observers is calculated. The video file size is immediately available.

$$MOS_k = \sum_{i=1}^{observers} \frac{mark_{k,i}}{observers}$$

k is the sequence number of current codec
 $mark_{k,i}$ is score given for codec k by observer i
 $observers$ is the overall number of observers

Formula 1. Mean Opinion Score formula.

4. Results

Results have been subdivided by quality and then by what is being evaluated. For the standard quality rank you can refer to *Table 3* and *Table 4*, and for the high quality rank *Table 5* and *Table 6* are provided.

Codec	M.O.S.	Codec	Size (MB)
H.264	3.4	ProRes 422	120.67
VP9	2.9	Dirac	47.79
VP8	2.7	Theora	44.62
ProRes 422	2.4	H.264	44.37
H.265	2.3	VP8	39.53
MPEG-2	2.1	MPEG-2	35.60
Theora	1.7	VP9	27.28
Dirac	1.5	H.265	16.51

Table 3. Standard quality MOS rank

Table 4. Standard quality size rank

As it can be seen, in the standard range of quality, H.264 is the best option. The gap is very noticeable, half a point from the second. A good choice is also VP8 or, a better one, VP9, if the size of video files produced is considered. H.264 scores only fifth in this rank, whilst VP8 is fourth and VP9 second.

The relatively new technology, H.265, and the older one, MPEG-2, score quite well in size rank, first and third, but for standard quality video files their paired quality is not as good. ProRes 422, due to its nature of archive and exchange video format, scores quite well in MOS, but it is incredibly heavy in size. Theora and Dirac are the worst options in both cases.

Codec	M.O.S.	Codec	Size (MB)
H.264	4.7	ProRes 422	561.74
H.265	4.6	Dirac	298.80
VP8	4.1	MPEG-2	180.95
VP9	3.8	Theora	166.64
ProRes 422	3.8	H.264	113.41
MPEG-2	3.7	VP8	105.43
Dirac	3.3	VP9	79.24
Theora	3.2	H.265	53.14

Table 5. High quality MOS rank

Table 6. High quality size rank

Surprisingly, H.264 scores the best also in the high quality segment. It cannot be said the same for its

file size, in fact it scores only fifth in the other rank. H.265 scores quite as well as H.264, only one tenth of point less, but it is also the lightest of all video files. It surely is the best choice for a high quality video sequence. VP8 and VP9 are about one point behind their MPEG counterpart, but they score second and third in the size rank. ProRes 422 reaches an average mark, but, as in the previous example, it is the heaviest. MPEG-2 shows its limits being very big in size and also obtaining a poor mark. As happened for the standard quality sequences, Theora and Dirac are a step behind the others, both in MOS and in size ranks.

In the end, H.264, H.265, VP8 and VP9 are the best options when it comes to choose a codec. They produce remarkable results on the quality side, adapting both to standard and high quality range, but also remain restricted in a small file size, especially in the case of H.265, marking tiny file sizes.

A possible explanation of this result is hidden in how this codec have come to exist. First of all the developers groups behind all this four is composed by teams of experts, who have gathered all the best compression techniques, i.e. motion detection and discrete cosine transform, when for example Theora does not use such technologies. In the second place, they are backed, intellectually and economically, by organizations such as Google or the MPEG/VQEG consortium, whilst projects like Theora or Dirac are tied to open source contribution.

5. References

- [1]. “Joint Video Team.” ITU. Web. 02 Oct. 2015.
<<http://www.itu.int/en/ITU-/studygroups/com16/video/Pages/jvt.aspx>>.
- [2]. “WebM: An Open Web Media Project.” The WebM Project. Web. 02 Oct. 2015.
<<http://www.webmproject.org/>>.
- [3]. “Theora Video Compression.” *Theora.org*. Web. 02 Oct. 2015. <<http://www.theora.org/>>.
- [4]. “Dirac” *BBC News*. BBC. Web. 02 Oct. 2015.
<<http://www.bbc.co.uk/opensource/projects/dirac/>>.
- [5]. “About ProRes” *Apple.com*. Web. 02 Oct. 2015. <<https://support.apple.com/en-us/HT202410>>.
- [6]. “FFmpeg.” *FFmpeg*. Web. 02 Oct. 2015. <<https://www.ffmpeg.org/>>.

- “VQEG FTP”. Web. 2 Oct. 2015. <ftp://vqeg.its.bldrdoc.gov/HDTV/SVT_MultiFormat/>.
- “P.913: Methods for the Subjective Assessment of Video Quality, Audio Quality and Audiovisual Quality of Internet Video and Distribution Quality Television in Any Environment”. Web. 02 Oct. 2015. <<https://www.itu.int/rec/T-REC-P.913/en>>.
- “BT.500: Methodology for the Subjective Assessment of the Quality of Television Pictures”. Web. 02 Oct. 2015. <<https://www.itu.int/rec/R-REC-BT.500/en>>.
- Vatolin, Dmitriy, Alexander Parshin, Oleg Petrov, and Artem Titarenko. *MSU Subjective Comparison of Modern Video Codecs* (2006): 1-35. Web. 2 Oct. 2015.
<http://www.compression.ru/video/codec_comparison/pdf/msu_subjective_codecs_comparison_en.pdf>.
- “YUV Pixel Formats”. Web. 02 Oct. 2015. <<http://www.fourcc.org/yuv.php>>.