FORUM SBTVD
BRAZILIAN DIGITAL TERRESTRIAL TV FORUM

COM

RNP
ORGANIZAÇÃO SOCIAL DO MCTI

UnB

TV 2.5 Project - Real-Time Video Coding Subjective Quality Assessment

---

13 March 2024

Brazilian Digital Terrestrial Television System Forum

Brazilian Ministry of Communications

Brazilian National Education and Research Network (RNP, *Rede Nacional de Ensino e Pesquisa*)

University of Brasília (UnB, *Universidade de Brasília*)

Table of Contents

1 Introduction

The SBTVD Forum was created by the Brazilian Presidential Decree # 5 820 / 2006, to advise the Brazilian Government regarding policies and technical issues related to the approval of technical innovations, specifications, development, and implementation of the Brazilian Digital Terrestrial Television System (SBTVD). The SBTVD Forum is composed of representatives of the broadcasting, academia, transmission, reception, and software industry sectors, and has the participation of Brazilian Government representatives as non-voting members.

Free-to-air terrestrial television is the main audiovisual distribution platform in Brazil, covering almost all Brazilian households and used in more than 70% of them. It secures to most of the Brazilian population a free-of-charge, universal and democratic access to information and entertainment, made by Brazilians for Brazilians. It is, therefore, an important social cohesion, and national/cultural identity factor.

For its first generation Digital Terrestrial Television system, after thorough testing and careful studies, the Brazilian Government adopted in June 2006 the ISDB-T standard, incorporating technological innovations that were deemed relevant, such as MPEG-4 AVC (H.264) video coding, MPEG-4 AAC audio coding, an appropriate closed caption character set for the Brazilian Portuguese, and a new middleware for interactive applications (Ginga).

The SBTVD Forum developed the first SBTVD standards, which were published in 2007, allowing the official opening of transmissions in that same year. Since then, the standards have been continuously revised and updated by the Forum. The technological innovations proposed by Brazil were incorporated into the International ISDB-T standard, which is currently adopted by 20 countries.

In 2016, Brazil started a safe and gradual analog TV switch-off process that was designed to ensure that no one would be deprived of the terrestrial free-to-air TV. The process was divided into three stages.

In the first stage (2016 to 2019), the analog television switch-off was performed in all the state capitals, metropolitan areas, and other areas where it was required to release the 700 MHz band. During the first stage, 1 366 cities in 47 different clusters were impacted, accounting for nearly 128 million people (62% of the population). More than 12 million Digital TV reception kits were distributed to low-income families. The analog switch-off had no significant impact on the free-to-air terrestrial TV audience.

In the second stage (2019 to 2023), the analog television switch-off was performed in 2 939 municipalities that, on 01 September 2020, had no digital terrestrial television stations in operation. In this stage, 1 565 municipalities were selected for a government-sponsored DTTB expansion program called "Digitaliza Brasil", which provided shared DTTB transmission infrastructure in each municipality and distributed additional 1.5 million Digital TV reception kits for low-income families.

In the third stage (2024 to 2025), the analog television switch-off will be performed in the remaining 1 265 municipalities up to June 2025. After the implementation of Digital Terrestrial Television Brazil adopted an industrial policy that determined that all flat-panel TVs manufactured from 2012 must have an integrated Digital TV receiver and from 2013 no more CRT TVs were manufactured. Therefore, it is anticipated, based on the expected product lifetimes, that after 2023 Brazil will have nearly all its TV sets already equipped with an integrated Digital TV receiver, thus facilitating the analog television switch-off without additional Digital TV reception kits distribution.

As the Brazilian digital television switch-over began, the SBTVD Forum started considering the next steps for the evolution of Brazilian Television. The analog TV (that we conventionally call "TV 1.0"), which started in Brazil in 1950, was black and white with monophonic sound. Then, some backward-compatible improvements (that we conventionally call "TV 1.5"), such as color (in the 1970s), stereo sound and closed caption (in the 1980s) were added to it. In 2007, the first generation of Digital Terrestrial Television (that we conventionally call "TV 2.0") was introduced in Brazil, bringing high-definition video, surround sound, mobile reception, and interactivity. Since then, the technological landscape has changed a lot. The rhythm of development and introduction of innovations is increasingly accelerated. These innovations create new consumption habits and increase the expectations of technological service users regarding the quality and convenience of these services. Since the introduction of SBTVD, new immersive audio and video formats have emerged, and are already present in the new TV sets available in the market. The TV sets currently available have resolution and contrast greater than those supported in the first-generation SBTVD standard. That is the opposite of the market situation when Digital TV was launched in Brazil, as the HDTV sets were not as common at the time. The availability and the speed of Internet access in Brazil, especially in metropolitan areas, increased significantly, enabling the consumption of on-demand audiovisual content. This connectivity is already in use by TV sets (Smart TVs) and by broadcasters' Over-The-Top (OTT) offers. However, in the first generation SBTVD standard, there was not an integration between the broadcasting service and the Internet content offer. Furthermore, new techniques for signal coding, transport, and modulation were also developed, allowing greater efficiency in audiovisual transmission. Many Digital Terrestrial Television systems have also been evolving, including in this evolution not only enhancements in quality and efficiency but also new convergent services between broadcast and broadband. Based on this technological landscape, the SBTVD Forum recognized the necessity to evolve the SBTVD. It also acknowledged that changing the physical layer, the transport layer, and/or audiovisual coding would not be backward-compatible. Nevertheless, the transition to a new generation of Digital Terrestrial Television is a long process, based on the investments required for both broadcasters and consumers and the expected life span of TV transmitters and receivers. It was, therefore, deemed necessary to increase the lifespan of the existing Digital Terrestrial Television system as much as possible through a backward-compatible evolution (a project we called "TV 2.5") and to start the development of the next generation Digital Terrestrial Television system (the project we called "TV 3.0").

The "TV 2.5" project comprised two aspects: broadcast-broadband integration and audiovisual quality. The first aspect involved the development of a new receiver profile for the middleware Ginga (receiver profile D, a.k.a. "DTV Play"), addressing use cases such as on-demand video, synchronized companion device, audiovisual enhancement over the Internet, and targeted content. The second aspect was addressed through the introduction of three new optional immersive audio codecs (MPEG-H Audio, E-AC-3 JOC, and AC-4) while retaining MPEG-4 AAC main audio for backward compatibility, and through the introduction of two new optional HDR video formats (SL-HDR1 dynamic metadata and HLG "preferred transfer characteristics" signaling) while keeping MPEG-4 AVC (H.264) / 8-bit / BT.709 / 1 080i for backward-compatibility. The revision of the SBTVD standards containing both "TV 2.5" aspects has already been published (available at https://forumsbtvd.org.br/legislacao-e-normas-tecnicas/normas-tecnicas-da-tv-digital/english/).

For the "TV 3.0" project, the SBTVD Forum, after agreeing on its requirements (use cases and corresponding technical specifications), decided to release a Call for Proposals (available at https://forumsbtvd.org.br/wp-content/uploads/2020/07/SBTVDTV-3-0-CfP.pdf) for any interested organization to submit its proposed candidate technologies for any of the system components or sub-components. Considering the results available of the TV 3.0 project testing and evaluation, as well as market and intellectual property aspects of the candidate technologies, the SBTVD Forum already selected technologies for most of the system components and subcomponents (see https://forumsbtvd.org.br/tv3_0/#panel-phase2). Among these, Low Complexity Enhancement Video Coding (MPEG 5 Part 2 – LCEVC) was selected as the scalable video coding technology.

In the context of TV 3.0, LCEVC will be used in conjunction with Versatile Video Coding (MPEG-I VVC / H.266) as the base layer codec. But LCEVC is agnostic to the base layer video codec and could also be employed in conjunction with MPEG-4 AVC (H.264), used in TV 2.0/2.5, to provide a backward-compatible enhancement. Furthermore, LCEVC could be combined with the TV 2.5 optional HDR formats. Considering this potential application of LCEVC, the SBTVD Forum decided to investigate the possibility of including it as another optional backward-compatible video enhancement technology into TV 2.5 specifications, potentially enabling the scaling of a 1 080i / 29.97fps / 8-bit video (maximum quality possible with H.264 under TV 2.0/2.5 specifications) to 1 080p / 59.94fps / 10-bit on compatible receivers, with a relatively small overhead.

To evaluate the possible quality gain of using LCEVC in TV 2.5, as well as the possible quality loss in receivers not compatible with LCEVC as a result of reducing the bitrate of the H.264 base layer to include the LCEVC enhancement layer, the SBTVD Forum decided to perform a real-time video coding subjective quality assessment, comparing the maximum quality achievable with real-time video coding for TV 2.0 (H.264 / 1 080i / 29.97fps / 8-bit / SDR @ 14 Mbps), and TV 2.5 (H.264 / 1 080i / 29.97fps / 8-bit / SL-HDR1 @ 14 Mbps), with what would be the resulting quality if LCEVC was included in TV 2.5 specifications, both on LCEVC-compatible receivers (H.264 / 1 080i / 29.97fps / 8-bit / SL-HDR1 base layer + LCEVC / 1 080p / 59.94fps / 10-bit enhancement layer, with total bitrate of 14 Mbps) and on TV 2.0 receivers (decoding only the H.264 / 1 080i / 29.97fps / 8-bit base layer - that would be using a bitrate somewhat lower than 14 Mbps - in SDR, disregarding the SL-HDR1 dynamic metadata and the LCEVC enhancement layer).

The TV 2.5 Real-Time Video Coding Subjective Quality Assessment was carried out by the University of Brasília from June 2023 to February 2024, in parallel with the TV 3.0 Phase 3 Real-Time Video Coding Subjective Quality Assessment, under the coordination of the SBTVD Forum and funded by the Brazilian Ministry of Communications through the Brazilian National Education and Research Network (RNP, *Rede Nacional de Ensino e Pesquisa*).

This document contains the results of the TV 2.5 Real-Time Video Coding Subjective Quality Assessment to assist the SBTVD Forum's decision-making process on the possible inclusion of LCEVC as an additional optional backward-compatible video enhancement technology into TV 2.5 specifications.

2 Research Team

This report was developed by the following team of the University of Brasília:

Eduardo Peixoto Fernandes da Silva, Universidade de Brasília - http://lattes.cnpq.br/7720922538480924

Pedro Garcia Freitas, Universidade de Brasília - http://lattes.cnpq.br/0669244312752630

Mylene Christine Queiroz Farias, Universidade de Brasília - http://lattes.cnpq.br/4465619366143200

José Edil Guimarães de Medeiros, Universidade de Brasília - http://lattes.cnpq.br/2307490672260330

Fernando Cardoso Guimarães, Universidade de Brasília - https://lattes.cnpq.br/8587888778545035

Felipe Sidrone da Silva Dionísio, Universidade de Brasília - http://lattes.cnpq.br/4451878253778940

Gabriel Correia Lima da Cunha e Menezes, Universidade de Brasília - http://lattes.cnpq.br/0543326076269968

Victor Orfeu Melo, Universidade de Brasília - http://lattes.cnpq.br/2846252961723018

We also had the collaboration of SBTVD Forum members, in particular:

Luiz Fausto – Chair, Technical Module, SBTVD Forum

Carlos Cosme – Chair, Audio & Video Coding Workgroup, SBTVD Forum

2.1 Partners

We want to thank the companies that provided laboratory equipment for the tests:

- Globo
- MainConcept
- V-Nova
- Philips / InterDigital

3 Glossary

1 080i Interlaced video with a resolution of 1 920 x 1 080 pixels

1 080p Progressive video with a resolution of 1 920 x 1 080 pixels

4K Progressive video with a resolution of 3 840 x 2 160 pixels

AAC Advanced Audio Coding

AC-4 Audio Compression 4

AVC Advanced Video Coding

BT.2020 Recommendation ITU-R BT.2020: Parameter values for ultra-high-definition television systems for production and international programme exchange

BT.709	Recommendation ITU-R BT.709: Parameter values for the HDTV standards for production and international programme exchange
CRF	Constant Rate Factor
CRT	Cathode Ray Tube
E-AC-3 JOC	Enhanced Audio Compression 3 Joint Object Coding
fps	frames per second
H.264	Recommendation ITU-T H.264: Advanced video coding for generic audiovisual services
H.266	Recommendation ITU-T H.266: Versatile video coding
HDMI	High-Definition Multimedia Interface
HDR	High Dynamic Range
HDR10	High Dynamic Range 10-bit
HDTV	High-Definition Television
ISDB-T	Integrated Services Digital Broadcasting-Terrestrial
ITU-R	International Telecommunications Union - Radiocommunication sector
ITU-T	International Telecommunications Union - Telecommunication standardization sector
kbps	kilobits per second
LCEVC	Low Complexity Enhancement Video Coding
MaxCLL	Maximum Content Light Level
MaxFALL	Maximum Frame Average Light Level
Mbps	Megabits per second
MOS	Mean Opinion Score

MPEG	Motion Picture Experts Group
OLED	Organic Light-Emitting-Diode
OTT	Over-The-Top
RNP	*Rede Nacional de Ensino e Pesquisa* (Brazilian National Education and Research Network)
SBTVD	*Sistema Brasileiro de Televisão Digital* (Brazilian Digital Television System)
SDI	Serial Digital Interface
SDR	Standard Dynamic Range
SL-HDR	Single Layer HDR
SMPTE	Society of Motion Picture and Television Engineers
TV	TeleVision
VUT	Video Under Test
VVC	Versatile Video Coding

4 TV 2.5

TV 2.5 should offer backwards compatibility with the existing TV 2.0 system, while still enhancing the image quality, particularly regarding dynamic range. Since TV 2.0 is already deployed and it is being used by millions of users, TV 2.5 does not involve changes in the physical layer of the broadcast stream. In addition, the video stream transmitted in TV 2.5 must be compatible with a TV 2.0 decoder, even if the decoder cannot decode the video stream in full quality. The situation is similar to the TV 1.5 system, which added color images to the black-and-white stream that was transmitted in the TV 1.0 (the old analog TV). At the time of deployment, most homes only had a black-and-white TV, which still worked (and continued working for decades) even though the stream transmitted had color components as well, which were ignored by TV 1.0 decoders.

The current TV 2.0 system uses a 14 Mbps channel, and the video stream is encoded with H.264/AVC at resolution 1 080i (1 920 x 1 080 at 29.97 fps) and using a BT.709 SDR color transfer function. Therefore, the main requisites for a TV 2.5 stream are:

1. The complete bitstream must be constrained to 14 Mbps;
2. A base layer compatible with H.264/AVC 1 080i SDR must be available; and
3. This base layer should offer a good quality compared to the existing TV 2.0 system.

Two different alternatives were proposed to the forum. The first alternative comprises using the H.264/AVC codec, which is already being used by TV 2.0, and adding SL-HDR1 metadata to the bitstream. A TV 2.0 decoder can just ignore the added metadata, which uses a negligent amount of bitrate, and decode a compatible TV 2.0 video. A TV2.5 decoder can use these metadata to reconstruct HDR from SDR, which significantly improves the dynamic range. This alternative will be referred to in this document as “H.264 + SL-HDR1” and it is already standardized in ABNT NBR 15602-1 since 2020.

The second alternative uses the H.264/AVC+SL-HDR1 codec with an added layer encoded by the LCEVC codec. This added layer will be ignored by a legacy TV 2.0 decoder, which would decode just the H.264/AVC base layer. A TV 2.5 decoder, however, would also decode the LCEVC enhancement layer, which improves the frame rate (from 59.94 fields per second to 59.94 frames per second), the bit depth (from 8 to 10-bit) and the dynamic range through the use of a SL-HDR1. However, since the total bitrate should still be transmitted through the TV 2.0 channel, the total bitrate of the H.264/AVC base layer and the LCEVC enhancement layer should be constrained to 14 Mbps. The bitrate added by SL-HDR1 is negligible. This alternative will be referred to in this document as “H.264 + LCEVC + SL-HDR1”.

5 TV 2.5 Real-Time Video Coding Subjective Quality Assessment Results

The scope of the current tests is to evaluate the overall subjective quality of each version when working in a real-time encoding and decoding scenario. Only the overall quality of the video streams is being evaluated - participants are not asked to detail why they have rated a video in a certain way.

Subsection 5.1 introduces the experimental setup arranged for the subjective quality evaluation experiments and details the experimental protocol each subject is submitted during the experiments, subsection 5.2 describes the analysis protocols executed to understand the data generated during the experiments, and subsection 5.3 detail the definition of the specific tests the team performed, their results and analysis.

5.1 Experimental Setup

This section details the contents used, the laboratory conditions and the interface of the experiments.

5.1.1 Experimental Stimuli

The video dataset used for the experiments was prepared by the SBTVD Forum specifically for the tests. It comprises five different video streams, mostly camera-generated content with some computer-generated graphical overlays and high contrast colorful scenes, each one with a few scene cuts. Some content contains smooth gradients between similar colors to assist in detecting possible banding artifacts. Some of the content have scenes commonly transmitted through Brazilian television, such as football matches and carnival scenes. The scenes were chosen due to their diverse range of spatial and temporal information. The details of the video streams are shown in Table 1.

The spatial and temporal information for the scenes is shown in Figure 1, while Figure 2 shows the file size for each content when encoding with libx265 at CRF = 28. This is used as an indicator of the complexity of encoding each content. It can be seen that content globo01 needs almost four times the bitrate compared to content philips03 when using the same Quantization Parameter. A short description and a sample frame for each content are shown in Table 2.

Table 1. Details of the contents used for the tests

Resolution (horizontal x vertical)	3 840 x 2 160
Aspect ratio	16:9 (square pixels)
Sampling ratio	YC_BC_R 4:2:2
Bit depth	10 bits
Frame rate	59.94 fps
Scan	Progressive
Transfer characteristics	Perceptual Quantization (max 1 000 cd/m^2)
HDR Metadata	HDR10 (SMPTE ST 2086, MaxFALL and MaxCLL)
Color Gamut	Rec. ITU-R BT.2020
Duration	30 s
File format	Uncompressed MOV file

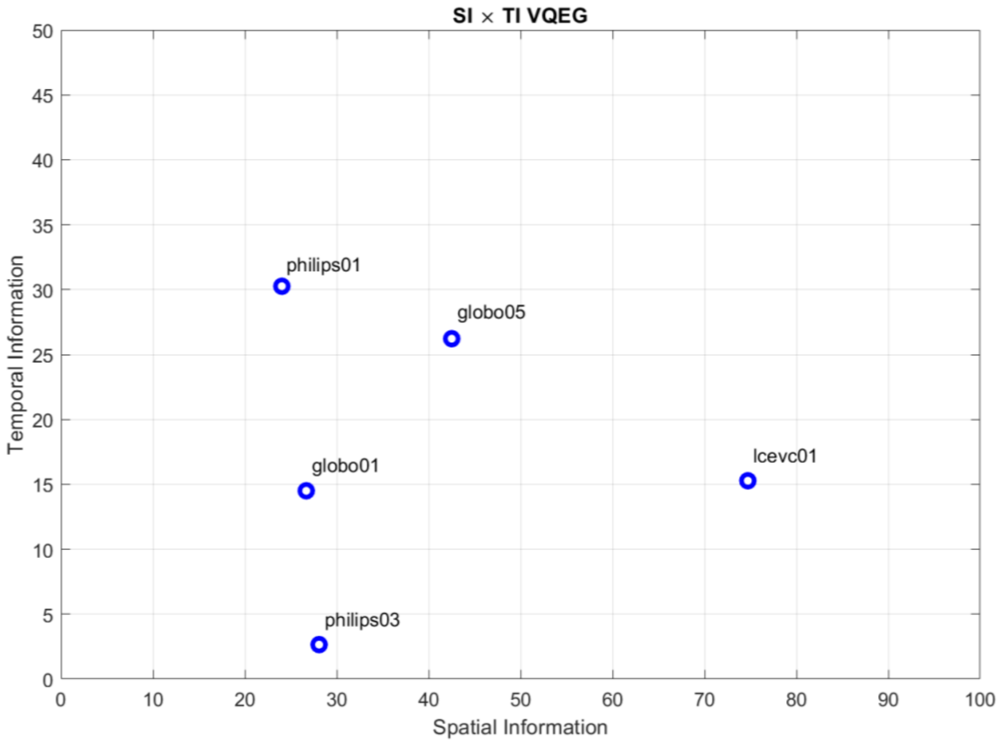


Figure 1. Spatial and Temporal Information for the test contents used

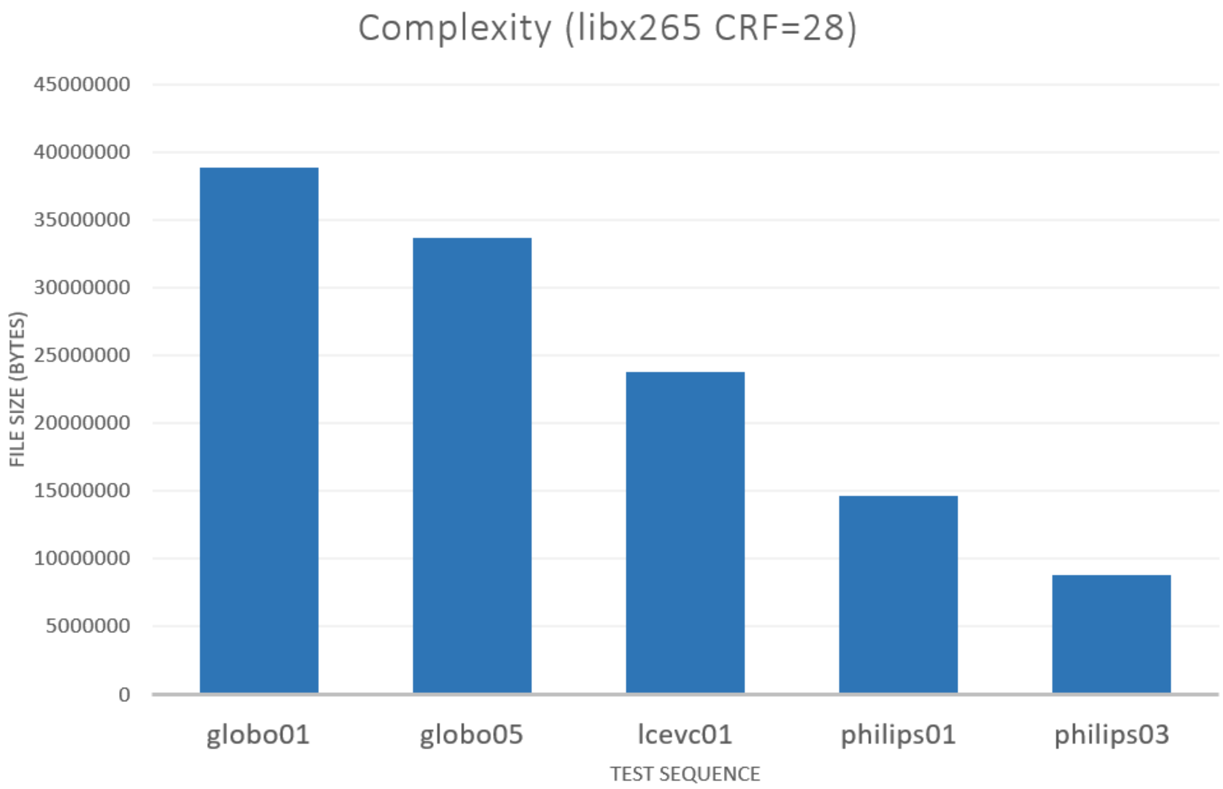
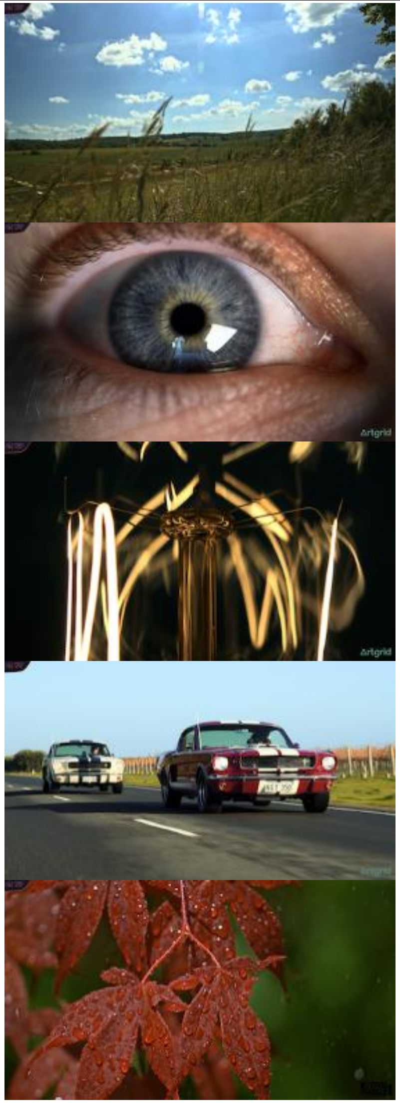


Figure 2. Content file size when encoded with libx265 at CRF=28

Table 2 - Short description of the contents used for the tests

globo01		Excerpts from the 2022 Rio de Janeiro samba school parade. This is a high-movement colorful scene. There is a lot of content with high spatial information, including many human faces close to the camera.

globo05		Excerpts from the 2019 Conmebol Copa America (South America Football Championship) final match. There is a lot of high-speed movement and a large number of people moving around in the background.
lcevc01		A collection of short clips of different landscapes with direct sunlight and graphics overlay, including a few sunrise and sunset scenes, and a sunflower field with bees. There are also some Formula 1 results graphics overlay in the left part of the screen which contributes to a lot of high-frequency content.

philips01		A collection of short clips, including a wheat field, a close-up of a human eye, a close-up of an incandescent lamp, fast-moving vintage cars on a highway, and some bright red leaves with raindrops.

philips03		A collection of short clips, including images of a canal in the Netherlands, a close-up of a teenage girl's face, and a pier with moving water.

5.1.2 Experimental Conditions

The experiments were performed on a 465cm x 490cm laboratory located on the first floor of the SG-11 building at the University of Brasília Darcy Ribeiro Campus. The lab was available exclusively for the TV 3.0 Phase 3 experiments.

Figure 3 shows the relative position of the subject in relation to the screen. In this figure, H is the vertical size of the screen, which is 81 cm. Therefore, the viewing distance from the screen was around 131 cm.

The equipment setup included two sets of the following equipment:

1. Mac Studio (2022 model) desktop equipped with an Apple M1 Ultra and 128GB of RAM running MacOS Ventura 13.2.1 operating system and a Blackmagic Design Media Express 3.8.1 software;
2. Blackmagic Design UltraStudio 4k Mini; and
3. LG 65" OLED TV set model OLED65C2PSA.CWZQLJZ (firmware version 03.33.85) with a screen area of approximately 81cm x 142cm.

The LG TVs are consumer-grade products; thus, a limited set of configurations was available. Table 3 shows the configurations used for both TV sets. Table 4 shows a summary of the tests that were performed.

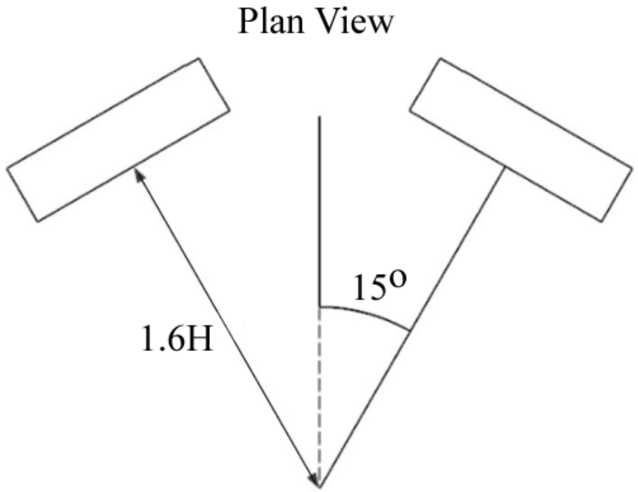


Figure 3. Laboratory set-up for the subjective tests

Table 3. Screen settings

	Parameter	Setting
Brightness	OLED Pixel Brightness	80
	Adjust Contrast	95
	Black Level	50
	Auto Dynamic Contrast	Medium
	Dynamic Tone Mapping[1]	On
	Peak Brightness	High
	Gamma	2.2
	Video Range	Auto
	Motion Eye Care	Off

[1] “Dynamic mode” on is automatically converting an SDR input signal to HDR based on a proprietary manufacturer algorithm (here LG high end TV set) for the reference and VUT 3.2.1. When an HDR signal is input, then no such processing is applied (VUT 3.1 and VUT 3.2.2) by the TV set.

Color	Color Depth	55
	Tint	0
	Color Gamut	Auto Detect
Color/Fine Tune	Color Adjustment	Off
Balance	Color Temperature	0
	Method	2 Points
	Point	High
	Red	0
	Green	0
	Blue	0
Clarity	Adjust Sharpness	25
	Super Resolution	Off
	Noise Reduction	Off
	MPEG Noise Reduction	Off
	Smooth Gradation	Off
	Real Cinema	Off
	TruMotion	Off

5.1.3 Experimental Protocol

We have performed one test with 30 subjects. For each test we have used 5 different contents, provided by the forum and detailed in Section 5.1.1. For all tests, the reference video used is the H.264/AVC bitstream that is being used today for TV 2.0. This is also the base layer for the "H.264 + SL-HDR1" alternative. We have then defined three videos under test (VUTs):

- VUT 3.1 - The H.264 stream encoded with the SL-HDR1 metadata, called "H.264 + SL-HDR1" here (compatible only with a current TV 2.5 decoder).
- VUT 3.2.1 - The base layer for "H.264+LCEVC+SL-HDR1" (compatible with a TV 2.0 decoder). For VUT 3.2.1, only the H.264 base layer is decoded.
- VUT 3.2.2 - The enhancement layer for "H.264+LCEVC+SL-HDR1" (compatible only with a potential future TV 2.5 decoder that also includes LCEVC and SL-HDR1).

The experimental session comprises two distinct segments: training and subjective evaluation. In the initial training phase, the conductor outlines the experiment's objectives and procedures, gathers demographic information from participants, and conducts a training session to get them acquainted with the experimental tasks. Before beginning the experiment, subjects are asked to read a 4-pages document and fill in some information. The first two pages of the document contain an explanation of the experimental protocol and the risks associated with it. Subsequently, participants are requested to sign a consent form and provide details about their age, gender, technical expertise, and familiarity with visual impairments.

In the training session, each subject is shown 3 video samples with varying quality levels that correspond to different encoding parameters. The goal of the training session is to provide the subjects an opportunity to familiarize themselves with the experimental task and to understand the quality rating scale. The participants are not instructed on what they should look at, and they are never asked to justify the rates given. After the training, the experimenter may answer questions from the subject. Then, the experimenter takes note of the VUT information as well as the pseudo random sequence number (used to analyze the data) and starts the subjective evaluation phase.

During the subjective evaluation phase, participants assess a series of 15 test video sequences, each lasting around 30 seconds. To record their quality scores, participants utilize paper forms and pens. The evaluation form contains a 7-level quality scale, as depicted in Table 4, for each test sequence to be evaluated. Prior to the presentation of each video, a 3-second preparation countdown is displayed on a gray screen. Following each video, a gray screen prompts participants to complete the evaluation form. The pace at which the succession of the video clips is presented is manually controlled by the experimenter, who adjusts the experiment's speed based on the capacities of the participants.

For all tests, the reference was always displayed on the left screen, and this was known by the participant. For each VUT, we have used three different pseudo-random sequences to display the 15 videos of each session.

Table 4. Subjective evaluation scale

Textual Scale (in Portuguese)	Textual Scale	Numeric Scale
Muito Pior	Much Worse	-3
Pior	Worse	-2
Pouco Pior	Slightly Worse	-1
Igual	Equal	0
Pouco Melhor	Slightly Better	1
Melhor	Better	2
Muito Melhor	Much Better	3

5.2 Data Analysis Protocol

This section details how the experimental data gathered during the experiments is processed. In the first part, the subjective scores for each VUT are processed through an outlier removal protocol, detailed next.

5.2.1 Outlier Removal and Mean Opinion Score determination

In conducting the statistical analysis of the experimental data for this study, we used the following protocol to ensure the reliability and validity of the obtained results. The protocol involves several key steps, starting with the identification and removal of outliers. Preliminary data inspection used statistical measures, such as the computation of Z-scores, to flag potential outliers. In this work, outliers are defined as data points lying outside predetermined thresholds and are an indicator of measurement errors. The outlier detection step was based on Recommendation ITU-R BT.500. Following the recommendation, the thresholds were used as 2 standard deviations. The subsequent removal of these outliers was performed systematically to prevent their undue influence on the subsequent statistical analysis phases. Following outlier removal, a critical step of the statistical analysis is the transformation of the collected experimental scores. Then, the results were further analyzed by finding the mean opinion scores (MOS) and confidence intervals (CI), as described in Recommendation ITU-T P.1401.

5.3 VUTs

The details on how each VUT is encoded are described in this section. It also details the characteristics of each video stream, such as resolution, color transfer function, frame rate, among others.

5.3.1 H.264 Reference Video

The reference video was encoded with an H.264 encoder, and it has the same characteristics as the videos being broadcasted today in Brazilian Television. The general workflow used to encode this video is shown in Figure 4. The original 4K video is transmitted to a cobalt pre-processor through SDI connection. The cobalt pre-processor downscales the video to 1 080i in real-time and transmits it to the encoder also using an SDI connection. The encoder encodes the video using H.264/AVC with the SL-HDR1 metadata. The bitstream is then transmitted in real time through a local network. The decoder, which is the MX Player app running on a NVIDIA Shield, decodes the bitstream, ignoring the SL-HDR1 metadata, and displays the decoded video through an HDMI output. This output is then recorded by our video recorder (a Blackmagic device), which is then stored for display during the tests. All steps are performed in tandem.

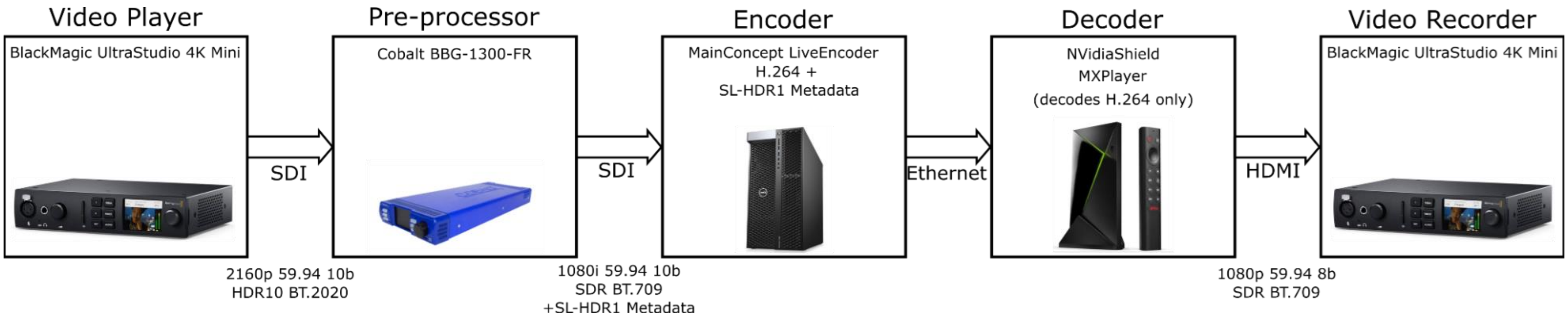


Figure 4. Workflow for the reference video

Note that the NVIDIA Shield decodes the 1 080i video and its built-in format converter converts the signal to 1 080p before forwarding over HDMI.

The complete details about the encoding and decoding setup of the workflow shown in Figure 4 are found in Table 5. Please note that the parameters were set following the manufacturer instructions.

Table 5. Reference Video encoding details

	Reference Encoding and Decoding Setup
Encoder input	1 080i
Resolution of the encoder input	1 920 x 1 080
Frame rate of the encoder input	59.94 fields per second (interlaced)
Color Transfer function of the encoder input	BT.709 (SDR) + SL-HDR1 Metadata
Codec	H.264/AVC
Profile	High
Encoder	MainConcept Live Encoder v0.0.0.961
Encoder type	Real-time
Channel Capacity of TV 2.0	14 000 kbps
Bitrate Mode (encoder parameter)	VBR
Target bitrate (encoder parameter)	12 000 kbps
Max bitrate (encoder parameter)	14 000 kbps
Keyframe (IDR) Interval (encoder parameter)	2 seconds
Decoder	MX-Player running on a NVIDIA Shield
Decoder type	Real-time
Resolution of the decoded output	1 920 x 1 080
Frame rate of the decoded output	59.94 frames per second (progressive, deinterlaced by the NVIDIA Shield decoder)
Color Transfer function of the decoded output	BT.709 (SDR)

5.3.2 VUT 3.1 - H.264 + SL-HDR1

The general workflow used to encode the video for VUT 3.1 is shown in Figure 5. The original 4K video is transmitted to a cobalt pre-processor through SDI connection. The cobalt pre-processor downscales the video to 1 080i in real-time, and it also adds the SL-HDR1 metadata. This new stream is transmitted to the encoder also using an SDI connection. The encoder encodes the video using H.264/AVC with the SL-HDR1 metadata. The bitstream is then transmitted in real time through a local network. The decoder, which is the “SLhdrExample player” app running on a NVIDIA Shield, decodes the bitstream, using the SL-HDR1 metadata, and displays the decoded video through an HDMI output. This output is then recorded by our video recorder (a Blackmagic device), which is then stored for display during the tests. All steps are performed in tandem.

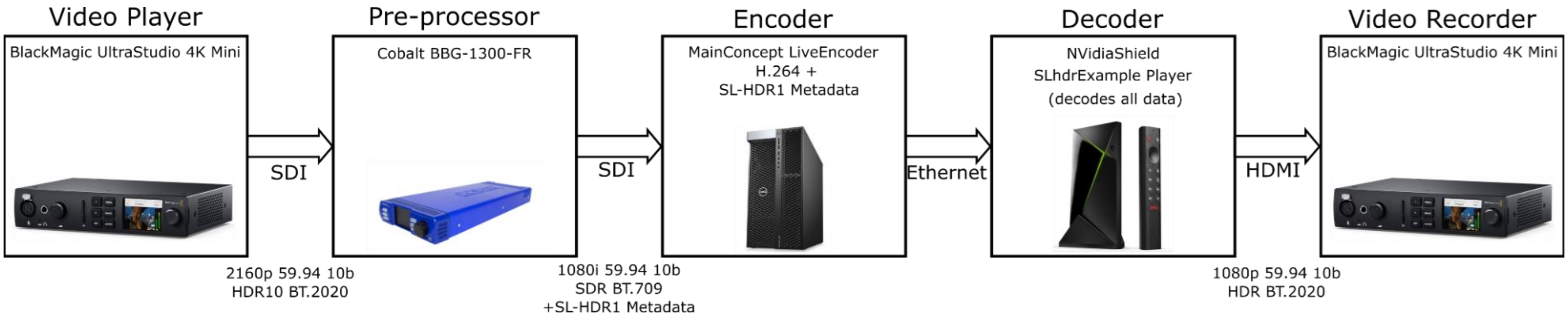


Figure 5. Workflow for VUT 3.1

Note that the NVIDIA Shield decoder generates 1 080i, 10-bit HDR from 8-bit resolution & BT.2020 color space constrained to BT.709 colors (both with SL-HDR1) video and its built-in format converter converts the signal to 1 080p before forwarding over HDMI.

The complete details about the encoding and decoding setup of the workflow shown in Figure 5 are found in Table 6. Please note that the parameters were set following the manufacturer instructions.

Table 6. VUT 3.1 H.264+SL-HDR1 encoding details

	VUT 3.1 H.264 + SL-HDR1 Encoding and Decoding Setup
Encoder input	1 080i
Resolution of the encoder input	1 920 x 1 080
Frame rate of the encoder input	59.94 fields per second (interlaced)
Color Transfer function of the encoder input	BT.2020 (SL-HDR1)
Codec	H.264/AVC
Profile	High
Encoder	MainConcept Live Encoder v0.0.0.961
Encoder type	Real-time
Channel Capacity of TV 2.0	14 000 kbps
Bitrate Mode (encoder parameter)	VBR
Target bitrate (encoder parameter)	12 000 kbps
Max bitrate (encoder parameter)	14 000 kbps
Keyframe (IDR) Interval (encoder parameter)	2 seconds
Decoder	SLhdrExamplePlayer running on a NVIDIA Shield
Decoder type	Real-time
Resolution of the decoded output	1 920 x 1 080
Frame rate of the decoded output	59.94 frames per second (progressive, deinterlaced by the NVIDIA Shield decoder)
Color Transfer function of the decoded output	BT.2020 (HDR)

5.3.3 VUT 3.2.1 H.264 Base Layer for H.264+LCEVC+SL-HDR1

The general workflow used to encode the video for VUT 3.2.1 is shown in Figure 6. The original 4K video is transmitted to a cobalt pre-processor through SDI connection. The cobalt pre-processor downscales the video to 1 080p in real-time and transmits it to the encoder also using an SDI connection. The cobalt pre-processor also adds the SL-HDR1 metadata. The encoder encodes the video using H.264/AVC and LCEVC. The encoder downscales the 1 080p video to 1 080i. The 1 080i stream is encoded with H.264 and an enhancement layer that upscales it to 1 080p is encoded with LCEVC+SL-HDR1. The bitstream is then transmitted in real time through a local network. The decoder, which is the MX Player app running on a NVIDIA Shield, decodes the bitstream, ignoring the LCEVC enhancement layer, and displays the decoded video through an HDMI output. This output is then recorded by our video recorder (a Blackmagic device), which is then stored for display during the tests. All steps are performed in tandem.

Note that in this configuration, the progressive to interlace conversion is necessarily performed in the video encoder rather than in the pre-processor, as it is the case for the reference video and VUT 3.1.

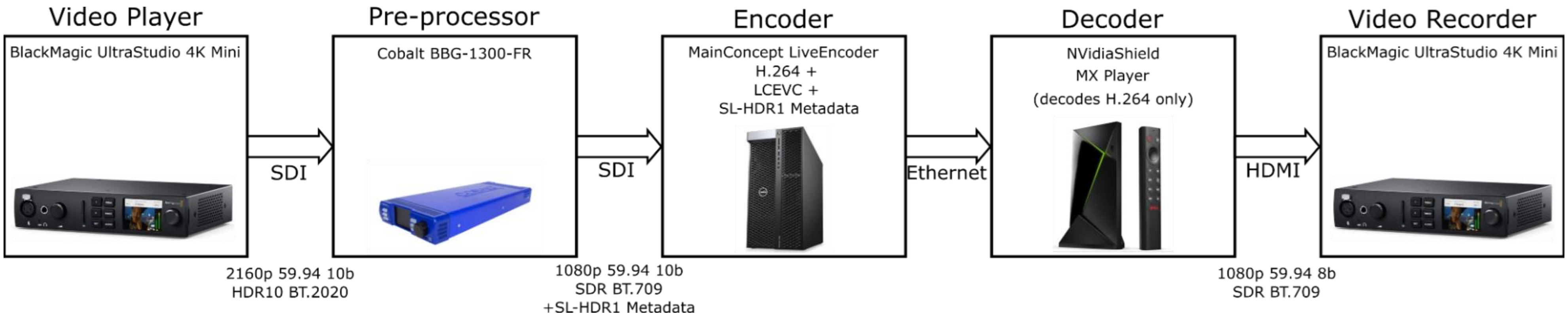


Figure 6. Workflow for VUT 3.2.1

Note that the NVIDIA Shield decodes the 1 080i video and its built-in format converter converts the signal to 1080p before forwarding over HDMI.

The complete details about the encoding and decoding setup of the workflow shown in Figure 6 are found in Table 7. Please note that the parameters were set following the manufacturer instructions.

Table 7. VUT 3.2.1 H.264 base layer encoding details

	VUT 3.2.1 H.264 base layer Encoding and Decoding Setup
Encoder input	1 080p
Resolution of the encoder input	1 920 x 1 080
Frame rate of the encoder input	59.94 fields per second (interlaced)
Color Transfer function of the encoder input	BT.709 (SDR) + SL-HDR1 Metadata
Codec	H.264/AVC
Profile	High
Encoder	MainConcept Live Encoder v0.0.0.961
Encoder type	Real-time
Channel Capacity of TV 2.0	14 000 kbps
Bitrate Mode (encoder parameter)	VBR
Target bitrate (encoder parameter)	13 000 kbps
Max bitrate (encoder parameter)	16 000 kbps
Keyframe (IDR) Interval (encoder parameter)	2 seconds
Decoder	MX-Player running on a NVIDIA Shield
Decoder type	Real-time
Resolution of the decoded output	1 920 x 1 080
Frame rate of the decoded output	59.94 frames per second (progressive, deinterlaced by the NVIDIA Shield decoder)
Color Transfer function of the decoded output	BT.709 (SDR)

As the bitstreams are not recorded, the exact bitrate for the base layer cannot be determined. However, based on the manufacturer’s information, this encoding set-up should yield 85% of the average bitrate to the H.264 base layer and 15% to the LCEVC enhancement layer. Although the capacity for the channel is 14 Mbps, the applied configuration used should fit in the channel capacity for the complete bitstream (H.264 + LCEVC+SL-HDR1).

5.3.4 VUT 3.2.2 H.264+LCEVC+SL-HDR1

The general workflow used to encode the video for VUT 3.2.2 is shown in Figure 7. The original 4K video is transmitted to a cobalt pre-processor through SDI connection. The cobalt pre-processor downscales the video to 1 080p in real-time and transmits it to the encoder also using an SDI connection. The cobalt pre-processor also adds the SL-HDR1 metadata. The encoder encodes the video using H.264/AVC and LCEVC. The bitstream is then transmitted in real time through a local network. The decoder, which is the ExoPlayer app running on a NVIDIA Shield, decodes the bitstream, using the LCEVC enhancement layer, and displays the decoded video through an HDMI output. This output is then recorded by our video recorder (a Blackmagic device), which is then stored for display during the tests. All steps are performed in tandem.

Note that in this configuration, the progressive to interlace conversion is necessarily performed in the encoder rather than in the pre-processor.

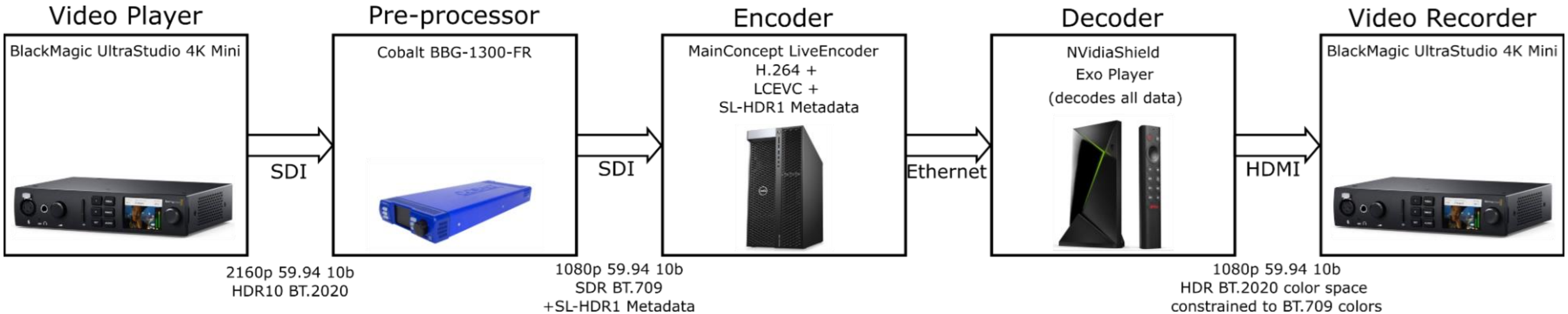


Figure 7. Workflow for VUT 3.2.2

Note that the NVIDIA Shield decodes a bitstream composed of a 1 080i SDR BT.709 H.264 base layer and a LCEVC+SL-HDR1 enhancement layer. It then combines these layers first into a 1 080p 10-bit SDR (with LCEVC), and then into a 1 080p 10-bit HDR & BT.2020 color space constrained to BT.709 colors (with SL-HDR1), which is then output over HDMI.

The complete details about the encoding and decoding setup of the workflow shown in Figure 7 are found in Table 8. Please note that the parameters were set following the manufacturer instructions.

Table 8. VUT 3.2.2 H.264+LCEVC+SL-HDR1 encoding details

	VUT 3.2.2 H.264 + LCEVC + SL-HDR1 Encoding and Decoding Setup
Encoder input	1 080p
Resolution of the encoder input	1 920 x 1 080
Frame rate of the encoder input	59.94 frames per second (progressive)
Color Transfer function of the encoder input	BT.709 (SDR) + SL-HDR1 Metadata
Codec	H.264/AVC
Profile	High
Encoder	MainConcept Live Encoder v0.0.0.961
Encoder type	Real-time
Channel Capacity of TV 2.0	14 000 kbps
Bitrate Mode (encoder parameter)	VBR
Target bitrate (encoder parameter)	13 000 kbps
Max bitrate (encoder parameter)	16 000 kbps
Keyframe (IDR) Interval (encoder parameter)	2 seconds
Decoder	ExoPlayer running on a NVIDIA Shield
Decoder type	Real-time
Resolution of the decoded output	1 920 x 1 080
Frame rate of the decoded output	59.94 frames per second
Color Transfer function of the decoded output	BT.2020 (HDR)

As the bitstreams are not recorded, the exact bitrate for the base layer cannot be determined. However, based on the manufacturer's information, this encoding set-up should yield 85% of the average bitrate to the H.264 base layer and 15% to the LCEVC enhancement layer. Although the capacity for the channel is 14 Mbps, the applied configuration used should fit in the channel capacity for the complete bitstream (H.264 + LCEVC+SL-HDR1).

5.4 Results

As detailed in Section 5.1.3, all 15 videos (3 VUTs, each one with 5 different contents) have been randomized and displayed to the participants. Table 9 shows the expected results of the experiment, considering the characteristics of the contents being analyzed.

Table 9. Expected results

	Average Score Target	Worst Case Target
VUT 3.1	Slightly Better	The Same
VUT 3.2.1	Slightly Worse	Slightly Worse
VUT 3.2.2	Slightly Better	The Same

5.4.1 VUT 3.1 Results

The results for VUT 3.1 are shown in Figure 8 and Figure 9. In these figures, the boxplot shows the locality, spread and skewness of scores through their quartiles and the violin shaped figures on each plot show the probability distribution of the collected subjective scores.

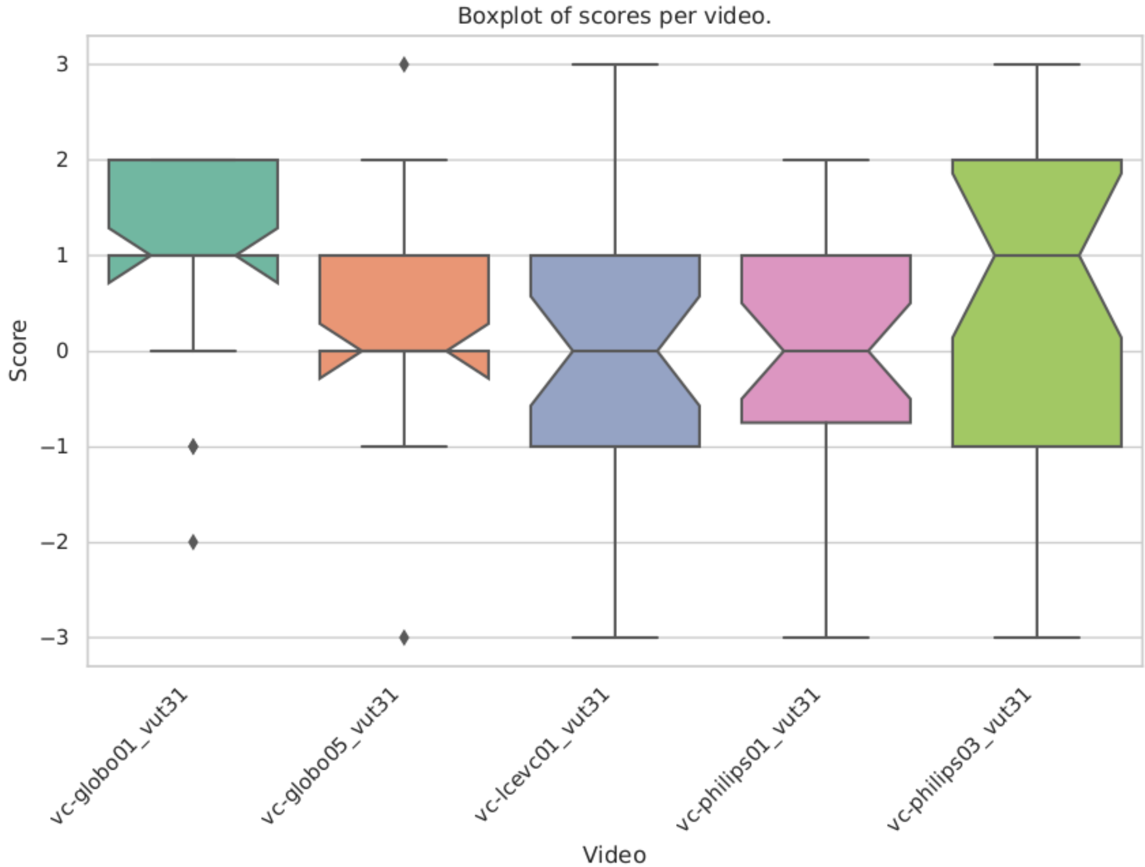


Figure 8. Boxplot for VUT 3.1

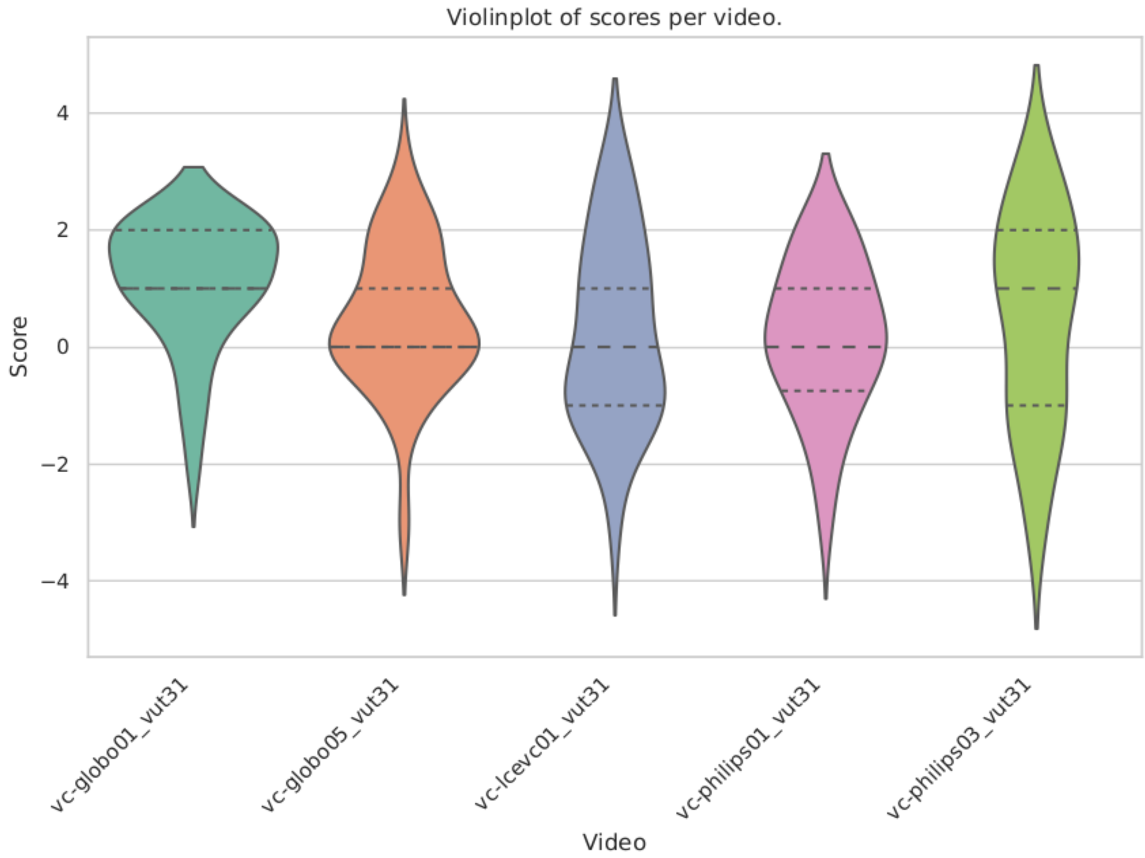


Figure 9. Violin plot for VUT 3.1

Table 10. Average MOS for VUT 3.1

Content	MOS Score
globo01	1.37 ± 0.68
globo05	+0.50 ± 0.96
lcevc01	+0.34 ± 1.47
philips01	+0.27 ± 1.16
philips03	+0.50 ± 1.79

We can see from Table 10 that VUT 3.1 has been rated higher, considering the MOS score, compared to the reference video for all contents. The best results are shown for globo01 video. Considering Table 9, which shows the expected results, the results of the experiment are on par with the expected results.

It is reminded that the reference SDR video is upconverted by the LG TV set to HDR (see asterisk below table 3).

5.4.2 VUT 3.2.1 Results

The results for VUT 3.2.1 are shown in Figure 10 and Figure 11. In these figures, the boxplot shows the locality, spread and skewness of scores through their quartiles and the violin shaped figures on each plot show the probability distribution of the collected subjective scores.

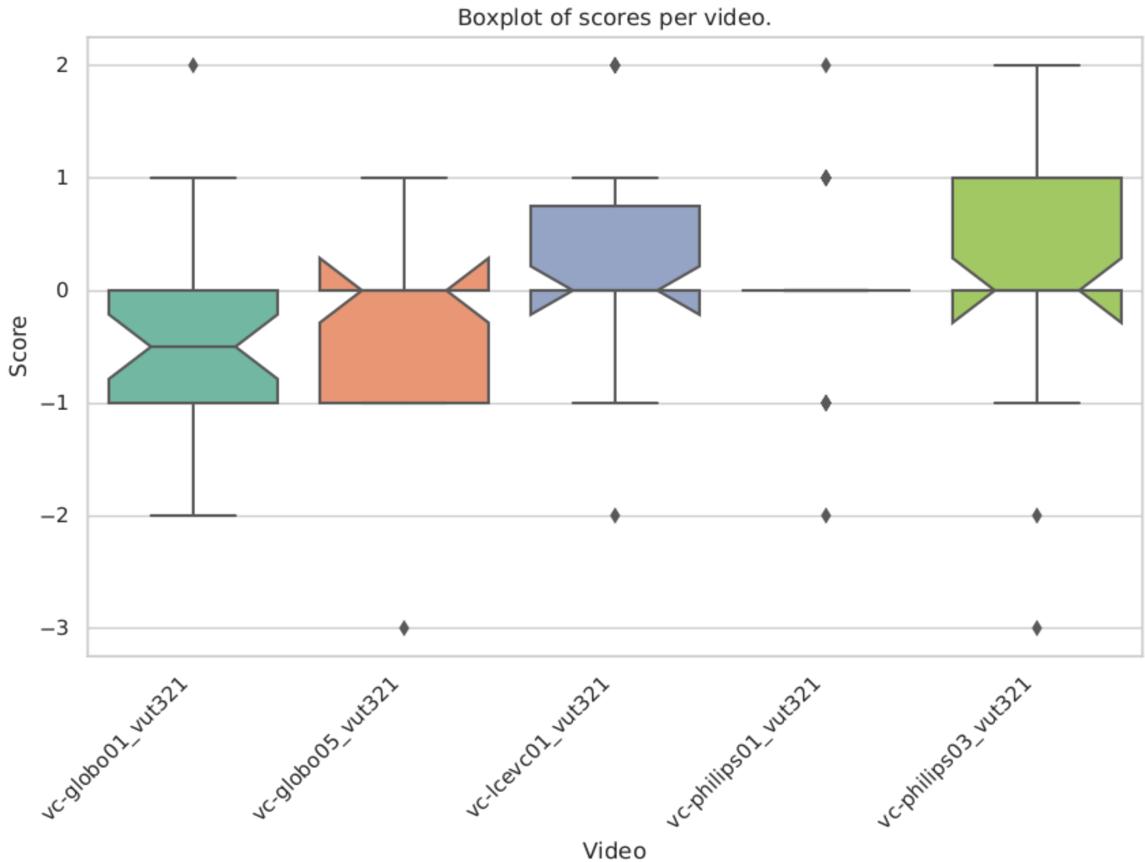


Figure 10. Boxplot for VUT 3.2.1

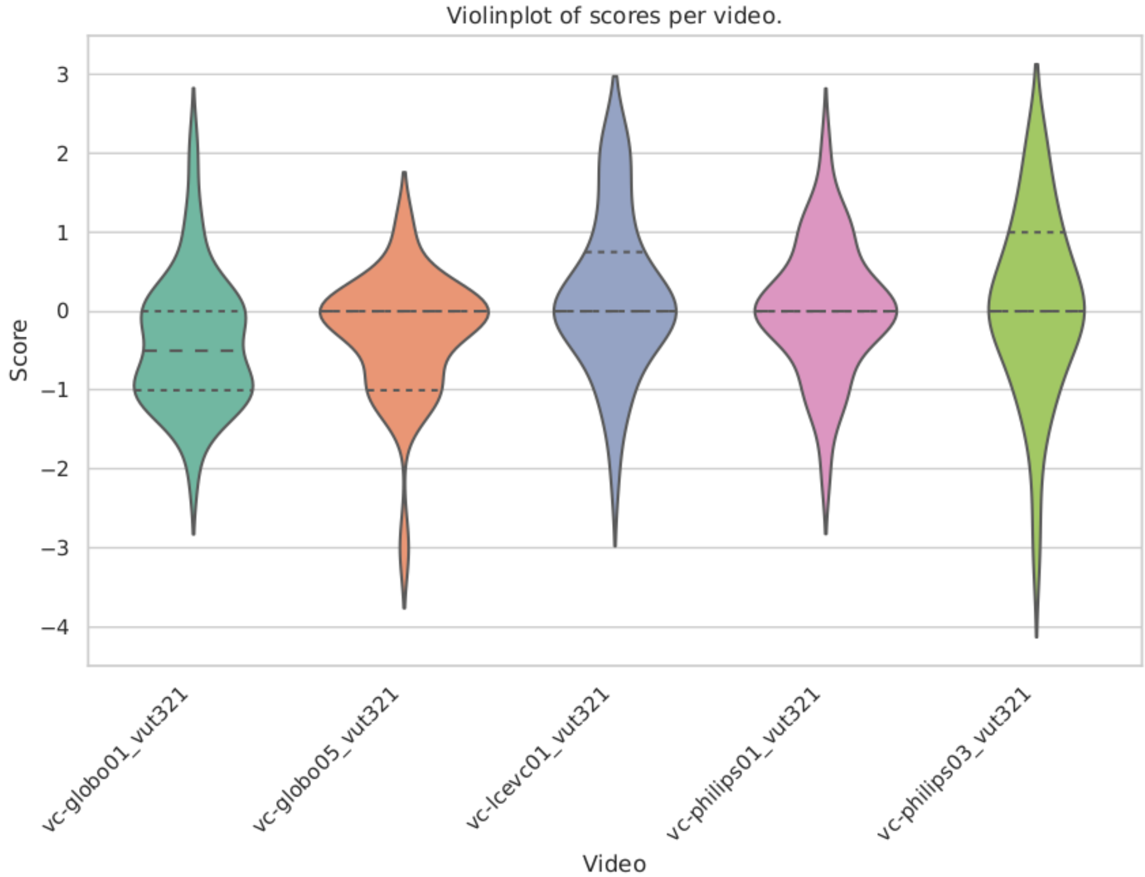


Figure 11. Violin plot for VUT 3.2.1

Table 11. Average MOS for VUT 3.2.1

Content	MOS Score
globo01	-0.48 ± 0.68
globo05	-0.20 ± 0.55
lcevc01	+0.27 ± 0.88
philips01	+0.03 ± 0.63
philips03	+0.17 ± 0.96

We can see from Table 11 that four out of the 5 contents have been on average rated as similar to the reference video, with the average close to 0. For globo01 video, however, there was a perceived difference, with VUT 3.2.1 performing slightly worse than the reference. This is not unexpected, as VUT 3.2.1 uses the same encoder as the reference video but targeting a smaller bitrate. However, considering Table 9, which shows the expected results, the result of the experiment exceeds the expected results, with an output MOS closer to 0 (same quality), instead of closer to -1 (slightly worse quality).

5.4.3 VUT 3.2.2 Results

The results for VUT 3.2.2 are shown in Figure 12 and Figure 13. In these figures, the boxplot shows the locality, spread and skewness of scores through their quartiles and the violin shaped figures on each plot show the probability distribution of the collected subjective scores.

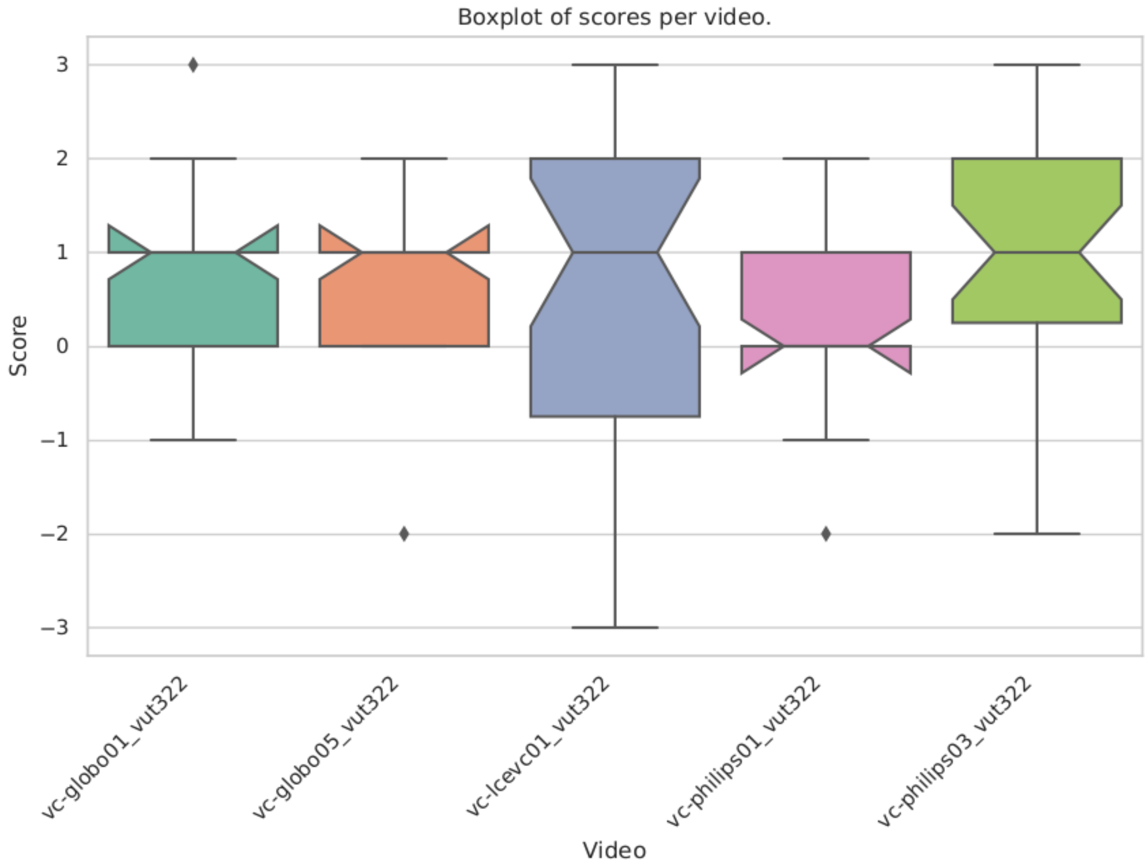


Figure 12. Boxplot for VUT 3.2.2

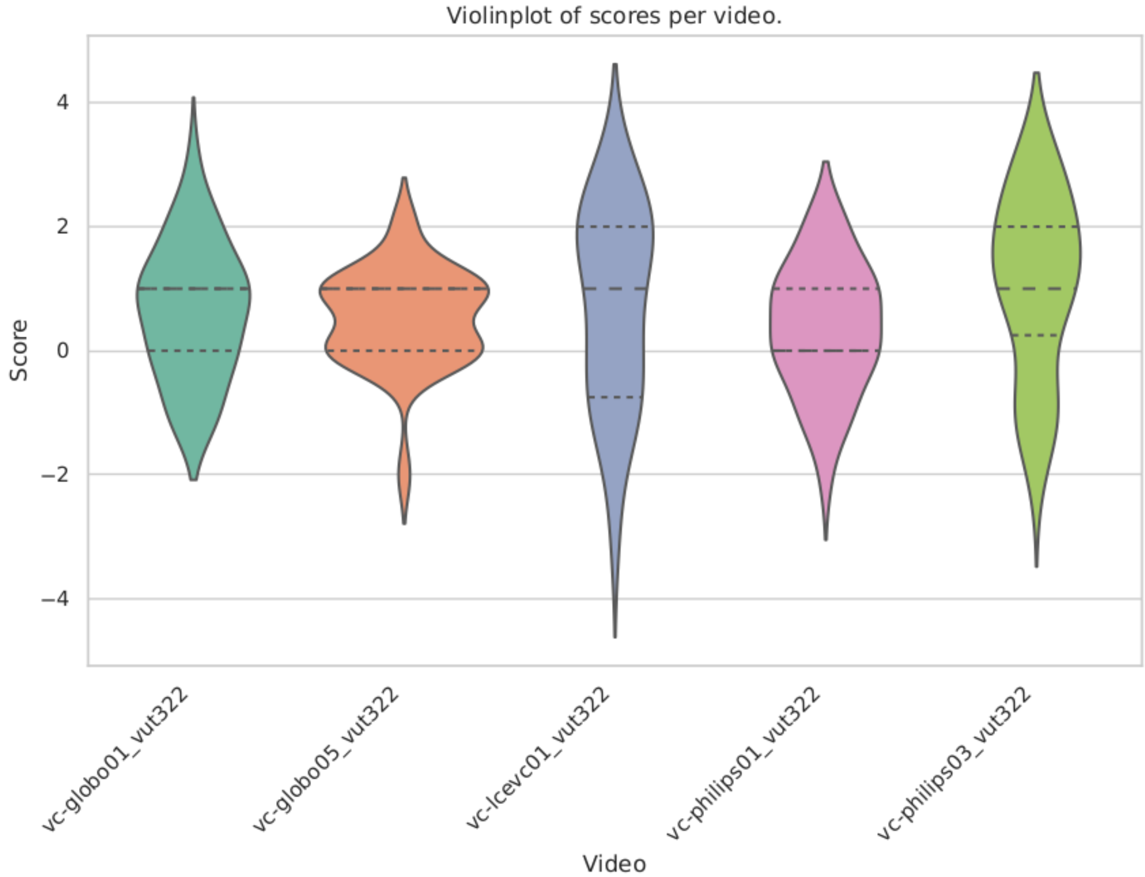


Figure 13. Violin plot for VUT 3.2.2

Table 12. Average MOS for VUT 3.2.2

Content	MOS Score
globo01	+0.55 ± 0.98
globo05	+0.62 ± 0.62
lcevc01	+0.86 ± 1.45
philips01	+0.44 ± 0.95
philips03	+1.17 ± 1.36

We can see from Table 12 that VUT 3.2.2 has been rated higher, considering the MOS score, compared to the reference video for all contents. The best results are shown for philips03 video. Considering Table 9, which shows the expected results, the results of the experiment are on par with the expected results.

5.5 Summary of the Results

The results of the previous sections are summarized in Table 13.

Table 13. Summary of the results

	VUT 3.1	**VUT 3.2.1**	**VUT 3.2.2**
Description	H.264 + SL-HDR1	Base layer for H.264+LCEVC	H.264+LCEVC+SL-HDR1
Expected Average Result	Slightly Better (+1)	Slightly Worse (-1)	Slightly Better (+1)
Expected Worse Result	The Same (0)	Slightly Worse (-1)	The Same (0)
globo01	+1.37 ± 0.68	-0.48 ± 0.68	+0.55 ± 0.98
globo05	+0.50 ± 0.96	-0.20 ± 0.55	+0.62 ± 0.62
lcevc01	+0.34 ± 1.47	+0.27 ± 0.88	+0.86 ± 1.45
philips01	+0.27 ± 1.16	+0.03 ± 0.63	+0.44 ± 0.95
philips03	+0.50 ± 1.79	+0.17 ± 0.96	+1.17 ± 1.36

We can see that both codecs, VUT 3.1 and VUT 3.2.2, have been rated with higher MOS scores compared to the reference video. However, VUT 3.2.2 performs better for two sequences (lcevc01 and philips03), slightly better for two sequences (globo05 and philips01) and worse for globo01 sequence.