Pro-AV CODEC Landscape Report

7 October 2024 v1.0

Version History

Version	Date	Remarks
0.1	23 October 2023	Early draft for review at VSF Face to Face in New York
0.5	21 December 2023	Draft for board review as direction check and to seek feedback from select individuals outside the VSF on direction.
0.6	27 February 2024	Approved draft version for circulation outside VSF
0.7	29 July 2024	Final draft for review #1
0.8	13 August 2024	Final draft for review #2

Contributors

The following companies and individuals were the primary contributors to the report:.

Company	Contributors	Email
ASPEED Technology	Bruce Chang	bruce_chang@aspeedtech.com
Intel	Ben Cope	ben.cope@intel.com
intoPIX	Tim Bruylants	t.bruylants@intopix.com
Kakadu Software	Michael Smith	miksmith@gmail.com
Macnica Americas	Andrew Starks	andrew.starks@macnica.com
Matrox Video	Alain Champenois Andre Testa	achampen@matrox.com atesta@matrox.com
Media Links	John Dale	jdale@medialinks.com

Thank you to others from the above-mentioned companies who provided one-off input through the stated contributors but are not named.

The following additional companies, with lead representatives stated, were part of the activity group that authored this report:

Company	Representative	Email
Amagi	Marko Horvat	marko.horvat@amagi.com
Appear TV	Andrew Rayner	andrew.rayner@appear.net
AT&T	Nicholas Nicas	nn7686@att.com
Comprimato	Jiri Matela	jiri@comprimato.com
Evertz	Paulo Francisco	pfrancisco@evertz.com
Net Insight	Per Hultin	per.hultin@netinsight.net
Nvidia	Thomas True	ttrue@nvidia.com
Packetstorm	Jack Douglass	jack@packetstorm.com
Sencore	Aaron Doughten	aaron.doughten@sencore.com
Sipradius	Sergio Ammirata	sergio.ammirata@sipradius.com
TAG Video Systems	Paul Briscoe	paul@tagvs.com

Acronyms

Acronym	**Description**	**Remarks**
AGIC3	Advanced Graphic and Image Codec Gen3	A video codec specifically tailored for ProAV applications and developed by ASPEED Technology.
AVC	Advanced Video Coding	Also known as H.264 or MPEG-4 Part 10
CDEF	Constrained directional enhancement filter	In-loop filter based on a non-linear low-pass filter
CODEC	CODEC is a portmanteau of "CODer" and "DECoder”	In the context of this report, only video CODECs are considered.
FHD	Full High Definition	Also known as 1080p
HEVC	High Efficiency Video Coding	Also known as H.265 or MPEG-H Part 2
HTJ2K	High-Throughput JPEG 2000	Also known as T.814 or ISO/IEC 15444-15 (a part of JPEG 2000)
JCT-VC	Joint Collaborative Team on Video Coding	A collaboration between the ISO/IEC MPEG and ITU-T Study Group 16 VCEG
JT-NM	Joint Task-force on Network Media	A self-coordinating group of industry bodies working together on the development of IP technologies for professional media systems
JPEG XS	JPEG eXtra Small	Also known as ISO/IEC 21122-1/-2
MPEG	Moving Picture Experts Group	An alliance of working groups established jointly by ISO and IEC that sets standards for media coding
SMPTE	Society of Motion Pictures and Television Engineers	Global organization developing standards for the film, television, and digital media industries.
UHD	Ultra High Definition	Also known as 4K or 2160p or UHD-TV1
VSF	Video Services Forum	An association of service providers, users and manufacturers dedicated to interoperability, quality metrics and education for media networking technologies.
VCEG	Video Coding Experts Group	A working group of the ITU concerned with standards for compression coding of video, images, audio and other signals.
VVC	Versatile Video Coding	Also known as H.266 or MPEG-I Part 3

Table of Contents

1. Introduction and Context

The Professional AV market is transitioning from traditional AV strategies to IT-oriented ones, driven by an increased interest in software-defined solutions. A key factor in this shift is the adoption of standard IP networks, moving away from specialized networks such as HDBaseT or SDI.

A similar transition is occurring in the Broadcast Studio market, where a set of standards under SMPTE ST 2110 and based on the original Technical Recommendation documents from the Video Services Forum (VSF) provide an open standard framework for video transport over IP networks.

Building on the Technical Recommendations that led to SMPTE ST 2110, the VSF initiated several workstreams, including the IPMX activity group, which aims to bring SMPTE ST 2110 to the Pro-AV market, mirroring the progress seen in the Broadcast Studio market.

A key question for the IPMX activity group is determining how to compress high-resolution video streams to fit within common Ethernet networks (e.g., 1G, 2.5G). The use cases and key performance indicators (KPI) in the Pro-AV market differ significantly from those in the Broadcast Studio market, necessitating a specific assessment within the VSF regarding IPMX.

As a forum to advance the discussion and understanding of video compression, a CODEC activity group was formed within the VSF in June 2023. The group's problem statement is looking to address the assertion: 'There are many CODEC approaches to AV over IP solutions today, with no established standard to guide the selection of a CODEC for specific use cases. While the group does not aim to endorse a specific CODEC or engage in benchmarking, it aims to present findings and observed trends that may assist others in making informed choices. The outcomes are expected to interest a broad audience in the Pro-AV industry, extending beyond the IPMX scope.

This document serves as a landscape report detailing the ongoing work of the CODEC activity group. It is structured as follows: The report begins by discussing use cases in the Pro-AV sector that establish CODEC requirements (Section 2). It then outlines the key decision factors involved in selecting the appropriate CODEC (Section 3). This is followed by a survey of pertinent CODECs (4) and a compilation of suitable test vectors for their validation (Section 5). Subsequently, an assessment of different CODECs for these criteria and scenarios is made (Section 6). The document concludes with key findings (Section 7) and suggests directions for future work (Section 8).

2. Use Cases in Pro-AV

Understanding the specific use cases is crucial for contextualizing the KPIs relevant to the suitability of a CODEC for one scenario over another. This section will provide a summary of the use cases that were considered in compiling this report.

2.1 List of Use Cases and Common Themes

As a recognized industry reference, the starting point for our analysis was the use cases outlined in the report by the Joint Task- Force on Network Media (JT-NM) [1], which also served as input for the VSF IPMX Activity Group. Building on this foundation, the CODEC activity group made two additional contributions:

1. The JT-NM Use Cases identify common themes of requirements. To make these more tangible, a lead use case was added for each theme, and in some instances, additional leading use cases were included to represent varying requirements.

2. The JT-NM Use Cases were reviewed for completeness, focusing on factors influencing CODEC choice. New common themes, like the need for mathematically lossless compression in the medical sector, were identified as important for analysis. These new themes have been incorporated back into the report of the JT-NM [1] (pull request #3) and highlighted with an asterisk in the table.

The full list of lead use cases relevant to this report is summarized in Table 1.

Lead Use Cases	**Common Themes**
KVM AV networking, Conference room, Post production	Excellent Desktop Quality @ 1GbE with 4K60 Subframe Latency
KVM AV networking, Conference room, Post production	Excellent Desktop Quality @ 2.5GbE with 4K60 Subframe Latency
Control room, AV networking, Education (presenter, lab)	Excellent Desktop Quality @ 1GbE with 4K60 Frame latency
AV networking to video wall displays (retail, theme park)	Excellent Video Quality @ 1GbE with 4K60
AV networking to synchronized quad video wall displays (retail)	Excellent Video Quality @ 1GbE with 4x FHD *
AV networking to video wall displays (live events, security)	Excellent Video Quality @ 1GbE with 4K60 Frame latency
Esports (distribution, players)	Subframe Latency
Operating Room Transmission (Medical)	Mathematically lossless coding * Subframe Latency
Diagnostic review of MRI scans (Medical)	Mathematically lossless coding * Frame latency
In-house video production at Corporate Enterprise	Software Based Workflow

Large-scaled or antiquated facilities with no wired ethernet (rail stations, farming, historical location), Mobile device integration to corporate AV networking	Limited Connections (wifi) *
Retrofit of installation with existing AV cameras or display tech	H264/H265 Ecosystem Support *
Low cost directive to IT manager, Pro-sumer social media, Amateur technology trial	Easily Accessible *
AV distribution in a sports venue with proxy stream, thumbnail for video production/switching	Multiple Res or Bandwidth of the Same Source's Content *
Local live production with overlaid CGI graphics for compositing	Carry outbound/inband signal (e.g. alpha) * Minimal stream switching delay *
LED tile video wall	High Bit Depth *

Table 1: Lead Pro-AV use cases and common themes
(*) New themes incorporated back to JT-NM [1] report

2.2 Explanation of the Common Themes

The following is an explanation of the 12 common themes we've identified and their application in Pro-AV use cases.

Excellent Desktop Quality
When working with computer-generated content, maintaining high fidelity with the original content is essential for the user experience.

Cost considerations and the need to integrate seamlessly with existing networks dictate that this quality should be attainable through commonly available Ethernet ports, which currently operate at 1G. The target display resolution is 4K60 (UHD-TV1), aligning with the standard prevalent on professional displays today.

A common issue encountered with high-frequency computer-generated content, such as desktop UI, is chroma bleed. This underscores the necessity for the image to utilize maximum color sampling, in either YUV 4:4:4 or RGB format. Additionally, given that these scenarios often involve user interaction, such as manipulating desktop content, achieving subframe latency is important.

In the future, as seen in the case of supporting display resolution of 8K60 (UHD-TV2), there is an expectation for Ethernet rate requirement to rise to 2.5G. However, the challenges from a CODEC perspective are sufficiently comparable to those encountered with 4K60 at 1G, making it a suitable proxy for this anticipated future demand. Utilizing 2.5G Ethernet today can also address desktop/CGI concerns related to the 1G INTRA frame mezzanine codec, ensuring smoother performance and higher quality outputs.

Excellent Video Quality
The need for excellent video quality for AV over IP solutions is already well covered within the JT-NM use cases [1]. While desktop content demands 4:4:4 chroma sampling, 4:2:2 or 4:2:0 is generally deemed sufficient for video content.

As well as video walls, satellite and surveillance systems are other applications with such requirements. It is noted that video walls can also have the requirement for excellent desktop content quality, for example, an LED video wall at the front of a corporate meeting room.

A dedicated, leading use case covering synchronized 4x FHD streams as tiled versions of a single 4K stream is included to address the specific challenge posed by this mapping. Apart from increasing coding complexity by splitting the spatial information, such setups require very high quality to prevent banding artifacts at the shared display edge. This becomes even more crucial when quadrants overlap, especially when using a projector display.

Subframe and frame latency
Subframe latency is defined as latency within a fraction of a frame, usually ¼, ⅛, or less. Frame latencies cover anything between 1 and a few frames, usually less than 5.

Subframe latency was already mentioned as important in scenarios where a user is interacting with desktop content, such as in a corporate meeting room or KVM environment. Further examples include e-sports, interactive gaming, and operating room transmission. For operation rooms, achieving low latency is extremely important for surgeons when their hands are making decisions based on what's transmitted to a screen from a camera.

A slightly more relaxed frame latency requirement can be pertinent for many use cases when a latency of one frame is acceptable versus subframe latency. This is particularly relevant in scenarios such as live event streaming or video conferencing, where minor delays do not significantly impact the user experience.

Mathematically Lossless Coding
Some applications necessitate absolute accuracy of the video, where visually lossless quality is sufficient. Mathematically Lossless identifies this level, where the compressed information remains identical to the source. Medical use cases serve as an excellent showcase for this requirement. For instance, diagnostic review involves distributing video to share medical data such as MRI scans and images on a screen, eliminating the potential for codec artifacts that could lead to misdiagnosis.

While mathematically lossless coding in CODECs is common for archival purposes, where size is a concern rather than latency, achieving mathematically lossless coding poses a significant challenge for fixed-bitrate streaming. Current mezzanine CODECs can normally achieve compression ratios of 1:2 to 1:4, but such ratios are not guaranteed and depend on the content's entropy. While acceptable compromises can be made, such as accepting minimal losses for extreme and unrealistic entropy, supporting variable-bitrate allowing mathematically lossless coding to be reached over multiple frames, mathematically lossless compression still presents a challenge for 4K60 streaming due to the non-deterministic compression ratio and the low latency requirement for decoding.

On the subject of CODEC options for medical applications, two studies have been published examining their use in processing medical image scans (ex: MRI/CT/PET). The first study [2] evaluates performance, specifically decoding latency and compression rate, while adhering to mathematically lossless requirements. The second study [3] evaluates the use of visually lossless compression (i.e., lossy compression with no noticeable visual artifact). From these studies, it is evident that there is no consensus on using visually lossless compression in medical applications.

Side note: In assessing the JT-NM use cases, the activity group members believed that medical applications were not fully represented. The operating room transmission and diagnostic review use cases exemplify significant codec requirements for quality and latency. Additionally, limited connection (Section 2.6) is emerging as a feature of interest in these applications.

Software Based Workflow
In-house production for live events demonstrates the need for software-based workflows. This is important when integrating AV over IP into common collaboration tools like Microsoft Teams, which aim to use high-quality video camera sources. The software implementation of the CODEC and protocols offers flexibility, allowing for easy adjustments to the number, placement, and configuration of cameras and displays, as well as the choice of AV over IP methods. Different approaches with various CODECs can be easily mixed using software transcoding. The key requirement is that the chosen CODEC must be suitable for software implementation, meeting the desired resolution and complexity KPIs.

Limited Connections
Limited connectivity use cases arise from various factors, driven by the need for portability, such as user devices in a corporate environment, or due to the absence of wired infrastructure, as seen in digital signage within aging transport hubs or historical sites like castles. These scenarios have been incorporated into Table 1 and the JT-NM use cases to represent the requirement for limited connections. Further details on these scenarios are covered below. For users connecting via WiFi, the following considerations are typically reported:

- The relevant use cases typically involve screen sharing and video applications.
- Ensuring that audio and video remain uninterrupted and free from jitter or display artifacts is of utmost importance. Any disruptions can be distracting and give the impression of a malfunction.
- Latency and quality expectations are somewhat subdued compared to wired connections at present. However, it is anticipated that these expectations will increase with future WiFi standards.
- Latency should be sufficiently low for a smooth conversation with someone using the same technology.
- While not essential, achieving visually lossless quality is considered desirable.
- The type of video typically transmitted over WiFi is the same as that for wired connections.

- Users operating on a single connection that must share bandwidth with AV over IP are generally willing to sacrifice some quality and tolerate some latency to conserve bandwidth.

For example, consider a director managing a small live production with three cameras on a laptop who needs wireless connectivity. With only a 1GB connection, the director must handle three camera feeds and four channels of audio. Despite the limited budget, which prevents access to the best equipment, quality and latency remain crucial. Therefore, the chosen CODEC must emphasize bandwidth efficiency over latency and potentially quality, providing a practical solution within the given constraints. This scenario underscores the importance of finding the right balance between performance and resource availability in real-world applications.

H264/H265 Ecosystem Support

An important consideration, particularly regarding CODEC selection, is the ability to support AV over IP solutions through software upgrades to existing installed products. For instance, many endpoint cameras and system-on-chip (SoC) based displays are equipped with hardware support for H.264 and H.265 CODECs. However, their compute capabilities may pose challenges for implementing the mezzanine CODEC options discussed later. Real-world use cases might include lecture capture systems or other collaboration solutions. This scenario is listed as integration with existing cameras and display technology in Table 1.

For example, consider a lecture capture system that relies on H.264 PTZ cameras and displays with SoC-based hardware. Users and manufacturers want to maximize the existing ecosystem of cameras, embedded devices, and PCs with accelerated H.264/H.265 support. Within the IPMX activity group effort, there is a requirement for IPMX/SMPTE ST 2110-2x to support H.264 and/or H.265 CODECs, to ensure compatibility and efficiency in allowing for seamless integration with current technologies.

Easy to use and accessible

When retrofitting AV over IP solutions to existing PTZ camera or display hardware, accessibility to trial AV over IP becomes an important consideration. Two key factors in this regard are the cost of entry and the ease of access to different CODEC options, such as whether they are available freely or not. Further explanation of the expected user personas and their needs for the first of these considerations is provided below.

For example, consider a curious professional, amateur, or prosumer who wants to explore AV over IP solutions:

- They currently use a free (without charge) proprietary AV over IP solution.
- They want to try an open solution approach.
- They are looking for the fastest and most free way to try the standard.
- They are very aware of proprietary solutions that are free and expect similar options.

Another example is a person creating an event and wanting to share content with guests:

- They understand the need to pay more for high-quality encoding.

- They need a barrier-free way to share their content with guests.

The implied requirement from these personas is that anyone should be able to watch produced content without paying a royalty for anything associated with the transport.

Multiple Resolution Support

A further use case impacting CODEC suitability is the need for transmitting multiple resolutions within the same distribution. Two use cases are summarized below and included as new entries in Table 1.

For example, consider a director or network operations center employee using a multi-viewer to monitor sources on large displays. The challenge is that supporting 16 or 24 inputs requires significant bandwidth and expense. To minimize bandwidth, a smaller/lower bandwidth version of the real feed can be used as a preview. Timing needs to be synchronized with the full feed to make quick decisions based on the observed content.

Another example involves a sports venue AV system designer managing feeds that originate as 4K flows for a production system. These feeds also need to stream to FHD displays, such as digital signage systems and press boxes. It is preferred not to duplicate these flows solely to support FHD. A typical use case would include 4K, FHD, and qHD (quarter High Definition) available from a single compressed stream on the network.

From the manufacturers' perspective, a method is required to provide multiple resolutions of a single source of content without significantly increasing network bandwidth consumption beyond that of the highest resolution content. It is essential that each version of the flow is in sync with every other version.

Carry Out of Band or InBand information

For local live production with overlaid artificial graphics for compositing or green screen use cases, it is important to transport alpha channel information along with the color data. For this reason, a use case requirement was added to Table 1, detailing that this alpha channel could be transmitted in-band as part of the compressed code stream or out-of-band as a separate stream.

For example, in a live podcast production, hosts might use green screens to display dynamic backgrounds, sponsor logos, or real-time audience interactions. To achieve seamless integration of these graphics, the alpha channel, which defines transparency, must be accurately transmitted alongside the video feed. This ensures that the composited video appears natural and professional without any visual artifacts.

Another example is in virtual production for film and television, where actors perform in front of a green screen, and complex virtual backgrounds are added in real-time during the production. Having the ability to carry alpha channel information either in-band or out-of-band allows for greater flexibility and precision during the live viewing, ensuring that the final output seen by the audience meets high-quality standards.

Minimal stream switching delay
The ability for a receiver to rapidly switch from one transmission feed to another, immediately upon request and on a frame boundary, can be a critical requirement, as highlighted in the Table 1 use case for local live production.

For example, consider a live interview broadcast with multiple guests, where the production team needs to switch between different camera angles quickly to capture reactions and interactions. Low switching latency ensures that these transitions are smooth and seamless, providing a professional viewing experience for the audience. Without low latency, delays in switching can lead to awkward pauses and disrupt the flow of the broadcast.

Another example is in live sports broadcasting, where capturing the action from various angles is essential. Quick switching between feeds allows the production team to follow the play closely, providing viewers with the best possible perspective of the event. This capability is crucial during fast-paced moments, such as a goal in soccer or a touchdown in football, where immediate switching can enhance the excitement and engagement of the broadcast.

High Bit Depth
One of the key features of advanced AV applications is the support for high bit depth in video streams, especially when dealing with direct-view LED video walls. High bit depth allows for more precise color representation and smoother gradients, which is critical for achieving high-quality visual output. This requirement is particularly relevant for LED wall use cases, as outlined in Table 1.

To illustrate the LED tile video wall use case, consider the following scenario:

As an AV integrator working in retail, transportation, or simulation environments for high-value goods, I need to distribute video signals for multiple (10-20) LED cabinet-based digital signage onto the existing standard Ethernet network infrastructure, which runs at 1-10G bandwidth. In this scenario, achieving low bandwidth and low cost is more important than low latency or visually lossless compression. For ease of deployment, the solution must be software-friendly and capable of efficiently managing high bit depth video streams.

In [4], Intel considers the use of SMPTE 2110 or IPMX for direct view LED video walls, treating each LED cabinet as a separate video stream. This scenario extends the 4x FHD stream use cases, but with even greater implications for coding efficiency. A crucial requirement for LED walls driven in this manner is the ability to transmit pixels at very high bit depths, such as 12 to 16 bits per channel, to accurately drive the LED tiles.

3. Key Industry Metrics for Choosing a CODEC

This section summarizes the working group's collective knowledge on how to evaluate CODECs in relation to the identified use cases. The intent is to provide an overview of "what we hear from the industry," acknowledging that opinions may vary but aiming to present observations that are broadly useful.

3.1 Metrics List

To establish the context for comparing metrics, the group determined that it is important to set some representative target video attributes for evaluation, as follows:

Resolution Target: Specified as an ITU-T standard number or simply <width> x <height>
Pixel Format: Bit depth, BT.2020 vs. BT.709, SDR or HDR
Color Space: RGB or YUV
Frame Rate: In frames per second
Bit Rate: In Gbps

Against these representative attributes, the following comparison metrics were analyzed, with the first three metrics discussed in more detail in subsections 3.2, 3.3, and 3.4.

Metric	Remarks
Visual Quality	Subjective assessment of artifact visibility: 5 - Imperceptible (Excellent): No visible artifacts. 4 - Slightly Perceptible (Good): Minor artifacts, hardly noticeable. 3 - Perceptible (Fair): Noticeable artifacts, but acceptable. 2 - Clearly Perceptible (Poor): Obvious artifacts, affecting quality. 1 - Annoying (Bad): Disturbing artifacts, unacceptable quality. See Section 3.2 for more details.
Glass-to-Glass Latency	Objective measure: A. Subframe Range: Latency within fractions of a frame. B. Frame(s) Range: Latency measured in frames (up to 5) C. Second(s) Range: Latency measured in seconds. Measures the latency from capture by the transmitter to display by the receiver under normal network conditions. Divided into three groups for simplicity since having too many categories makes comparison difficult. This is not a scientific measurement but a general guideline. See Section 3.3 for more details.
Stream Switching Delay	Objective measure of the duration for a clean connection to a stream (in frames). Measures the worst-case duration before reaching an artifact-free state when connecting to an already running stream. See Section 3.4 for more details.
Pixel Format	Supported Resolution, Bit Depth, Color Space, Alpha Channel. Lists the range and capabilities of supported video specifications.

Bitrate Range	Bitrate operating range for typical usage Bitrates normally used (in bpp) to achieve visually acceptable quality to the user.
BItrate Control	CBR, VBR, Frame-Based, Slice-Based, Byte Exact, etc. Describes the bitrate control methods and granularity options available.
Hardware IP complexity	Complexity in logic gates of the hardware to implement the CODEC High Complexity: High number of logic gates and/or memory; ASIC implementation is necessary. Medium Complexity: Moderate number of logic gates; either ASIC or large FPGA. Low Complexity: Low number of logic gates; can be implemented in a smaller FPGA. The complexity level determines the feasibility and cost of hardware deployment.
Software IP complexity	Real-time performance on which CPU, GPU, or SoC platforms, presence of HW acceleration Indicate which CPU, GPU, or SoC platforms can support real-time performance and whether hardware acceleration is available. Lists the platforms required to enable the software to run in real-time.
Offline Container	Supported file formats (e.g., Quicktime, MXF, MPEG2 TS, HEIF, MP4) Only list major file formats when many are supported.
Availability: Implementation	HW and SW IP availability, Indicate the accessibility and availability of hardware and software intellectual property (IP), including open-source implementation.
Availability: Licensing	Patent license information (e.g. RAND/Royalty free) Indicates what conditions are required to use the codec.
Error Resilience	Capability to correct, detect, recover, and, hide errors (CODEC specific only) Exclude transport or container mechanisms not specific to the CODEC.
Multi-resolution Support	Resolution range/flexibility for the same stream (e.g., 2^n, thumbnail) Indicates the CODEC's ability to support multiple resolutions within a single stream, including scaled-down resolutions and thumbnails.
Multi-generation Quality Loss	Quality impact upon re-encoding for 5 generations Indicate if the visual quality is impacted when the same stream is re-encoded multiple times.
Other Notable Features	Important attributes not categorized above (e.g., long-term losslessness) Includes additional features of the CODEC that don't fit into the other categories but could be useful or pertinent for ProAV users.

Table 2: List of metrics for CODEC analysis

3.2 Subjective Video Quality Metric

In the context of this document, it's important to clarify our approach to evaluating video quality. Subjective quality refers to the visual quality of a video as perceived by viewers during playback. This concept is crucial for three reasons:

1. Preference for Subjective Assessments: We prioritize subjective perceptions of quality over objective measurements. While objective data can provide initial insights, the subjective experience of viewers is paramount for assessing end-user satisfaction. Objective metrics serve as a starting point but cannot fully capture the nuances of the viewer experience.
2. Variability in Quality Perceptions: Objective metrics can fall short because video quality can vary significantly across different content types. Fast-paced action scenes, animated graphics, and slow-moving landscapes can be perceived differently, even when encoded with the same technical parameters. On the other hand, subjective evaluation assesses quality within the context of the content type and its use. This dissonance between objective and subjective metrics is particularly noticeable for computer-generated content like desktop UI, where any minor deviation in quality can have a huge impact on subjective evaluation.
3. Focus on Playback Experience: Unlike static images, videos are experienced as sequences of moving images. Frame-by-frame analysis can identify specific artifacts, but it does not represent the viewer's experience accurately. The true measure of video quality emerges during normal playback, highlighting the importance of evaluating videos as dynamic, moving sequences.

To formalize this assessment, this report employs a simplified methodology based on the Expert Viewing Protocol (EVP) presented in Part 2 Annex 8 of ITU-R BT.500 [5], with the following five-point scale:

5 - Imperceptible (Excellent) : No visible artifacts
Quality issues are virtually undetectable, even by discerning viewers.

4 - Slightly Perceptible (Good) : Minor artifacts, hardly noticeable.
Minor flaws exist but do not significantly impact the viewing experience.

3 - Perceptible (Fair) : Noticeable artifacts, but acceptable.
Defects are noticeable but may be acceptable in certain contexts.

2 - Clearly Perceptible (Poor) : Obvious artifacts, affecting quality.
Quality issues are obvious and detract from the viewing experience.

1 - Annoying (Bad) : Disturbing artifacts, unacceptable quality.
Defects are highly distracting and greatly impair the video experience.

Using a five-point scale allows for a quick assessment of a codec's potential video quality. We recognize the limitations of this simplified evaluation, which serves only as a broad outline and not a basis for detailed in-depth analysis. A more formal and complex process would be required for such purposes, which is beyond the scope of this report.

To address this limitation, references and encoded sequences of each reviewed CODEC are provided for download (see Annex C). Readers can therefore review, analyze, or compare CODECs using their preferred methodology. For further information on subjective video quality assessment, see Annexes A and B.

3.3 Glass-to-Glass Latency Metric

Glass-to-glass latency refers to the total time taken for a video signal to travel from the input source to the output display. This measurement includes both the encoding and decoding processes, using ideal transmission, capture, and display conditions. This end-to-end measurement provides a comprehensive view of the actual overall latency experienced by the end user and can be used to evaluate the overhead over an uncompressed scenario. To represent different groups of use cases, it is essential to assess various latency requirements and conditions.

We have defined three ranges of values for glass-to-glass latency to represent the three types of use cases encountered:

A. Subframe Range: For applications requiring ultra-low latency, such as interactive gaming and remote surgical procedures, where every millisecond counts.

B. Frame(s) Range: Suitable for live broadcasting and video conferencing, where a small delay is acceptable but should not disrupt the natural flow of conversation or action. Must be 5 frames or less.

C. Second(s) Range: For non-interactive applications, such as video streaming services and lecture captures, where higher latency can be tolerated without significantly impacting the user experience.

3.4 Stream Switching Delay Metric

Stream switching delay refers to the time it takes for a receiver to switch from one transmission feed to another and achieve maximum quality (artifact-free) viewing. This delay can vary depending on the alignment of refresh rates between the transmitter (TX) and the receiver (RX). When a switch occurs, the stream might initially display at a lower quality and progressively improve until it reaches the highest quality. However, this improvement doesn't always happen smoothly or predictably. For the purpose of evaluating all CODECs, we focus on the delay to reach maximum quality, assuming the worst-case alignment of refresh rates. This approach ensures a consistent and fair evaluation of the time required for each CODEC to deliver optimal viewing quality.

4. Survey of Related CODECs

Below is a summary of CODECs that the activity group members are familiar with as applicable to AV over IP in the Pro-AV market. The focus is on CODECs that are based on an industry standard, with capabilities cited relevant to the standard rather than a vendor-specific implementation. Where publicly disclosed information is available, non-industry standard CODECs are also mentioned.

This summary table includes the CODECs considered within the scope of this report. Given the numerous CODEC options, the regular introduction of new CODECs, and the existence of many proprietary CODECs, the authors do not claim this list to be exhaustive.

CODEC	Originating Group	Technology Information
AGIC3	proprietary	ASPEED Technology owned
AV1	AOMedia	AV1 Bitstream & Decoding Process Specification
AVC	ISO/IEC | ITU-T	ISO/IEC 14496-10 | ITU-T H.264
AVS3	IEEE	IEEE 1857.10
Colibri	proprietary	Audinate owned [7]
HEVC	ISO/IEC | ITU-T	ISO/IEC 23008-2 | ITU-T H.265
HTJ2K	ISO/IEC | ITU-T	ISO/IEC 15444-15 | ITU-T T.814
JPEG 2000	ISO/IEC | ITU-T	ISO/IEC 15444-1/-2 | ITU-T T.800/801
JPEG LS	ISO/IEC | ITU-T	ISO/IEC 14495-1 | ITU-T T.87
JPEG XS	ISO/IEC	ISO/IEC 21122-1/-2 [8]
NVX	proprietary	Crestron owned
SpeedHQ	proprietary	NDI owned [9]
VC-6	SMPTE	SMPTE ST-2117
VP9	Google	VP9 Bitstream & Decoding Process Specification
VVC	ISO/IEC | ITU-T	ISO/IEC 23090-3 | ITU-T H.266

Table 3: Non-exhaustive list of existing CODECs

The next Sections (4.1, 4.2, 4.3) divide these CODECs into categories: high compression, Intra-frame mezzanine, and Inter-frame mezzanine. Each category highlights features that support the comparison metrics from the previous section.

The entries are intended as high-level directional statements and should not be used to make fine-grained comparisons to determine if one CODEC is slightly superior to another. Instead, the intent is to view the CODECs from a broad perspective, identifying which types of use cases each CODEC is best suited for.

4.1 Intra-frame Mezzanine CODECs

Intra-frame mezzanine CODECs compress each video frame independently, ensuring high quality for applications like professional editing in the intermediate stage of video production. They typically achieve compression ratios ranging from 1:5 to 1:12, enabling efficient streaming while preserving editing flexibility. "Mezzanine" indicates these codecs are used for high-quality coding.

	JPEG 2000	**JPEG XS**	**HTJ2K**	**SpeedHQ ***	**AGIC3 Intra Mode**
Visual Quality	Visually lossless with traditional video content and screen content using high bitrate. Supports MLS.	Visually lossless with traditional video content and screen content using high bitrate. Supports MLS.	Visually lossless with traditional video content and screen content using high bitrate. Supports MLS.	Visually lossless with traditional video content	Visually lossless with traditional video content and screen content using high bitrate. Supports MLS.
Glass-to-Glass latency	2 frame latency is common, but tiled Ultra-Low-Latency (ULL) mode achieves subframe latency, see tradeoffs in [10]	< 32 lines, can be configured for better quality resulting in more latency	33 lines, can be configured for better quality resulting in more latency, see latency tradeoffs in [10]	1 to 2 frames	12 horizontal active lines
Stream switching latency	Next frame	Next frame	Next frame	Next frame	Next frame
Pixel format	Supports an arbitrary number of components (more than 4). Up to 38-bits.	Supports up to 4 components (RGB and YUV with alpha). Up to 12-bits. Supports direct raw Bayer/CFA up to 16-bits.	Supports an arbitrary number of components (more than 4). Up to 38-bits.	Only YUV with alpha. Up to 10-bits.	YUV 4:4:4, 4:2:2, and 4:2:0. Up to 12-bits.
Bitrate range	1.0 to 3.0bpp	1.25 to 3.0bpp	1.25 to 3.0bpp	1.5 to 2.0bpp	1.7 to 2.5bpp
Rate control	Byte-exact CBR, or VBR (Visually Lossless or Mathematically Lossless)	Byte-exact CBR, or VBR (Visually Lossless or Mathematically Lossless)	Byte-exact CBR, or VBR (Visually Lossless or Mathematically Lossless)	VBR	VBR with frame based bandwidth upper bound
Hardware IP complexity	Medium	Low	Low	Low	Low
Software IP complexity	Medium	Low	Low	Low	Low
Offline Container	MXF, MPEG2 TS, MP4, HEIF	MXF, MPEG2 TS, MP4, HEIF	MXF, MPEG2 TS, MP4, HEIF	MPEG2 TS	None currently
Availability: Implementation	SW & HW widely available, Open Source Software available	SW & HW available from various vendors for CPU, GPU, FPGA and ASIC, Open Source	SW & HW available, Open Source Software available [13]	SW implementation from various vendors available for CPU (Intel and Arm) and FPGA, Open Source	ASPEED ASIC

		Software available [12]		Software available	
Availability: Licensing	Royalty-Free	RAND License	Royalty-Free	Royalty-Free	Proprietary IP requiring a license for usage.
Error resilience	RTP format includes resync	RTP format includes resync	RTP format includes resync	No RTP, but custom UDP	Auto recover in next frame
Multi-resolution support	Yes, depending on the number of vertical wavelet transforms, start at 2, typically factor of 4 or more. RTP format includes resolution information (RES) in each packet, enabling data belonging to unneeded higher resolutions to be dropped at the packet level.	Yes, factor 2 and 4 depending on the wavelet configuration	Yes, depending on the number of vertical wavelet transforms, start at 2, typically factor of 4 or more. RTP format includes resolution information (RES) in each packet, enabling data belonging to unneeded higher resolutions to be dropped at the packet level.	Yes	No
Multi-generation quality loss	No	No	No	No	Yes
Other notable features		XS provides multi-generation robust coding (prevents quality degradation for multiple rounds of encoding decoding)		Part of a full suite AV protocol (NDI)	Support Dolby Vision standard mode.

Table 5: Overview of Intra-frame mezzanine CODECs metrics

(*) Publicly available information

4.2 Inter-frame Mezzanine CODECs

Inter-frame mezzanine CODECs are video compression algorithms used in professional video production. Unlike intra-frame codecs, which compress each frame independently, inter-frame codecs analyze and reuse data from previous frames to achieve better compression. They typically achieve compression ratios ranging from 1:5 to 1:24 depending on the type of content.

	HTJ2K Code Block caching	**JPEG XS Temporal Differential Coding (TDC)**	**Colibri ***	**AGIC3 Inter Mode**
Visual Quality	Visually lossless with traditional video content and screen content. Supports MLS.	Visually lossless with traditional video content and screen content. Supports MLS.	Visually lossless with traditional video content and screen content.	Visually lossless with traditional video content and screen content. Supports MLS.
Glass-to-Glass latency	33 lines, can be configured for more latency and better	< 32 lines, can be configured for better quality resulting in more	< 32 lines for visually lossless quality	12 horizontal active lines

	quality, see latency tradeoffs in [10]	latency		
Stream switching latency	1 to 32 frames (depends on bitrate and width)	1 to 32 frames (depends on refresh setting, bitrate and width)	1 to 120 frames (depends on refresh setting, bitrate and width)	4 frames
Pixel format	Supports arbitrary number of components (more than 4). Up to 16-bits.	Supports up to 3 components (RGB and YUV). Up to 12-bits.	Only 3 component YUV support. Up to 10-bits.	YUV 4:4:4, 4:2:2, and 4:2:0. Up to 12-bits.
Bitrate range	1.25 to 2.0bpp	1.25 to 3.0bpp	1.5 to 4.0bpp	1.0 to 1.7bpp
Rate control	Byte-exact CBR, or VBR (Visually Lossless or Mathematically LosslessIs RGB)	Byte-exact CBR, or VBR (Visually Lossless or Mathematically Lossless)	Byte-exact CBR, or VBR	VBR with frame based bandwidth upper bound, advanced GDR Intra Refresh
Hardware IP complexity	Low	Low	Low	Low
Software IP complexity	Low	Low	Low	Low
Offline Container	MXF, MPEG2 TS, MP4, HEIF	MXF, MPEG2 TS, MP4, HEIF	-	None currently
Availability: Implementation	SW available, HW planned availability	SW & HW available	HW available (single vendor)	ASPEED ASIC
Availability: Licensing	Royalty-Free	RAND License	Proprietary IP requiring a license for usage.	Proprietary IP requiring a license for usage.
Error resilience	RTP format includes resync	RTP format includes resync	Recovery on a slice alignment (32x8).	4 frames to recover
Multi-resolution support	Yes, depending on the number of vertical wavelet transforms, start at 2, typically factor of 4 or more. RTP format includes resolution information (RES) in each packet, enabling data belonging to unneeded higher resolutions to be dropped at the packet level.	Yes, depending on the number of vertical wavelet transforms, start at 2, typically factor of 4 or more. RTP format includes resolution information (RES) in each packet, enabling data belonging to unneeded higher resolutions to be dropped at the packet level. XS TDC profiles provide group and band refresh mechanisms.	Yes	No
Multi-generation quality loss	No	No	No	No
Other notable features			Screen Content Coding with long GOP structure.	Support Dolby Vision standard mode. Mathematically Lossless mode supports either on each frame or over multiple frames.

Table 6: Overview of Inter-frame mezzanine CODECs metrics.

(*) Publicly available information

4.3 High Compression CODECs

High Compression CODECs use inter-frame compression techniques to significantly reduce video streaming requirements, achieving compression ratios typically between 1:100 and 1:200 while maintaining visual quality. This capability is crucial for efficient streaming in bandwidth-limited environments.

	AVC	**HEVC**	**VVC**	**VP9**	**AV1**
Visual Quality	Lossy high-quality, In-loop deblocking filter	AVC features + Screen Content Coding, Sample adaptive offset	HEVC features + improved transform and quantization, Adaptive Loop Filtering	Lossy high-quality, Deblocking filter	VP9 features + CDEF and loop restoration filters
Glass-to-Glass latency	Ranges from 1 to many frames depending on following features: Gradual Intra Refresh, Intra-frame only, slice based	Ranges from subframe to many frames depending on following features: AVC features + tile support, wave front processing	Same as HEVC + Enhance tile and slice structure, improved WFP	Ranges from 1 to many frames depending on following features: Intra-frame only, slice based	Ranges from subframe to many frames depending on following features: VP9 features + tile support
Stream switching latency	Next frame or multi-frame depending on following features Intra-frame mode for lower, standard GOP is higher	Sams as AVC	Same as HEVC + Reference Picture Resampling	Next frame or multi-frame depending on following features Intra-frame mode for lower, standard GOP is higher	Same as VP9
Pixel format	Only 3 component YUV support. No formal 8K mode, HDR added in 2017 [14]. Up to 14-bits.	Only 3 component YUV support. Up to 16-bits.	Only 3 component YUV support. Up to 16-bits.	Only 3 component YUV support. No formal HDR or 8K mode. Up to 12-bits.	Only 3 component YUV support. Up to 12-bits.
Bitrate range	0.1 to 0.5 bpp	0.05 to 0.25 bpp	0.02 to 0.1 bpp	0.1 to 0.5 bpp	0.05 to 0.25 bpp
Rate control	GDR Intra Refresh, CBR/VBR, CRF	Same as AVC	Same as AVC	CBR/VBR, CRF	Same as VP9
Hardware IP complexity	High	High	High	High	High
Software IP complexity	Low when done hardware via GPU or SOC	Same as AVC	Same as AVC (hardware support still very limited)	Low when done hardware via GPU or SOC	Same as VP9
Offline Container	Wide support	Wide support	Wide support	WebM support	WebM support
Availability: Implementation	SW & HW widely available x264	SW & HW widely available x265 SVT-HEVC [15]**	Emerging (early SW & HW available)	SW & HW widely available WebM-VP9 [16]	SW & HW widely available AOM-AV1 [17] SVT-AV1 [18]**
Availability: Licensing	RAND License	RAND License	RAND License	Royalty-Free	Royalty-Free

Error resilience	None (must be done in software)	None (must be done in software)	None (must be done in software)	None (must be done in software)	None (must be done in software)
Multi-resolution support	SVC profile	SVC profile	SVC profile	Yes	Yes
Multi-generation quality loss	Yes	Yes	Yes	Yes	Yes
Other notable features	3D, Stereoscopic		VR360		

Table 4: Overview of high-efficiency CODECs metrics

(*) note color space and other capability support varies between venders

(**) Performance models resulting from Intel Scalable Video Technology (SVT) project [19]

5. Suitable Test Video Sequences

The working group observed that there is a sufficient supply of test sequences for consumer 4:2:0 high-efficiency CODECs, similar to those made available through Xiph [20]. However, there are noticeable gaps in test sequences that are easily accessible for AV over IP scenarios, especially concerning RGB (YUV 4:4:4) content.

As part of this report, we are making public the test sequences that have been gathered, modified, or developed by our group (see Annex C). We aim for this collection of test sequences to become a standard in the industry, helping others determine the most suitable CODEC for their specific needs.

The long-term goal for these test video sequences is to provide reference encoded outputs for all assessed CODECs (using practical, real-world hardware or software) to facilitate straightforward visual analysis and evaluation.

5.1 Tests Video Sequences Overview

The test video sequences accompanying this report have been curated from existing public databases, with minor editing (time cropping and resolution or frame changes), or have been created from scratch to meet our requirements. Besides selecting content pertinent to modern IPMX use cases, our technical goal was to cover a wide range of formats. This included UHD and FHD resolutions, with pixel formats in 4:2:0, 4:2:2, and RGB. YUV content is available in both 8-bit and 10-bit SDR and HDR, while RGB content is 8-bit SDR.

We standardized the duration of each sequence to 10 seconds to allow for useful and efficient subjective quality assessments. While objective quality assessments are usually done on very short sequences (e.g., 2 seconds), subjective assessments require a minimal length to enable viewers to properly identify, understand, and evaluate the content without being too long to become cumbersome to view repeatedly.

To ensure comprehensive testing, we included both static and dynamic content, ranging from medium to high spatial complexity. This variety helps to better stress the CODECs and provides a thorough evaluation of their performance under different scenarios.

All sequences have copyright licenses compatible with use for CODEC evaluation, with most of them using a form of Creatives Commons by attribution license CC-BY or CC-BY-SA. A copyright license file is provided for each of the sequences.

Table 7 shows the details and characteristics of the 20 sequences chosen for this report, totaling 200 seconds. Each sequence includes the pixel format, resolution, type of content, and source of the content.

Sequence name	Pixel format	Resolution	Content type	Content source
nocturnedance	4:2:0 10-bits	UHD60	natural	Netflix Open Content
meridianroad		UHD60	natural	Netflix Open Content
neon1224	4:2:0 8-bits	UHD30	natural	AV2 Test suite
bbbremix		UHD60	CG	Blender foundation
foodmarket		FHD60	natural	Netflix Open Content
aov5		FHD60	CG	AV2 Test suite
waterfall	4:2:2 10-bits	UHD60	natural	SVT 2022 Test suite
midnightsun		UHD60	natural	SVT 2022 Test suite
smoothskater		UHD30	natural	AV2 Test suite
sparks		UHD60	natural	Netflix Open Content
smokesauna		FHD60	natural	SVT 2022 Test suite
timelapse		FHD60	natural	Public domain
wallcomposite	RGB 8-bits	UHD60	desktop	VSF
medicalscan		UHD60	desktop	VSF
slideshow		UHD60	desktop	VSF
stresstest		UHD60	desktop	SVT 2022 Test suite
spreadsheet		FHD60	desktop	VSF
widget		FHD60	desktop	VSF
webpage		FHD60	desktop	VSF
inspection		FHD60	desktop	VSF

Table 7: List of sequences use for subjective assessment

5.2 Test Video Sequences Description

SVT 2022 Test Suite

midnightsun	smokesauna	stresstest	waterfall

Four sequences were taken from the SVT 2022 suite. These high-quality sequences were useful for covering 4:2:2 high-quality natural content (in SDR or HDR), and they also introduced the only synthetic test of the suite. These sequences were shortened to 10 seconds and downscaled using a high-quality scaler to the desired resolution and pixel format. For more information on the sequences, refer to the PDF document from SVT 2022 [21].

Netflix Open Content

foodmarket	meridianroad	nocturnedance	sparks

Over the years, to encourage experimentation and innovation within entertainment, technology, and academic circles, Netflix has developed test titles in documentary, live action, and animation, with open-source assets available for use [22].

Four sequences derived from this content have been included to represent high-quality cinematic content, covering either 4:2:0 or 4:2:2 in SDR or HDR.

AV2 Test Suite

aov5	neon1224	smoothskater

Three sequences from the AV2 test suite candidate [23] were used unmodified. One sequence captures gaming content in 4:2:0, while two others capture live events typically recorded by phone cameras.

Desktop Test Suite

inspection	medicalscan	slideshow	spreadsheet
wallcomposite	webpage	widget	

One type of content notably absent from public test suites covers desktop-related UI, an area vital for professional ProAV applications, which commonly exhibit very high spatial complexity combined with low temporal complexity.

A series of seven sequences were created using lossless methods to capture the HDMI output of a computer. Special care was taken to ensure that the content has licenses compatible with public testing of CODECs.

Miscellaneous Content

bbbremix	timelapse

The iconic animated short, Big Buck Bunny [24], has been included. To increase complexity and enhance the testing scenario, a composite video using multiple panels and sequences was created. Additionally, a timelapse created from a set of pictures was included to provide a different type of content for evaluation.

6. Pro-AV CODECs Assessment

This section details the methodology and process used to assess a selected group of Pro-AV CODECs. Each identified lead use case (Section 2, Table 1) was evaluated using the metrics previously outlined (Section 3, Table 2).

Assessment Goals

The primary aim of these assessments is to showcase the potential of different CODECs using a common set of metrics. By employing a well-defined and documented methodology, we ensured that the assessment was conducted using consistent criteria. This structured approach helps mitigate bias and provides a more objective evaluation of the CODECs.

The goal is not to select a winning CODEC or by that measure, compare one CODEC versus another, but to present an assessment of each CODEC and profile in isolation.

Assessment Methodology

The suitability of CODECs for each use case is evaluated by tabulating the use cases (common themes) as columns and the metrics as rows. With this format, the group's work aims to provide a clear visual summary, offering a high-level view of each CODEC's main attributes and enabling quick understanding of their strengths and suitability for specific use cases. This approach ensures that stakeholders can efficiently assess the potential of various CODECs for specific use cases of interest.

CODECs Selection

For this assessment, multiple CODECs were selected, and contributors were identified to do a self-assessment and provide the necessary information for this report. Contributors, chosen for their expertise in CODECs, are members of the VSF participating in the working group. They were responsible for compiling information for each CODEC profile, which was then reviewed for uniformity and integrity by the working group.

These assessments were not exhaustive. Instead, we focused on CODECs that showed potential for Pro-AV use cases, were open in nature through a standardized or publicly available specification, and had a readily available implementation.

CODECs Profile Identification and Contributor

Each CODEC was evaluated across one or more profiles. A profile is a set of CODEC parameters optimized for specific use cases. Although we aimed to limit the number of profiles, multiple profiles were defined to adequately cover various use cases and demonstrate the CODECs' flexibility. These profiles, including both their names and definitions, were proposed by contributors and refined with working group feedback, and should be used within the scope of the report and not be viewed as official or definitive.

The assessed CODECs, their profiles, and the contributors conducting the self-assessments are listed in Table 8.

Assess CODEC	CODEC Profiles	Contributor
AGIC3	INTRA INTER	Bruce Chang, ASPEED Technology
H.264	INTRA_UltraLowLatency INTER_LowLatency INTER_LongGOP	Alain Champenois, Matrox Video
H.265	INTRA_UltraLowLatency INTER_LowLatency INTER_LongGOP	Alain Champenois, Matrox Video
HTJ2K	INTRA CODEBLOCKCACHING	Michael Smith, Kakadu Software
JPEG2000	INTRA	Michael Smith, Kakadu Software
JPEG XS	XS_HIGH XS_TDC	Tim Bruylants, intoPIX

Table 8: List of assessed CODEC

CODEC Assessment Deliverables

Each CODEC profile assessment includes:

- **Metrics Spreadsheet:** A comprehensive table displaying the metrics assessed values of the CODEC profile for all identified lead use cases.
- **Encoded Test Sequences:** A set of 20 reference test sequences, encoded and decoded by the CODEC profile, provided in the same RAW format as the reference sequence.

The Metrics Spreadsheets provide a quick overview of each CODEC profile's capabilities and suitability for different use cases. In parallel, the encoded sequences enable readers to independently conduct their own analysis, allowing for more in-depth evaluation and comparison using their own metrics.

For details on how to fill out the spreadsheet, create the encoded test sequences, or download the assessment results, refer to Annex C.

7. Report Summary

This report summarizes the efforts of the VSF CODEC activity group in assessing Pro-AV CODECs for various use cases. The group's objective was to address the lack of established standards for selecting CODECs. While not endorsing any specific CODEC, the group aimed to present findings and trends to assist the Pro-AV industry in making informed decisions.

We first discussed use cases in the Pro-AV sector to establish CODEC requirements (Section 2) and outlined key decision factors involved in selecting the appropriate CODEC (Section 3). A comprehensive survey of selected CODECs was conducted (Section 4), followed by the compilation of suitable test vectors for validating the CODECs (Section 5). Finally, an assessment of different CODECs based on the identified criteria and scenarios was performed (Section 6), with all the resulting data provided to the readers (Annex C).

With this report, the VSF CODEC activity group has consolidated valuable knowledge into a single document. Key contributions include:

- **Enhanced Pro-AV Use Case Framework:** Substantial pre-existing know-how was identified, such as the JT-NM Use Cases. Specific improvements were made to evaluate CODEC suitability for AV over IP implementations like IPMX, with new themes and lead use cases added to the JT-NM Use Cases.
- **Curated Pro-AV Test Sequences:** The group identified and curated 20 test sequences specifically tailored for Pro-AV use cases, including the creation of eight desktop-oriented RGB sequences, a type of content uncommon in other sequence sets. These sequences were made publicly available to increase Pro-AV content usage outside VSF.
- **Detailed Pro-AV CODEC Assessment Metrics:** Comprehensive metrics were developed for assessing CODEC suitability for Pro-AV applications. These metrics, along with defined use cases, were used to summarize candidate CODECs and conduct a tabulated analysis. This approach provides valuable guidance for vendors in self-testing future CODECs.
- **Broad Pro-AV CODEC Assessment:** Five CODECs with 11 profiles were assessed using these metrics, and the encoded sequences have been made available for review.

We hope this report will be useful and equip stakeholders with the necessary knowledge and resources to make informed decisions about CODEC implementation in Pro-AV environments.

8. Future Work

This report offers a snapshot of the currently considered CODEC options. As new CODECs emerge, the authors anticipate that the presented approach can be used to evaluate them. Future work will involve applying this methodology to new CODECs, ensuring that the assessment process remains relevant and comprehensive alongside existing results.

Some key technical aspects were identified during the development of this report but were left unexplored on purpose because they fell outside the scope of this report:

Standard profiles for all CODECs
Technical definitions of standard profiles for each CODEC were not always available, and generic profiles were used in some cases instead during the assessment.

Requirements for CODEC Acceptance
Criteria for CODEC acceptance were not addressed.

Network Requirements
Detailed explanations of network requirements and behaviors influencing CODEC performance were not explored.

Additional Metrics
Specialized metrics such as energy efficiency, security, and support for non-standard resolutions (e.g., VR, 3D) were not investigated.

Additional Use Cases
New Pro-AV use cases that impact CODEC suitability could be added in the future.

Annex A: ITU-T Subjectives Assessments Methods

The ITU-T (International Telecommunication Union Telecommunication Standardization Sector) provides several recommendations concerning the subjective assessment of video quality.

The two most relevant documents from ITU-T for video assessment are:

- **ITU-R BT.500 [5]**
  Methodologies for the subjective assessment of the quality of television images
  This recommendation covers various assessment methods for different types of content and transmission methods.
- **ITU-T P.910 [6]**
  Subjective video quality assessment methods for multimedia applications
  This recommendation specifically covers multimedia applications, including video conferencing and telemedicine.

While both recommendations were created nearly 30 years ago, they have been regularly updated since then. These recommendations provide methodologies for the assessment of video quality, including general testing methods, the grading scales used during assessments, and the recommended viewing conditions for carrying out assessments.

Included in this annex is a detailed summary of the primary test methods covered in ITU BT.500, providing an overview of the procedures and techniques used for assessing the quality of television pictures, which are also valid for modern display technologies.

A.1 Subjective Assessments Methods

Single-stimulus (SS) methods
Single stimulus methods involve viewing an impaired sequence and assigning it a grade. This method is also referred to as Absolute Category Rating (ACR). A common variation, ACR-HR (Absolute Category Rating with Hidden Reference), includes grading the unimpaired sequence without informing the observer.

Single stimulus continuous quality evaluation (SSCQE)
Similar to single stimulus methods, continuous stimulus methods involve viewing a sequence and providing a continuous grade throughout the sequence using a slider. This method allows for longer sequence viewing and more detailed feedback.

Double-stimulus impairment scale (DSIS)
Double-stimulus methods involve viewing the unimpaired sequence followed by the impaired sequence, and then giving a grade to the impaired sequence relative to the unimpaired sequence. This method is preferred by the EBU (European Broadcasting Union).

Double-stimulus continuous quality-scale (DSCQS)
In this method, the unimpaired and impaired sequences are shown one after the other in an unknown order, and both sequences are graded using a continuous scale. The difference in scores between the unimpaired and impaired sequences is then used for evaluation.

Stimulus-comparison methods
This method is used to compare two sequences by grading one against the other. Viewing can be done alternately on a single display or simultaneously on two similar displays. This method is also known as the Pair Comparison method.

Simultaneous double stimulus for continuous evaluation (SDSCE)
The unimpaired and impaired sequences are viewed simultaneously, and a continuous grade is given to the impaired sequence. When the fidelity is perfect, the slider should be at the top of the scale range (coded 100). When the fidelity is null, the slider should be at the bottom of the scale (coded 0).

Subjective assessment of multimedia video quality (SAMVIQ)
This method was created specifically to address multimedia applications, where content and viewing conditions can differ greatly from television content. It involves viewing impaired sequences (one or many) and continuously grading each of them. Multiple viewings are possible, and grades can be changed. The reference unimpaired sequence can be viewed at any time. Each observer moves a slider on a continuous scale graded from 0 to 100, annotated with five quality levels arranged linearly: excellent, good, fair, poor, and bad.

Expert viewing protocol (EVP)
This method is fine-tuned for use by a small number of expert reviewers. The unimpaired sequence is followed by one or more impaired sequences, which are then graded in comparison to the unimpaired sequence.

A.2 Test Methods Comparison Table

Table 9 summarizes the characteristics of each subjective assessment methods as define in BT.500

Method	Description	Scale	Advantages	Disadvantages
Single-Stimulus (SS)	View an impaired sequence and assign a grade. ACR-HR includes grading the unimpaired sequence without notice.	5-level scale	Simple, quick	Limited detail
SSCQE	Continuous grading throughout the sequence using a slider.	Continuous scale	Detailed feedback, longer viewing	More complex

DSIS	View unimpaired then impaired sequence, grade the impaired relative to unimpaired. Preferred by EBU.	11-level scale	Effective for major impairments	Requires clear identification
DSCQS	View unimpaired and impaired sequences in unknown order, grade both, use score difference for evaluation.	Continuous scale	Evaluates minor impairments well	More time-consuming
Stimulus-Comparison Methods	Compare two sequences by grading one against the other. Viewing can be alternate or simultaneous.	Pair comparison	Direct comparison	Requires two displays or alternation
SDSCE	View unimpaired and impaired sequences simultaneously, give a continuous grade to the impaired sequence.	Continuous scale (0-100)	Accurate fidelity assessment	More complex setup
SAMVIQ	View impaired sequences, continuously grade each, multiple viewings and grade changes allowed. Reference viewable at any time.	Continuous scale (0-100)	Flexible, detailed feedback	More complex, time-consuming
Expert Viewing Protocol (EVP)	Unimpaired sequence followed by one or more impaired sequences, graded in comparison to the unimpaired sequence.	Various scales	High precision, expert judgment	Requires expert viewers

Table 9: BT.500 subjective assessment methods overview

Annex B: Subjective Evaluation Case Studies

This annex includes information on several subjective video assessment case studies and evaluations conducted over the past two decades.

B.1 Subjective Evaluation of Next-Generation Video Compression Algorithms

In 2010, the Joint Collaborative Team on Video Coding (JCT-VC) conducted a subjective video assessment of 27 coding technologies under evaluation for the HEVC standard [25]. Based on the ITU-R BT.500 recommendations, this assessment used two methods: the Double Stimulus Impairment Scale (DSIS) and the Double Stimulus Continuous Quality Scale (DSCQS).

Double Stimulus Impairment Scale (DSIS):

- **Methodology:** Pairs of sequences, stimulus A and B, are sequentially presented to the subject. The subject is informed that stimulus A is the reference video, expected to have the best quality, and is asked to rate the level of annoyance of the visual defects observed in stimulus B.
- **Usage:** This method is useful for assessing the quality of test material with major impairments.
- **Rating Scale:** The rating scale ranges from 0 to 10, from very annoying to imperceptible.

Double Stimulus Continuous Quality Scale (DSCQS):

- **Methodology:** Pairs of sequences, stimulus A and B, are presented twice sequentially to the observer. The subject is not informed that stimulus A is the reference video. The subject is then asked to rate the quality of both stimuli.
- **Usage:** This method is useful for assessing the quality of test material with minor impairments.
- **Rating Scale:** The rating scale goes from 0 to 100, bad to excellent.

These methodologies provided a structured approach to evaluating the coding technologies, ensuring consistent and reliable assessments of video quality.

B.2 Evaluation of HEVC Screen Content Coding Extensions

In 2017, the JCT-VC ITU-T/ISO joint working group, responsible for developing the High Efficiency Video Coding (HEVC) standard, conducted a subjective assessment for the Screen Content Coding (SCC) extension [26]. The SCC extension was developed to improve compression for videos containing a significant portion of rendered (moving or static) graphics, text, or animation, in addition to camera-captured video scenes.

Expert Viewing Protocol (EVP) Structure:

1. A mid-grey screen showing the letter “A” in the middle (1 second).
2. The uncompressed source video clip (original, not coded).
3. A mid-grey screen showing the letter “B” in the middle (1 second).
4. The coded video clip to evaluate.
5. A mid-grey screen showing the message “Vote N”, where N is a progressive number indicating the BTC number to vote.

The method used was the Degradation Category Rating (DCR) with an 11-grade impairment scale, ranging from '0' (lowest quality) to '10' (highest quality).

This structured approach allowed for a consistent and reliable evaluation of the SCC extension, ensuring that the improvements in compression for screen content were accurately assessed.

B.3 Comparison of Subjective Methods for Quality Assessment

In 2019, a research group from CNRS (French National Centre for Scientific Research) produced two papers [27] [28] evaluating methods from BT.500 for subjective video assessment of 3D models viewed using a VR headset. Three methods were evaluated:

Absolute Category Rating with Hidden Reference (ACR-HR): Impaired sequences and the unimpaired reference sequence are shown without informing the observer about the reference. Each sequence is graded on a 5-level scale.

Double Stimulus Impairment Scale (DSIS): The unimpaired and impaired sequences are shown one after the other (each properly identified). The observer grades the impaired sequence using an 11-level scale. For the study, both sequences were shown side by side.

SAMVIQ (Subjective Assessment of Multimedia Video Quality): The unimpaired and impaired sequences are presented one at a time. The observer can review each video multiple times and modify the quality score. The impaired sequence is graded using a continuous quality scale (0-100).

These papers conclude that having an explicit reference is better when evaluating the quality of material unknown to the observer. The DSIS method is identified as the most accurate and efficient method.

B.4 New subjective Quality Evaluation using screen-splitting matrix

A 2019 paper [29] presents a new concept for video presentation in subjective quality evaluation methods with a reference given in parallel. The idea is to split a full-resolution video into an n-picture matrix and encode each cell differently to compare aspects of processing, such as encoding parameters or lossy compression algorithms.

Key Findings screen-splitting matrix:

- **Accuracy:** Experiments conducted show that this method yields more accurate results.
- **Efficiency:** The time required for evaluation is shortened by at least half.
- **Cost-Effectiveness:** The experimental setup is less complicated and cheaper.

This innovative approach allows for more efficient and precise subjective quality evaluations, enhancing the ability to compare various encoding techniques and compression algorithms.

B.5 New Subjective Quality Evaluation using triple stimulus

A 2008 paper [30] proposes a novel method for the assessment of picture quality called the Triple Stimulus Continuous Evaluation Scale (TSCES). This method allows for the direct comparison of different HDTV formats.

Key Features of TSCES:

- **Triple Stimulus:** The method involves three sequences: one reference and two test sequences.
- **Continuous Evaluation:** Observers provide continuous quality ratings throughout the viewing period.
- **Direct Comparison:** This approach facilitates the direct comparison of different HDTV formats by allowing simultaneous evaluation against a reference.

The TSCES method aims to improve the accuracy and reliability of subjective quality assessments for HDTV formats, providing a more robust framework for evaluating picture quality.

B.6 New Subjective Quality Evaluation SAMVIQ

A 2005 article from EBU [31] evaluated the performance of the newly created SAMVIQ (Subjective Assessment Methodology for Video Quality) evaluation methodology. This methodology was recently used during B/VIM’s Phase 2 subjective evaluations of four codecs designed for internet use.

Key Features of SAMVIQ:

- **Flexibility:** Allows observers to view and re-evaluate sequences multiple times.
- **Continuous Scale:** Uses a continuous quality scale from 0 to 100.
- **Reference Viewing:** Observers can view the reference video at any time.

The SAMVIQ method is effective in discriminating between various quality levels, from low to high, and is simpler, faster, and more user-friendly than traditional subjective evaluation methods.

Annex C: Submission Assessment Information

C.1 Submission Process for CODEC Assessment

A. Contact the VSF to request permission and guidelines for doing a CODEC submission.
B. Download the template metrics spreadsheet and test video sequences.
C. Complete the template metrics spreadsheet.
D. Receive feedback from VSF on the metrics spreadsheet and make adjustments if needed.
E. Generate the encoded sequence outputs for each profile.
F. Upload both the metrics spreadsheet and encoded sequences to the VSF cloud storage.

Once a submission is completed, the VSF Pro-AV CODEC Landscape report will be updated with a reference to the new submission.

C.2 Inputs for CODEC Assessment

Two types of input files exist and will be provided when making a submission.

Template Metrics Spreadsheet
The spreadsheet is composed of an introduction sheet providing more information about the filling process, a CODEC information page, and a template metric sheet. For more information about the metrics, see Section 3.

Test Video Sequences
Twenty test video sequences are provided in RAW format. Information concerning the pixel format, resolution, framerate, and dynamic range of each sequence can be found in the filename. For more information about the metrics, see Section 5.

See Table 8 and 9 for mapping the filename keywords.

Format Name	Pixel Format	Pixel Format Description
yuv420p	YUV 4:2:0 8-bits	planar 8-bits
yuv420p10le	YUV 4:2:0 10-bits	planar 16-bits (10-bits LSB)
yuv422p10le	YUV 4:2:2 10-bits	planar 16-bits (10-bits LSB)
rgb24	RGB 8-bits	packed 24-bits (blue is LSB)

Table 10: Test sequence format list

Resolution name	Resolution Size (pixels)	Framerate (Hz)
UHD60	3840x2160	60
UHD30	3840x2160	30
FHD60	1920x1080	60

Table 11: Test sequence resolution list

C.3 Deliverables for CODEC Assessment

When doing an assessment, the contributor needs to define one or more pertinent profiles for a CODEC submission. A profile represents a set of defined parameters to be used by the encoder. Having more than one profile serves to better showcase the CODEC for different scenarios, but it is not a requirement.

Metrics Spreadsheet
The contributor will fill out the CODEC information sheet with details about the profile parameters and the environment used for the assessment encoded sequence generation. For each profile, the contributor will complete a metrics sheet detailing the general behavior of the CODEC across different use cases. This is not a detailed technical description of the CODEC but rather a high-level view of its capabilities.

Encoded Sequences Outputs
For each profile, the contributor will generate encoded sequences at a bitrate of 1.7 bpp using the test video sequences. The encoded sequences will be provided in decoded RAW form (.yuv or .rgb), compressed into a .zip file. The pixel format of the encoded sequences will match the test video sequences, maintaining the same number and order of frames. The original sequence filename will have a suffix added to denote the codec and profile name.

Ex: Input sequence midnightsun_UHD60_yuv422p10le.yuv
Encoded sequence: midnightsun_UHD60_yuv422p10le_CODECNAME_PROFILENAME.yuv.zip

The encoded sequences can be created using software or by performing a hardware capture of the decoded images. The goal of the encoded sequences should be to represent the normal quality level of the CODEC.

C.4 VSF Cloud Storage

Test Video Sequences download path:
https://www.dropbox.com/scl/fo/jl27mywo4ats6xkvdzhc4/APv9wAtPJAO9oiF_IgEzaQk?rlkey=v89m20yha3qsddbg3yzbtkj07&st=sp5aouip&dl=0

Encoded Sequence Outputs download path:
https://www.dropbox.com/scl/fo/9xem7qxepc9ngaouv9oxa/AJMIyzB7DZbedqoNOih_wJY?rlkey=019uql3g0u2brwrgqaqgnyb01&st=r8kkrv47&dl=0

References

[1] JT-NM-ProAV-User-Requirements. Retrieved from https://github.com/AMWA-TV/pro-av-user-requirements

[2] Zia, B., Guruvareddiar, P., Chua, Y. T., Guo, X., Jian, S., Lattanzio, L., & Knoplioch, J. (2023). "Low-Latency Lossless Compression White Paper." Intel. Retrieved from https://ieeexplore.ieee.org/document/10125475

[3] Liu, F., Hernandez-Cabronero, M., Sanchez, V., Marcellin, M. W., & Bilgin, A. (2017). "The Current Role of Image Compression Standards in Medical Imaging." Information (Basel), 8(4), 131. doi:10.3390/info8040131. Retrieved from https://www.mdpi.com/2078-2489/8/4/131

[4] Cope, B., Kai, B. J., Ching, Y. C., & Qiang, L. Z. (2023, June). "Open Standard Software Defined AV over IP Framework for LED Video Wall System." In SID Symposium Digest of Technical Papers (Vol. 54, No. 1, pp. 425-428).

[5] ITU-R BT.500-15 (05/2023). "Methodologies for the Subjective Assessment of the Quality of Television Images." International Telecommunication Union. Retrieved from https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.500-15-202305-I!!PDF-E.pdf

[6] ITU-R BT.500-15 (05/2023). "Methodologies for the Subjective Assessment of the Quality of Television Images." International Telecommunication Union. Retrieved from https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-P.910-202310-I!!PDF-E

[7] "Audinate to acquire Silex Insight video business.". Retrieved from https://www.audinate.com/news/press-releases/audinate-to-acquire-silex-insight-video-business

[8] Burrows, D. (2023). "Evaluating the Suitability of NDI SpeedHQ for High-Quality Broadcast Workflows." BT Wholesale. Retrieved from https://pages.btwholesale.com/rs/827-CTB-381/images/MB-Evaluating-SpeedHQ-Whitepaper.pdf

[9] Descampe, A., et al. (2021). "JPEG XS—A New Standard for Visually Lossless Low-Latency Lightweight Image Coding." Proceedings of the IEEE, 109(9), 1559-1577. doi:10.1109/JPROC.2021.3080916. Retrieved from https://ds.jpeg.org/whitepapers/jpeg-xs-whitepaper.pdf

[10] Edwards, T., & Smith, M. D. (2021). "High Throughput JPEG 2000 for Broadcast and IP-Based Applications." SMPTE Motion Imaging Journal, 130(4), 22-35. doi:10.5594/JMI.2021.3066183. Retrieved from https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9424052

[11] Richter, T., Fößel, S., Descampe, A., & Rouvroy, G. (2021). "Bayer CFA Pattern Compression With JPEG XS." IEEE Transactions on Image Processing, 30, 6557-6569.

doi:10.1109/TIP.2021.3095421. Retrieved from https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9488288

[12] SVT-JPEG-XS Encoder and Decoder. Retrieved from https://github.com/OpenVisualCloud/SVT-JPEG-XS

[13] Open-source implementation of HTJ2K. Retrieved from https://github.com/aous72/OpenJPH

[14] ITU-T H.264 Specification. (2017). International Telecommunication Union. Retrieved from https://www.itu.int/rec/T-REC-H.264-201704-S

[15] SVT-HEVC Encoder. Retrieved from https://github.com/OpenVisualCloud/SVT-HEVC

[16] The WebM Project. Retrieved from https://www.webmproject.org/code/

[17] AOMedia AV1 Reference Software. Retrieved from https://aomedia.googlesource.com/aom

[18] SVT-AV1 Encoder and Decoder. Retrieved from https://github.com/BlueSwordM/SVT-AV1

[19] Intel SVT Software for Video Codec. Retrieved from https://www.intel.com/content/www/us/en/developer/articles/technical/scalable-video-technology.html

[20] Xiph.org Test Sequences. Retrieved from https://media.xiph.org/

[21] SVT Open Content Test Suite 2022 PDF. (2022). Retrieved from https://www.svt.se/open/docs/SVT_Open_Content_Video_Test_Suite_2022_Natural_Complexity_v1-3-reduced.pdf

[22] Netflix Open Source Content. Retrieved from https://opencontent.netflix.com/

[23] AV2 Candidate Test Media. Retrieved from https://media.xiph.org/video/av2/

[24] "Big Buck Bunny." Blender Foundation. Retrieved from https://peach.blender.org/

[25] De Simone, F., Goldmann, L., Lee, J.-S., & Ebrahimi, T. (2010). "Subjective Evaluation of Next-Generation Video Compression Algorithms: A Case Study." Proceedings of SPIE - The International Society for Optical Engineering. doi:10.1117/12.863485. Retrieved from https://www.researchgate.net/publication/45657741_Subjective_Evaluation_of_Next-Generation_Video_Compression_Algorithms_A_Case_Study

[26] Baroncini, V., Yu, H., Joshi, R., Liu, S., Xiu, X., & Xu, J. (2017). "Verification Test Report for HEVC Screen Content Coding." ISO/IEC JTC1/SC29/WG11/N16882. Retrieved from https://mpeg.chiariglione.org/standards/exploration/future-video-coding/n16882-verification-test-report-hevc-screen-content-coding

[27] Nehme, Y., Farrugia, J.-P., Dupont, F., Le Callet, P., & Lavoué, G. (2019). "Comparison of subjective methods, with and without explicit reference, for quality assessment of 3D graphics." ACM Symposium on Applied Perception, Sep 2019, Barcelona, Spain, 1-9. doi:10.1145/3343036.3352493. Retrieved from https://hal.science/hal-02297220v2/document

[28] Nehme, Y., Farrugia, J.-P., Dupont, F., Le Callet, P., & Lavoué, G. (2020). "Comparison of Subjective Methods for Quality Assessment of 3D Graphics in Virtual Reality." ACM Transactions on Applied Perception, 18(1), 1-23. doi:10.1145/3427931. Retrieved from https://www.researchgate.net/publication/348156299_Comparison_of_Subjective_Methods_for_Quality_Assessment_of_3D_Graphics_in_Virtual_Reality

[29] Ostaszewska-Liżewska, A., & Kłoda, R. (2019). "Subjective Quality Evaluation of a Full Resolution Video with Eligible Number of Reference Pictures." Pomiary Automatyka Robotyka, 23(1), 17-20. doi:10.14313/PAR_231/17. Retrieved from https://www.par.pl/content/download/14881/182956/file/SubjectiveQualityEvaluationFulResolutionVideoEligibleNumber.pdf

[30] Hoffmann, H., Itagaki, T., Wood, D., Hinz, T., & Wiegand, T. (2008). "A Novel Method for Subjective Picture Quality Assessment and Further Studies of HDTV Formats." IEEE Transactions on Broadcasting, 54(1), 1-13. doi:10.1109/TBC.2008.916833. Retrieved from https://www.researchgate.net/publication/3042038_A_Novel_Method_for_Subjective_Picture_Quality_Assessment_and_Further_Studies_of_HDTV_Formats

[31] Kozamernik, F., Sunna, P., Wyckens, E., & Pettersen, D. I. (2005). "Subjective Quality of Internet Video Codecs - Phase 2 Evaluations Using SAMVIQ." EBU. Retrieved from https://tech.ebu.ch/docs/techreview/trev_301-samviq.pdf

