AVHD-AS Project Synopsis

Version 3.0

March 3, 2016

|  |  |
| --- | --- |
| Contacts: |  |
| Margaret Pinson, NTIA | mpinson@its.bldrdoc.gov |
| Chris Schmidmer, OPTICOM GmbH | cs@opticom.de |
|  |  |

Editorial History

|  |  |  |
| --- | --- | --- |
| Version | Date | Nature of the modification |
| 1.0 | Sep. 15, 2015 | Initial Draft, created by C. Schmidmer |
| 3.0 | March 2, 2016 | Editing at VQEG meeting, C. Schmidmer |
|  |  |  |

Contents

[1 . Introduction 1](#_Toc444807812)

[2 . Project Synopsis 2](#_Toc444807813)

[2.1 Objectives and Application Areas 2](#_Toc444807814)

[2.2 Model Types 2](#_Toc444807815)

[2.3 Source Signal Video Properties 2](#_Toc444807816)

[2.4 Target Distortions 2](#_Toc444807817)

[2.5 Model Input 3](#_Toc444807818)

[2.6 Model Validation 4](#_Toc444807820)

[2.7 Subjective Test Method 4](#_Toc444807821)

[2.8 AVHD-AS Data Analysis 5](#_Toc444807822)

[2.9 General Procedure 5](#_Toc444807823)

[2.10 Model Disclosure 5](#_Toc444807824)

[2.11 Relation to other Standardization Activities 5](#_Toc444807825)

[2.12 Duties and Responsibilities 5](#_Toc444807826)

[2.13 Schedule 5](#_Toc444807827)

[2.14 AVHD-AS Fees 6](#_Toc444807828)

[2.15 Detailed Documentation 6](#_Toc444807829)

[2.15.1 AVHD-AS Project Synopsis (this document) 6](#_Toc444807830)

[2.15.2 VQEG Definitions 6](#_Toc444807831)

[2.15.3 AVHD-AS Detailed Description of Test Conditions 7](#_Toc444807832)

[2.15.4 AVHD-AS Subjective Test Method 7](#_Toc444807833)

[2.15.5 AVHD-AS Model Requirements 7](#_Toc444807834)

[2.15.6 AVHD-AS Test Plan 7](#_Toc444807836)

[2.15.7 AVHD-AS Validation Test Designs 7](#_Toc444807839)

# . Introduction

This document contains an executive summary of the VQEG Adaptive Streaming project (AVHD-AS). All content is plain informational. A more detailed description of the individual aspects of the project is contained in separate documents. A description of these documents as well as their current versions is contained in this synopsis in section 2.13. In case of ambiguities those documents supersede the project synopsis presented here. It is planned to perform the project in phases. The first phase is described in here and may be followed by a characterization phase where the validity of the models for applications which were not previously tested can be explored. In any case, a second phase is planned for entirely new applications, like e.g. UHD. Planning for future phases shall start in parallel to the work of already ongoing phases.

# . Project Synopsis

##  Objectives and Application Areas

The goal of the AVHD-AS project is to validate objective methods for the assessment of adaptive bitrate streaming services like YouTube, Netflix, Hulu or Amazon Instant Video from an end users perspective.

The models benchmarked are for video only; audio will be excluded from all sequences. Audio-visual integration is subject of a different VQEG project.

Targeted scenarios include, but are not limited to streaming of video sequences over fixed and mobile networks, with a focus on consumer entertainment. The objective models are meant to be applied to sequences of up to 5 min duration. The models should be based on general perceptual concepts and not be depending on specific protocols which may change very often. It is assumed that the protocols used by the services of interest are packet loss free on the application layer. It is also assumed that access to the unencrypted video signal is available on a pixel level.

##  Model Types

Model types submitted for evaluation may comprise no-reference (NR), reduced reference (RR) as well as full reference (FR) methods. Hybrid models may also be evaluated if the time frame permits this and a reasonably abstract but still relevant definition for the bitstream can be found (tentative agreement, details to be defined).

##  Source Signal Video Properties

Video resolutions under study will be 1080p. Deinterlacing 1080i to 1080p is allowed in order to broaden the available footage, as long as no strong artifacts are introduced by this. The video frame rate shall be between 23 and 60 fps. The color formats used are YUV420 format. As a general rule of thumb, SRCs should have a quality of “good” or “better” as scored by expert viewers. The duration of the source sequences shall be 1 to 5 min.

##  Target Distortions

The models shall be able of handling a wide range of distortions typical for the application. These include changes of bitrate, resolution and framerate as well as initial buffering and stalling (rebuffering). Quality changes due to bitstream switching may occur not more than once per second. Note that bitrate and resolution may change simultaneously, but the coding scheme used will remain the same within each video sequence. Coding schemes which are currently discussed for use in this study are H.264, H.265 and VP9. Example video bitrates and resolutions used are shown in Table 1 below. The video frame rate may switch between 12 and 60 fps.

|  |  |  |  |
| --- | --- | --- | --- |
| **Bitrate [Mbit/s]** | **Resolution** | **Framerate** | **Codec** |
| **10** | **1920x1080** | **24** | **H.264** |
| **6** | 1280x720 | 24 | H.264 |
| **2.5** | 1280x720 | 24 | H.264 |
| **2** | 720x404 | 24 | H.264 |
| **1.6** | 1280x720 | 24 | H.264 |
| **0.75** | 682x384 | 24 | H.264 |
| **0.45** | 512x288 | 24 | H.264 |
| **0.25** | 426x240 | 24 | H.264 |
| **0.1** | 256x144 | 12 | H.264 |
|  |   |   |   |
| **1.8** | 1920x1080 | 24 | VP9 |
| **0.9** | 1280x720 | 24 | VP9 |
| **0.45** | 854x480 | 24 | VP9 |
| **0.25** | 640x360 | 24 | VP9 |
| **0.13** | 426x240 | 24 | VP9 |
| **0.07** | 256x144 | 12 | VP9 |

Table , examples for combinations of bitrates and resolutions

The total stalling duration shall be limited to 25% of the source sequence duration. Initial buffering should not exceed 30 s.

##  Model Input

Input to the models will be:

* The source video sequence (FR and RR models only), or in the case of data captured from real services, the stream representation with the highest available quality (this quality must be “good” or better as scored by expert viewers.).
* The decoded video sequence as it is seen by the observer in the subjective experiment (the PVS), but without the periods of stalling or initial buffering.
* A file with information on initial buffering and stalling events.
* Additional input to Hybrid models are the coding scheme and a file containing for each GOP the resolution (height, width), size (in bytes), and the duration in seconds.

##  Model Validation

The scores produced by the models will be compared to MOS scores acquired in subjective tests. The subjective test designs and SRCs will be secret and unknown to the model developers until all proposed models have been submitted. Subjective test design and SRC selection will be controlled by the ILG. It is under discussion if subjective tests shall be conducted by proponents and/or ILG members. In case that proponents are responsible for subjective experiments, they are free to outsource subjective testing.

##  Subjective Test Method

The subjective test method to be used is ACR according to P.913 using a reasonably controlled environment which can be reproduced. Especially light sources should be controlled by adhering to a minimum and maximum luminance level. Deviations from P.913 may be acceptable, but must be discussed with other ILG members and proponents. The final decision whether such a deviation is acceptable shall be up to the ILG.

Acceptable viewing devices are

* TVs with a diagonal between 40’’ and 55’’ diagonal
* Computer Monitors between 27’’ and 32’’
* Smart phones / tablets with screens between 4” and 10”
* Use of notebook computers with screens between 13’’ and 17’’ is under discussion for phase 2.

It is assumed that subjective results for viewing using a Monitor and TV are sufficiently similar. It is therefore not foreseen to perform separate validations for these two classes of viewing devices. It is also not foreseen to investigate the impact of different monitors.

The use of notebook computers as viewing devices is planned to be investigated either in phase 2, or in a separate characterization phase.

The test conditions should be reasonable for the display size and resolution. As a guideline, bitrates between 75 and 500 kbps are more appropriate for small screens (e.g., smart phones & tablets) than large screens (e.g., 40” to 55” TV); and bitrates between 2 and 10 Mbps are more appropriate for large screens than small screens. Content coded at such bitrates is less likely to be streamed to the other devices. Some excursions beyond these limits may be considered for inclusion in tests. [Editor’s note: The number of monitors and test conditions must be determined when the number of subjective tests has been decided (or at least estimated).]

##  AVHD-AS Data Analysis

The final report from VQEG will list all KPIs which are defined in P.1401 in so far, as they are valid for the applied subjective test method. Emphasis shall be on RMSE based metrics, which means that models should be designed in order to minimize rmse and not to maximize correlation. Other, supplementary metrics may be provided, but must be agreed upon between proponents and ILGs.

##  General Procedure

The model development will be performed in a competitive way, but proponents will cooperate in order to validate the models.

##  Model Disclosure

One clear objective of VQEG is that the benchmark shall lead to the standardization of one or more of the tested models by standardization organizations (e.g. ITU). This may involve the need for each proponent to fully disclose its model and to license its related IPR under reasonable and non-discriminatory terms when it is accepted for standardization.

##  Relation to other Standardization Activities

ITU-T SG 9, ITU-T SG 12, ITU-R WP6C

##  Duties and Responsibilities

It is desired that model development and model validation will be handled by the proponents under supervision of an independent lab group (ILG). The procedure will be designed in a way that no proponent will have an advantage over the other proponents. Test design and scene selection will be controlled by the ILG.

##  Schedule

It is planned to finish all documents related to the test plan by March 2016. A rough schedule is provided below. Documents containing critical parameters will be given priority for writing and approval (i.e., so that proponents can adjust their models appropriately). Model submission will be scheduled shortly after the documents have been finalized.

At the end of the project, a detailed report will be published that includes the results of the model analysis and an explanation of the evaluation procedure.

Preliminary schedule, Mile Stones (Mx):

1. AVHD-AS Detailed Description of Test Conditions, AVHD-AS Model Requirements and AVHD-AS Subjective Test Method documents as well as this synopsis document are found to be mature enough. Rough amount of databases to be provided is known. At this point in time
	* proponents will be requested to confirm participation
	* ILGs will be requested to confirm participation
	* VQEG will issue a liaison to other standardization bodies.
2. Deadline for final confirmation of participation (M1+4 weeks). Number of required databases per participant can be decided.
3. SRC pool finalized
4. Model submission deadline ((M2 +12 weeks) &&M3) [Ed. NOTE: The 12 weeks are yet to be confirmed by proponents.]
5. Start generation of validation databases (For proponents after M4, for ILGs after M3)
6. Databases ready, model validation by ILGs and proponents (M5+8 weeks)
7. Statistical analysis ready (M6+4 weeks)
8. Final report ready (F2F meeting after M7)

##  AVHD-AS Fees

Proponents may have to pay a reasonable fee to the ILGs in order to pay for their efforts. These fees are discussed and arranged between proponents and ILGs only.

##  Detailed Documentation

The details of the project will be laid out in a series of documents, all focusing on individual aspects the project. Currently the following documents are available:

|  |  |
| --- | --- |
| Document | Status |
|  |  |

###  AVHD-AS Project Synopsis (this document)

This document provides a high level overview of the project.

###  VQEG Definitions

VQEG uses specific terminology. Important definitions and abbreviations are described in this document. [Editor’s note: insert name of document draft]

Note: This document is not specific to the AVHD project!

###  AVHD-AS Detailed Description of Test Conditions

This document describes the applications, resolutions, temporal structure and distortions we are looking at. Also, the procedures for PVS creation are mentioned here. Limits of source sequences (SRCs), processed sequences (PVSs) and test conditions (HRCs) are described. This is the “handbook” for data base generation and model designers. This document is designed cooperatively during telemeetings of at least the group of proponents. Editor in charge is Shahid Satti

###  AVHD-AS Subjective Test Method

This document describes the subjective method used to generate subjective tests. [Editor’s note: document to be written. Methods under consideration: ACR, SSCQE, content immersive evaluation of transmission impairments (CIETI) from University of Madrid]

###  AVHD-AS Model Requirements

This document describes the input and output of the models as well as other aspects which are important with regard to the implementation of the models. Only implementation related aspects are mentioned. This document will be designed cooperatively during telemeetings of at least the group of proponents. Editor in charge is Shahid Satti.]

###  AVHD-AS Test Plan

[Editor’s note: This document describes the procedure how models shall be developed, submitted and validated, as well as the schedule, division of labor, and other organizational aspects. This is a joint effort and will be worked on as soon as the other documents are finalized. For the time being, this synopsis serves as reference. Document “VQEG\_AVHD\_2015\_111\_ILG Proposal to AVHD.docx” from the Glasgow 2015 VQEG meeting will serve as a starting point for discussions on division of labor among proponents and ILG. ILG points of contact are the ILG Co-Chairs.]

### AVHD-AS Validation Test Designs

[Editor’s note: This document will define the validation test designs. It will be written by the ILG in consultation with proponents and VQEG. This document will be published after model submission. This document may be revised after model submission, to fix problems.]

[Editor’s note: To be mentioned somewhere, but not as separate documents: AVHD-AS Source Scenes]