Rough Outline of P.NATS

And it’s relation to VQEG AVHD

1.0

2015-02-17

|  |  |
| --- | --- |
| Contacts: |  |
| Name: Christian Schmidmer, OPTICOM GmbH | Email: cs@opticom.de |
| Name: Shahid Mahmood Satti | Email: ss@opticom.de |

Editorial History

|  |  |  |
| --- | --- | --- |
| Version | Date | Nature of the modification |
| 1.0 | February 17, 2015 | Initial Draft, created by Christian Schmidmer |
|  |  |  |

Contents

[1 . P.NATS Scope 1](#_Toc411958946)

[2 Model Structure 2](#_Toc411958947)

[3 Model Type 3](#_Toc411958948)

[3.1 Track 1 3](#_Toc411958949)

[3.2 3](#_Toc411958950)

[3.3 Track 2 3](#_Toc411958951)

[4 HRCs, sequences… 4](#_Toc411958952)

[5 From the ToR 5](#_Toc411958953)

[6 The Niche for VQEG (Personal Opinion!) 6](#_Toc411958954)

# . Scope of P.NATS and Scope of this Document

##  This Document

The general purpose of this document is to summarize the most important facts around P.NATS as a basis for VQEG to avoid overlaps between the VQEG Adaptive Streaming project and P.NATS. It is not meant to be an exhaustive description.

##  Scope of P.NATS

P.NATS is currently under study by Q14 of ITU-T SG12. The description in here is by no means complete. Readers interested in the details of P.NATS should refer to the documentation available from the ITU. The main objectives for P.NATS are:

* Audio-visual quality estimation for adaptive HTTP streaming systems.
* Primary goal is a bitstream-based parametric no-reference model (Track 2 only partial)

# . Model Structure

Individual blocks for:

* Audio quality
* Video quality
* Impact of stalling and initial buffering
* Audio-visual and temporal integration

Proponents are free to submit models for any of these blocks. All submitted blocks will be tested in all combinations. It is currently under discussion if and how the individual outputs of each block shall be mapped using an individual optimization of the mapping formula.



# . Model Type

The development is split into two “Tracks”. A full mixing and matching of the submitted modules will be performed in Track 1 only. In Track 2 only the video quality estimation module of the best found combination for Track 1 will be replaced.

##  Track 1

Only strict parametric, no-reference, bitstream based models will be developed. Only partial decoding of the bitstream is allowed. The amount of data which may be decoded is depending on the mode as given by the table below:

|  |  |  |  |
| --- | --- | --- | --- |
| **Mode** | **Encryption** | **Input** | **Complexity** |
| 0 | Encrypted media payload and media frame headers | Meta-data | Low |
| 1 | Encrypted media payload | Meta-data and frame header information | Low |
| 2 | No encryption | Meta-data and up-to 2% of the media stream | Medium |
| 3 | No encryption | Meta-data and any information from the video stream | Unlimited |

**Some restrictions apply:**

**Restriction 1:** Applicable to Mode 2 and 3 models

* The video signal must not be transformed into another domain (e.g. into the pixel domain) before calculating parameters. Therefore the inverse block transform is not allowed.

**Restriction 2:** Applicable to Mode 2 models

* The underlying algorithm used by a model should not perform any overly complex analysis. For example the transform coefficient analysis should be limited to O(N) complexity where “N” is the number of coefficients. For example, analysis complexity of O(N^2)  and O(N log(N)) are not acceptable.

##  Track 2

The parametric NR video quality estimation module of the best Track 1 model combination is replaced by an FR module. No retraining of the remaining modules will happen. The video module has to produce one MOS score every 1 seconds for a duration of sliding window of 10s duration (tentatively). The video quality only block *Pv* will not see any stalling or delay changes.

# . HRCs, SRCs

Note: The current P.NATS Terms of Reference describe a wide range of applicable conditions. Due to time constraints however, only a very small set will be really used. The ToR document will be corrected accordingly.

* Sequences may have from 30 s up to 5 min duration
* H.264, high10 profile only (this means that YouTube is not covered by P.NATS, since it would typically use the VP9 codec)
* Min 3, max. 5 different video resolutions per subjective test (all <=HD)
* Three well defined audio qualities (AAC @ 32, 64 or 128 Kbit/s) Audio quality varies with video quality. Audio quality never drops if Video quality doesn’t drop.
* Videos are split in chunks of exactly five seconds and quality changes or stalling happens at exactly those boundaries only.
* Bitrates for each chunk may be chosen as desired.

# . Excerpt from the P.NATS ToR

*The primary application for this model is the monitoring of transmission quality for operations and maintenance purposes*

*…*

*The effects of audio level, noise, delay, (and corresponding similar video factors) and other impairments related to the payload are not reflected in the scores computed by this model. Therefore, it is possible to have high scores with this model, yet have a poor overall stream quality.*

* This excludes real end-to-end measurements

# . The Niche for VQEG (Personal Opinion!)

The first point which can clearly be identified is that that VQEG should fill the end-to-end measurement gap. This field is intentionally completely left blank by P.NATS.

Furthermore, looking at the original objectives for P.NATS and comparing those to what remains realistically, we can definitely cover a much broader range of HRCs with regard to codecs and codec settings.

Regarding the length of video sequences, the initial plan for VQEG was to cover very long durations as well. This is still very desirable, but we should consider reaching that goal in steps. From P.NATS we can already learn that long sequences drastically reduce the amount of possible HRCs and SRCs in a single test. Instead of reducing the amount of codec variation in the rest for the sake of sequence duration, we should better have a very good model which was tested on a broad range of video qualities and which is not restricted to e.g. H.264.

Audio-visual testing should certainly remain a major topic for VQEG AVHD. However, we should think twice before we include it in the AS project. The way it is handled in P.NATS, the impact of audio will be close to zero, but still it is one parameter more which needs to be validated. It may be better to start with a very good and well validated video-only model and then, in a second step integrate that with the audio quality. Here we can learn from our AV project and recycle a lot.

Generally, the VQEG model should not be parametric, quite the opposite, it should extract all parameters which it requires for processing from either the bitstreams or the video signal(s). The only exception could/should be information on stalling events. It is almost impossible to estimate this correctly from the bitstream only. Detecting it from the pixels is fairly easy if a still picture is displayed in such cases. If, however, a rebuffering indicator like e.g. a spinning wheel is overlayed over the image, then at least the position and size of that indicator must be known. On the other side, some players allow to generate events whenever they change between rebuffering and playing state. As far as the initial buffering is concerned, that piece of information should in our opinion definitely be an input to the model.