# VOEG MEETING NOTES 02-27-2015

## Old Business

Margaret went through the notes for the minutes for the previous day and the minutes were approved.

## AVHD

Remote Participants:

\* Shahid Mahmood Satti

\* Iheanyi Caleb Irondi

Marcus set up the bridge, and Shahid gave a presentation remotely.

SShahid gave an overview on the P.NATS activity. it was on the scope and its connection to VQEG. P.NATS is under study by Q14 of the ITU-T SG12. their goal is a bitstream-based paramametric no-reference model and they're interested in audio-visual quality estimation of adaptive HTTP streaming.

P.NATS have agreed on a new architecture for their algorithm. There are separate audio and video quality scores and it measures short term scores. It takes into account stalling quality impacts.

P.NATS will be conducting a subjective test with sequence lengthes from 30 seconds to 5 minutes, encoded using h.264, high profile. There were three different video resolutions and three distinct audio quality levels. Quality changes happen on 5 second intervals.

VQEG should fill the end-to-end measurement gap. Proposal: start with one minute evaluations now and move towards longer evaluations.

Quan opened the floor for questions.

\* Margaret Pinson thanked Shahid for the overview of P.NATS.

\* Arthur asked about the scope of the AVHD test as it relates to P.NATS.

\* Margaret Pinson remarked that AVHD won't be using bitstream paramaters.

\* Margaret Pinson asked which methodology was used, and the answer was ACR without hidden reference.

\* Margaret Pinson asked for an update when testing is finished, how the test methodology was received in the context of adaptive streaming.

Florence Agboma gave a presentation on over the top audiovisual quality.

Sky has 20 million customers in 5 different countries. UK, Ireland, Italy, Germany. It's direct to your house—there's a STB in your living room. But they've extended towards OTT streaming. They have several channels, movies, sports, etc.

But there are some challenges to OTT. There are different systems, different packaging, different libraries, ads must be inserted, etc. There are lots of challenges, but let's start with a step at a time. But how do you compare two OTT services? Treat the internet as a black box.

Different catalogs cause problems comparing similar content. Maybe use complexity to pick similar scenes? But the act of filming the content is different enough, along with picture in picture things, both cause problems. Hard to do a side-by-side test also. Also, is ACR an appropriate test methodology?

For the experiment design, they got a hold of manifest files. There were 16 seconds for each profile, there were two groups of test subjects, and test profiles were randomized among each session per group. They used ACR.

Quan and Lucjan commented about being emotionally connected to a particular sports team.

Florence continued by mentioning that feelings towards service providers may also affect scores.

A research firm was recruited to recruit football fans. The subjects were pre-screened for football fandom and familiarity with streaming services. There were 47 football fans, 75% male, 25% female, aged 16-65 years old. They used Seven Samsung Galaxy Note 10. and used the default earbuds. At the end they conducted a survey. They rented a house in which to conduct the test. The room illumination was changing throughout the day, but the changes were documented.

\* Naeem Ramzan asked if the videos were watched same time same room? yes. but quality can be affected by the emotions of other people.

\* Ioannis Katsavounidis asked about the actual configuration of the seating arrangement.

\* Arthur Webster asked if they were watching the same thing or different things. A group watched the same sequence, a different group watched a different sequence.

\* A group discussion ensued.

Provider A and Provider B were not usually statistically different, except for few cases, at least as stated on the bar charts. ACR seems to have worked for this test. Florence opened the floor for questions.

\* Lucjan asked to see the results again. He asked if if the hard sequences were actually easy to code or if there was some other effect that made people vote higher. Florence answered that provider B had a higher quality.

\* Ioannis mentioned that perhaps the difficulty of the scenes may have been different. He also asked if the bitrates for the video were the same. He mentioned that the only comparison metric would be total bitrate.

\* Naeem Ramzan gave a suggestion for a future test—he suggested that those participating at the same time may affect the results of the test. Again.

Iheanyi Caleb Irondi gave a presentation remotely.

The study was about DASH. Because it works over HTTP, it works through firewalls. It adapts to the network conditions of the clients. There's no subjective test methodology that's approved for use evaluating DASH. They set up a DASH web server, sent traffic through a network emulator and then viewed video on a client. They emulated packet loss ratios of 1, 3, 5%, delays of 50ms, 100ms, 150ms. They investigated the different between different segment sizes. The test was conducted in a controlled environment on a 22" monitor and ACR was used.

The results showed that greater bandwidth improved quality. For the most part, the differences between segment size were not statistically significant. Same for delay. Increased packet losses generally decreased measured quality.

They have developed a testbed for DASH and conducted a subjective test with a few different parameters. Dash has problems with initial delay, stalling during playback, and flickr.

Quan opened the floor to questions.

\* Ioannis asked why packet loss was considered for an HTTP protocol. The goal was to investigate the nature of dash—does packet loss actually influence this?

\* Lucjan asked what the length of the sequences were. There was a 60 second clip used. Lucjan asked if impairments towards the end of the clip were influencing the quality more than impairments at the beginning of the clip. Flicker and startup delay were often more annoying.

Quan and margaret then closed the session.

## Second session: 3DTV

Marcus opened the session to discuss the definitions of 3DTV—there was not much interest. Marcus proposed to not go through the definitions during the session.

Kjell asked how to make progress with the offline work effectively. Marcus suggested giving email reminders and doing conference calls.

Margaret had a discussion with SGN9 and would like to go forward by sending the most recent documents to the rapporteurs.

3DTV was closed.

## VQEG ELETTER

Margaret said that Naeem would work with Quan and Michelle would work on the eletter. Ioannis said that he may be able to submit something.

Naeem added that if you would like to organize an issue following the VIME issue, contact Naeem so they can organize topics.

Margaret will look for open research problems that are of value to industry.

Glenn said that he would be taking on as an editor for the eletter.

### OTHER BUSINESS

Liasons:

Margaret displayed the list of liasons.

Naeem agreed to write a liason on UHD Service Quality Iussues and Measurement Tools.

Marcus agreed to write a liason on 3D recommendations.

The method is to write the text and send it to the cochairs of VQEG. After the liasons are reformatted, they'll distribute to the board and give one week for a reply.

Kjell will write two new liasons about Qualinet server support, and one to SG12 to consider Type 1 error in amendment P.1401.

There were no objections to the liason plan.

Next bit of business is to approve the last set of minutes.