**Charles Poynton**
tel +1 416 535 7187
charles @ poynton.com
www.poynton.com



# Poynton's Vector 10 Video quality (VQEG)

Video Quality Experts Group, <http://vqeg.org>

Winkler, Stefan (2005), *Digital Video Quality – Vision Models and Metrics* (Chichester, U.K.: Wiley)

Wu, Hong Renand and Rao, K.R. (2005), *Digital Video Image Quality and Perceptual Coding* (Boca Raton, Fl.: CRC Press)

Wang, Zhou and Bovik, Alan C. (2006), *Modern Image Quality Assessment,* Synthesis Lectures on Image, Video, and Multimedia Processing (San Rafael, Calif.: Morgan & Claypool, Publishers)

ITU-T Rec. J.247 (2008), *Objective perceptual multimedia video quality measurement in the presence of a full reference*

If you're a video enthusiast – especially if you're interested in home theatre, or you're a home theatre calibrator – you're interested in video quality. A group called *Video Quality Experts' Group* (VQEG) sounds promising: VQEG has, for about 15 years, been studying objective estimation of video quality. There are hundreds of research papers, several books, and a few international standards.

The main concern of VQEG is the evaluation of the perceptibility of MPEG-class coding errors. We can use a panel of observers and perform a subjective test to evaluate MPEG compression, transmission, and decompression – for example, according to the protocol of BT.500 – but that's time-consuming and expensive. We can use "golden eyes" – trained observers – but that's also time-consuming and expensive. VQEG seeks what I call an "MPEG-o-meter."

The MPEG-o-meter, in one form, has two video inputs. One input presents a pristine, original "reference" video sequence. Another input presents the result of compression, transmission or recording, decompression, and perhaps some processing – the "test" sequence. On its (metaphorical) front panel is a meter that estimates the quality of the reproduction, compared to the original. More specifically, the device attempts to estimate the quality impairment that a human observer would report in a subjective test session. This technique, where the comparison has access to the original video and the impaired video, is called *full-reference* (FR). A full-reference algorithm starts by subtracting, pixel-by-pixel, the test from the reference. The difference reflects errors, which are then processed in a manner intended to mimic the visual system's sensitivity to image features.

If video transport or decompression shifts the picture six samples to the right and two lines down, a significant visual difference would be estimated by such an analysis algorithm. Similarly, if a compression system offsets luma +10 codes, scales luma down by 0.92, and scales chroma up by 1.08, full-reference analysis would start with a significant image difference. The MPEG-o-meter would report poor quality. I say, "Correct!" Picture shifting, scaling, and offsetting are errors. There's no good reason that transport or decompression should distort those aspects of the signal; such processing is properly deemed erroneous, and the estimated opinion score should be penalized.

However, VQEG offers bad news: The VQEG algorithms "normalize" – or, in what I consider to be a bizarre choice of words, "calibrate" – out such image data modifications! An active participant

in the VQEG process told me that such "calibration" is necessary because many encoders reposition the image and introduce luma or chroma gain or offsets. VQEG experts apparently think video looks ok when subject to such changes; they take steps to null-out such modifications in their tests because if such changes were allowed to influence their results, their "measured" quality level would go way down.

People familiar with approval and mastering of high-value content know that scaling and levels are carefully controlled. Broadcast-grade compression systems don't arbitrarily reposition the picture, and they don't arbitrarily scale or offset luma or chroma. Studio engineers and home theatre enthusiasts are disturbed when such modifications take place in transport or at decompression. Take the notorious "DVD chroma upsampling error": Chaos ensued when certain consumer DVD player manufacturers decoded in a manner nonconformant with encoding standards. VQEG fails to penalize systems that introduce comparable errors.

It seems to me that we want to encourage natural – or should I say technological – selection: We want poor picture quality to cause selection pressure for improvement. If poor processing isn't penalized by poor scores, how will compression system engineers learn to do processing correctly? So far, VQEG's efforts have been used in multimedia, IP video, and teleconferencing; however, their efforts have – too put it bluntly – fallen flat in high-end space. Perhaps this is why.

VQEG is also evaluating "no reference" (NR) techniques, where quality assessment is attempted without access to the original material. With no reference, any VQEG NR method will inevitably estimate *Blair Witch Project* by as having horrible quality. As far as I'm concerned – and content creators presumably agree – the "no reference" idea is a non starter: If you remove the noise from *Blair Witch Project,* you destroy the movie! Many movies have unusual visual features. Content creators want flexibility to create whatever visual stimulus they like. From the point of view of telling a story through visual means, it seems to me that there can *never* be an algorithmic measure of what constitutes a "high quality" picture.

ITU-T Rec. BT.500-12 (2009), *Methodology for the subjective assessment of the quality of television pictures*

There's further bad news. The ITU BT.500 standard specifies subjective testing protocols and test conditions, but BT.500 prescribes display and viewing conditions wildly different from those used today to master high-quality program material. BT.500 specifies contrast ratio of 50:1 or 100:1, but approval and mastering today is typically performed at around 1000:1. BT.500 specifies a surround ratio of 0.15, but approval and mastering typically has a surround ratio of 0.01 or 0.02. BT.500 specifies ambient illuminance of around 200 lx (!), but approval and mastering facilities are typically illuminated between 1 and 5 lx. These are factor-of-10 or factor-of-100 differences!

Perhaps we can convince VQEG that "normalizing" luma and chroma is a relic of analog interfaces; that such a practice is not only unnecessary but flat-out wrong in the digital age. Perhaps we can convince VQEG to use, for its subjective tests, viewing conditions that are representative of the way in which high-quality material is mastered today. In the mean time, use your instruments, but take measured readings with a grain of salt. Learn to evaluate pictures visually; trust your eyes. I welcome your comments and suggestions!