VIDEO LAYER QUALITY OF SERVICE: UNPRECEDENTED CONTROL AND THE BEST VIDEO QUALITY AT ANY GIVEN BIT RATE

Ron Gutman,
Marc Tayer

Imagine Communications, Inc.

Abstract

Spurred by DirecTV's 2007 declaration that it will be the world's first television service provider to reach 100 HD channels, cable operators are moving rapidly to create additional bandwidth not only to carry dozens more linear HD channels, but also to provide hundreds, and eventually thousands, of HD-VOD titles. The video quality bar is simultaneously rising due to the mass consumer adoption of large HDTV displays and the growing popularity of Blu-ray.

This paper discusses the fundamental and elusive paradox of how to cost-effectively increase bandwidth efficiency without sacrificing video quality. Leveraging a concept from IP networking, Video Layer Quality of Service (Video Layer QoS) involves creating minimum and maximum video quality values at the service level, while adding the technical dimension of true Variable Bit Rate (VBR) constant quality video coding.

Similarly, Video Layer QoS allows the optimization of bandwidth efficiency while guaranteeing the quality of service in a sustainable manner throughout the various switching, multiplexing and splicing stages of video communications networks. The paper also discusses the human visual perceptual system as well as related video processing and delivery aspects for a variety of digital video services.

INTRODUCTION

The North American market for video subscribers is becoming increasingly competitive and fragmented, with cable, DBS, telco and Internet service providers all jockeying to gain a bigger piece of a growing pie. After a long gestation period, the HDTV market is finally hitting its stride. The most successful service providers will offer libraries of virtually unlimited content delivered conveniently and with the highest possible video quality.

An important emerging element of this infrastructure is Video Layer QoS, defined as the establishment of video quality levels at the content origination or delivery site, combined with the process of sustaining these levels all the way to the consumer viewing environment in the most bandwidth efficient manner possible.

A Video Layer QoS solution provides:

1. Excellent MPEG-2 video quality at 3:1 HD and 15:1 SD (per 256 QAM channel) on an end-to-end basis, from content origination all the way to the set-top box.
2. Consistency (equalization) of a service provider's video quality across the Digital Broadcast, SDV, VOD and Network PVR (e.g., Start Over) categories.

3. The ability to assign different quality levels to different classes of assets (e.g., HD-VOD PPV), or even individual assets (e.g., the Super Bowl), at the discretion and under the control of the content provider or operator.
4. Sustenance of the pre-calibrated video quality levels in a cost effective, non-disruptive and backward compatible manner throughout the various multiplexing stages, including local and addressable ad insertion.

THE HUMAN VISUAL SYSTEM AND PERCEIVED VIDEO QUALITY

A logical place to begin a discussion of video quality is the area of human visual quality perception. The visual and perceptual system can not merely be construed in the context of resolution, frame rate and bit rate since these factors alone do not explain the phenomenon in which two streams with equivalent parameter settings can appear very differently to the human eye. The two streams may look quite similar most of the time, but the majority of subjects in a typical focus group will still select one sequence over the other.

When standing close to two identical screens positioned side-by-side, a trained set of eyes can begin observing the traditional compression artifacts. To mention a few notorious examples, many of us have observed blockiness at facial edges, in sky-dominated backgrounds, and during scene changes; random noise on football field grass or basketball courts; lack of detail, softness or the absence of a "pop" effect in a complex or colorful image; or tiling around logos or scrolling text areas. In many cases, the discerning viewer may need to wait for a period of high activity in the video stream in order to see artifacts, such as rapid motion, panning, scene changes, fades or flashes. This instability and unpredictability of quality over time can be quite annoying, and is highly correlated to consumer complaints regarding delivered video quality.

The following images show the same picture compressed with three different methodologies. In the first image, a "Compression by Quality" technique is used, in which an objective video quality measurement system is involved to minimize compression artifacts relative to the source. In the second image, a typical MPEG-2 encoder is used. In the third image, pre-processing and high frequency pixel filtering techniques are used in addition to the traditional compression methods.

The first image appears noticeably sharper and cleaner than the other two, showing neither the blocking artifacts of the second image nor the blurriness or loss in detail of the third image. All other things being equal, such an improvement in perceived quality can be made possible if the video quality has been exhaustively analyzed and measured as part of the video processing and multiplexing solution.

Figure 1 – Compression by Quality

Figure 2 – Typical Compression

Figure 3 – Compression by Pre-Processing and High Frequency Pixel Filtering

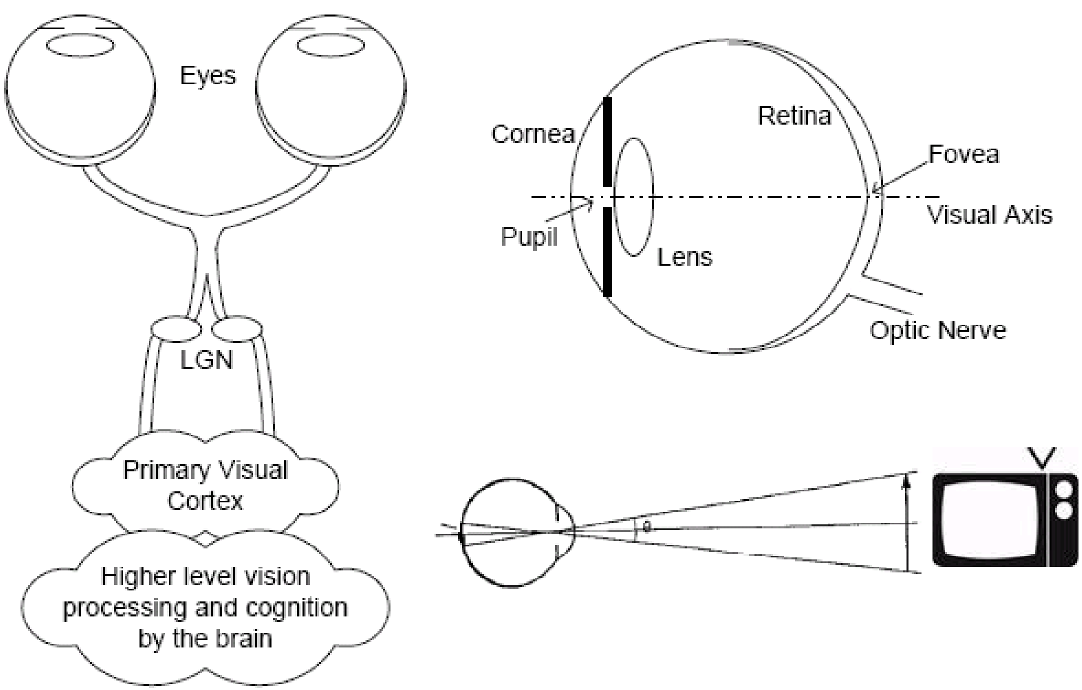

Figure 4 – Human Visual System (HVS)

In this manner, it is finally possible to guarantee the Video Layer QoS at any second, at any frame and even at any pixel. But before jumping more deeply into video quality measurement, we must first briefly discuss the anatomical and psycho-visual features of the human visual system (HVS).

Visual stimuli, in the form of light and images coming from a TV screen, are focused by the optical components of the eye, including the cornea, pupil, lens and eye fluids. These stimuli are then translated into electrical signals by the neurons and photoreceptor cells in the retina, before being received and organized by the lateral geniculate nucleus (LGN) in the brain. The LGN then transmits these signals to the primary visual cortex which tunes and processes them into spatial and temporal frequencies, orientations and motion. Higher levels of visual processing, cognition and memory associations subconsciously analyze these complex information streams while the viewer relaxes comfortably on his or her couch watching television.

Each layer of visual processing or compression removes unnecessary levels of informational redundancy, forming the video signal into data that are essential for human interpretation, i.e., entertainment or viewing satisfaction.

The more redundant information that can be removed during the video compression process, the fewer bits are required to be delivered over a communications network or stored on a storage medium. The ubiquitous audiovisual coding standards of MPEG-2, and increasingly MPEG-4 AVC (H.264), are designed specifically for this purpose. However, with TV screens becoming increasingly larger, compression-related

Human Visual System (HVS) Feature	**Compression & Quality Measurement Guidance**
Eye optics modeled by a low-pass point spread function (PSF)	Caveat: making tradeoffs against picture sharpness is a risky area
Non-uniform retinal sampling	"Compress the perimeter more" trick has some utility but is not working as well as in previous video quality tests
Luminance masking	Potentially ripe area; extreme darks and lights can be compressed more
Spatial frequency, temporal frequency and contrast sensitivity functions	Another ripe area, but needs adaptation to specific MPEG-2 and MPEG-4 AVC compression impairments
Masking and facilitation	Some image components do a good job of masking the visibility of others. Very difficult area to model and compute.
Neural pooling (cognition layer)	Perceptible distortion is more annoying in some areas of the scene (human faces, text, sea or sky background) than in others.

Table 1 – Compression Tricks Relative to the Human Visual System (HVS)

artifacts which may previously have been relegated to the category of acceptable or imperceptible marginal noise are now distinctly observable and in many cases even annoying to ordinary consumers.

Table 1 shows some known characteristics of the human visual system, and then comments on their potential effectiveness with respect to video quality measurement and image compression.

VIDEO PROCESSING AND MULTIPLEXING BY QUALITY

In this section we describe a method that allows "closing the loop" with respect to objective video quality measurement systems, significantly increasing the signal quality (and bit rate efficiency), and providing Video Layer QoS through the re-processing, re-multiplexing, VBR to CBR conversion (or vice versa) and splicing stages.

Step 1: Select or devise a video quality measurement technique

There are several subjective video quality testing methods that are accepted by industry professionals, such as the Double Stimulus Impairment Scale, the Double Stimulus Continuous Quality Scale, and other methods described in ITU-R BT.500-11. In contrast, objective video quality measurement methods, by attempting to correlate as closely as possible to subjective test results, are very elusive by definition. For this reason, through the history of digital video, subjective video methods have been heavily relied upon, with objective video quality method serving more as a sanity

Objective Video Quality Measurement Method	Computational Complexity	Correlation with Subjective Test Results
PSNR Peak-Signal-to-Noise-Ratio Most common. Based on mean squared error (MSE).	Simple	Poor , even imperceptible pixel errors contribute negatively to the measured result
MPQM Moving Pictures Quality Metric	Complex	Mixed. Certain parameters are incorporated.
VQM Video Quality Metric ANSI T1.801.03-2003	Very Complex	Good, measures perceptible impairments such as blurring, jerkiness and distortion.
SSIM Structural Similarity Index [3]	Complex	Fair, uses a structural distortion measure instead of error.
ICE-Q™ Interchangeable Compressed Elements-Quality	Very Complex	Excellent. Accounts for numerous visual impairments; designed and optimized specifically for MPEG-2 and MPEG-4 AVC (H.264)

Table 2 – Objective Video Quality Measurement Systems

check or rationale for adding certain compression tools to a standard. In other words, subjective video quality testing has been the litmus test up until now.

All objective video quality measurement methods use some form of HVS modeling. The more successful methods are backed by correlation with subjective video quality test results and have endured long periods of tuning. They are also very computationally intensive, in effect representing a form of artificial intelligence. Table 2 shows a summary of the known objective video quality measurement methods and their correlation

to subjective tests results based on personal experience as well as available information:

Note that all of these objective video quality measurement methods involve the comparison of two signals. For example, they may involve an uncompressed source vs. a compressed/decompressed signal, or a satellite-received compressed signal vs. a re-encoded signal. This remark becomes more relevant and important in subsequent stages of a signal path, in which the stream is re-multiplexed, potentially multiple times, before arriving at the consumer's set-top box.

Step 2: Video Processing or Encoding

For this step, a video processing device is required capable of "closing the loop" with the selected objective video quality measurement method. It requires processing of every frame and every macroblock of every frame, as part of the selection of a constant video quality requirement, level or "grade." Once the decisions are made, by iteratively comparing the re-processed options to the source, using the objective video quality measurement system as the arbiter, the resultant reconstructed signal is essentially guaranteed to be a constant quality signal at levels or grades which are known in advance. This signal is Variable Bit Rate (VBR) by definition since the activity and complexity vary over time. High complexity scenes will automatically be processed at higher bit rates than low complexity scenes, with both types of scenes being coded at the same measured quality level, hence the notion constant quality.

In great contrast to today's encoding or rate-shaping methods, the video processor is configurable to a pre-calibrated quality level rather than a maximum, minimum, or average bit rate. A recommended method to guide this process is to use a mathematical scale, such as 1 to 100, rather than more crude or subjective groupings such as "good," "bad," or "average."

Step 3: Calibration

During this stage, all of the available signals or video assets need to be processed (i.e., intelligently compressed using an effective objective video quality measurement system), using the video processor from Step 2, employed at various selected quality grades. The system should be calibrated in such a way that the service provider is reasonably comfortable with the constant quality experience at any grade. If this is not the case, then the previous steps should be repeated. One can define a minimum of two quality grades in a similar fashion to the QoS utilized in IP networks as follows:

1. QG_a – target average video quality grade, for example "96"
2. QG_b – Guaranteed or minimum allowed video quality grade, for example "90"

In some deployments, QG_a can be defined as "just noticeable difference" (JND), which means even expert viewers (i.e., “golden eyes”), cannot see substantial differences from the source. Then, QG_b can be defined as the quality level or grade at which, the vast majority of the time, ordinary viewers can’t discern differences from the source.

A good practice for delivering the signals, including packing density, suggests a target of no more than 1% of the time the video stream will contain QG_b.

Step 4: Statistics

Process all of the target channels or video assets at QG_a and QG_b and gather statistics for at least 24 hours, or preferably for one week. Measure the respective bit rates per second and create two vectors, one for each quality grade.

$B_a(t)$ – bit rate measured per second at QG_a
$B_b(t)$ – bit rate measured per second QG_b

Per channel, calculate your global (time tested) average bit rate at QG_a and your global maximum bit rate at QG_b.

BA_a = Average ($B_a(t)$)
BM_b = Maximum ($B_b(t)$)

BQ = Maximum (BA_a, BM_b) – defined as the channel effective bit rate for lineup allocation, utilized statically for digital broadcast and dynamically for VOD, SDV and Internet video.

It is also possible to correlate the bit rate statistics to time of day or type of program. Interestingly enough, the quality requirements during prime time are generally higher than average. In other words, the average bit rate of QG_a and the maximum bit rate of QG_b are higher in prime time; therefore, BQ should be calculated during this time window.

Step 5: Lineup

Determining the digital service combination per multiplex contains a goal of providing QG_a quality on average and never less than QG_b. The following equations can help optimize the multiplex lineup using the bit rate measurement statistics first. For example in 3:1 HD within a 256 QAM channel at 38.8Mbps:

$$\sum_{c=1}^{3} (BQ)_c <= 38.8\text{Mbps}$$

In order to guarantee the quality it is possible to simulate the statmux by repeating this calculation for every second in the database

$$\sum_{c=1,t}^{3} (B_a(t), B_b(t))_c <= 38.8\text{Mbps}$$

Select $B_b(t)$ only when needed and by measuring $B_b(t)$ usage at less than 1%.

Because of the natural statistical behavior of constant quality signals, it is advisable to have the largest number of signals per mux as possible.

Step 6: Statistical Multiplexing

Using the lineup as defined in Step 5, it is now time to actively statistical multiplex the streams. The encoders should be able to encode at multiple quality grades in real time and the statmux should choose the highest quality grade possible under the maximum channel bit rate constraint. The grades are expected to extend to the entire range between QG_b (“90”) or even lower, through QG_a (“96”), and up to “100.” The proportion of null packets should be very close to 0% at any grade under “100.”

The statmux device should report the eventual quality grades utilized in the stream. Some of the channels may change their content type over time. HD channels currently using upconverted SD content will use an increasing proportion of native HD content over time. Certain movie channels may rarely show concerts or sports events that are more difficult to compress, while other channels may alternate between movies, sports and concerts. It is important to monitor the average and instantaneous video quality grade for every mux, including the percentage of time the system is running at a grade under QG_a (expected to be less than 1%) and the percentage of time the system is running at a grade under QG_b (expected to be less than 0.1%).

A service provider may also choose to completely skip Steps 4 and 5 and base the lineup selections entirely on quality statistics rather than on the bit rate statistics. In this case, the process involves adding or subtracting one SD channel at a time to output muxes that are over or under the video quality requirement, respectively.

Although the statmux uses the entire mux bit rate to provide the highest quality, it is possible to assess the effective available bit rate according to the desired calibrated thresholds. If a certain mux consistently has average grades above the QG_a, there may be some available bandwidth for other services. The available bit rate can be computed by monitoring BA_a and BM_b in real-time even when the statmux is selecting other quality grades.

Guaranteed available mux bit rate = 38.8Mbps - $\sum_{c=1}^{3}(BA_a, BM_b)_c$

Average available bit rate (opportunistic data) = 38.8 Mbps – Average ($\sum_{c=1,t}^{3}(B_a(t), B_b(t))_c$)

The statmux device can also calculate in advance what it would take to convert any of the streams to a CBR. In this case, the minimum CBR rate would be BM_b under the CBR buffer model calculation, but when converting the signal into CBR the percentage of time at which it is running under quality grade QG_a might be significantly higher than 1%. It is possible to iteratively and heuristically determine the optimal CBR rate for QG_a and QG_b. Note that this bit rate is significantly higher than BQ, which is the effective bit rate in VBR. Since the CBR rates are generally expected to be 3.75Mbps for SD and 15Mbps HD, it is possible to calculate, in advance, the average quality and percentage of time at which streams are running at quality grades under QG_a and QG_b.

THE BOTTOM LINE RESULT: VIDEO LAYER QoS

Video Layer QoS provides an unprecedented level of control for a system operator or content provider, all the way from content origination to the set-top box. Assuming the IP and MPEG-2 transport layers are intact, this capability opens up new possibilities for ensuring video quality, not available with previous digital or analog delivery solutions. Technically, Video Layer QoS means maintaining the pre-determined quality requirements (QG_a and QG_b) through the communications delivery network, including sustainability through the various re-multiplexing, splicing, encryption, edge statistical multiplexing, and VBR to CBR conversion for services such as Start Over and SDV.

In order to take advantage of this capability, the statmux device from Step 6 needs to convey the following information per service:

1. QG(t) – instantaneous quality per frame
2. QG_a – target average quality grade, for example "96"
3. QG_b – Minimum allowed quality grade, for example "90"
4. QCBR – target CBR rate in a multi-rate CBR switched environment, the rate that

will support QG_a on average and QG_b no more than 1% of the time.

5. BQ – channel effective bit rate for VBR lineup allocation in real time

The importance of sending QG(t) indications is crucial for maintaining ultimate video quality. As noted above, objective video quality measurement techniques compare two signals and it will be impossible to compare the target to the original stream at a receive site at the terminal of the network. Given the instantaneous quality per frame QG(t), it becomes possible to keep the quality within the target range, where it requires re-multiplexing, by repeating Steps 4-6.

AD INSERTION

There are two main approaches for ensuring video quality of advertisements during ad insertion. The first approach involves pursuing the highest quality possible for the ad, even at the expense of the underlying digital services not containing ads at the same time. In this approach, during the splicing period (the ad avail), the other streams are constrained to being multiplexed at QG_a and not higher.

The second approach involves equalizing the ad quality to the underlying stream quality to the extent possible. In this case, the ad is multiplexed at QG_a and not higher, or at the eventual average quality grade of the primary stream. In any case, the ads should be processed and stored on the ad server at the maximum possible bit rate and quality level, providing downstream flexibility. A third approach is to provide the advertiser with QG_a and QG_b on a per asset or group of assets basis.

CBR FOR VOD AND SDV

Applying Video Layer QoS to the conversion of VBR signals to CBR (for SDV and nPVR applications) is relatively straightforward, with the quality levels being calculated in advance at the content origination site as discussed in Step 6.

In some cases, due to content complexity and also the inherent nature of CBR, the selected CBR rates may need to be higher than the standard SD and HD rates of 3.75 Mbps and 15Mbps, respectively. With respect to VOD, since VOD assets are originally encoded in CBR, it is possible to insert the QG(t) information into the stored stream for downstream edge statistical multiplexing.

EDGE STATMUX

A state-of-the-art edge statistical multiplexer can increase, by up to 50%, the number of streams per QAM channel without quality degradation for VOD and SDV applications.

In order to simultaneously maintain the Video Layer QoS and optimize the bandwidth efficiency, it is important to also involve the Edge or Session Resource Managers (ERM/SRM). A brute force method involves simply allocating 15 SD “blocks” of 3.75 Mbps each per QAM channel (or 3 HD “blocks” of 15 Mbps each), i.e., tricking the system into thinking each QAM channel has available up to 56.25 Mbps.

A more intelligent design can allocate the service bandwidth according to each service’s effective bit rate (BQ) as suggested in Step 4, and then load balancing the quality across the switched QAM channels, thereby guaranteeing Video Layer QoS. This method ensures the best quality at any given bit rate for edge and switched applications including VOD, SDV, nPVR, Switched Unicast and addressable ad insertion. The effective video quality will be significantly higher than today’s capped quality at 3.75 Mbps. SDTV CBR and overall network efficiency will be 50% better allowing 15 SD VBR streams per edge QAM channel.

INTERNET VIDEO

Using this approach in conjunction with the standard IP QoS mechanism ensures constant video quality and Quality of Experience (QoE) for video services over the Internet and to mobile device. The IP QoS guaranteed bit rate should be set to BQ and the maximum required bit rate for the service should be set to QCBR. In some preliminary assessment, it is shown that this approach not only provides the best video quality at any bit rate, but it also consumes 25% less bandwidth and storage.

CONCLUSION

The cable industry is in the midst of a dramatic transformation toward an increasingly competitive and complex environment. Multiple categories of digital television services will co-exist on a unified platform, including digital broadcast, VOD, SDV, nPVR, and Internet video, each of which will encompass standard definition and high definition signals.

This evolving comprehensive suite of services and architectures must be presented in a transparent and convenient manner to consumers, who now have multiple choices for their service provider. In this new environment, a key consideration and a competitive differentiator is the ability to provide true Video Layer QoS, combining control and optimal video quality across all categories with the utmost in bandwidth efficiency.

References:

1. *Objective Video Quality Assessment, by Zhou Wang, Hamid R. Sheikh and Alan C. Bovik , Department of Electrical and Computer Engineering , The University of Texas at Austin*
2. *Survey of Objective Video Quality Measurements, by Yubing Wang , EMC Corporation Hopkinton, MA 01748, USA*
3. *Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, Apr. 2004*

NCTA, The Cable Show
May 18-20, 2008
New Orleans

Ron Gutman, CTO and Co-Founder
ron@imaginecommunications.com

Marc Tayer, SVP Marketing & Business Development
marc@imaginecommunications.com

Imagine Communications, Inc.
2053 San Elijo Ave
Cardiff-by-the-Sea, CA 92007
(760) 230-0110