telchemy

Application Note

Title **Understanding IP Video Quality Metrics**

- **Series** Understanding IP Video Performance
- **Date** September 2020 (orig. February 2008)

Overview

Video streaming, IPTV, Internet TV and Video on Demand provide a range of exciting revenue opportunities for service providers. This Application Note describes some of the typical issues and problems affecting IP video service quality, the testing methods used to measure quality, and the primary video quality metrics and what they represent.

Contents

Introduction

Video services delivered over IP networks (including streaming video, IPTV and videoconferencing) are, like VoIP, highly susceptible to quality degradation from various types of impairments. Degradation can occur when the video is encoded, during transmission of the packets across the IP network, and/or during decoding and playback.

This Application Note describes some of the factors that can contribute to IP video quality degradation, the various methods used to evaluate video quality, and the most common metrics derived from objective testing methods and what they mean.

Factors that Affect Video Quality

Video is even more sensitive to IP impairments than voice, and even low rates of packet loss can cause severe degradation in perceived quality. However, all occurrences of packet loss do not have an equal impact on perceptual quality, in part because of the way video frames are structured during the encoding process, and in part due to subjective factors—such as the degree to which perception is affected by the levels of motion and detail in the video sequence, the natural delay in human reaction time to impairments, and the recency of the impairment event.

GoP and Frame Structure

Today's video codecs use a Group of Pictures (GoP) frame structure, which consists of an independently-encoded reference frame (an "I" frame) followed by a sequence of "B" and "P" frames, in which only the motion changes from previous frame are encoded.

When packet loss occurs, it can lead to decoding errors in one or more of these frame types on the receiving end of the video stream. An error occurring in an "I" or a "P" frame will propagate through all the remaining "B" and "P" frames, and thus be more likely to cause a visible impairment that may last up to several seconds. An error in a "B" frame, however, does not propagate to subsequent frames and may not even be noticeable to the viewer.

Figure 1 shows how errors propagate through the different frame types in a GoP.

Content

The content of the video itself is a contributing factor in how noticeable or annoying a particular loss event will be to the viewer. Some impairments, such as "blockiness" and frame freeze, tend to be much more noticeable during sequences containing a lot of motion (for example, during a televised sporting event) than during relatively static scenes (such as footage of a TV news anchor).

Time-varying Impairments

Perception of video quality by viewers is affected somewhat by temporal phenomena, such as the delay in viewer reaction time that occurs when quality changes from good to bad or vice versa. When impairments occur, viewers do not react immediately to the change; similarly, when quality improves after a period of degradation, there is a delay before the viewer perceives the return to "normal."

Also playing a role is the "recency" effect, in viewers tend to judge recent impairments more

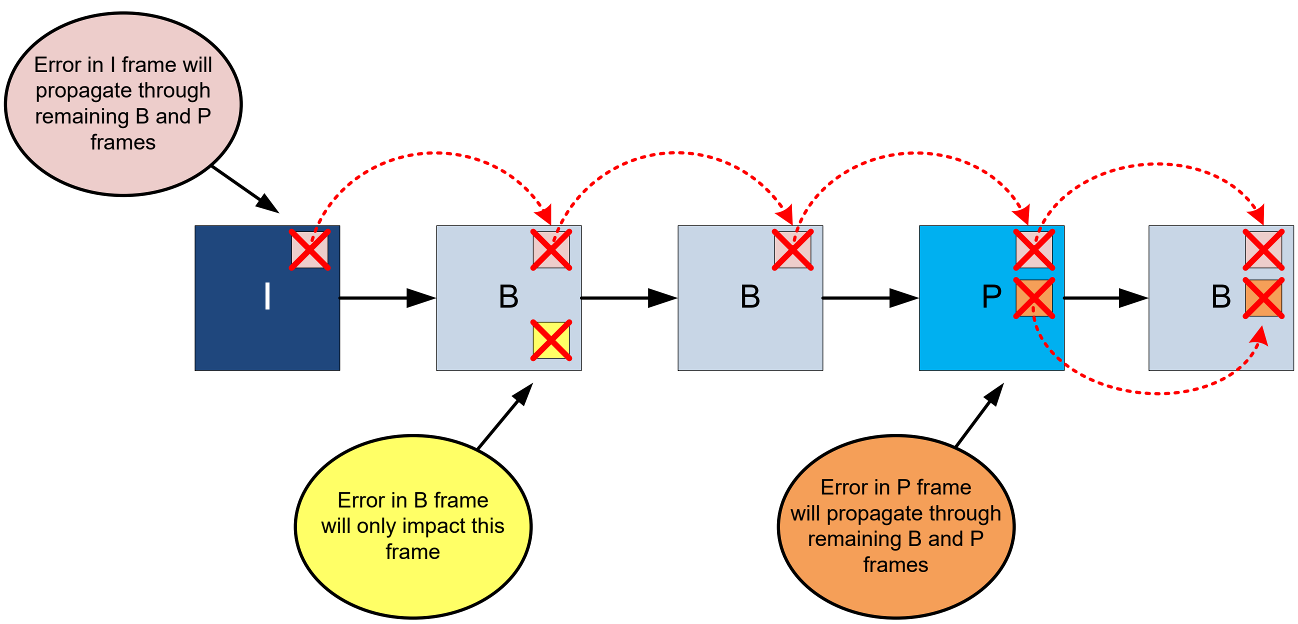


Figure 1. Impact of Frame Structure on Error Propagation

severely, and to "forgive" impairments to some extent after time has passed—i.e., the same impairment viewed near the end of a video sequence may cause a user to perceive overall quality as being lower than if the same impairment had been observed early in the video sequence, followed by a period of relatively good quality.

Defining Video Quality

IP Video quality degradation can manifest itself in a number of ways, including jerkiness, freezes or gaps in playback, and image-related impairments such as "blocky" video, blurred edges, and pixelation. As IP video is (in most cases) an audiovisual medium, there can be additional problems related to audio quality and synchronization of the audio and video—and in the case of videoconferencing, potential complications (such as delay and echo) that can affect conversational quality.

In measuring video quality, it is therefore useful to address separately the various components:

- Picture Quality – a measurement of the viewing quality of the video image
- Audio Quality – a measurement of the quality of the audio stream
- Audiovisual (Multimedia) Quality – a measurement comprising picture quality, audio quality, and audio-video synchronization

In addition to these perceptual quality metrics, the following should be considered:

- Transmission Quality – a measurement of the network connection's ability to reliably transport video. This measurement reflects the network service level rather than the quality of the specific video stream, and is independent of the type of codec used.

Metrics obtained from video quality measurement may involve one or more of these categories, and can be derived using subjective or objective testing methods—that is, by surveying human test subjects or by using algorithms to calculate estimated quality scores from objective test data.

Subjective Testing

Subjective testing is a traditional, well-proven method of evaluating video quality that provides good results; however, it can be very expensive, time-consuming, and impractical for many testing applications. One of the most commonly used test methodologies, and one that is also widely used for VoIP quality testing, is the Absolute Category Rating (ACR) test.

In an ACR test, one or more video sequences are played for a pool of viewers, who are asked to rate the video quality on a scale of 1 to 5, as follows:

5 Excellent

4 Good

3 Fair

2 Poor

1 Unacceptable

From the resulting scores, an average, or Mean Opinion Score (MOS) can be determined for each video sequence. Because the test is based on subjective opinion, the results can of course vary from test to test, but with a large enough pool (16 or more test subjects) the scores tend to stabilize. It is also important to conduct the test under controlled conditions, preferably in a dark and quiet room.

Objective Testing Algorithms

As mentioned previously, the costly and time-consuming nature of subjective testing make it an impractical method of evaluating video quality in many situations, e.g., for day-to-day service level monitoring of IPTV or IP videoconferencing services. A number of algorithms have been developed that use objective test metrics to calculate estimated video quality scores, which are intended to correlate as closely as possible with quality scores obtained from subjective survey methods such as ACR.

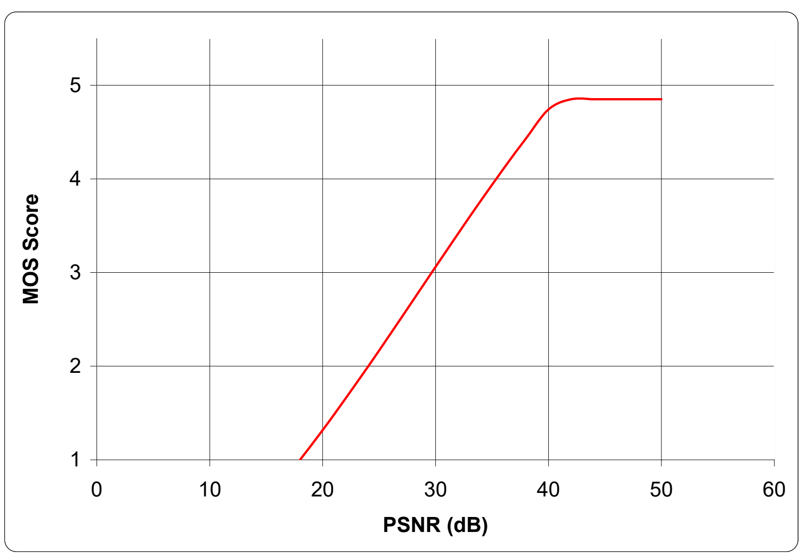
Algorithms used to estimate video quality from objective test data are of three general types:

- *Full reference* algorithms compare the output video stream to its input and determine the level of distortion that has occurred.
- *Zero reference* algorithms analyze only the output stream.
- *Partial (reduced) reference* algorithms extract specific parameters from the input stream and compare them to the same parameters extracted from the output stream.

Full Reference Algorithms

Full reference algorithms perform a detailed comparison of the input and output video streams, including per-pixel processing and temporal/spatial alignment. Although this method can derive quality scores that correlate accurately with subjective test data, the process is very computationally intensive and thus appropriate only for specific applications such as lab testing, pre-deployment testing, and troubleshooting.

The earliest and most widely-used full reference algorithm for image/video quality measurement is PSNR (Peak Signal-to-Noise Ratio), which measures the mean error between input and output as a ratio of the peak signal level expressed in dB. A PSNR value of 35db is generally considered "good," with values below 20db considered unacceptable. The chart in Figure 2 shows a comparison of PSNR and MOS.



■ *Figure 2. Chart comparing PSNR (in dB) to MOS*

A number of other full reference algorithms have also been developed, including the ITS lab's VQM (Video Quality Metric), which was incorporated into the ITU-T's Recommendation J.144, and Opticom's PEVQ (Perceptual Evaluation of Video Quality).

Zero Reference Algorithms

Unlike full reference algorithms, zero reference algorithms can be applied to the output video stream without requiring access to the input stream, making them suitable for a wider range of performance monitoring applications including day-to-day service level monitoring. This type of algorithm is less computationally intensive, and can be integrated into a range of network and test equipment.

Media-stream-based algorithms such as Telchemy's VQmon® analyze the IP stream and video transport protocols and assess video quality, which is expressed as a perceptual quality score. VQmon includes a number of advantages that improve the accuracy of its quality scores, including:

- Frame structure/GoP detection – VQmon identifies I, B, and P frames in both unscrambled and encrypted video streams, and determines GoP length and the rate and distribution of packet loss in each frame.
- Per-frame quality computation – VQmon calculates the quality in each frame using the frame type, frame size, codec type, bandwidth, and packet loss data. For P and B frames, VQmon models the loss propagated from earlier reference (I or P) frames.
- Bandwidth estimation – the bandwidth used by certain types of video frames is analyzed in order to estimate the quantization level applied by the video encoder.
- Content type detection – VQmon performs high-level content analysis, distinguishing the level of motion and detail in the video sequence, even when the video stream is scrambled.

In calculating quality scores, VQmon's perceptual model considers per-frame metrics as well as content type, adjusting for typical user reaction times to time-varying impairments and the "recency effect"—a phenomenon whereby impairments occurring earlier in a video sequence tend to be "forgiven" by the viewer as time passes, while those occurring more recently (i.e., nearer the end of the sequence) tend to have a greater negative impact on the perception of video quality.

Partial Reference Algorithms

Like full reference algorithms, partial reference (a.k.a. reduced reference) algorithms perform a comparison of the output and input video stream and calculate the level of distortion that has occurred. Partial reference algorithms do this by comparing only certain parameters from the "before" and "after" streams, rather than a detailed analysis of the entire stream, which helps reduce the complexity of the calculations and the time/processing power required to obtain results.

Video Quality Metrics

Objective testing algorithms such as VQmon produce a variety of metrics addressing different aspects of the perceptual and transmission quality, which correspond to the categories described earlier in the "Defining Video Quality" section of this document.

Mean Opinion Score (MOS)

The following estimated Mean Opinion Score (MOS) values correlate to MOS scores obtained from subjective testing methods such as ACR:

- MOS-V – Video MOS, a 1-5 score that considers the effects of the video codec, frame rate, packet loss distribution, and GoP structure on video quality.
- MOS-A – Audio MOS, a 1-5 score that considers the effects of the audio codec, bit rate, sample rate, and packet loss on viewing quality.
- MOS-AV – Audio-Video MOS, a 1-5 score that considers the effects of both picture and audio quality and the audio-video synchronization on the overall user experience.

Table 1 below shows the relationship between MOS and user satisfaction. A good target for acceptable video quality is a MOS of 4.0 or higher; lower scores indicate significant impairment and a decreasing number of satisfied viewers.

Absolute and Relative MOS

When comparing MOS values, it is important to consider that some types of video inherently produce a higher level of quality than others—for example, HDTV delivers a higher resolution image than regular SDTV, so all other factors excluded, the MOS for an HDTV video stream will be higher than the MOS for the same sequence delivered over SDTV.

Relying solely on absolute MOS values can be misleading when comparing these dissimilar types of video service, as viewers tend to form expectations of quality based in part on the perceived capabilities of the medium. For example, a video viewed on a cellular handset might receive an absolute MOS of 3.1 when little or no quality degradation is evident, while for an HDTV video sequence, a MOS of 3.1 would suggest that there were significant impairments present.

To facilitate comparing quality in different video service types, VQmon provides both Absolute and Relative MOS-V:

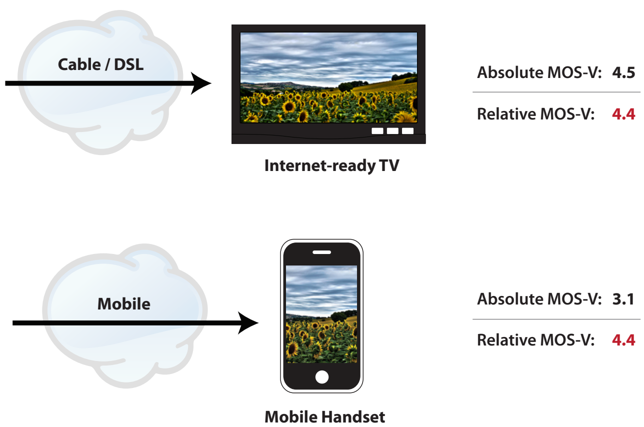
- **Absolute MOS-V** – considers the impact of frame resolution, frame rate, codec, compression level, transmission

User Opinion	MOS
Very Satisfied	4.3 - 5.0
Satisfied	4.0 - 4.3
Some Users Satisfied	3.6 - 4.0
Many Users Dissatisfied	3.1 - 3.6
Nearly All Users Dissatisfied	2.6 - 3.1
Not Recommended	1.0 - 2.6

■ ***Table 1. Relation between MOS and User Opinion*** impairments and frame loss concealment on video quality.

- **Relative MOS-V** – considers the impact of all of the factors used to determine Absolute MOS-V *except* frame resolution, producing a MOS relative to the ideal for the current video format.

For example, a program viewed on a high definition 4K television might receive an Absolute MOS-V of 4.5, while the same program on a lower resolution mobile handset might receive an Absolute MOS-V of 3.1. Both programs might receive a Relative MOS-V of 4.4, which indicates that in both cases, the video quality is close to the best possible for that format.



■ ***Figure 3. Absolute and Relative MOS-V***

Transmission Quality Score

In addition to MOS values (which are codec-dependent), VQmon reports the following quality metric:

- VSTQ – Video Service Transmission Quality, a 0-50 codec-independent score that measures the ability of the network to reliably transport video.

I/B/P Frame Statistics

As explained in "GoP and Frame Structure" earlier in this document, packet loss in the video stream may or may not be apparent to viewers, depending on whether encoding errors affect I, B, or P frames in the Group of Pictures. To form an accurate assessment of user quality of experience, it is therefore necessary to know not just overall rates of packet loss/discard, but also which frame types were affected.

VQmon reports detailed statistics for each frame type, including the number of received, lost, and discarded I, B, and P frames and the proportion of each frame type impaired by packet loss and discard. These metrics can be useful for troubleshooting and can help determine which GoP type and length should be used to obtain the best performance from the video service.

Acronyms

FR	Full Reference
GoP	Group of Pictures
IETF	Internet Engineering Task Force
IP	Internet Protocol
LAN	Local Area Network
MDI	Media Delivery Index
MOS	Mean Opinion Score
NTIA	National Telecommunications Information Administration
PLC	Packet Loss Concealment
PSNR	Peak Signal-to-Noise Ratio
QoS	Quality of Service
RFC	Request for Comments
RTCP XR	RTP Control Protocol Extended Reports
RTP	Real Time Protocol
VoIP	Voice Over Internet Protocol
VQM	Video Quality Metric
VSTQ	Video Service Transmission Quality

Summary

Video performance management technologies such as VQmon provide a convenient and practical way to evaluate user quality of experience levels using objective test data. MOS scores provide a familiar, easily understood numeric representation of picture, audio, and overall audiovisual quality, and metrics such as Relative MOS and VSTQ can be useful in accurately comparing video quality across different service types.

References

[1] ATIS-0800008 - QoS Metrics for Linear Broadcast IPTV (August 2007)

[2] ETSI TR 101 290, Digital Video Broadcasting (DVB); Measurement Guidelines for DVB Systems

About Telchemy, Incorporated

Telchemy® is the global leader in analytics technology for real-time applications and multimedia IoT with its VQmon®, Embiot®, DVQattest®, SQprobe® and SQmediator® families of service quality monitoring and analysis products. Telchemy pioneered the use of embedded analytics technology and the application of big data for VoIP and Video performance management, and is positioned to be a leading provider of analytics technology for the emerging IoT market. Founded in 1999, the company has products deployed worldwide and markets its technology both directly and through many leading networking, test and management product companies. For more information, please visit **www. telchemy.com**.

Download application notes:
www.telchemy.com/application.html

Telchemy Application Notes

Series **Understanding VoIP Performance**
- **Title** Voice Quality Measurement
- **Title** Impact of Delay on VoIP Services
- **Title** Impact of Echo on VoIP Services
- **Title** Data and Fax Modem Performance on VoIP Services
- **Title** Voice Quality Estimation in Wireless & TDM Environments

Series **VoIP Performance Management**
- **Title** Managing Enterprise IP Telephony
- **Title** Managing Cloud VoIP Services
- **Title** Managing Cable Telephony Services
- **Title** Managing Wireless LANs and Wi-Fi Services
- **Title** Six Steps to Getting Your Network Ready for Voice over IP

Series **Understanding IP Video Performance**
- **Title** Understanding IP Video Quality Metrics
- **Title** TVQM Video Quality Metrics

Series **IP Video Performance Management**
- **Title** Managing IPTV Performance

Information

Email
sales@telchemy.com
info@telchemy.com
Phone
+1-678-387-3000

Telchemy Incorporated
105 Nobel Ct
Suite 100
Alpharetta, GA 30005
USA