

Why Monitor Quality?

Executive Summary

Measuring audio and video quality has taken on great importance as video becomes an increasingly bigger part of how we communicate. The goal of quality assessment is to determine whether the quality corresponds with human observers' opinions. Considerable research has been conducted in an attempt to provide objective solutions to this subjective problem.

Service providers are continuously adding new video services — IPTV, video on demand, peer-to-peer video streaming — in order to offer customers more options. What happens if one of these services has poor video quality? Simply identifying that the data was received with errors is not sufficient. We need to define the effect on the video quality, which depends on categorizing the severity and placement of the errors.

Moreover, the viewer is the ultimate judge of video quality and will immediately call customer support if the quality degrades or continually goes below acceptable levels. A service provider who is insensitive to viewer comments will eventually lose subscribers, and therefore significant revenue is at stake. Thus, it is imperative to use comprehensive video monitoring and analysis capabilities before and during content delivery.

This paper will discuss a methodology for evaluating video quality by comparing two points within the network using a full-reference technique.

Introduction

Most video is digitized and compressed at a data rate allowing it to be transmitted over existing transmission paths: satellite, microwave, fiber, and Internet. The only exception is when analog video is transmitted, which happens less often. The digital video formats defined by MPEG and the Video Quality Experts are the de facto standards for entertainment video. They are popular because:

- There are no restrictions on the implementation of the video encoder.
- The capabilities of the video decoder (e.g., set-top box, PC, smartphone, tablet) are fully defined based on levels and profiles.
- The standards include video, audio, transport, and timing functions.

These video formats include MPEG-1, MPEG-2 (DVD), H.263 (video surveillance), MPEG-4/H.264, JPEG (still pictures), JPEG-2000 (archival), and MPEG-H/H.265 (HEVC). With the exception of JPEG-2000, all of them are lossy (info is lost during compression so the quality after encoding/decoding is not as good as the original). JPEG-2000 can be lossy or mathematically lossless.

Even if a suitable encoder has been chosen and the settings are properly configured, real-time errors can still occur.

In practice, all lossy encoders generate artifacts (areas of unfaithful visual/audible reproduction). If the encoder is designed well and the data rate is high enough, then these artifacts will be virtually invisible. The quality of the encoder and the selection of appropriate settings can be checked offline (as opposed to in real time) using a quantitative video quality analysis device such as the ClearView video picture and audio quality measurement system.

Even if a suitable encoder has been chosen and the settings are properly configured, real-time errors can still occur due to:

- Real-time compression
- Ad insertion
- Statistical multiplexing
- Re-encoding
- Transmission systems



Transformed and decoded reference, 15Mbps @ 480x270

Decoded IP video signal, 480x270 @ 350kbps

Figure 1: Source Feed is Transformed in Real-time for Comparison to Downstream IP Video Transmission

Real-Time Compression

Real-time compression is needed for live transmissions (or retransmissions). The compression device (encoder) runs, creating A/V streams of the highest possible quality. Two ways exist for encoding:

- 1) Constant bit rate (CBR)
- 2) Variable bit rate (VBR)

Video encoders that are based on inter- and intraframe compression (those mentioned above with the exception of JPEG) decrease bit rate by reducing redundant information within a frame and between one frame and the next. For slow-moving (highly redundant) scenes, they do quite well, but reducing the bit rate in high-motion video is more difficult. Video, by its very nature, is dynamic. In addition, if an IPTV distribution requires significant bit rate reduction (see Figure 1), such as a scene that may have only modest motion but a significant amount of high frequency picture information, there may be a point at which the picture becomes objectionable to the audience.

VBR produces better video quality because it can change the compression rate depending on the scene complexity. Of course, more bits require more bandwidth to stream. Most of the time, the streaming bandwidth is fixed over the network, so VBR is not an option.

Most people implement CBR for fixed-bandwidth applications, such as Internet delivery, cable TV, satellite TV, and IPTV. CBR is segmented into pieces, where the bit rate over time is constant, but the instantaneous bit rate is higher or lower depending on scene complexity. Buffers to smooth out variations in complexity are used to reduce the effects caused by complex scenes. This process is known as allocating headroom, and headroom must be pre-allocated sufficiently for the type of material.

For real-time compression, it is very important to allocate headroom. When the headroom is not sufficient, errors will occur.

Non-real-time encoding is termed “file-based encoding.” In this process, compressionists can take time to encode/re-

encode the material based on their expertise. Headroom problems are eliminated by the compressionist’s skill. The subsequent digital stream (file) is played out of a video server or written to a DVD.

Ad Insertion

Ad insertion is the process of inserting an advertising message into a stream. The ads can be inserted nationally, geographically, or demographically. Normally a digital tone (known as a cue tone) is generated, which tells an ad server to play the ad instead of the normal programming. Another tone signals a return to normal programming.

Problems can occur during the switch if:

- 1) The resolution or aspect ratio between the programming and advertising is different.
- 2) The advertising starts or stops early or late.
- 3) The advertising causes the real-time encoder to need more headroom.

Statistical Multiplexing

A broadcaster purchases a fixed amount of bandwidth and, to maximize the use, packs as many channels as possible into this bandwidth. The normal technique for doing this is called statistical multiplexing. Statistical multiplexing is a technique for combining a number of uncorrelated, bursty traffic sources together so that the sum of their peak rates does not exceed the link capacity.

A series of encoders are arranged so that their output can be combined by the multiplexer (combiner) into a single multiprogram transport stream (MPTS). Each encoder is assigned a target bit rate, and the multiplexer monitors the sum of the traffic. When an encoder encounters a complex scene, it requests more bits. The multiplexer steals bits from the other encoders and allocates more to the requesting encoder. If many of the encoders encounter a challenging scene concurrently, then problems will occur. The multiplexer will either deny the encoders’ requests or discard data (drop frames). Either way, the video quality is affected.

Statistical multiplexing is important when delivering video over a fixed pipe — as in satellite, microwave, and fiber transmission. The subscribed data rate is guaranteed, and so the viewer expects to use most, if not all, of bandwidth for which they subscribed/paid.

Re-encoding

Re-encoding is another approach to maximizing bandwidth and is similar to statistical multiplexing. However, re-encoding does not result in a full decode and encode. If a full decode is necessary, then it is better to use a statistical multiplexer.

Re-encoding modifies an existing compressed digital stream in real time without decoding. A rebroadcaster — i.e., a cable TV, satellite TV, or IPTV operator — might choose to re-encode when pulling programming from multiple sources, combining it, and sending it over their own fiber, satellite, or microwave channel.

Re-encoding parses the compressed syntax and removes some of the encoding details to fit multiple programs into a new MPTS. This technique is normally used in conjunction with a system multiplexer and when multiple MPTSs are groomed (that is, a new MPTS is formed by pulling programs out of multiple MPTSs).

Once again, complex scenes can cause a situation where oversubscription will happen. In this case, the video quality will be affected.

Transmission Systems

Video is transmitted over either a guaranteed service (microwave, satellite, or IP) or a controlled load service (IP). A controlled load service is a best-effort service, meaning that video will be delivered using the best available bit rate at a given moment based on network traffic (as opposed to delivering at a guaranteed level of service regardless of traffic). Translation: With a best-effort service, quality will vary. Due to the explosive growth of video on current IP networks, the controlled load method requires considerable data shaping.

Even in guaranteed service networks, bit errors happen. The streams are sent over many routers, and any one of them can delay the packets (causing jitter), reroute the packets (causing loss or reordering), or simply fail.

Monitoring for Real-Time Errors

The simple truism is that errors will occur. What is the effect on the video quality?

The answer depends on the type of compression. In general, for block-based algorithms — MPEG-x, H.26x — the frames are divided into three categories:

1. Intra frames (I) — a fully specified picture
2. Predicted frames (P) — holds the changes from the previous frame

3. Bi-predictive frames (B) — holds the differences between the preceding and following frames

If an I frame is lost or corrupted, then it affects the video quality until the entire picture is redrawn. If a P frame is lost, then the affected area has reduced video quality until it is redrawn by a subsequent P frame or I Frame. If a B frame is lost, then the effect is minimal.

How do you know which type of frame was lost? I frames contain the largest amounts of data, followed by P then B, so some algorithms attempt to look intelligently at the size of the packets. Others do a deep packet analysis and read the stream syntax. Deep packet analysis, of course, takes the most time, and the fact that broadcasters encrypt their services renders deep packet analysis impossible.

Set-top boxes (STBs) are computerized devices that receive compressed digital signals, decrypt/decode them, and convert them to either an analog or digital format to be shown on your TV. The STB can be either an external box, built into the TV, a PC, a gaming console, etc. Regardless of its form, an STB makes it possible to receive and display TV signals, connect to networks, play games, and surf the Internet. One of its primary functions is to detect errors and fix or conceal them. It does this by:

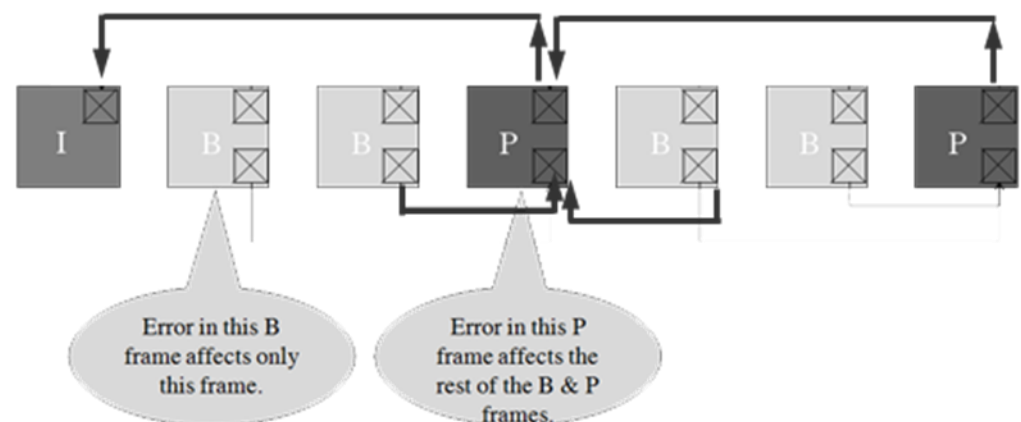
1. Holding previous frame/partial picture.
2. Asking for retransmission.

Some set top boxes do an exceptional job of hiding errors, which is why monitoring must be done after the set top.

Why Bother?

The answer is simple — competition. A poor-quality service will affect sales to future customers and reduce satisfaction for existing customers. Monitoring video quality can identify issues before they affect the bottom line. Furthermore, testing new equipment during provisioning can save data rates for newly added programming. Applying new encoder technologies can further increase channel count and enable higher-quality and higher-resolution formats, such as 4K with high dynamic range (HDR) and new color spaces.

Figure 2: Potential Error Locations in MPEG Frame Sequence



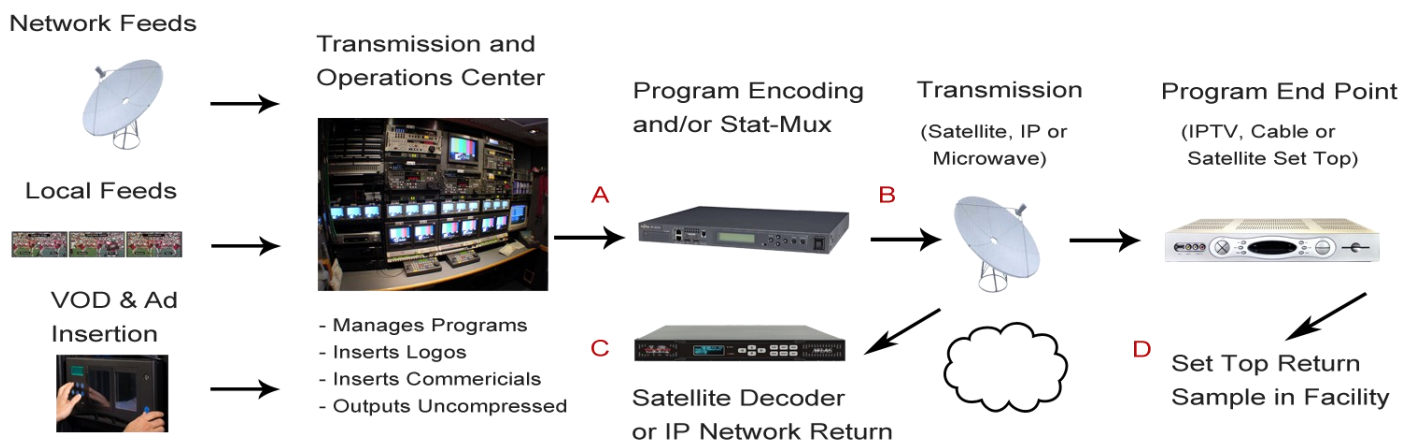


Figure 3: Content Transmission Network Simplified Flow Diagram

Monitoring should return three basic data points:

- Knowledge that an error has occurred
- The effect on the end customers' perception
- Placement of the error – which points caused the error
- Important measurement statistics for perceptual video quality, audio quality, lip sync, and loudness.

Armed with this knowledge the service provider can fix the current error and prevent future errors. For these reasons, the best place to monitor is everywhere. But since monitoring everywhere is impractical, the monitor should be placed after the:

- Ad insertion (master control)
- Real-time encoder
- Statistical multiplexer
- Re-encoder
- Transmitter

Monitoring in the early phase can give a deeper understanding of the effect of errors. If the monitoring device saves the error states, then deeper analysis can occur to solve the error. In the end, a well-devised monitoring system will cut costs, reduce customer churn, and help define a better-performing video-delivery solution.

Video Clarity RTM Solution

RTM compares any source to any point in the delivery chain fed back to the transmission center, alerts when errors cause visual, audio, and ancillary data glitches, reports audio-video offset (Lip-sync), and saves the error segments. RTM also saves performance trend data for immediate analysis.

To summarize RTM

- Measures source versus live video/audio feeds or files
- Continuously monitors and logs the audio and video quality
- Monitors the full program loudness and VANC data
- Calculates and logs A/V sync (lip sync)
- Saves uncompressed video and audio of the errors and quality score data for offline analysis

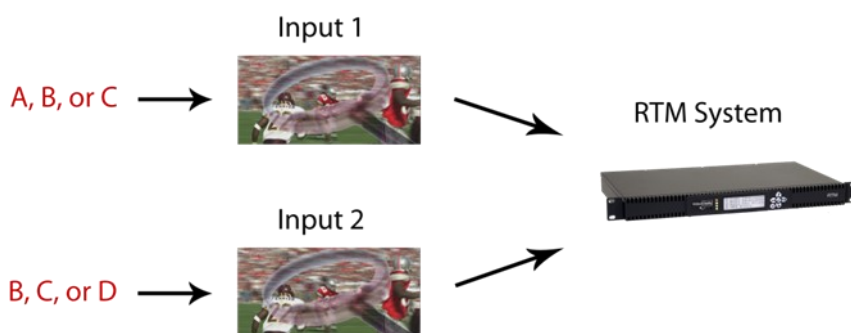


Figure 4: A Source Point is Compared to a Downstream Feed in Real-Time

Appendix: Terms Defined

Full-reference testing — Comparing the original video to the processed video when evaluating video quality. This method measures the quality difference as opposed to guessing at what the quality should be.

Headroom — Encoders are allowed to allocate more bits than the average data rate if the scene is difficult to compress. Headroom refers to pre-allocated extra space just in case this happens.

Lossless — A compression algorithm that faithfully restores the original audio and video.

Lossy — A compression algorithm that does not attempt to faithfully restore the original audio and video

MPEG – Moving Pictures Experts Group — The informal name of ISO/IEC JTC1/SC29 WG11, responsible for standardizing MPEG-x.

MPTS — Multiprogram Transport Stream — Multiple programs (streams) that are combined so that they can be sent as a combined single stream when a certain data rate has been pre-allocated.

Video quality — This term refers to both the image (picture) and the audio (speech).

Figure 1 image comparison courtesy of AT&T Laboratories.