◆ Analysis and Realization of IPTV Service Quality

Bilgehan Erman and Elissa P. Matthews

Digital television (TV) service over Internet Protocol (IP) networks is becoming a crucial element in a network provider's portfolio. However, the limitations of the existing communication networks and the complexity of the service impose many operational, performance, and scalability barriers. Analytic discussion of service quality is non-trivial because of the system complexity as well as the subjective nature of video quality assessment. This paper provides an analysis of IP video service quality metrics with a primary focus on the need for correlation of viewer perception with network performance and operation. It gives a top-down impact analysis from the user's perspective and then categorizes a set of key quality indicators (KQIs). The paper further describes a computational model, called the video quality analyzer, which both verifies and augments the top-down analysis with a bottom-up experimental realization of the KQI definitions and a refinement of the theoretical results. The discussion addresses a set of techniques for measuring video quality with respect to network performance indicators, network configuration, digital video attributes, and video content. © 2008 Alcatel-Lucent.

Introduction

Introducing new services based on new technology into carrier grade networks is a complex and a challenging task. Part of this challenge is the task of ensuring that the quality of services offered continuously meets user expectations and fulfills the demands of competitive pressures. The work described in this paper is aimed at expanding the definitions and accuracy of measurements for Internet Protocol (IP) video quality and IP video impairment, with an eye to incorporating the new standards as they evolve.

There is no commonly accepted set of performance metrics at a level of detail that supports technical operations activities. The need for a standardized set of video metrics is clear from several ongoing industry initiatives: some standards bodies are beginning to publish early results, and many probe vendors are introducing techniques for elementary video quality monitoring. Still, the industry has only made limited progress so far. One weakness in the currently available work is the reliance on representing video quality as a single score—e.g., Mean Opinion Score (MOS) or Media Delivery Index (MDI). Using a scalar number to represent all aspects of the video quality impairments has shortcomings—one of these is the lack of causal information available for follow-on error analysis. Another is the inability of current tools and standards to integrate subjective viewer perception with network performance.

This paper discusses the definition of a set of metrics to fill this gap and describes a computational

Bell Labs Technical Journal 12(4), 195–212 (2008) © 2008 Alcatel-Lucent. Published by Wiley Periodicals, Inc.

Panel 1. Abbreviations, Acronyms, and Terms

ATM—Asynchronous Transfer Mode
BL—Bell Labs
CBR—Constant bit rate
DSL—Digital subscriber line
DVB—Digital Video Broadcasting
ETSI—European Telecommunications Standards Institute
FEC—Forward error correction
GUI—Graphical use interface
IEC—International Engineering Consortium
IP—Internet Protocol
IPTV—IP television
ISO—International Organization for Standardization
ITU—International Telecommunication Union
ITU-R—ITU Radiocommunication Sector
ITU-T—ITU Telecommunication Standardization Sector
KPI—Key performance indicator
KQI—Key quality indicator
MDI—Media delivery index
MOS—Mean opinion score
MPEG—Motion Picture Experts Group
PLR—Packet loss ratio
QoE—Quality of experience
RTP—Real Time Transport Protocol
STB—Set-top box
TV—Television
UDP—User Datagram Protocol
VBR—Variable bit rate
VQ—Video quality
VQA—Video quality analyzer
VQC—Video quality calibration
VQI—Video quality indicator
V-QoS—Video quality of service

model, called the Bell Labs Video Quality Indicator (BL-VQI), which both verifies and augments the top-down analysis with a bottom-up experimental realization of the key quality indicators and a refinement of the theoretical results.

Background

The service providers now beginning to implement new video services are encountering a lack of standards and knowledge about quality assurance techniques. The metrics and measurements required to drive this quality control effort undergo a maturation process in step with that of the technology being measured. For example, though a single MOS score is applicable and useful in a voice environment, it does little to capture the complexity of the video end user's quality of experience (QoE), and nothing at all toward helping isolate or troubleshoot any problems that might occur. Industry norms and standards for monitoring and managing services based on new video technologies and IP-based next generation networks are still just beginning to emerge.

Many studies have been published on modeling the impact network errors have on the quality of digital video delivery. Early studies were based on asynchronous transfer mode (ATM) networks, whereas more recent studies base their validation on Internet video at bit rates under 500 Kbps. Video bit rate is an important factor in modeling how network errors impact video quality [4]. The study by Tao, Apostopolos, and Guérin [8] offers a practical application of this model, even though its analysis is not based on high bit rate video. Meanwhile, Reibman and co-authors [7] provide a detailed analysis of how packet loss will impact subjective video quality for high resolution video, although they do not provide a complete measurement model. Our work provides an application model and an empirical analysis of the impact of all network errors on subjective quality of broadband video.

Despite the lack of analytic models, in pace with the growth of IP video service offers, more and more vendors are announcing new video quality measurement models. One of the few publicly documented models includes the Media Delivery Index (MDI) from IneoQuest* [9]. The MDI model does not attempt to measure video quality but rather gives a measure of the two key performance indicators: jitter and media packet loss. Two other video quality measurement models whose equations have not been published are V-Factor* from QoSmetrics and VQI from Brix Networks. Although the video quality measurements they provide are related to the work discussed in this paper, based on the public information surveyed, neither of these models captures true subjective user assessment using discrete service key quality indicators (KQIs) as classified by this work.

This document contains three sections:

1. *IP Video Service Quality Analysis: The Top-Down Approach* briefly discusses the target quality of experience metrics used for the subsequent work, and the method by which those targets were determined.
2. *IP Video Network Quality of Service Analysis: The Bottom-Up Approach* discusses the quality of service network metrics, their definitions, properties, and the detailed description of how they impact quality of experience.
3. *Bell Labs Video Quality Analyzer Measurement Model* provides a description and some early results from the implementation of a prototype tool to capture subjective response to video quality and correlate that response with concurrent network health.

IP Video Service Quality Analysis: The Top-Down Approach

In order to integrate subjective viewer perception with objective network performance, it is crucial to begin with a clear view of those factors most critical to the human viewer. A report on IPTV quality published by Multimedia Research Group (MRG) in February 2007 [2] provides an excellent definition of the three layers of video service that a network provider needs to address:

- *Quality of experience* refers to the overall IPTV user experience, including application responsiveness, functionality, usability, and the service context that surrounds it. Unlike video quality (VQ) and video quality of service (V-QoS), which are each subject to measurement and conformance to specific metrics, QoE is ensured by using a combination of objective, testable criteria and subjective, anecdotal criteria that reflect the performance of the entire IPTV delivery ecosystem.
- *Video quality of service* refers to the error-free video delivery from the operator's facilities to the customer's premises.
- *Video quality* refers to the video content itself.

The work reported in this paper began by defining these nested quality levels from the top down, that is, beginning with the customer's QoE, moving next to identifying the V-QoS components, and finally isolating the VQ performance indicators that have the greatest ability to impair the V-QoS and therefore the QoE. This approach is more fully described below.

Service Performance Key Quality Indicators

The service performance key quality indicators (KQIs) are those metrics that best evaluate the performance of the video service, from the point of view of the end user. The process of selecting a manageable set of metrics which are both crucial and relevant to network management is the key to guiding the subsequent stages of research. In this work, the scope of investigation was restricted in two ways:

1. We limited our research focus to that portion of service under the control of an Alcatel-Lucent customer of average size. Specifically, we looked at the processing that occurs in the network, between the encoder and the set-top box (STB). The content as delivered to the encoder was taken as a reference value against which subsequent results could be normalized if needed. Normalization is not discussed in this paper. At the terminus, the customer premise configuration is not addressed; however, where it is possible to obtain STB measurements, these metrics are incorporated into the video quality indicators, as discussed in the video quality analyzer description which follows.
2. The paper represents the initial phase of an ongoing research effort, in which this work was restricted to address only those KQIs related to program quality, as opposed to reliability, availability, or ease of use. **Figure 1** shows the scope of context of the work and the scope of dynamic error analysis.

Analysis of both emerging standards in digital broadcasting such as European Telecommunications Standards Institute (ETSI) Digital Video Broadcasting (DVB) TR101 290 and International Telecommunication Union Telecommunication Standardization Sector (ITU-T) Rec. J, as well as existing standards for perception of video quality, such as several recommendations in the ITU Radiocommunication Sector (ITU-R), formed the categorization base for the KQI definitions used in this research. The resulting set of

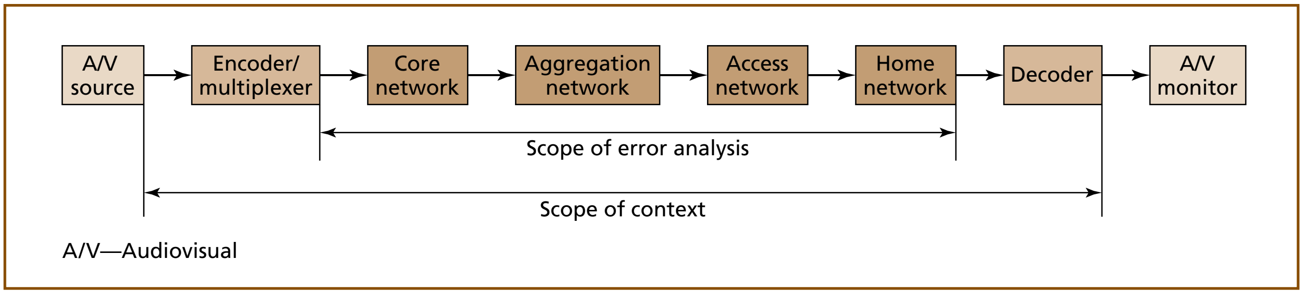


Figure 1.
Scope of modeling.

service performance KQIs is shown in **Table I**. Key definitions appear below:

- *Image element loss* refers to the loss of macro blocks or stripes, leading to the appearance of dark or randomly colored square blocks spread around the image. Image element loss may also include blanking of the image or loss of large sections of the image.
- *Image jerkiness* implies that the sequence of images does not have a smooth transition but rather has jumpy progression. Longer breaks in motion will appear as freezing.
- *Image blockiness* is the texturing of the image with small, square patterns. Affected areas of the image still preserve the original image color and luminance levels.
- *Audio impairments* include the chirping and clipping associated with impairment in digital audio signals.

IP Video Network Quality of Service Analysis: The Bottom-Up Approach

As described in the section titled Service Performance Key Quality Indicators, the service KQI set is determined by top-down analysis, taking into consideration what a television viewer experiences. A similar analytic method was used to determine the initial set of network key performance indicators (KPIs) and then tested by experimental techniques. The findings are described in the sections following.

Technical descriptions throughout this work make frequent references to Motion Picture Experts Group

Table I. Service performance KQIs.

Service attribute	KQI
PICTURE QUALITY Measured via gathering of subjective customer ratings	• Image element loss • Image jerkiness or progressive freezing • Image blockiness • Non IP-affected impairment: color, luminosity problems, ghosting, letterboxing, aspect/zoom ratio formatting, horizontal/vertical formatting
AUDIO QUALITY Measured via gathering of subjective customer ratings	• Noise: static, chirping, clipping, distortion • Volume level fluctuating or inconsistent between channels • Sound quality naturalness • Audio channel (mono/stereo/quad) problems • Wrong language
SYNCHRONIZATION Milliseconds difference between audio frame and corresponding video frame or between video frame and corresponding caption frame	• Audio + video (lip synch) • Video + program control (e.g., closed captioning)

KQI—Key quality indicator

(MPEG)-2 digital video compression. MPEG-2 is an international standard—International Organization for Standardization/International Engineering Consortium (ISO/IEC) 13818 and ITU-T H.222 and H.262. Poynton provides a good, high-level explanation of digital video compression and MPEG-2 [6].

Network Key Performance Indicators

An important outcome of the work was the identification of the KPIs that have the most significant impact on IP video impairments. To measure IP video service quality accurately, it is important not to overlook any variable that may have significant impact on quality. At the same time, the overall quality equation should not become too complicated by including variables that have little noticeable impact on video quality. The KPIs are grouped into three categories: network performance-related KPIs, network configuration-related KPIs, and content and service profile-related KPIs.

Determining the KPI set first involved logical reasoning using theoretical knowledge of digital video properties, IP network transmission properties, and the interplay of the two. The set was then verified with lab observations and subjective experiments. The resulting set of KPIs identified is shown in **Table II**. The majority of the KPIs are dynamic measures; the others (such as transport protocol, shown in parentheses in Table II) determine the operation of the model.

The scope of this modeling includes characteristics of the source audio/video and characteristics of encoders and decoders. The measurement model does not target real time analysis of errors that may result from these components, as can be seen in Figure 1. For example, problems with video streams that arrive at the video headend must be isolated as a different domain of analysis. For a second example, configuration of video and entertainment equipment in the home network may have high impact on video quality but lies outside the scope of this work.

Table II. Key performance indicators.

Category	Key performance indicators
Network performance	Packet loss
	Bursty jitter
	Reordering
	Bit error
	(Latency)
Network configuration	Decoder buffer size
	Decoder type
	(Transport protocol)
	(Access line bandwidth)
	(Access technology)
	(CBR/VBR)
Content and service profile	Video frame size
	Coding protocol
	Motion
	Video bit rate

CBR—Constant bit rate
VBR—Variable bit rate

Network performance. KPIs in the network performance group have the highest direct impact on video quality when compared to the other two groups of KPIs and therefore are discussed in detail here.

- *Packet loss.* Almost all instances of packet loss will have a measurable impact on image quality. The visible impact on the result will depend in part on the error concealment algorithm of the decoder. With little or no motion in the video, the error concealment process will hide impairments effectively. Blockiness will be seen at the edges of moving objects as the decoder tries to recover from lost packets. Depending on the decoder type, a panning video scene with no other movement of objects will typically result in jerkiness of the image motion. Without error concealment, the primary service KQI resulting from packet loss will be image element loss. Error recovery mechanisms such as forward error correction (FEC) will allow recovery of lost packets to a certain extent. With FEC, the system will tolerate a higher packet loss ratio (PLR) before video quality is impacted.
- *Jitter.* Video impairments caused by jitter will typically appear as jerkiness and progressive freezing of the image. It is frequently mentioned and documented that the IP packet jitter should be below 10 to 50 milliseconds for the delivery of high quality IP video (e.g., ETSI's DVB guidelines recommend a maximum jitter of 40 milliseconds [1]). However,

in practice, the level of jitter that will impact video quality will depend on the size of the video buffer and the bit rate of the video. As an example, a decoder can typically be configured to buffer more than 150 milliseconds of video flow at a 6 Mbps bit rate. Such a system can tolerate jitter levels much higher than commonly reported—provided it is bursty in nature so the cumulative delay does not exhaust the decoder video buffer. In comparison, without Real Time Transport Protocol (RTP) encapsulation of the IP packets, the system will not tolerate even a few milliseconds of jitter if there is possibility of packet reordering.

- *Reordering.* In carrier-grade IPTV networks, packets traverse a pre-determined virtual path end-to-end. Packet reordering will occur during transient events such as a protection switch or when segments of the network are configured to retransmit lost packets, and the impact of this re-ordering is significant. The nature of the impact depends on whether the MPEG transport streams are encapsulated with User Datagram Protocol (UDP) or with RTP. With UDP, packet reordering will result in packet loss at the decoder. Even though the out-of-order packets will eventually arrive at the decoder, the decoder will not have any means to use them and will therefore eliminate them. However, if RTP is used, the information in the sequence number field of the protocol header will allow reordering of the out-of-sequence packets, resulting in the ability to use some portion of the reordered packets. Note that some will exceed certain internal buffering or timing constraints and will therefore still be unusable.
- *Bit error.* Copper lines, often found in the last mile, are the most common source of bit errors in broadband networks. As the use of optical links increases in the network, bit errors will become less common.
- *Latency.* Because of the unidirectional nature of IP video service, latency with a small variance will not have a significant quality impact. However, large, instantaneous fluctuations in latency (such as those greater than 150 milliseconds) will impact quality. A steady level of latency will also affect the quality of experience because of its impact on signaling traffic that is required for channel change. As a KPI, latency only determines the operational boundary of the model; it is not interpreted as a dynamic variable for KQI calculations.

Network configuration. The configuration of the network will influence the degree to which network performance impacts the IP video service quality. For example, the level of bit errors will become important only if the customer's home network is connected to the access network via an older copper digital subscriber line (DSL) in the "last mile."

Key elements in a video delivery network and their potential impact to video service quality are described below:

- *Decoder type.* The image impairment artifacts the user will see will depend strongly on the decoding process and the error concealment algorithms implemented by that decoder. Therefore, in this work, the decoder type has been the first qualifier for subjective data analysis.
- *Decoder buffer size.* The impact of bursty jitter on the viewer depends on the size of the decoder buffer, as the buffer allows the decoder to compensate for changes in the packet arrival time. Excessive jitter may cause these buffers to overflow or underflow, which will typically result in image jerkiness or progressive freezing.
- *Transport protocol.* As in the earlier discussion of packet reordering, the transport protocol has a direct influence on whether decoders can recover from certain network errors and, more importantly, has great influence on the measurement capabilities of the configuration. If MPEG messages are encapsulated with RTP, then the measurement of bursty packet loss becomes easier and more accurate, and the decoder will be able to re-order packets that are out-of-sequence.
- *Access line bandwidth.* The BL-VQI model takes access line bandwidth into consideration as a boundary condition. Considering the current network deployments, the model assumes a minimum of 2 Mbps DSL line capacity. This assumption determines the boundary condition for the subjective experiments.
- *Access technology.* Similar to the access line bandwidth, the access technology is also a qualifying

factor in the context of the BL-VQI model. DSL and optical links are the only access technologies that are considered in the current model; the model does not include wireless and narrowband connections. Copper DSL and fiber will have different quality properties; in a DSL segment, a quality model should give significant emphasis to bit error analysis. Use of FEC or proprietary error recovery techniques will also modify how errors impact video quality. The "access technology" is not a dynamic measure but a property that determines the operation of the model.

- *CBR/VBR.* The dynamic bit rate envelope of the IP video packet streams affects the overall quality of video coding. If the video system is configured to use constant bit rate (CBR) (or capped CBR), then the encoder will lower the video resolution, generally in sections of the video with high levels of motion. With variable bit rate (VBR), those sections of the video will be kept at the same resolution level, but the encoder will start sending large bursts of data into the network. Even if the encoder is submitting packets to the network under CBR, any segment of the network can be configured differently. Because of this effect on the reference video quality, the BL-VQI model takes CBR/VBR as a qualifier of model calibration.

Content and service profile. Video content has many properties, some of which are configurable and some of which represent the fundamental nature of the video. The configurable properties include attributes such as the bit rate, frame size, and frame rate. Fundamental characteristics of the video include motion level, complexity of objects, color, and image texture. These properties are described here:

- *Video frame size.* The width and the height of a video frame, in conjunction with the bit rate, determine the compression ratio of the video, as the amount of coding information that will be generated and transported per image frame will be proportional to the frame size. Therefore the video frame size will be a factor in how significantly a lost or damaged packet will impact video quality: the larger the frame size, the smaller the percentage of payload each packet will carry.
- *Coding protocol.* There are many commonalities among the different digital video coding standards. However, building a video quality model requires incorporating specific characteristics of a targeted video coding standard.
- *Motion.* The amount of movement of the image has significant impact on how well video content can be encoded, assuming constraints on the amount of bandwidth available for use. If there is a large amount of motion in a section of the video, either much more data has to be generated to encode that section or the resolution of the video has to be lowered to stay within the bit rate budget. Motion also impacts how well error concealment can be performed by the decoder, as objects span a greater number of pixels from frame to frame. If there is not much movement of the objects in the video, then the error concealment is simplified by reuse of temporally redundant image data. Since the coding density per video image frame changes as motion increases, the video quality for a given transmission error will also change.
- *Video bit rate.* The overall quality of the compressed digital video is affected by the coding bit rate. Here, in the context of network-induced impairments, the video stream at the ingress of the network is considered as the reference video. For the purpose of this paper, the quality of the reference video is assumed to be the baseline against which comparison is made and will not be discussed further. Video coded at different bit rates will exhibit different quality impairments when exposed to the same network error. The type of decoder and its error concealment algorithm will again be a factor in what image artifacts will be visible on the video monitor. With lower bit rate video, spatial damage to the image will be greater; however, since the image resolution is lower, the decoder's error concealment effort will also be more effective. At the same time, with lower bit rate video, since the reference image resolution is not very good to start with, the viewer's perception of error-induced impairments may be partially masked.

Merging the Top-Down and the Bottom-Up Approaches

Service KQIs are those measurements that evaluate the quality of the experience as perceived by the end customer. The root cause of any impairment viewed by the customer, however, lies in the transport network and the application infrastructure underlying the particular service. Understanding the relationship between a service KQI and the network level key performance indicators that can impact that service KQI is important for:

- Troubleshooting end customer-affecting problems as this will require knowledge of the root causes of any impairment.
- Modifications to the network can only be made with an understanding of the impact these changes will have on the end-customer experience.
- Understanding the impacts affecting different types of IP video impairment will drive selection of the best set of network KPIs to monitor in order to maintain service quality.

Mapping Service KQIs to Network KPIs

Figure 2 shows how service KQIs are mapped to network KPIs. The identification and application of KPIs for IP video will be the subject of the section titled Video Quality Analyzer: The Bottom-Up Approach.

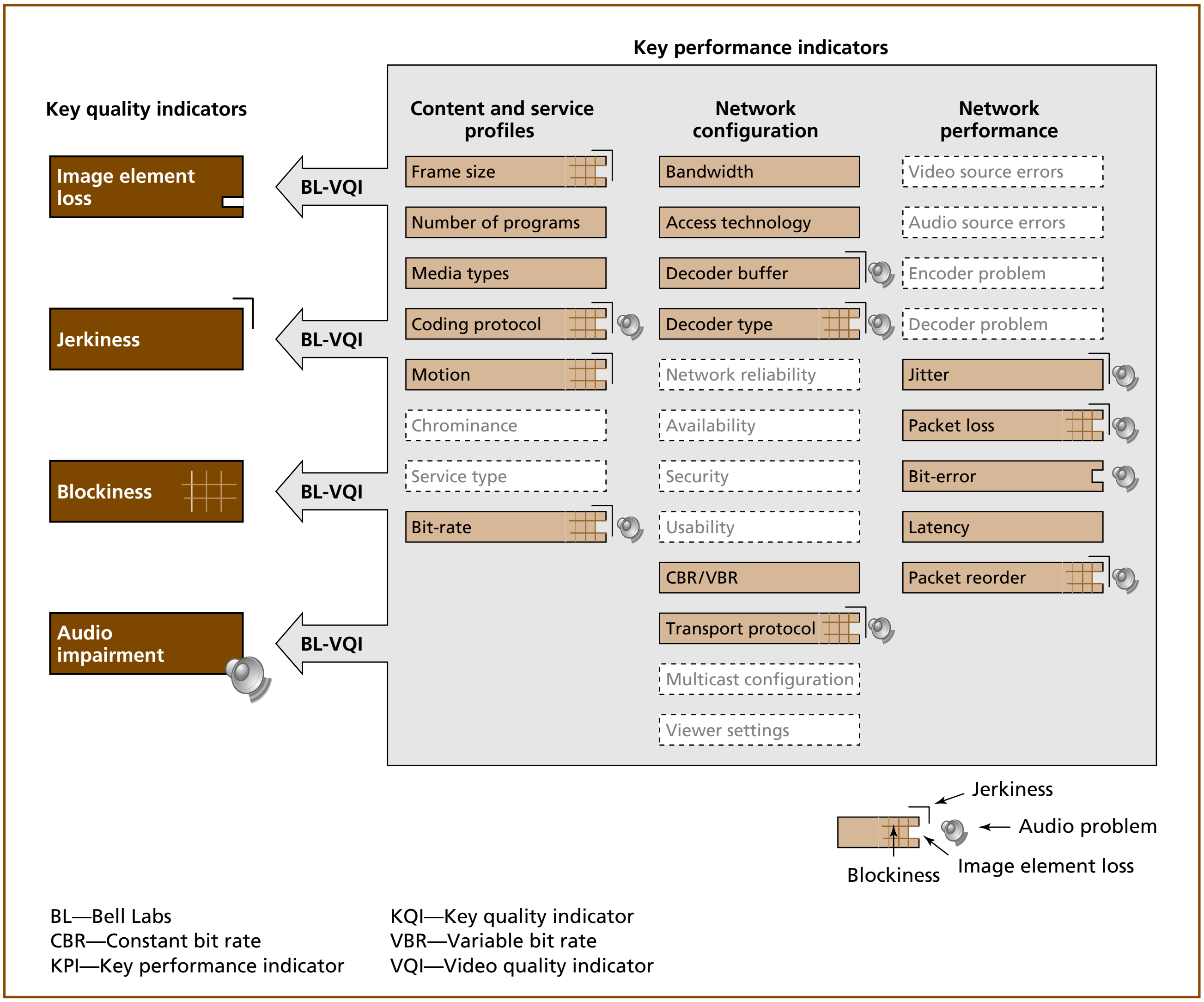


Figure 2.
KPI/KQI mapping.

- Each service KQI is impacted by more than one network KPI.
- Not every network KPI will impact a service KQI. Those network KPIs that do not affect a service KQI appear in white in Figure 2.
- Each service will have its own network KQI mapping.

Empirical Analysis of KQI-to-KPI Mapping

In order to empirically validate the mapping between service KQIs and network KPIs, an experiment was run to determine a) that each type of network impairment produced the expected image quality degradation, b) that no network impairment produced unexpected image quality degradation, and c) that levels of impact could be calibrated. That is, the experiment was designed to measure the immediate effect of errors, both to identify high and low user perception tolerances and to observe the KPI/KQI relationships.

A low threshold and a high threshold were defined. The low threshold identified the point below which the subjects did not identify errors as noticeable. The high threshold identified the point above which the video was considered unwatchable—that is, the point at which they would have changed the channel if they were watching the program by choice, not the point at which no information whatsoever was discernable. Note that each subject defined this high threshold differently; however, this phase of the work was attempting to measure qualitative not quantitative response.

Some of the findings of these initial experiments are described in the Results section. The learnings from the initial experiments helped the design of the video quality measurement computational models described below.

Experimental Conditions for Empirical Analysis of KQI-KPI Mapping

This experiment included three sessions with five subjects in each session. Experimental method was improved after each session. A summary of the results of the last session is shown in **Table III**. Results in Table III were tabulated by calculating the 10 percent trimmed average of the recorded scores.

Table III. Video quality impairment observation thresholds.

Network error	Low	High
Bursty jitter (msec)	44	216
Packet loss (PLR)	0.0004	0.0064
Bursty loss (PLR, 20)	0.005	0.022
Bit error (ratio)	1×10^{-7}	8×10^{-6}
Bursty bit-error (ratio)	4×10^{-7}	4×10^{-6}
Re-order (ratio)	0.0008	0.0032

msec—Millisecond
PLR—Packet loss ratio

Sessions were conducted by a controller. As the subjects watched a video, the controller injected network errors that caused impairments on the video. During the first session, network errors were injected at equal time intervals, first to find the low-threshold and then to find the high-threshold regions. Then, during the last session, the time interval for injecting network error was set at much smaller increments in the target regions. The smallest increment for packet loss ratio was 0.0001; for jitter, it was 10 milliseconds; for bit error, it was 5×10^{-8}; and for packet reordering ratio, it was 0.0005.

The controller increased the network error every 12 seconds. Subjects were asked to observe the video in that interval, and as the network errors increased, they were asked to identify where they first perceived impairment (the low-threshold) and where they would have "changed the channel" (the high-threshold). Subjects were not told the type of network errors that were being injected.

The subjects watched the video on a 30-inch TV from a distance of 6 to 7 feet in a room with office lighting. The sample video was a panning, zooming mountain view in color, with complex texture, and a view of the sky. The audio was classical guitar music.

Bell Labs Video Quality Analyzer Model

The video quality analyzer (VQA) system is the set of computational models that was designed to measure video quality. These models were specifically designed to address a weakness in other quality measurement models current in the industry today: the need to incorporate concrete, objective data with the subjective, user perception of experienced quality,

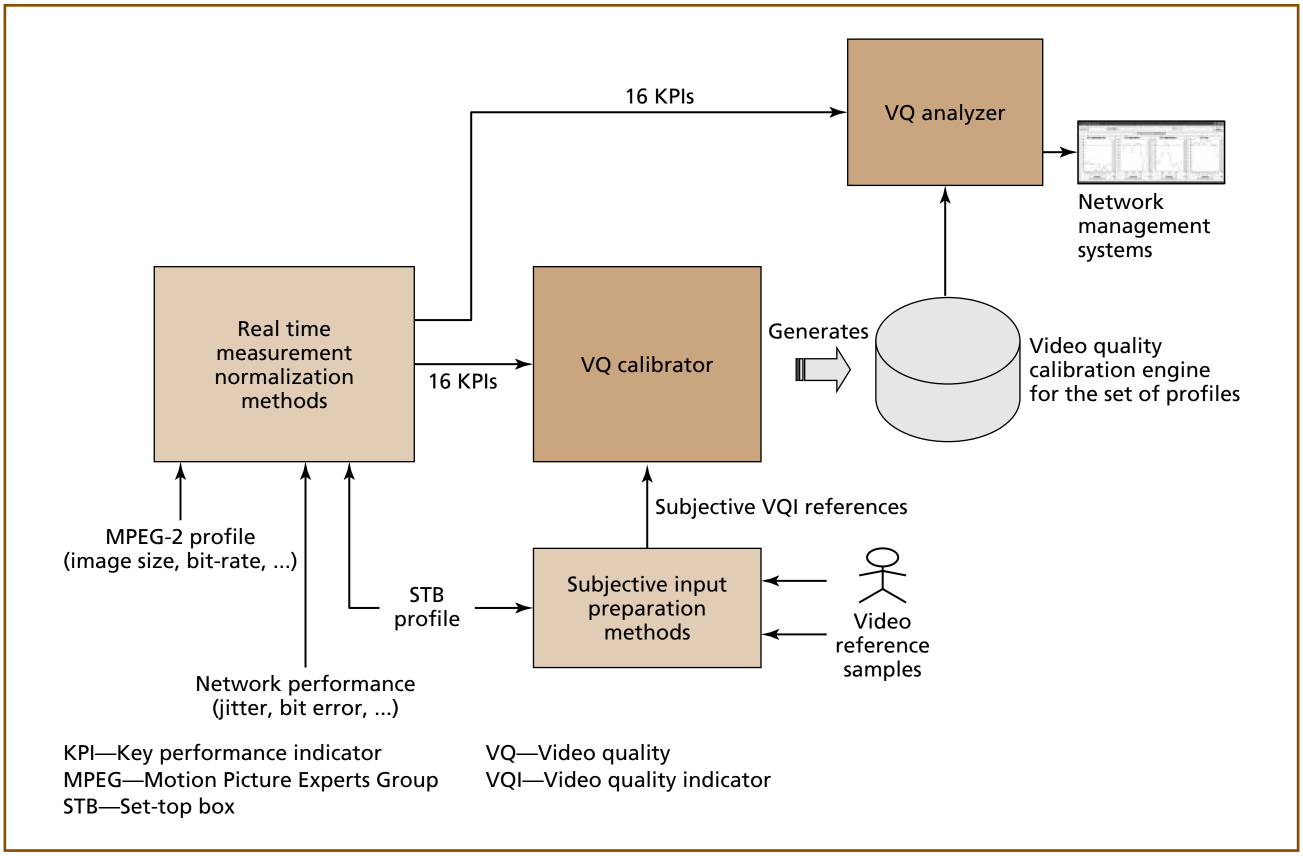


Figure 3.
IP-video measurement system.

and the characteristic instances of different network configurations. Besides the conceptual design of the model, the study also includes an experimental, prototype implementation that was built to demonstrate the feasibility of the subjective measurement techniques.

As shown in **Figure 3**, the design of the model incorporates both subjective and objective input and calibrates them relative to each other. This calibration can then be used, in conjunction with real time monitoring of specific network parameters, to predict the perceived quality of a real time viewing experience.

Figure 3 shows both the calibration and the analysis components of the model. The calibration component in the middle receives two sets of inputs: 1) the subjective assessment of the user, pictured at bottom right, and 2) the real time performance measurements on the left. By processing both input sets, the calibration component generates the calibration engine, pictured at middle right. The figure also shows the analyzer component, which measures video quality during field operation, at the upper right. The analyzer uses the same calibration engine data set and the same set of real time measurements as that of the calibration component.

The VQA system consists of two models:

- The Bell Labs video quality indicator model, and
- The video quality calibration (VQC) model.

The Video Quality Indicator Model

The BL-VQI model defines the transfer function that maps the network KPIs to the BL-VQI values, which are the quantitative representations of the corresponding KQIs. Figure 2 illustrates this KQI mapping for many measurable KPIs. Some of the entities shown on the right side of the figure dynamically

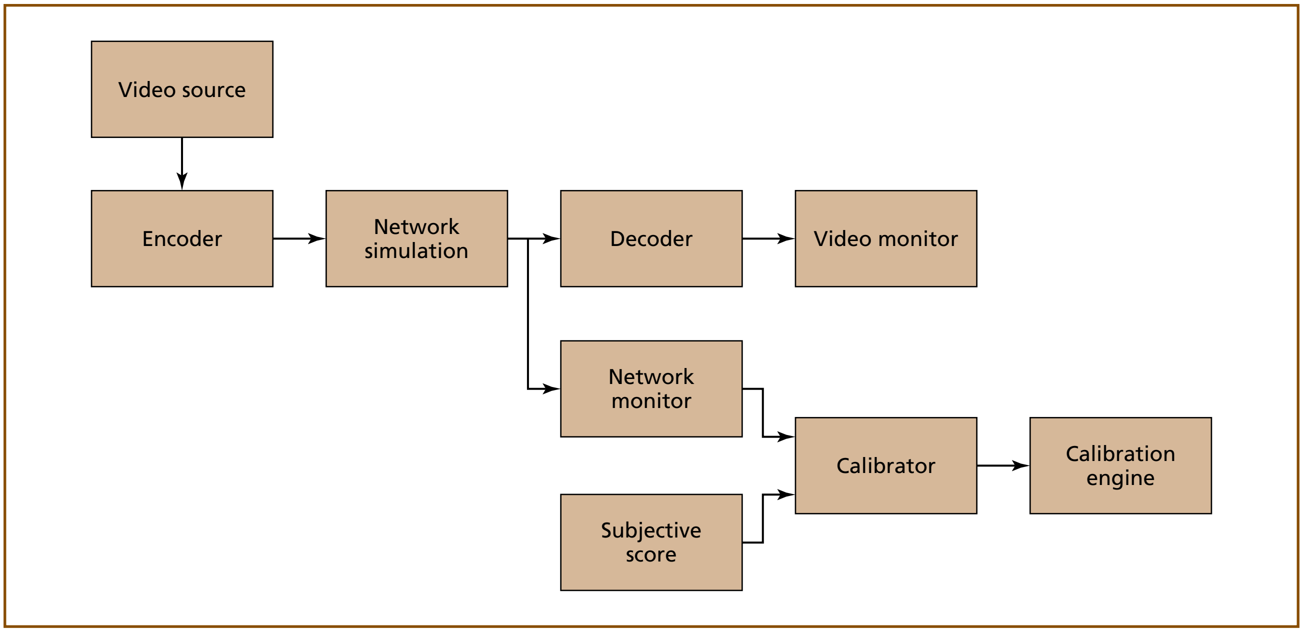


Figure 4.
Calibration model.

contribute to the KQI calculations. A significant portion of the subjective component that impacts viewer experience is not feasible to measure in real time. However, the BL-VQI component of the model incorporates this subjective score by using an innovative correlation technique, captured in the VQC.

The output of the BL-VQI component is a scalar value from 0 to 100, where 0 indicates the lowest video impairment level (highest quality) relative to the condition of the reference video, and 100 indicates the highest level of video impairment. Somewhere inside that range, a low threshold defines the level of quality impairment which is perceived to be negligible and a high threshold which identifies the quality impairment level above which the service is not considered usable.

Video Quality Calibration Model

The purpose of the calibration model is to create instances of the data set used to correlate the objective and subjective data against the known video segments. This data set is called the calibration engine. Each calibration engine instance is built for a specific decoder profile using subjective human scoring as one of the key inputs.

As shown in **Figure 4**, the calibration model is made up of nine components. At the one end, the output of the model is the calibration engine. The other end is a video source sending video streams into the simulated network by an encoder. The network is made up of devices that simulate transmission errors in the sample video stream. The decoder picks up the impaired video stream on the network and decodes and displays the video on a video monitor. Simultaneously, a real time network monitor listens to the same video stream on the network. The network monitor compiles a KPI data set from the measured values and sends it to the calibrator. The calibrator computes and fuses together the KPI values and the human subjective scores to build the calibration engine.

Video Quality Analyzer Design

The models and the KPI sets described in the earlier sections have been successfully implemented as an operational prototype. As shown in **Figure 5**, the VQA system has three main components: the video quality monitor (VQ monitor), the video quality calibrator (VQ calibrator), and the video quality analyzer (VQ analyzer).

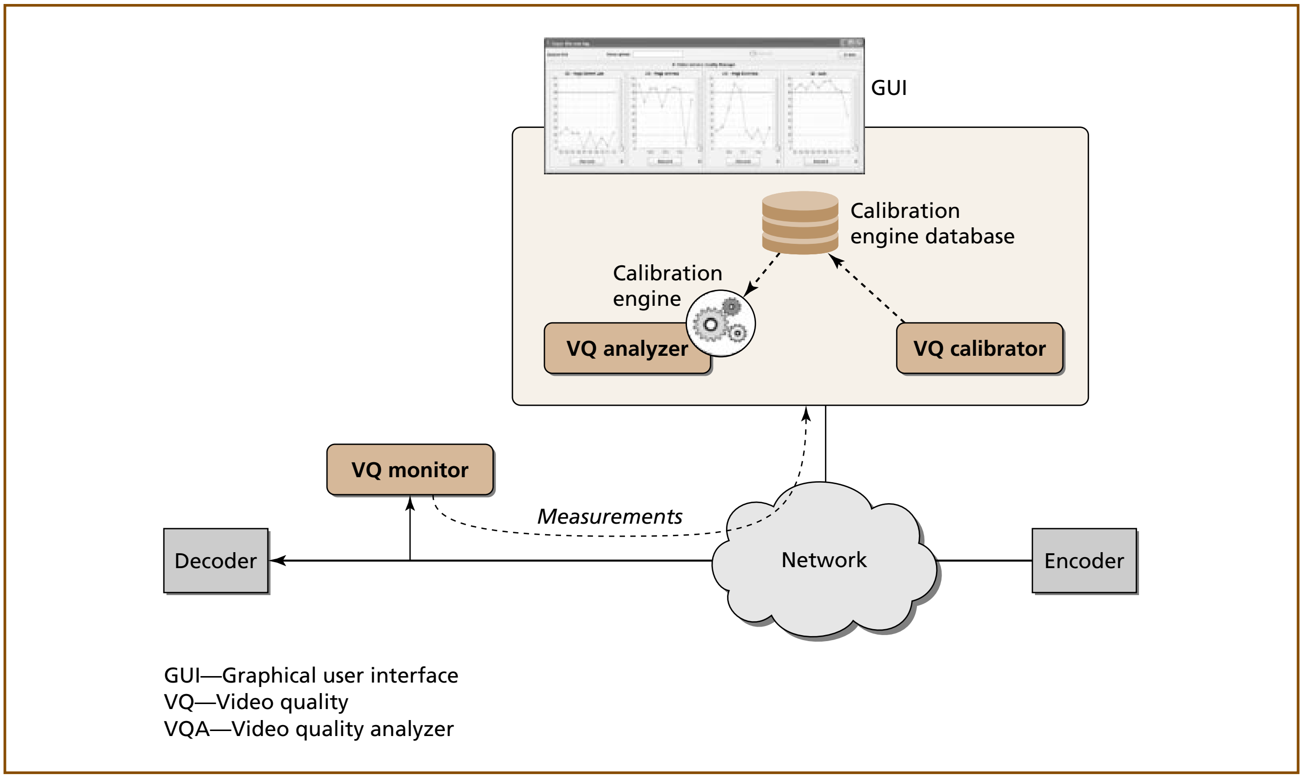


Figure 5.
VQA implementation.

VQ Monitor

The primary role of the VQ monitor module is measurement of network performance and digital video stream analysis. Commercial implementations of this module are commonly known as probes. The VQ monitor captures IP packets that carry digital video content and parses them to extract performance measurement information. Periodically, the VQ monitor calculates statistical results of the last measurements interval. At the end of the statistical calculation, an instance of the KPI value set is transmitted to the VQ analyzer.

VQ Calibrator

The keystone of the VQA system is the VQ calibrator. It is the module that creates the calibration. The calibration engine is built based on real time viewer feedback during subjective scoring sessions. The VQ calibrator builds the calibration engine by fusing the KQI value instances and auxiliary data with the viewer's quality assessment score. The constructed data structure becomes an instance of the calibration data vector:

$$v = \{ K, A, w, q \}$$

where K is an instance of the KPI set, which includes elements such as packet loss and jitter; A is an instance of the auxiliary data set, which includes elements such as the timestamp, description field, KQI identifier, and data version; w is the multiplicity of the instance vector, the weight; and q is the VQI score of the subject.

The calibration engine is comprised of a large collection of calibration data instance vectors:

$$C = \{ v_{i,} \, i \in \mathrm{N} \}$$

A new instance, a new version of the calibration engine, is created at the end of each subjective scoring session. Post-session processing includes preparation of the calibration engine data for use by the VQ analyzer. The preparation process includes putting the tuples of the calibration data vector in a specific order and sorting them. The calibration data vector is expressed as an ordered set of tuples:

$$v_i = \{ v_{i1}, v_{i2}, v_{i3}, \ldots, v_{ik} \}$$

where k is the maximum number of tuples in the vector.

If C_S is the collection of new instance vectors generated in a new calibration session, then the new instance of the calibration engine will be:

$$C' = C_{j-1} \cup C_S, j \in \mathrm{N}, j > 1$$

$$C_j = \mathrm{sort}(C')$$

where j denotes the instance of the calibration engine.

When the calibration engine reaches a maturity point for a given set-top box (STB) profile with enough data points, then it is used by the VQ analyzer to generate KQI values from real time KPI measurements.

Calibration process. The calibration process used to create an instance of the calibration engine captures viewers' assessment of video quality and stores it for later use. As the KQI-response-profile for a given KPI input is different for each STB type, every calibration session must be performed for a target STB profile. The process proceeds through the following steps:

1. *Training.* The viewers are given a) the definitions of image element loss, image jerkiness, image blockiness, and general audio problems; b) audio/video samples of each type of impairment; and c) explanation and practice using the capture tool graphical user interface (GUI). Viewers are not given any instructions on what KQI levels might be associated with which level of impairment artifacts on the video screen.
2. *Video display.* Each subject (the viewer) is shown a random segment of a randomly selected video. As the subject views the video, errors are injected to the network, but the viewer is not told when the errors are injected or what type of error is injected. There is no target level of precision for error rate increments; the higher the precision the more accurate the calibration engine becomes. However, the higher the precision, the longer the subjective sessions will be, increasing the cost of the calibration process.
3. *Subjective scoring.* The viewer uses a special purpose scoring tool GUI to express his or her assessment of the noticability of the video impairments in terms of KQI values. Assessment is given for each segment of video display, signaled via audio-visual indicators on the scoring tool. The viewer records his or her quality assessment, on a scale from 0 to 100, by moving slider bars that correspond to each KQI.

VQ Analyzer

The VQ analyzer module is the unit used in the field to provide video quality measurements. The VQ monitor monitors the target network and periodically sends real time KPI measurements to the VQ analyzer. When the VQ analyzer receives the KPI values, for each KQI, it executes a query on the calibration data set to find a matching vector instance from which the KQI value can be determined. Once the value is determined, the VQ analyzer displays the value to the network operator. If no exact match is found, a linear interpolation is done to generate the KQI value.

Video Quality Analyzer Prototype Implementation

The models and the KPI sets described in the earlier sections have been successfully implemented as an operational prototype. The purpose of the prototype and the accompanying experiments was to demonstrate the feasibility of building a video quality measurement system that could incorporate subjective response.

VQ Monitor Implementation

The VQ monitor implementation was based on a 2004 proof-of-concept implementation by our colleagues Mark Smith and Suryanarayan Perinkulam.

The VQ monitor module captures IP packets on Ethernet links using an open source library module called libpcap for the Linux* platform and WinPcap* for the Microsoft platform.

The UDP packet loss analyzer module of the VQ monitor used a special-purpose algorithm to compute bursty loss using only 4 bits of the continuity counter in the MPEG message (typically there are 7 MPEG transport stream messages in one IP packet, so the value wraps around within the third message.)

Typically, the VQ monitor was configured to perform statistical calculations at 10 second intervals. At the end of the interval it transmitted the measurement results to VQ analyzer (and to VQ calibrator during calibration.)

VQ Calibrator Implementation

VQ calibrator was implemented as a Java* application. It provided a complex graphical interface that allowed the recording capability during subjective scoring sessions. The implementation used a flatfile as the repository for the calibration engine data. The VQ calibrator listened to the VQ monitor on an IP port to receive objective, real time measurements.

VQ Analyzer Implementation

The VQ analyzer implementation shared the same Java platform with the VQ calibrator. The same graphical interface also served both the analyzer and the calibrator. Its interface had the capability to display the BL-VQI results as graphic charts. The VQ analyzer used the calibration engine data generated by the calibrator. It also received objective measurements from VQ monitor via IP messages.

Experimental Conditions

Sample videos included commercial films such as *Amadeus*, *Shrek**, and *Star Wars**, as well as music videos. The bit rate of the stream was approximately 6 Mbps Capped VBR. Encoding was MPEG-2 transported using UDP protocol. Sessions lasted 20 to 30 minutes. During the sessions, a random segment of a randomly selected film was played continuously. The statistical calculation interval was 10 seconds. Subjects watched a 30-inch TV from a distance of 2.5 feet while using the VQ calibrator computer program.

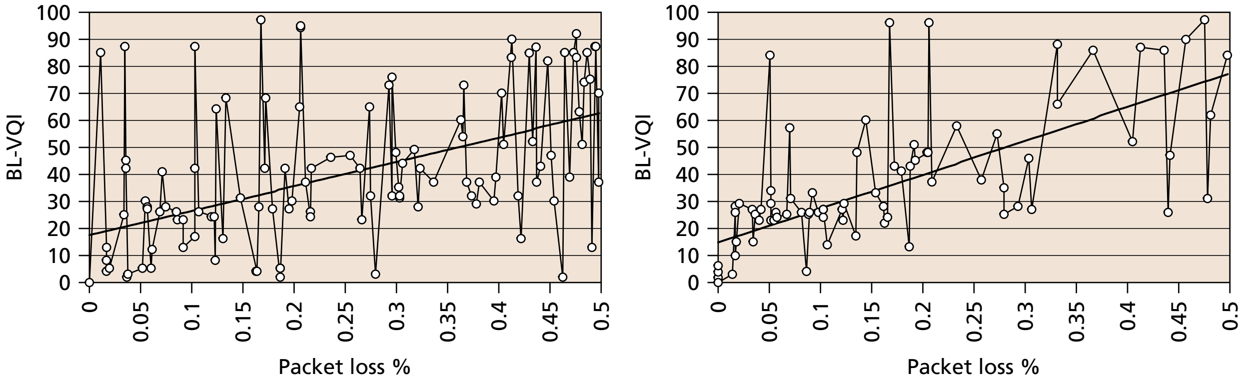
In the lab configuration, a custom-built application and a commercial product made up the network simulation unit. The application was capable of generating fixed packet loss, bursty packet loss with fixed or variable loss length, packet loss with specific MPEG frame damage with control of payload or header damage, packet reordering errors, bursty jitter with or without video stream flow control, and packet loss at specific points of a video stream. One important feature of the custom application was the automatic scheduler with graphical controls, making it possible to conduct subjective scoring sessions in a short time with minimal need for a session facilitator. The application was configured to change the level of network errors on a continuous scale. Since the application generated impairment based on statistical models, in multiple sessions, any random error level could be generated. The PLR was processed at 0.00001, but the timing accuracy of the application was limited by the general purpose operating system platform.

Results

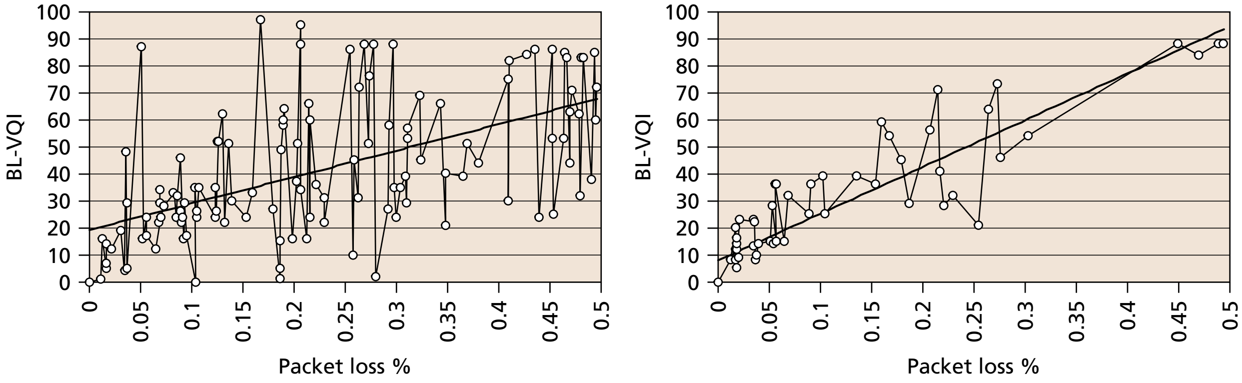
While digitally compressed video may have many subtle artifacts [10], and the accurate subjective testing of these artifacts requires careful control of many variables [3], the video quality impairments caused by network transmission errors are typically of gross granularity. At the macro block level they are not subtle image errors, such as color bleeding or blur, and therefore it is easier to conduct subjective and objective tests to detect these types of errors.

Noticeability Threshold Measurement Results

The experiments on the noticeability thresholds showed that there is a negative correlation between burstiness for a given PLR and video quality impairment. Bursty packet loss will have less impact on video quality than non-bursty packet loss at the same loss rate. (In the measurement model, burstiness was interpreted as a loss length greater than one in a loss period.) Since, on average, each IP packet will carry several micro blocks of one frame of the MPEG-2 video stream, multiple packet losses with a small loss distance will result in loss of consecutive macro blocks. Since image errors such as those resulting from bursty packet loss are temporally and spatially localized, the resulting image impairment will be less perceptible. With the loss of 20 consecutive packets, the experiment measured viewer tolerance as accepting up to 10 times more average packet loss than the packet loss of single, scattered packets. Analytic and objective models may produce an outcome predicting that burstiness increases image distortion, as in



(a) Image loss **(b) Jerkiness**



(c) Blockiness **(d) Audio impairment**

BL—Bell Labs
VQI—Video quality indicator

Figure 6.
BL-VQI outputs for packet loss.

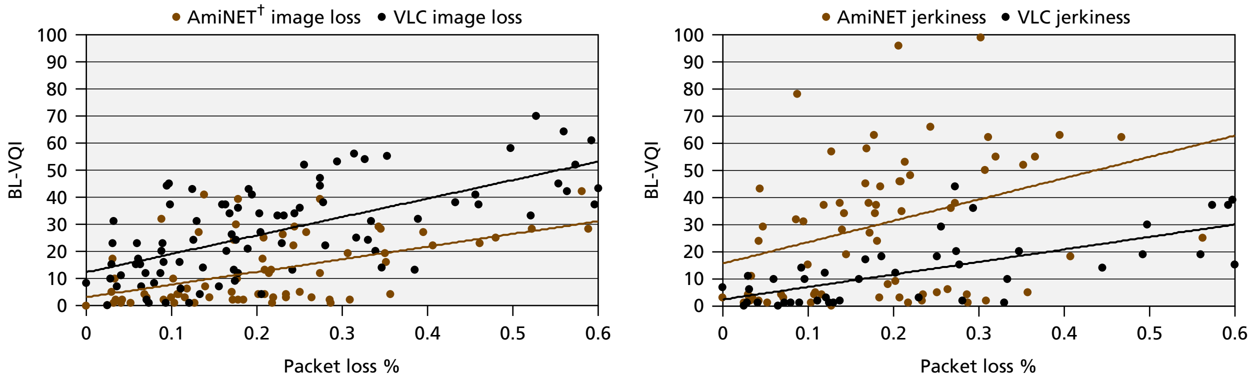
Liang's 2003 study [5]. However, within the bounds and limitations of the relatively simple testing environment, experiments showed that burstiness significantly decreases the perceived level of impairments.

Another observation the threshold experiments provided was that bit error impact on the MPEG payload (as opposed to the header) is one of the rare error types that may slip through the error concealment step, unless a decoder is configured to verify error checksums of the IP packet and the MPEG video payload. Note that not all decoders actually perform this function. Decoders provide some level of error concealment to repair image data loss resulting from network errors. However, where error concealment does not occur, bit errors will appear as image element loss (missing stripes and macro-blocks) on the video monitor.

VQA Prototype Implementation Results

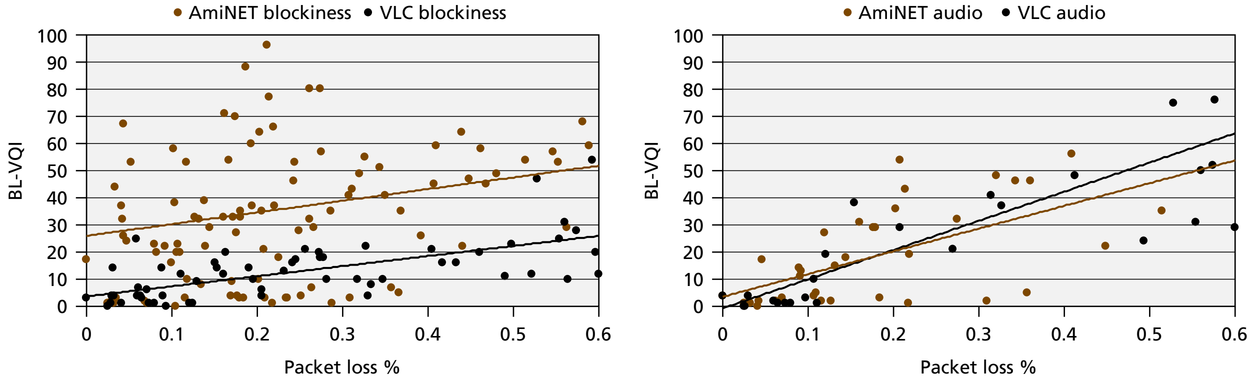
Results from experiments conducted with 25 subjects are shown in **Figure 6**. The figure shows the sample BL-VQI output set for packet loss values between 0 and 0.005 PLR. The black line in the diagram shows the linear trend line. All of the results show the expected positive correlation between packet loss and all of the KQIs.

Results from the observation of a single subject viewing two different decoders are shown in **Figure 7**. One of the decoders is an AmiNET110* STB, and the



(a) Image loss

(b) Jerkiness



(c) Blockiness

(d) Audio

BL—Bell Labs
LAN—Local area network
VLC—VideoLAN cross-platform media player
VQI—Video quality indicator

†Trademark of Amino Holdings Limited

Figure 7.
BL-VQI comparison for different decoders.

other is the software-based VLC media player (VLC). The data shows the BL-VQI output versus PLR between 0 and 0.006, with all other KPIs kept within a small range. As expected, the linear trend lines reflect differences between the two decoders. The VLC's linear constant for image element loss is about 15 percent higher than that of the STB, due to the fact that the STB's error concealment function eliminates image element loss or makes it appear as regional blockiness. Readings for jerkiness also show about a 20 percent difference between the two linear trends. This is because certain panning video scenes displayed considerable jerkiness as the STB tried to recover lost image components. By comparison, readings for blockiness show the reverse trend—the VLC does not perform image processing to recover from errors, and thus no regional blockiness is observed. Audio impairment has a very small average difference from the linear trend constants because audio processing is similar in both decoders.

Large variance in the data sets is seen due to several factors. First, real time results displayed by the VQ analyzer are driven by more than 16 KPIs processed simultaneously as a vector; plotting BL-VQI output values for a single KPI (as in Figures 6 and 7) requires filtering a considerable amount of experimental data. For example, there will be a wide range of BL-VQI output values for a given packet loss value, depending on the level of motion in the video. Second, the accuracy of the KPI measurements sent by the experimental probe prototype affects the BL-VQI output accuracy. Third, the instructions given to early test subjects and the lack of experience with the measurement tool affected the consistency of their responses.

However, the main goal of the experimental implementation of the BL-VQI model was not to determine concrete subjective levels, but to demonstrate the feasibility of building a quality measurement model based on subjective techniques. Despite the limitation of the experimental platform, it was possible to see the expected correlation between the KPIs and the KQIs that were identified in the analytical study. The results also demonstrated how different decoders handle network errors and how this can impact the video quality in different ways.

Summary and Conclusions

The purpose of this work was to provide a theoretical basis for a mature video quality measurement model, and to test that model in practice. The results of the experimental phase show both intuitive and non-intuitive results—for example, the subjective response to jitter was in line with expectations; the subjective response to bursty packet loss was not.

There are numerous factors in play in determining what the user experiences. Only part of what the viewer sees can be measured by objective techniques; a large part of perceived quality is based on subjective assessment. In this study, some inroads were made in capturing the subjective components of this complex ecosystem. It was possible to demonstrate the feasibility of implementing a system that can model key components such as the STB and the characteristics of video content, as well as the network performance and quality.

While expected trends were present, the overall level of variance in many of the results indicates the need for a larger and better controlled subjective sampling. The video model that was presented with this work will be further refined as real world data is observed and collected in the field, working with large service providers and additional test subjects. Moving forward, efforts will focus on improving the accuracy and feasibility of subjective testing, and enhancing the model with advanced objective video measurements.

Acknowledgments

The authors would like to thank the many people in Bell Laboratories who helped them with this work: first and foremost, Vanita Katkar for her excellent insights, guidance, and management; Suryanarayan N. Perinkulam, Mark Smith, Xue Li, Dorel Livescu, Harshad Padhye, Ashok Choubey, and Janet Pflederer for contributions to the original work effort; and Samrat Kulkarni, Jim McGowan, and Kim Matthews for their review efforts during the progress of this work.

*Trademarks

AmiNET is a trademark of Amino Holdings Limited.
IneoQuest is a trademark of IneoQuest Technologies Inc.
Java is a trademark of Sun Microsystems Inc.
Linux is a trademark of Linus Torvalds.
Shrek is a trademark of DreamWorks Animation L.L.C.
Star Wars is a trademark of Lucasfilm Entertainment Company Ltd.
V-Factor is a trademark of QoSmetrics, Inc.
WinPcap is a trademark of CACE Technologies.

References

[1] European Telecommunications Standards Institute, “Digital Video Broadcasting (DVB), Transport of MPEG-2 Based DVB Services Over IP Based Networks,” ETSI TS 102 034, v1.2.1, Sept. 2006, p. 51. <http://www.etsi.org>.

[2] S. Hawley and G. Schultz, IP Video Quality: QoS & QoE—Quarterly Technology & Content Report, IPTV Tracking Service, Multimedia Research Group (MRG), Feb. 2007.

[3] International Telecommunication Union, Telecommunication Standardization Sector, “Subjective Video Quality Assessment Methods for Multimedia Applications,” ITU-T Rec. P.910, Sept. 1999, <http://www.itu.int>.

[4] International Telecommunication Union, Telecommunication Standardization Sector, Focus Group on IPTV, “Working Document: Quality of Experience Requirements for IPTV,” ITU-T FG IPTV-DOC-0086, May 2007, p. 20, <http://www.itu.int>.

[5] Y. J. Liang, J. G. Apostolopoulos, and B. Girod, “analysis Of Packet Loss for Compressed Video: Does Burst-Length Matter?,” Proc. IEEE Internat. Conf. on Acoustics, Speech, and Signal

Processing (ICASSP ‘03) (Hong Kong, 2003), vol. 5, pp. 684–687.
[6] C. Poynton, Digital Video and HDTV: Algorithms and Interfaces, Morgan Kaufmann, San Francisco, CA, 2003, pp. 473–496.
[7] A. R. Reibman, S. Kanumuri, V. Vaishampayan, and P. C. Cosman, “Visibility of Individual Packet Losses in MPEG-2 Video,” Proc. IEEE Internat. Conf. on Image Processing (ICIP ‘04) (Singapore, 2004), vol. 1, pp. 171–174.
[8] S. Tao, J. G. Apostolopoulos, and R. A. Guérin, “Real-Time Monitoring of Video Quality in IP Networks,” Proc. Internat. Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV ‘05) (Stevenson, WA, 2005), pp. 129–134.
[9] J. Welch and J. Clark, “A Proposed Media Delivery Index (MDI),” IETF RFC 4445, Apr. 2006, <http://www.apps.ietf.org/rfc/rfc4445.txt>.
[10] S. Winkler, Digital Video Quality: Vision Models and Metrics, John Wiley and Sons, Chichester, West Sussex, Hoboken, NJ, 2005, pp. 42–47.

(Manuscript approved August 2007)

BILGEHAN ERMAN is a member of technical staff in Bell Labs’ Network Planning, Performance, and Economic Analysis organization in Murray Hill, New Jersey. He holds a B.S. degree in electrical engineering and an M.S. degree in computer engineering both from Middle East Technical University in Ankara, Turkey. His current interests include IP video quality modeling and large scale deployment architectures. His past professional expertise and activities include network discovery and data modeling, distributed applications, and real time switching software architectures. Mr. Erman has three patents pending.

ELISSA P. MATTHEWS is a distinguished member of technical staff in Bell Labs’ Network Planning and Performance Engineering Center in Murray Hill, New Jersey. She holds a B.S. and M.S. in mathematics from Union College in Schenectady, New York. Her area of expertise is telecommunications operations analysis and solution design. She was awarded a patent for numerical analysis of a new video-encoding algorithm, is the author of several papers in operations and telecommunications, and was the Chief Editor for the ITU SG2 standard E41.TOC: Operations Competency Metrics. Most recently, she has been creating a mathematical model of network management operations, for which her team received the Bell Labs President’s Gold Award in 2005. ◆