# Video Performance Requirements for Tactical Video Applications

Margaret H. Pinson, Stephen Wolf, and Robert B. Stafford
Institute for Telecommunication Sciences, Boulder, CO, 80305, 303-497-3579, mpinson@its.bldrdoc.gov

**Abstract**—*The Public Safety Statement of Requirements (PS-SoR) for Communications & Interoperability focuses on the needs of first responders to communicate and share information as authorized, when it is needed, where it is needed, and in a mode or form that allows the practitioners to effectively use it. PS-SoR Volume I defined functional communication and interoperability requirements. Published in September, 2006, PS-SoR Volume II identifies quantitative performance metrics, including minimum video performance requirements for public safety's tactical video applications. The goal was not to identify what is achievable with current technology but rather, looking towards the future, to investigate the minimum level of performance that first responders need in order to effectively use their video equipment.*

*On behalf of the SAFECOM Program and the Office of Law Enforcement Standards, the Institute for Telecommunication Sciences (ITS) conducted subjective video quality testing to estimate the level of video quality that first responders find acceptable for tactical video applications. This subjective testing utilized source video content that is typical of public safety operations in structured subjective viewing experiments with 35 first responders. The evaluations from these first responders, in viewing high quality video (original video) and purposefully degraded video (using video compression and transmission equipment), allowed determination of basic quality thresholds for public safety tactical video applications. These perceptual quality thresholds have been translated into technical parameters for use by video equipment designers, manufacturers, and customers. This paper summarizes those findings. Other testing to evaluate requirements for other public safety applications is underway.*

## 1. Introduction

First responders – police officers, fire personnel, emergency medical personnel – need to share vital information to successfully respond to day-to-day incidents and large-scale emergencies. Video applications provide significant assistance and will only increase in importance over time. The problem is to identify acceptable video characteristics.

On behalf of the SAFECOM Program[1], the Office of Law Enforcement Standards[2], and first responders around the United States, the Institute for Telecommunication Sciences (ITS) is developing a comprehensive Public Safety Statement of Requirements (PS-SoR). This involves identifying performance metrics and their associated values for communications equipment such that a guaranteed minimum level of quality is achieved. Different modes of communication (e.g., voice, video) and usage scenarios (e.g., telemedicine, surveillance) are identified as separate applications that may require different performance metrics and values. For each application, minimum performance levels indicate whether equipment will meet the functional needs of first responders. All stages of the PS-SoR involve working directly with first responders to identify their needs, solicit feedback, and validate study results.

The PS-SoR is unique in that it focuses on public safety requirements from a broad perspective. Operational and functional requirements are not based on a particular approach or technology, spectrum allocations, business models, or funding requirements. Rather, first responders are shown systems that could be deployed in the next 5 to 20 years and asked to identify systems that are acceptable to them.

The initial investigations into first responders' requirements for video equipment focused on real time video in tactical situations ("tactical video"). These video services are used during an incident by first responders to make decisions on how to respond to that incident. Technical specifications for tactical video equipment were determined by analyzing first responder feedback from a questionnaire and a subjective video quality test (i.e., where people assess the quality of television pictures, while sitting in a controlled viewing environment). This paper summarizes the tactical video equipment requirements that resulted from these investigations (see PS-SoR Volume II version 1.0 [1] for a more detailed description of these requirements). Some values may be adjusted slightly in future versions of PS-SoR Volume II, as additional results become available.

Figure 1 identifies the individual pieces of a real time tactical video system: camera, video coder, network, video decoder, and monitor. This paper focuses primarily on the video system (between points B and E) including a network as well as the video coder and decoder. Determining

[1] See http://www.safecomprogram.gov/SAFECOM/.
[2] See http://www.eeel.nist.gov/oles/.

performance characteristics for the cameras, monitors, and other video applications will require further research and development.

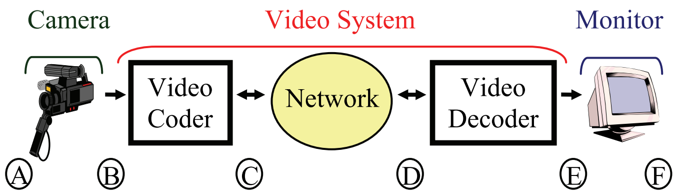


**Figure 1 –** Tactical video reference diagram.

## 2. Experiment Summary

Since the use of video in public safety applications is relatively new, a questionnaire was first used to gather fundamental information about user requirements. Eighteen first responders from around the United States were asked questions concerning acceptable values for certain characteristics, such as video delay, image size, frame rate, coding impairment, color fidelity, and focus distance. First responders also prioritized video applications, so that the most important application could be examined first.

As a result of the questionnaire, the highest priority video application was real time video used in tactical applications, where the features of interest occupy a relatively large percentage of the video frame ("zoomed in"). First responders were given the following examples: (1) a camera carried by a firefighter into a burning building to provide the incident commander with situation information, (2) a body-worn camera, and (3) an aerial camera following a subject on foot. This application ("narrow field-of-view tactical video") will be referred to as "tactical video" for the body of this paper.

The questionnaire results shaped the design of the first subjective video quality test. A controlled viewing environment was used to obtain accurate feedback on the level of quality required by first responders for tactical video. During the subjective test, first responders watched short video sequences and indicated whether they would find them acceptable or unacceptable for use in the field.

High-quality source video sequences (point B in figure 1) and high-quality monitors were used so as not to influence the outcome of the experiment. A wide variety of public safety video scenes were collected, including surveillance video of large incidents, underwater crime scenes, scenes where the camera follows police or firefighter personnel as if the camera were attached to another participant in an event, and in-car camera video. From this pool of high-quality video sequences, sixteen source video sequences were selected to visually portray what first responders might encounter in the field for tactical video applications. Figure 2 shows frames from four of these source video sequences.

**Figure 2 –** Frames from four sequences used in the subjective video quality test.

The source video was used to create destination video (point E in figure 1) that indicated fundamental characteristics of the video system. The video systems included simulated systems that demonstrated the effects of a pure reduction in image resolution and/or frame rate, as well as real MPEG-2 and H.264 codecs operating over a range of bit-rates and IP packet loss ratios. In all, 40 different video systems were examined. To reduce the total number of video scenes that had to be subjectively evaluated, the sixteen source scenes were divided into two subsets of eight each, and each video system was tested with only one subset of scenes.

First responders watched the source and destination video sequences in a controlled environment, using standard subjective testing practices [2]. The source and destination video scenes were presented in randomized order over four viewing sessions with breaks between sessions. After viewing each video sequence, first responders recorded their opinions of the video sequence's quality by choosing one of five quality levels: excellent, good, fair, poor, or bad. They also indicated whether that video sequence was acceptable or unacceptable. Data were collected from 35 first responders representing a variety of jurisdictions, disciplines, years of service, geographic locations, and ages.

To determine if a particular video system was acceptable for first responder use, the acceptability data were averaged over all viewers and scenes for that system. A system was judged "acceptable" if the average acceptability over all possible viewers and scenes was expected to exceed 70% (i.e., the lower bound of the 95% confidence interval for the average exceeded 70%). The 70% acceptability level was chosen as a reasonable tradeoff between having a perfect system and having a system where a large number of users would be dissatisfied.

Figure 3 compares system acceptability with system quality (averaged over viewers and scenes, so each point represents one video system). The horizontal red line depicts the 70% acceptance criteria. Below 70% acceptance, the data forms a straight line, indicating that increasing system acceptability is linearly related to increasing quality. Above 70%, increasing quality yields diminishing returns in acceptability. This seems to support the choice of using a 70% acceptability threshold. Points below the red line in figure 3 but above 0% acceptability imply that at least some of the viewers find these systems acceptable for at least some scene content. Conversely, points above the red line but below 100% acceptability imply that at least some of the viewers find these systems unacceptable for at least some scene content. In fact, none of the systems that were tested produced an acceptability of exactly 0% or 100%, including the original source video.

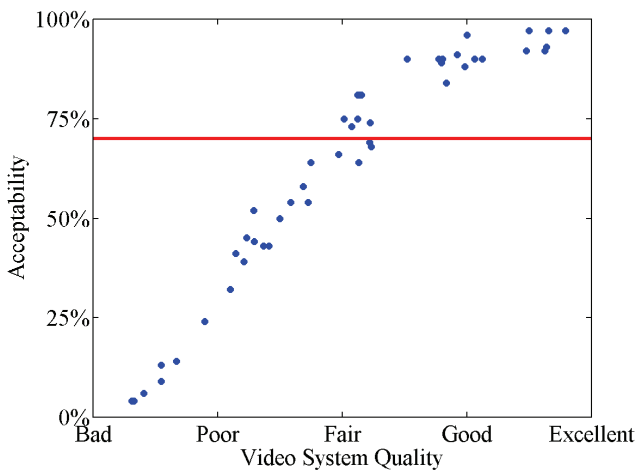


**Figure 3 –** Comparison of system acceptability and system quality.

## 3. VIDEO EQUIPMENT SPECIFICATIONS

This section summarizes the performance specifications for tactical video. The first three subsections (one-way video delay, frame rate, and luma image size) identify parameters that describe the performance of the entire video system (points A through F in figure 1). The next three subsections (lossless coding & transmission, coder type & bit rate, and packet loss & error concealment) describe parameters that characterize the video system (points B through E in figure 1). The last subsection (objective video quality) identifies a tool that can be used to estimate the acceptability of a video system. For brevity, PS-SoR Volume II performance specifications that were not determined by public safety first responders have not been included in this paper. An example specification of this type would be the maximum allowed horizontal spatial shift of the video picture.

### 3.1. One-Way Video Delay

One-way video delay is the length of time taken to send video through the entire video system (between points A & F in figure 1). Since coding, network transmission, and decoding can each add substantial delay, some compression algorithms and networks will not be suitable for public safety applications. Based on the questionnaire, a maximum one-way delay of 1 second is recommended.

### 3.2. Frame Rate

Frame rate is the rate at which a video system can produce unique consecutive images, called frames. Frame rate is measured in frames per second (fps). Broadcast television in the United States updates half of the picture every 1/59.94 seconds, which is approximately 30 fps. The frame rate of some video systems is constant, while the frame rate of other video systems changes over time. A minimum of 10 fps is recommended.

To determine acceptable frame rates, the source video was synthetically impaired, by reducing the frame rate only. The picture resolution was left at Standard Definition (SD), used by broadcast television. Figure 4 shows the acceptability of 30 fps, 15 fps, 10 fps, and 5 fps. The vertical red line depicts the 70% acceptance criteria.

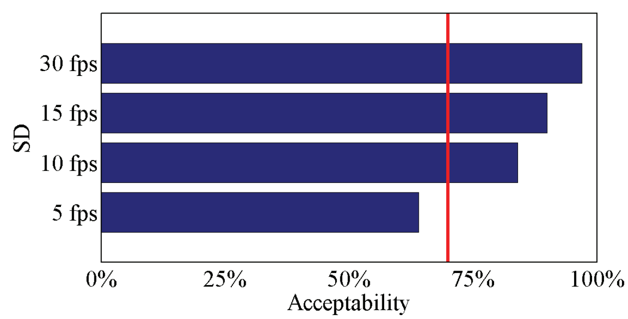


**Figure 4 –** Acceptability of pristine SD video at different frame rates.

### 3.3. Luma Image Size

The luma video signal is the black and white portion of the video picture. The luma image size establishes an upper limit of the usefulness of what could be achieved by an optimal video system. For example, the fact that a video system could produce 352 by 288 useful pixels does not mean that it actually does – every other pixel could be a pixel replication.

The minimum recommended luma image size for the video system is approximately half that of standard television, both vertically and horizontally. Vendors may refer to this as "compression resolution" and any of the following sizes are suitable: CIF (352 x 288), SIF (360 x 240), and QVGA (320 x 240). This is supported by the results presented in figures 5 and 6. Figure 5 shows the acceptability of 30 fps video as image size decreases from SD (720 x 486), to SIF, to QSIF (180 x 120). Figure 6 shows the acceptability of 10 fps video as image size decreases.

Before being displayed, CIF/SIF/QVGA video should be up-sampled by a factor of 2 in both the horizontal and vertical direction, to approximately the luma image size of a standard television. This up-sampling does not increase the inherent resolution of the image. However, preliminary display investigations indicate that first responders require the larger display size and resolution provided by a standard television monitor, or the equivalent portion of a computer monitor.

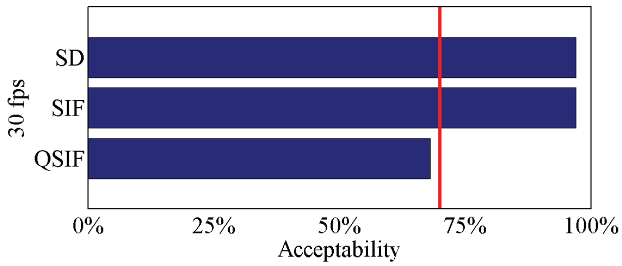


**Figure 5 –** Acceptability of 30 fps video at different image resolutions.

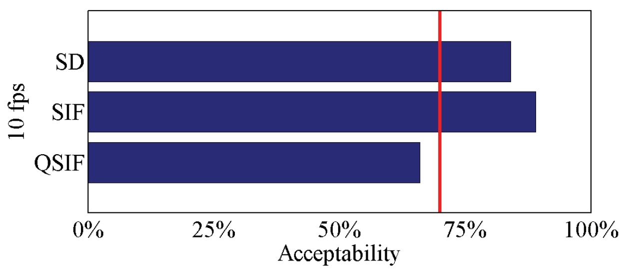


**Figure 6 –** Acceptability of 10 fps video at different image resolutions.

### 3.4. Lossless Coding & Transmission

Coding and decoding are a reality of today's digital video systems. Coding entails compression, which enables a video service to be transmitted using a bandwidth that cannot accommodate the non-compressed video signal. The flip side of this coin is that some of the quality of the video signal may be lost. Lossless video coding & transmission means that the video stream entering the video system (point B in figure 1) is bit-identical to the video stream leaving (point E in figure 1). Lossless video coding & transmission is *not* required for tactical video.

### 3.5. Coder Type & Bit Rate

Coder bit rate is the amount of information (in bits per second) output by the video coder to the network, excluding all network overhead (e.g., transport & protocol overhead). Different coding algorithms will have different minimum recommended coder bit rates. The following minimum coder bit rates are recommended for MPEG-2 and H.264 coders if they are encoding SD image sizes.

- Minimum of 1.5 Mbps for MPEG-2, a widely available, established coding scheme.
- Minimum of 768 kbps for H.264, also known as MPEG-4 Part 10, or AVC (advanced video compression). H.264 is a state-of-the art coding method that requires only one-half to one-third the bit rate of MPEG-2 for the equivalent quality. Thus, within the next several years, H.264 can reasonably be expected to become the coder of choice for many new services being deployed.

Older coders will likely require coder bit rates similar to or significantly higher than MPEG-2 (e.g., MPEG-1, MPEG-4 Part 2, H.263, Motion JPEG). Newer coders will likely require bit rates similar to or perhaps even lower than H.264. Encoding at CIF/SIF/QVGA resolution will likely produce acceptable video at lower bit rates than given above (under study).

### 3.6. Packet Loss & Error Concealment

Packet Loss Ratio is the fraction expressed in percentage of packets lost by the network. The maximum packet loss ratio that is acceptable depends upon the exact codec and network deployed. The following maximum packet loss ratios are recommended for MPEG-2 and H.264 codecs without error concealment and/or correction.

- Maximum of 0.1% for H.264 video codec.
- Maximum of 0.5% for MPEG-2.

When operating above these packet loss ratios, the quality and thus acceptability of video sequences drops rapidly as packet loss increases. Figure 7 depicts this rapid drop for system "384 kb/s" (a software H.264 codec using SD resolution and 384 kb/s) and system "1.5 Mb/s" (the same software H.264 codec using 1.5 Mb/s).

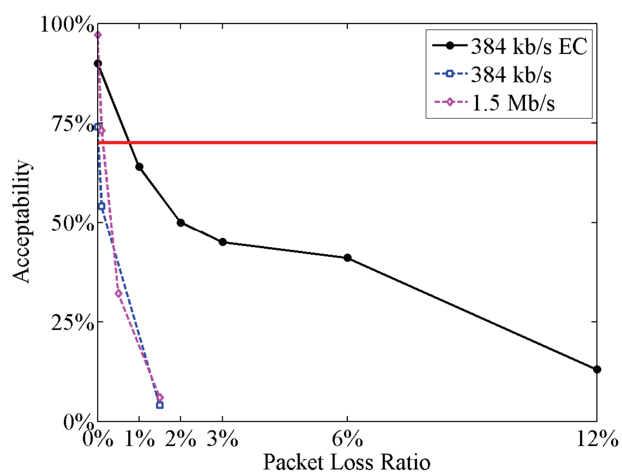


**Figure 7 –** Acceptability of H.264 systems with increasing packet loss ratios, with and without error concealment.

When error concealment and/or correction strategies are implemented, the quality and acceptability of video sequences drops off more gracefully with increased packet loss ratio. Also, significantly higher packet loss ratios can be tolerated. Based on preliminary data, the maximum packet loss ratio is expected to be around 1% to 2% for systems that perform error concealment. These results can be contrasted with the results of a parallel study on speech quality, which used G.711 speech coding with packet loss concealment. At the same 70% acceptance criteria, this speech quality study found packet loss ratios as high as 10% to 15% to be acceptable.

This more graceful drop of acceptability for systems with error concealment (EC) can be seen in figure 7 for system "384 kb/s EC" (a hardware H.264 codec that implemented error concealment, using SIF resolution and 384 kb/s). Figure 8 displays the same frame output by the "384 kb/s" system with 0% packet loss (upper left) and 1.5% packet loss (upper right); and the "384 kb/s EC" system with 0% packet loss (lower left) and 12% packet loss (lower right). The right-hand frames depict the decoders' typical response to packet loss.

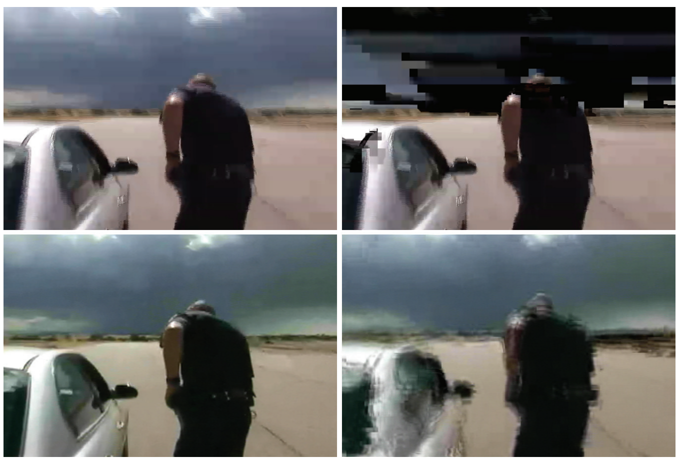

**Figure 8 –** Sample frames, depicting H.264 at 384kbits with & without EC.

### 3.7. Objective Video Quality

The above results for coder types, bit rates, packet loss rates, and error concealment strategies identify only a few of the many video systems available. Fortunately, the relationship between video quality and acceptability is a simple one (see figure 3). Thus, the suitability of other standard and non-standard codecs and networks can be determined by estimating the quality of the end-to-end video system.

The NTIA General Video Quality Metric ($VQM_G$) predicts what people would say about the quality of a video system, if subjective testing had been performed in a controlled environment. $VQM_G$ has been standardized by ANSI T1.801.03-2003 [3] and is available to all interested parties [4]. To utilize this software, users must inject source video into the system (at point B in figure 1), and record the resulting destination video coming out of the system (at point E in figure 1). Systems with integrated camera/coder or decoder/display cannot be measured if these systems do not have standard video inputs and outputs. Here is the recommended procedure to measure $VQM_G$:

1. Select eight video sequences of 8 to 12 seconds each that span the range of scene content for the application being deployed. These scenes should be of the highest possible quality, and each scene should contain content that is substantially different from the other scenes being used. Scenes should span the range of spatial detail, motion, color, and lighting levels for the user application being considered. Development of a standard set of scenes to use for this purpose will require further research and development.

2. Inject these scenes into the video system at point B in figure 1 and record the output at point E. The playing and recording of the source and destination scenes should be of the highest possible quality – uncompressed recording formats are recommended.

3. Compute $VQM_G$ for each pair of source and destination video streams from steps 1 and 2, respectively. $VQM_G$ outputs values from 0.0 (i.e., no perceptual impairment) to approximately 1.0 (maximum impairment).

4. Average the $VQM_G$ scores for all scene pairs. An average $VQM_G$ of 0.41 or less is recommended.

## 4. Video Quality Today

The above minimum recommendations were developed looking toward the future. Thus, the question naturally arises, to what extent does the video equipment available today meet these requirements? Equipment used for tactical applications falls into two broad configurations. Sometimes a camera is attached directly to a display by wires carrying uncompressed video (e.g., a camera on a pole). Equipment of this sort does not utilize compression and will typically meet or exceed the specifications in Section 3. Other video systems compress the video and then transmit this compressed information over a digital network. The state of these components today will be considered.

### 4.1. Video Compression Today

Today, there are many codecs available that meet or exceed the recommended minimum specifications for one-way delay, image size, and frame rate. Some tactical systems have a one-way delay that exceeds the recommended maximum delay. However, first responders indicated (via the questionnaire, in response to visual examples) that "more delay and a better picture quality" was preferable to "less delay and a worse picture quality" for tactical video.

The codecs on the market today utilize a wide range of coding standards, including proprietary codecs. Standardized codecs include Motion JPEG (M-JPEG), MPEG-1, MPEG-2, MPEG-4, and H.264. H.264 is newer and hence more difficult to find in common use today, but increasingly the codec is being deployed by manufacturers. Video systems typically have optional codec bit rate settings ranging from significantly below to somewhere above the recommended minimum bit rates. The acceptability of such equipment will depend upon how the equipment is set up.

Unlike audio codecs, most video codecs available today do not implement error concealment and/or correction and, as a result, the minimum requirements for packet loss are very stringent. This is probably the most problematic issue for today's first responders.

### 4.2. Networks Today

Current wired enterprise networks are often over-engineered and can achieve packet loss percentages that are a small fraction of one percent under light usage cases. However, under normal levels of usage, packet loss values between 1% and 2% in enterprise networks are not uncommon [5]. The network begins to experience queuing delays due to an increased traffic load, and the delayed packets arrive too late to be used.

Wireless data networks are considerably less robust, since they generally have only a fraction of the throughput capacity of their wired counterparts, and also suffer from impairments due to interference and radio propagation effects that wired networks do not contend with. To combat these effects, wireless protocols often send extra information for error correction, and may include some sort of limited redundant transmission. In spite of these mitigations, wireless 802.11 networks experience wide variations in packet loss percentages depending on signal strength and channel characteristics. A recent study of an 11-node stationary indoor network measured an average packet loss of 11% under good signal strength conditions, with no interference [6]. However, the introduction of an interference source similar to a cordless phone caused packet losses to approach 50% during some intervals.

These values indicate that transmission of usable public safety video data over wired networks is probably a realistic goal in the near term, but accomplishing the same goal with current wireless technology promises to be more problematic.

## 5. Conclusions

ITS continues to investigate first responders' requirements for video equipment. Investigations underway on tactical and surveillance video will further explore the interactions between codecs and networks. Wireless networks are anticipated to be particularly useful for first responders, and the propagation challenges of deploying ad-hoc wireless networks in emergency situations warrant further investigation. Further research and development will be needed to address first responders' minimum requirements for cameras, displays, recorded video, forensics video, telemedicine, infrared, and possibly HDTV video. Any first responders willing to support this effort by participating in subjective testing should visit www.its.bldrdoc.gov/psvq/.

Most of the PS-SoR Volume II specifications for tactical video can be met with equipment on the market today. The challenge for the future lies in the development and deployment of video codecs, networks, and in particular wireless ad-hoc networks that address the packet loss issues presented herein.

Part of the challenge ITS faced in developing PS-SoR Volume II was to identify applications that appropriately categorize first responder video needs. Based on discussions with first responders regarding the scenarios in PS-SoR Volume I, tactical video was split into two sub-categories: narrow field-of-view ("zoomed in") and wide field-of-view ("zoomed out"). These two sub-categories were likewise identified for observed surveillance video (which was defined as "video observed in real time to make decisions on whether a situation requires a response"). Recent subjective testing results (not described in this paper) indicate that first responders have statistically equivalent video quality requirements for all four of these real time applications. Thus, the next version of PS-SoR Volume II is expected to list video performance specifications for the joint application "tactical & observed surveillance video."

## References


[1] "Public Safety Statement of Requirements," Volume II: Quantitative, Version 1.0, October 2006. Available: http://www.safecomprogram.gov/SAFECOM/library/technology/1258_statementof.htm.

[2] ITU-R Recommendation BT.500, "Methodology for subjective assessment of the quality of television pictures," Recommendations of the ITU, Radiocommunication Sector. Available: www.itu.org.

[3] ANSI T1.801.03 – 2003, "American National Standard for Telecommunications – Digital Transport of One-Way Video Signals – Parameters for Objective Performance Assessment," American National Standards Institute. Available: www.ansi.org.

[4] M. Pinson and S. Wolf, "Video quality metric software, Version 2," NTIA Software/Data Product SD-03-396, Volumes 1-5, October 2002. Available: www.its.bldrdoc.gov/n3/video/vqmsoftware.htm.

[5] R. Pang, M. Allman, M. Bennett, J. Lee, V. Paxson, and B. Tierney, "A first look at modern enterprise traffic," in *Proc. 2005 ACM Internet Measurement Conf.*, pp. 14, 2005.

[6] M. Souryal, L. Klein-Berndt, L. E. Miller, and N. Moayeri, "Link assessment in an indoor 802.11 network," *Proc. WCNC 2006*, paper PHY30-1.