The Relationship Among Video Quality, Screen Resolution, and Bit Rate

G. Cermak, M. Pinson, and S. Wolf

Abstract—How much bandwidth is required for good quality video for a given screen resolution? Data acquired during two Video Quality Experts Group (VQEG) projects allow at least a partial answer to this question. This international subjective testing produced large amounts of mean opinion score (MOS) data for the screen resolutions QIF, CIF, VGA, and HD; for H.264 and similar modern codecs; and for many bit rates. Those data are assembled in the present report. For each screen resolution, MOS is plotted as a function of bit rate. A plot of all four data sets together shows the bit rate that would be required to achieve a given level of video quality for a given screen resolution. Relations among the four data sets are regular, suggesting that interpolation across screen resolutions might be reasonable. Based on these data, it would be reasonable to choose a bit rate, given a screen resolution; it would not be reasonable to choose a screen resolution given a bit rate.

Index Terms—Bit rate, CIF, H.264, HDTV, QCIF, quality, subjective testing, VGA.

I. INTRODUCTION

VIDEO is being offered on screens of all sizes from cell phones to stadium displays, and this trend appears to be increasing as more new devices with video capability enter the market, and as more powerful networks (e.g., 4G) are developed. Perceived quality of video depends on many factors such as bit rate, transmission errors, and coding method [1], [2], and the effect of some of these factors may be influenced or mediated by the size and resolution of the output display screen. In particular, the bit rate required to achieve a given level of video quality almost certainly depends on the characteristics (e.g., size, resolution) of the output display.[1]

Determining an adequate bit rate for a given size device can be done on a case-by-case basis in the lab (taking into account type of codec, type of material, and so on). However, one can also get an estimate of required bandwidth from test data that are in the public domain. In this report we assemble public-domain video quality data for screens of different resolutions at different bit rates. The data are presented in a single chart that shows the relationship between bit rate and judged video quality as a function of screen resolution.

The data are empirical rather than theoretical, and so are based on particular screens, screen resolutions, codecs, video material, and so on. Naturally, as technology changes then empirical curves such as the ones presented here will also change. However, the curves presented here should be useful in the near term, and the idea of the method for creating such curves should be useful in the longer term with new data sets.

Several other factors that are likely to affect the subjective value of video quality are beyond the scope of the current report; for example, the role of end-user expectations based on screen size and resolution, kind of video material viewed, price of a device and video service, and advertising. Still, actual perceived video quality must be at least an important component of the overall value of video quality, and perceived video quality is the subject of this report.

II. DATA AVAILABILITY

A. Experimental Data

To produce plots of the required bit rates for given levels of video quality for different screen sizes or resolutions, [2] two major requirements must be met: (1) video quality, bit rate, and screen size or resolution must be measured, and (2) bit rate and screen size or resolution must vary independently in a set of data. In the present case, video quality was measured by multiple panels of consumers making subjective judgments. When a sufficient number of consumers' rate video quality and the data are averaged, the resulting average rating or mean opinion score (MOS) is a stable and repeatable measurement [3]–[5] in the sense that data across labs correlate very highly (however, see caveats in Section IV). MOS is the "ground truth" against which objective algorithms for measuring video quality are judged. Bit rate and screen size were measured in the process of producing and displaying video samples; screen resolution was given by manufacturers' specifications.

The second requirement, statistical independence, affects how the data are interpreted. Statistical independence is important in two places: (1) in the relation between bit rate and screen resolution, and (2) in the relations among bit rate, screen resolution, and other factors that can affect video quality such as type of codec and transmission packet loss. Statistical independence between bit rate and screen resolution is satisfied because for any given screen resolution several bit rates are represented, and because the bit rates for the different screen resolutions overlap (see Fig. 5). Statistical independence of screen resolution and bit rate with respect to type of codec and presence of packet loss was achieved simply by analyzing only data from H.264 codecs in conditions of no packet loss. Because the datasets are

[1]Viewing distance influences the impact of screen size and resolution on video quality and is well controlled in subjective assessments. Viewing distance is normally chosen to maximize the perception of impairments (i.e., matched to the resolution of the human visual system) or to emulate the natural usage of the device for a chosen application.

[2]Screen size and screen resolution are different but correlated for the datasets being examined. In some parts of the text we use the terms interchangeably, but data are all given units of screen resolution (e.g., 640 pixels).

quite large, the statistically independent subsets presented here still contain enough data to be reliable. More detailed statistical analysis of these data is presented in [6] and [7].

B. VQEG Multimedia (MM) Test Data

The Video Quality Experts Group (VQEG) is composed of industry, government, and university groups concerned with video and multimedia quality and their measurement. VQEG conducts projects, most of which are aimed at developing and proving algorithms for objectively measuring video quality. The typical project contains two parallel evaluations of test video material. One evaluation is by panels of human observers (i.e., subjective testing). The other is by objective computational models of video quality. The objective models are meant to predict the subjective judgments. For present purposes, only the subjective data sets and corresponding information about the production of the test video material are of interest.

During the VQEG Multimedia (MM) Test [3], a collaboration of 17 labs produced 41 data sets spread nearly equally across three video resolutions, VGA, CIF, and QCIF. Each video treatment was defined by its scene, bit rate, encoded frame rate of 2.5 to 30 frames per second (fps),[3] packet loss level, codec, and other specialized video production parameters. Specifics of the testing, such as the H.264 encoding parameters and display equipment, can be found in the VQEG report [3] and documents contributed to the ITU and ATIS standards bodies [6], [7].

The subjective MOS scores were tested for consistency and repeatability by correlating the results across labs: For each screen resolution, a set of 30 video sequences was common across all 13 or 14 labs participating in the test. The minimum and median across-lab Pearson correlations were 0.94 and 0.97 respectively. These results indicate that the MOS part of the data was of good quality.

The labels used on the five-point rating scale employed in the VQEG MM test were:

>   5 = Excellent
>   4 = Good
>   3 = Fair
>   2 = Poor
>   1 = Bad [8]

Regardless of whether the labels meant the same thing to all viewers, or exactly what those labels meant, the VQEG MM test results show that video samples with better video production parameters (e.g., lower packet loss, higher bit rate) are consistently rated higher on this MOS scale than video with poorer video production parameters.

Tests at each of the three screen resolutions (QCIF, CIF, and VGA) were conducted on the same high-quality monitors. At each resolution, the display was at the native resolution of the monitor (i.e., each pixel of the video picture was mapped to a pixel on the screen without scaling).[4] Thus, the sizes of the displays for QCIF, CIF, and VGA scale according to their respective numbers of lines and pixels (see Table I). The monitor

---

TABLE I
RESOLUTION AND SCREEN SIZE FOR VQEG MM TESTS

Standard	Resolution, Pixels	Image Size on Monitor, Inches Diagonal	Viewing Distance, Picture Heights (H)
QCIF	176 x 144	2.2 – 2.6	6H – 10H
CIF	352 x 288	4.5 – 5.3	6H – 8H
VGA	640 x 480	7.9 – 9.3	4H – 6H

sizes ranged from 17 inches to 30 inches (diagonal), and their resolutions ranged from 1280×1024 to 2560×1600. However, the pixel density on the monitors in different labs was very nearly the same; the range in pixel density was 86.4 pixels/inch to 101.3 pixels/inch.

In this paper's primary analysis, for a given resolution and a given bit rate, only data at zero packet loss for H.264 are presented. These data are averaged across scenes (original video material), across the 24 human judges, across source frame rates (25 fps and 30 fps), across GOP structures, and across other processing parameters (e.g., differences in pre- and post-processing). A secondary analysis combines data from H.264 and the other modern codecs in the VQEG MM data sets (i.e., RV10, MPEG-4, and VC-1). [5] Thus, the results presented here reflect some of the diversity that characterizes the real world of video, but also are focused on a specific common codec and conditions of no transmission loss. Packet loss conditions were eliminated from this study due to lack of sufficient data to fully characterize changes in MOS with respect to packet loss, and since individual decoder responses to packet loss can vary significantly.

C. VQEG High Definition (HD) Video Test

The VQEG HD Phase I project [4] adds data at the high end of the resolution spectrum. Standard Definition (SD) data are also available for multiple bit rates for another VQEG dataset [5], but those data are all for MPEG2. By contrast, the VQEG HD test data include H.264, which allows direct comparison with the QCIF, CIF, and VGA data from the VQEG MM dataset.

Like the VQEG MM project, the purpose of the HD test was to evaluate objective models of video quality. The same five-point rating scale as in the VQEG MM test was used. Specifics of the HD test designs, encoding parameters, and equipment are given in [4]. An overview is that the VQEG HD test consisted of six experiments supported by 15 laboratories.

All video samples were displayed on monitors with 1920×1200 native resolution in controlled lab conditions meeting ITU test specifications [8]. The diagonal display sizes of the monitors varied from 24 inches to 47 inches and all the tests were conducted at a viewing distance of three times the picture height (3H). The monitors were all LCD, either high-end consumer TV or professional grade LCD monitors. The lab-to-lab correlations of MOS scores for a subset of video samples that were judged by all labs ranged from 0.94 to 0.99; as in the multimedia test, the averaged MOS scores are reliable enough to be considered meaningful data.

---

[3]By "encoded frame rate" we mean the number of unique frames per second output by the encoder, which may vary over time.

[4]In the VQEG MM test, the video resolution was smaller than the screen resolution (e.g., VGA displayed on a 1920×1200 monitor). However, the visual impact simulated a smaller screen (e.g., VGA displayed pixel-for-pixel with a small black border, surrounded by gray and viewed at a simulated hand-held distance).

[5]H.264, RV10, MPEG-4, and VC-1 are modern codecs that exhibit similar behavior regarding quality versus bit rate, and thus may be combined. The older codecs such as MPEG-2 have a different response curve and so were excluded from this secondary analysis.
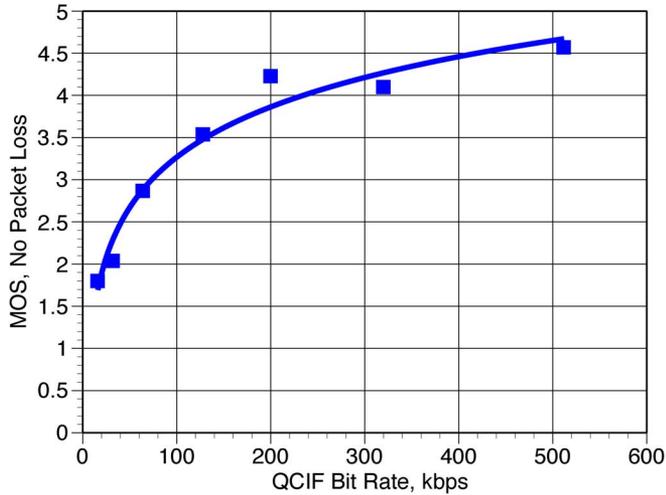
Fig. 1. QCIF results from VQEG MM project; H.264, no packet loss.
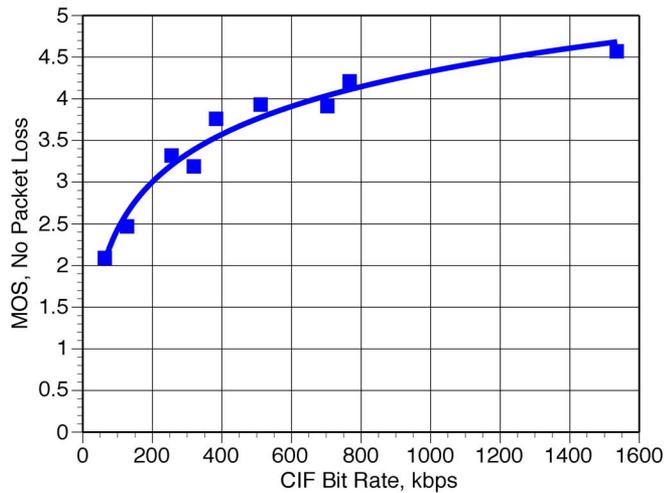
Fig. 2. CIF results from VQEG MM project; H.264, no packet loss.
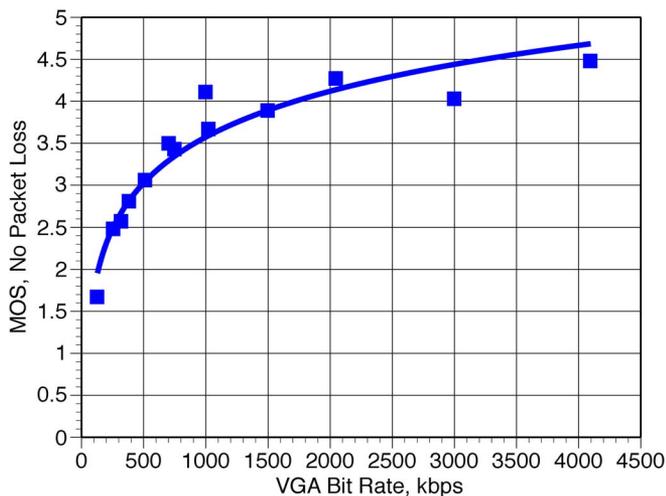
Fig. 3. VGA results from VQEG MM project; H.264, no packet loss.

The full VQEG HD data set includes data for video sequences from different codecs, different levels of packet loss, two source
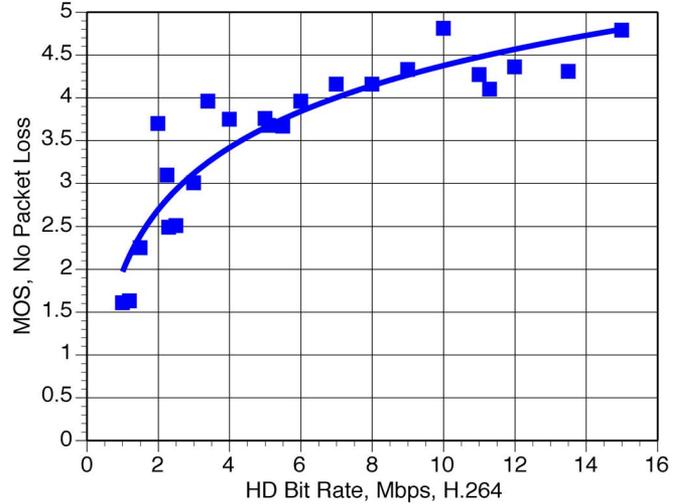
Fig. 4. HD results from VQEG HD project; H.264, no packet loss.

frame rates (25 and 30 fps), several different bit rates, different original video source material, and other specialized video production parameters. The level of aggregation of these data presented here is: Only data for H.264 codecs and no packet loss are included; the data are categorized by bit rate. [6] The data are aggregated across the two source frame rates, across differences in the original source video material, across individual judges, and across labs. Thus, the HD data points that follow consist of (MOS, bit rate) pairs for H.264 coding, displayed at 1920 × 1200 resolution in similar lab conditions. The data are averaged across other factors.

III. DATA

A. MOS vs. H.264 Bit Rate for Each Resolution

Figs. 1 –4 plot MOS as a function of bit rate for the VQEG QCIF, CIF, VGA, and HD datasets aggregated as described above (H.264, no packet loss). The fitted curves all have the same general shape that demonstrates diminishing returns from increasing the transmission bit rate to the right of the knee. For a given MOS level, the required bit rate does not quite increase linearly according to the number of pixels in the image $(\text{QCIF, CIF} = 4 * \text{QCIF}, \text{VGA} \approx 12 * \text{QCIF}, \text{HD} \approx 82 * \text{QCIF})$. Relatively fewer bits are required at the higher resolutions to achieve the same MOS.

B. H.264 Bit Rate for Desired Level of Quality for Different Screen Resolutions

By reading the bit rate at each level of MOS from the individual figures above, and plotting them for the different horizontal screen resolutions, Fig. 5 emerges.

For each of the levels of subjective video quality, MOS, the relationship between horizontal screen resolution and bit rate necessary for that MOS appears quite regular. It seems likely that the relationship between MOS and bit rate would also hold for horizontal screen resolutions interpolated between those shown in Fig. 5. This information may be useful in estimating the amount of bandwidth necessary to support video

[6]VQEG HD examined only the H.264 and MPEG-2 codecs.
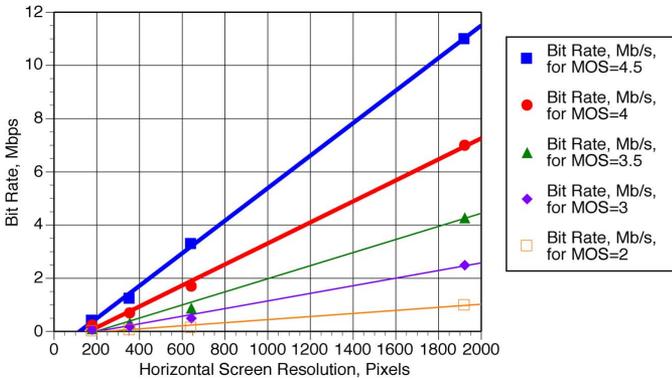
Fig. 5. H.264 bit rate required for desired level of MOS at different screen resolutions.

at given quality levels for devices with given horizontal screen resolutions.

C. MOS vs. Bit Rate for QCIF, CIF, and VGA, Using Modern Codecs

The VQEG QCIF, CIF, and VGA datasets can also be aggregated over all modern codecs (H.264, MPEG-4, RV10, and VC-1) with no packet loss. The fitted curves and data points are very similar to those seen in Figs. 1–3, which contain only data from H.264. The inclusion of these other codecs increases the data points available by approximately two times (an increase from 560 to 1432 for QCIF, 668 to 1640 for CIF, and 848 to 1512 for VGA).

IV. CAVEATS

While we believe the data shown here are valuable, and are certainly based on the largest and broadest video quality testing we know of, the data do have limitations, some of which we acknowledge here. The first limitation is not technical, it is a consumer behavior limitation: the video quality data do not necessarily translate directly into sales because video quality is only one of several factors that contribute to sales of video display devices, e.g., mobility, size, content availability, price, and design.

Among technical limitations, one is that these are 2007–2009 data, and technology has been changing quickly. The results presented here will require updating. Also, the results pertain to H.264 codecs and conditions of no packet loss. Further, the results were obtained in lab viewing conditions designed to maximize the observer's ability to detect small differences in video quality (e.g., close viewing distances to maximize visual acuity in ideal lighting). Real-world viewing conditions may include poor lighting, poor viewing angle, and distractions of many kinds. Thus, the VQEG data should represent viewers' judgments at their most critical. Therefore, the figures for bandwidth necessary for a given MOS may represent an upper bound.

All the data for any given viewer were obtained at a single screen size. Viewers did not compare the video quality at one screen size and bit rate with the quality at another size and the same bit rate. Suppose a device designer had a fixed bit rate budget to work with; would the designer use Fig. 5 to choose

the smallest possible screen resolution because it would have the greatest judged video quality? Probably not. The logic of the present data instead is that, given a fixed screen size or resolution, the designer and the network planner can use Fig. 5 to estimate the bit rate necessary to support a target level of video quality assuming H.264 encoding.

More generally, does a MOS of X for one screen size mean the same thing as an MOS of X for a different screen size? For screens of approximately the same size and function, MOS scores probably do mean the same thing. For three-inch screens on mobile devices compared to 50-inch stationary screens, probably not. However, what MOS values for screens of very different sizes do have in common is this: The MOS values provide an ordering of the various video segments. Abstractly, MOS for a three-inch mobile device may mean something different than MOS for a 50-inch stationary device, but in each case MOS provides an ordering of video segments in terms of their quality. Moreover, the set of videos being judged often includes examples of video at the highest quality achievable on a given screen. Thus, every video can be compared to the best quality video by means of their MOS values, three-inch videos compared to the best quality three-inch videos, and 50-inch videos compared to the best quality 50-inch videos.

Also, the scale range of MOS scores tends to be comparable across screens of different sizes. There is no logical or mathematical constraint that forces this fact, but it is often empirically true-certainly in the case of the VQEG datasets above (Figs. 1 –4). Therefore a MOS of 3 for a small screen means that the video is about as far from the quality of the best video at that screen resolution as is a video of MOS 3 that is shown on a large HD screen. That is, intervals on MOS scales for screens of different sizes/resolutions mean approximately the same thing regarding the relative relations of perceived quality for video samples. Absolute "quality" may actually mean something quite different for a 50-inch HD screen and a 3-inch QCIF screen, but MOS differences mean roughly the same thing for the two cases.

In addition, a video test set usually includes well over 30 videos. Because the MOS scores create an ordering of the videos, and the set being ordered is relatively large, correlations among sets of MOS scores for the same set of videos can be quite large (e.g., see [3], [4]) even though the sets of scores may differ by some linear transformation [9]. Returning to Fig. 5, the consequence of the fact that MOS provides an ordering of video quality within a given screen size, and that the video quality ratings are generally bounded between about 2.0 and 4.5, is that the slopes of the lines for the different MOS levels cannot be very far wrong, even if they are approximate. So, for example, considering the line in Fig. 5 for $MOS = 4.0$ and a horizontal screen resolution of 1200 pixels, realistically the corresponding bit rate must lie somewhere between 3 and 5 Mbps, and probably closer to 4.

The meaning of MOS from different labs also leads to logical uncertainties. However, as in the case of MOS for different screen sizes, the practical consequences for the VQEG data presented here are probably minor. It is known that MOS from different labs correlates highly, but that the actual scale values usually differ by a linear transformation [9]. So, what is the "correct" value of MOS for a given test video sequence that had

been judged by multiple labs? The answer from psychophysics [10] is that there is no "correct" MOS value. A more satisfying answer is that averaging the MOS from several labs, especially several diverse labs, is bound to approximate the MOS values from all labs. Such is the case with the VQEG datasets used here. To the extent that the approximation to "true" MOS values is inaccurate, the slopes of the lines in Fig. 5 will be inaccurate. However, given the interlocking constraints of the MOS scores having similar ranges, and ordering the test video sequences in essentially the same way, the "true" slopes of the lines in Fig. 5 could not be very different from the way they are drawn.

Corresponding to each of the caveats and uncertainties is an opportunity for further work on video quality: with improved technologies when they appear, with more processing and transmission scenarios under current technologies, and with other dimensions of video devices such as their portability, design, and access to content. A timely first step would be to *directly* compare the judged quality of video at different screen sizes and resolutions.

References

[1] DSL Forum, "Triple-play services quality of experience (QoE) requirements," Technical Report TR-126 2006 Digital Subscriber Line Forum.

[2] M. Pinson, S. Wolf, and G. Cermak, "Subjective quality of HDTV: H.264 vs. MPEG-2, with and without packet loss," *IEEE Trans. Broadcast.*, vol. 56, no. 1, pp. 86–91, Mar. 2010.

[3] K. Brunnstrom, G. Cermak, D. Hands, M. Pinson, F. Speranza, and A. Webster, "Final report from the video quality experts group on the validation of objective models of multimedia quality assessment, Phase I," VQEG, 2008.

[4] M. Pinson, F. Speranza, M. Barkowski, V. Baroncini, R. Bitto, S. Borer, Y. Dhondt, R. Green, L. Janowski, T. Kawano, C. Lee, J. Okamoto, R. Renaud, C. Schmidmer, N. Stalens, A. Takahashi, and Q. Thu, "Report on the validation of video quality models for high definition video content," VQEG, 2010.

[5] G. Cermak, P. Corriveau, M. Farias, T. Martin, M. Pinson, F. Speranza, and A. Webster, "Final report from the video quality experts group on the validation of objective models of video quality assessment, Phase II," VQEG, 2003.

[6] G. Cermak, "Multimedia video quality: Results from the VQEG multimedia test," ATIS PRQC-QoS Contribution, Aug. 18, 2008.

[7] G. Cermak, "Multimedia video quality: Results from the VQEG multimedia test," ITU-T SG-9 Contribution, Sep. 2008.

[8] "Subjective video quality assessment methods for multimedia applications," Geneva, Switzerland, ACR: ITU-T Recommendation P.910, 2007 [Online]. Available: www.itu.int

[9] M. Pinson and S. Wolf, "An objective method for combining multiple subjective data sets," in *SPIE Video Commun. Image Process. Conf.*, Lugano, Switzerland, Jul. 2003 [Online]. Available: http://www.its.bl-drdoc.gov/pub/n3/video/spie03obj.pdf

[10] R. N. Shepard, "Psychological relations and psychophysical scales: On the status of 'direct' psychophysical measurement," *J. Math. Psych.*, vol. 24, pp. 21–57, 1981.