Journal of Mobile Multimedia, Vol. 1, No. 4 (2006) 327–341

SUBJECTIVE QUALITY OF MOBILE MPEG-4 VIDEOS WITH DIFFERENT FRAME RATES

SHELLEY BUCHINGER
Institute of Distributed and Multimedia Systems
University of Vienna, Lenaugasse 2/8
Vienna, 1080, Austria
shelley.buchinger@univie.ac.at

HELMUT HLAVACS
Institute of Distributed and Multimedia Systems
University of Vienna, Lenaugasse 2/8
Vienna, 1080, Austria
helmut.hlavacs@univie.ac.at

Received July 4, 2005
Revised December 30, 2005

In this study we investigate the influence of the video frame rate on the subjective quality of digital video. MPEG-4 videos showing content of different type and frame rates, and having a resolution typically used in mobile environments, have been shown to a test audience, which then rated the subjectively perceived quality of the videos. The resulting mean opinion score (MOS) then indicates for given bitrates, which frame rate is optimal for the used videos. We show that in contrast to classical assumptions, the optimal frame rate often is as low as 10 or even 5 frames per second.

Keywords: Perceived quality, mobile video, MPEG-4 video, frame rate

Communicated by: I Ibrahim & G Kotsis

1 Introduction

The demand for mobile multimedia content, especially digital video, has increased drastically in the last years, and new efficient compression schemes together with rapid bandwidth growth have enabled services such as live broadcast and wireless video on demand. However, mobile video consumption is hindered by the typically limited bandwidth and unreliability of wireless access technologies like GPRS, UMTS, or soon mobile WiMAX. Additionally, mobile terminals like smartphones or personal digital assistants (PDAs) offer a much smaller display compared to, for instance, TV sets or small laptop computers. When using a specific codec, the reduction of bandwidth for a given video usually is done by one of the following ways: (i) reduce the image resolution of the video, (ii) decrease the image quality due to higher image compression rates (resulting usually in larger quantization), or (iii) reduce the frame rate, i.e., the number of frames sent per second (fps). If the video resolution is given by the terminal resolution, for instance Common Intermediate Format CIF (288 lines and 352 pixels per line) or Quarter Common Intermediate Format QCIF (144 lines and 176 pixels per line) [1], then only techniques (ii) and (iii) remain.

Any combination of (ii) and (iii), however, has a severe impact on the way human observers would judge the perceived quality of the videos. Image quality reduction affects the spatial information of the video, images get blurred or show tiles of, for instance, 8×8 pixels, image details no longer can be distinguished. Reducing the frame rate affects the temporal information of the video, videos with reduced frame rate get jerky, objects jump from one place to the other without smooth motion. If the video should be compressed for a fixed predefined target bitrate, then it has to be decided to which extent each of the aforementioned reduction principles should be used, since for instance a reduction of the frame rate will result in better image quality, a higher frame rate on the other side will result in lower image quality.

In principle, the quality of videos can be computed by using two ways. Objective metrics, for instance the Mean Squared Error (MSE) or the Peak Signal to Noise Ratio (PSNR) [2] compute mathematical metrics for the pixel by pixel difference, resulting in values which must be interpreted with care, since the connection between the metric and the quality perceived by humans is not obvious. Subjective metrics are derived by showing videos to a large audience and asking them for their subjective opinion. The main result then is given by the mean opinion score (MOS), i.e., the average over all ratings [3, 4].

Many researchers have investigated the effect of a reduction of frame rate on the subjectively perceived quality of videos. Reducing the frame rate usually is done for frame dropping. This means that the video has been compressed with fixed frame rate, for instance 25 or 30 frames per second, and is then streamed over an unreliable packet network with best effort service, limited bandwidth and unpredictable bandwidth variation. In case the available bandwidth suddenly drops, some of the video frames may be dropped in order to decrease the needed video bitrate [5, 6]. This method, however, does not affect the spatial information of the videos, since the sent frames are not altered. In [7] the authors focus on the relationship between raw quantization, i.e., image quality, frames per second, peak rate and MOS of MPEG-2 videos. In [8] the authors investigate the MOS-based optimal frame rate for sports videos showing high temporal information, i.e., fast camera and object movement. In [9] the authors derive the MOS-based optimal frame rate for different video types and codecs for videos compressed with the fixed bitrates 100 and 300 kbit/s, but without justifying their choice of encoding bitrate and restricting the parameter frame rate to 30, 15, and 10.

It is interesting to note that in all recent investigations about the influence of the frame rate on the subjectively perceived video quality, it has been found that if the available network bandwidth is limited, then the optimal frame rate usually is not the typical maximum like 25 or 30 frames per second, but lower. This is also the case for videos with high temporal activity, for instance sports or action sequences, contrasting previous assumptions that in such cases, the video compression should always trade image quality for a maximum frame rate. For achieving smooth motion we show in this paper that the optimum frame rate can even be as low as 5 frames per second.

This choice can be explained by the fact that a lower frame rate leaves more bits for the individual frames and thus results in better image quality (for less frames). To some extent human observers obviously prefer a jerky video with a good image quality to smooth motion with blurred images. Hence we aim at finding the optimum number of frames per second that gives the best subjective results when maintaining a certain image quality of the videos.

We chose the encoding bitrate b_i for source video i on the basis of image quality analysis first, thus ensuring that the image quality stays above some tolerance threshold. We then created videos having a resolution typically found in mobile terminals, with different frame rates but with fixed bitrate b_i for video i, and asked a number of human observers to rate the subjectively perceived quality of the videos and computed the MOS. Our findings are consistent with previous papers in the way that for limited bandwidth the frames per second should always be below the maximum.

2 Experimental Design

2.1 Used Standards

The methods used for our data collection and evaluation followed commonly used standards and settings to assure comparability and reproducibility of our results. Settings for realization of subjective tests elaborated by the MPEG-group were used according to published directives [10]. In this work, data was assessed by the SSCQM method (see Section 2.3).

Preprocessing and verification of obtained data were based on recommendations of the ITU [3] and on classic statistical methods [11]. Data analysis was performed following recommendations of the ITU [3].

2.2 Test Videos

For comparability and reproducibility of our results we followed standard rules for subjective video experiments [3, 4]. Choice of the test sequences is critical as their specific subjective ratings should lead to a valid evaluation for all the video clips. Hence we took the following parameters into account:

- Frame rate (frames per second)
- Resolution (width, height)
- Bitrate (kbit/s)
- Spatial activity (SA)
- Temporal activity (TA)

The test sequences were encoded in MPEG-4 (Chapter 2) using the Mpegable Broadcaster Version 2.1,[a] which offers the possibility to specify the parameters frame rate, resolution and bitrate. The frame rate was the central parameter of this investigation. For each setting of all other parameters, video clips were encoded for the frame rate values 5, 10, 15 and 25 frames per second. As subjectively perceived video quality also depends on the display size of the presentation [12], the video resolution was equally defined as parameter. In a first evaluation session, all the test sequences were encoded in CIF resolution (352×288). In a second step, only a subset of these sequences was encoded in QCIF resolution (176×144), in order to keep the number of necessary test runs as small as possible. The parameter spatial activity (SA) indicates the amount of color changes in a frame. The computation of its value was based on DCT analysis [13]. Temporal activity (TA) indicates the amount of color changes over time.

[a]http://www.mpegable.com/

We used a professional video tool for deriving the metrics SA and TA for the test videos.[b] The following test sequences have been selected according to their average temporal and spatial activity:

- Sequence *Conference* shows a talking head. It has nearly no SA and only little TA.
- Sequence *Comic* shows a typical animated scene as often seen in TV programmes made for children. It exhibits moderate SA and little TA.
- Sequence *Advertising* shows a shampoo commercial with fast movement and short sub-scenes. It can be described by moderate SA and TA.
- Sequence *Soccer* shows a fast sports scene of a typical soccer match. It has high SA and high TA.

Average SA versus average TA of the selected clips is shown in Fig. 1 and the respective standard deviations of SA and TA are listed in Table 1.

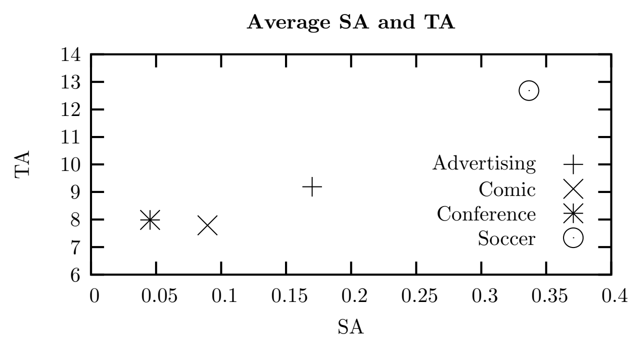

Fig. 1. Average SA versus average TA of selected video clips.

Table 1. Averages and standard deviations of SA and TA.

	Conference		Comic		Advertising		Soccer	
	avrg	std	avrg	std	avrg	std	avrg	std
SA	0.045	0.29	0.09	0.36	0.17	0.64	0.34	0.45
TA	7.98	0.33	7.79	0.42	9.19	3.64	12.69	3.71

In order to demonstrate the variability of SA and TA over time, their time-dependent behavior is shown in Fig. 2 and Fig. 3.

The video tool we used also implements an objective video metric mimicking subjective MOS scores for image quality called Digital Video Quality (DVQ) [13]. The used reference videos originally are encoded with very high bitrates and thus result in high DVQ quality parameter (qp) values, which denote best quality for value 100 and worst quality for value 0.

[b] http://www.rohde-schwarz.com/www/dev_center.nsf/html/111420frame

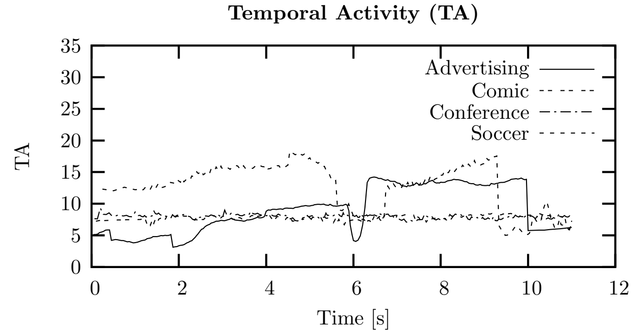

Fig. 2. Time-dependent TA variation of selected video clips

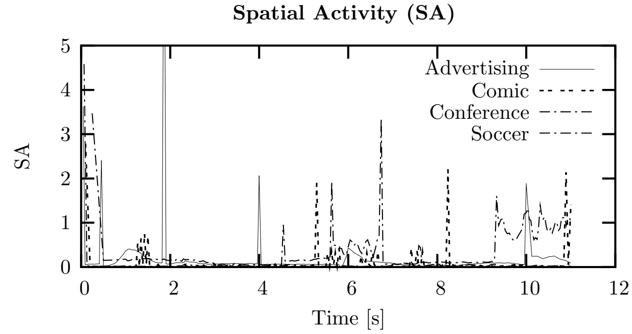

Fig. 3. Time-dependent SA variation of selected video clips

The original encoding bitrate must be regarded as limiting factor of possible picture quality. In a first step the new encoding bitrate was chosen such that the average value of the quality parameter qp obtained by DVQ evaluation ranged between 55 and 80. As the bitrate for the clip *Soccer* had to be set to 1500 kbit/s to achieve a quality parameter qp ranging within those boundaries, a much lower bitrate was investigated as well.

From Fig. 1 and Table 2 it can be concluded that sequences containing high TA and/or SA need a higher encoding bitrate to reach the same quality parameter qp as sequences with low TA and/or SA.

Table 2. Properties of reference sequences.

Video clip	Bitrate [kbit/s]	qp
Conference CIF	1923	86
Comic CIF	1986	80
Comic QCIF	433	94
Advertising CIF	2185	78
Advertising QCIF	341	89
Soccer CIF	5247	62

2.3 Test Environment

Subjective evaluation was achieved following the Single Stimulus Continuous Quality Method (SSCQM) [3, 4]. The data has been collected in two rounds at two different places, thus implying different environmental factors (light, position, etc.). In the first round, evaluations of test sequences in CIF resolution were performed, in the second round the sequences in QCIF resolution and additionally a lower bitrate (128 kbit/s) of the sequence *Soccer* in CIF resolution were assessed. The tests were performed on a notebook with 14.1" active matrix TFT display and Windows XP. The subjectively perceived video quality of our test sequences was assessed using software developed in Java. As the SSCQM was chosen to rate the video, the software contained a scrollbar allowing the test persons to change the rating continuously during video presentation (Fig. 4).

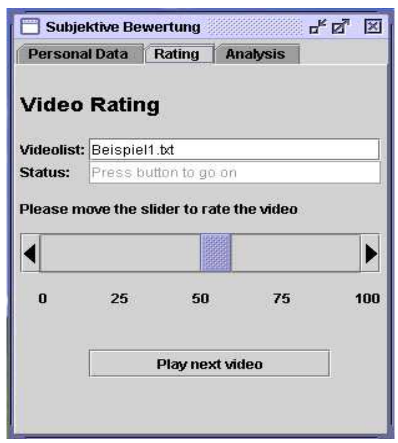

Fig. 4. Subjective video quality rating software.

The player ffplay[c] was used for video clip presentation. Each test sequence was evaluated by at least 15 non-experts. The video quality of all test sequences was rated by using DVQ as described above. In addition to the quality parameter qp, TA and SA, DVQ evaluated the parameter frame rate. Thus the quality rating per frame could be transformed into a quality rating by time unit, realized by calculating the mean quality parameter qp value of all the frames per time unit. Each test person was asked to accomplish one test run consisting of the presentation and rating of all selected video clips in random order.

2.4 Data Preprocessing

The collected data contained all ratings performed, including invalid scores and results of unreliable test persons that had to be filtered out before proceeding the analysis. Basically, Hartung [11] takes two types of measurement errors into consideration: (i) the systematic and (ii) the statistical error. As a measurement value consists of true value, systematic error and statistical error, detection and quantification of the latter are subject of the preprocessing phase. Furthermore different methods for identification of outliers are proposed by the ITU [3, 4].

[c]http://sourceforge.net/projects/ffmpeg/

A *systematic error* depends on the measurement tools that were applied. It can be reproduced by using the same settings and parameters. In the presented experiment a systematic error was detected by analyzing the software used to obtain the subjective rating.

The *statistical error*, however, is unpredictable. Its significance can be evaluated if multiple measurements with identical settings are available (i.e., repeated measurements), and is assessed using confidence intervals.

A single measurement can also accidentally be erroneous and must hence be considered as *outlier* that has to be removed. In our experiment, outliers typically resulted from a short distraction of a test person during the quality evaluation phase. Therefore, a test to identify such insignificant measurements was carried out before analyzing data.

Preprocessed collected data as will be pointed out in Sections 2.5, 2.6 and 2.7 was analyzed in three steps. First, by using the statistic tool *R*,[d] a simple analysis of variance [11] was performed to investigate whether the parameter frame rate influences the subjectively perceived quality of the test sequences. In a second step, the resulting data reflecting dependencies with respect to this parameter were diagrammed. These graphical figures then were used for further interpretation of results. In a third step the considered data finally was approximated by suitable models.

2.5 Scores

For each test sequence i and each test person j the mean rating score μ_{ij} over time was calculated. In other words, μ_{ij} denotes a single scalar subjective mean score that was given from person j to sequence i. Some test persons reacted more confidently than others during evaluations. They used the whole available range for their votes while more reserved persons only slightly moved the slide during the whole clip. As the same interval of quality parameter qp changes had different meanings for different persons, their votes were normalized. For each test person its minimum and its maximum rating value was calculated by

$$\mu_j^{\min} = \min_i \mu_{ij}, \; \mu_j^{\max} = \max_i \mu_{ij}.$$

Then the mean minimum and maximum rating value over all test persons was determined:

$$\mu^{\min} = \frac{1}{N}\sum_{j=1}^{N} \mu_j^{\min}, \; \mu^{\max} = \frac{1}{N}\sum_{j=1}^{N} \mu_j^{\max}.$$

These values were used to standardize the single ratings:

$$\hat{\mu}_{ij} = \frac{\mu_{ij} - \mu_j^{\min}}{\mu_j^{\max} - \mu_j^{\min}} \times \left(\mu^{\max} - \mu^{\min}\right) + \mu^{\min}.$$

As the overall quality estimation of the test persons differed, standardized scores were transformed into standardized-centered scores [3, 4]:

$$\mu_{ij}^{*} = \hat{\mu}_{ij} - \sum_i \hat{\mu}_{ij} + \sum_{i,\,j} \hat{\mu}_{ij}. \tag{1}$$

[d] http://cran.r-project.org/

2.6 Systematic Error

Systematic errors caused by the rating software were eliminated before data analysis. As illustrated in Fig. 4 a test person had to perform the following steps to rate a test sequence.

- The button “Play next video” had to be pressed to start the clip presentation.
- The mouse had to be moved to the scrollbar at its initial position.
- With mouse button pressed, the test person could start rating.

For each test sequence some time elapsed, until rating was performed. Therefore the average over time mentioned in Section 2.5 corresponds to the average over time minus the first 2.5 seconds per video clip and test person.

2.7 Outliers and Statistical Error

Ratings falsified by short distraction of the test person for example might have lead to an insignificant or invalid result. To identify such erroneous data, a David-Hartley-Pearson-Test [11] has been performed. Statistical errors were determined using two different methods: the β_2-test and the computation of the confidence interval. To verify whether a test person returned reliable ratings, a β_2-test for SSCQM was carried out [3]. This test equally indicated, whether the distribution could be considered as normal, a fundamental prerequisite for subsequent analysis of variance. Except for single seeds, the data resulted in being normally distributed. Reliability of results was assessed by computing a 95% confidence interval of raw and manipulated data (standardized, centered, rejection of outliers and unreliable observers).

2.8 Analysis of Variance

Analysis of variance was applied in order to investigate whether the parameter frame rate had an impact on the collected data. Each test sequence was considered with respect to the four different frame rates selected (5, 10, 15, 25 frames per second). Ratings related to one test sequence and frame rate were grouped. The p-value of the ratings of each clip distributed in four groups was calculated using the statistical tool *R*. Therefore the interpretation of the p-value has been extracted from [14], adapted to this particular study, and is shown in Table 3. The null-hypothesis was defined as follows: Frame rate changes of a test sequence have no influence on its subjectively perceived video quality.

Table 3. Interpretation of the p-value.

p-Value	Interpretation
$p \geq 0.1$	No indication against the null-hypothesis
$0.05 \leq p < 0.1$	Weak indication against the null-hypothesis
$0.01 \leq p < 0.05$	Moderate indication against the null-hypothesis
$0.001 \leq p < 0.01$	Strong indication against the null-hypothesis
$p < 0.001$	Very strong indication against the null-hypothesis

The null-hypothesis is considered as rejected if the p-value is smaller than 0.01. Table 4 shows the p-values for each test sequence. It can be concluded that the subjective ratings of all test sequences depend on the frame rate, which has a significant influence on the perceived video quality.

Table 4. p-Values of the ratings.

Video	Bitrate [kbit/s]	Resolution	p-Value
Conference	64	CIF	0.0003184
Comic	64	CIF	4.556e-09
Advertising	128	CIF	7.784e-06
Soccer	1500	CIF	0.0045
Soccer	128	CIF	7.08e-06
Comic	32	QCIF	4.521e-10
Advertising	48	QCIF	9.598e-06

The sequence *Comic* is highly affected by this parameter, followed by *Advertising* and *Soccer* at a low bitrate. The sequences with the highest score of quality parameter qp, *Conference* and *Soccer* encoded at 1500 kbit/s, yield the lowest dependency.

3 Subjective Results

The following figures give an overview of the subjective ratings collected. The single dots symbolize normalized and centered ratings from (1). The solid line represents the mean rating value of these scores and the dashed lines represent the boundary of the corresponding 95% confidence interval, here assuming a normal distribution (see Section 2.7). The median is not shown in these figures as the values are almost similar to the mean scores.

The results show a mean confidence interval width below 10.5 and a maximum confidence interval width of 20. This results in a mean error of around 5 and a maximum error of 10 quality parameter units with a probability of 95%. Furthermore as pointed out in [4, 10] interpretation of video quality is commonly done using the five quality classes *bad*, *poor*, *fair*, *good*, and *excellent*, each representing in this case 20 quality parameter units (bad if $qp \in [0, 20), \ldots,$ excellent if $qp \in [80, 100]$). Thus when mapping the MOS to these five classes, the correct class is chosen with high probability.

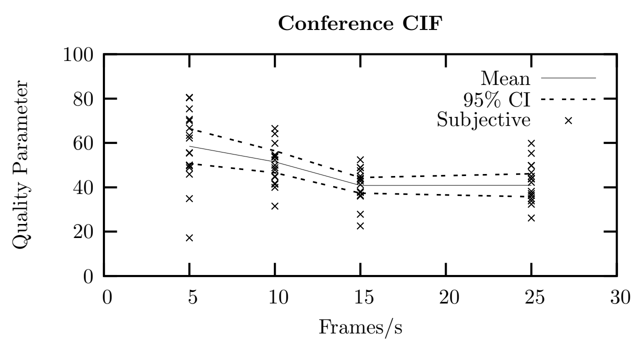

Fig. 5. Conference (CIF, 64 kbit/s)

The experimental results for the sequence *Conference* are shown in Fig. 5. The sequence contains both little SA and little TA. Especially since the TA is small, there is no need to smooth its motion, meaning that a high frame rate would be useless. Such a choice just affects the perceived quality in a negative way as there is less bitrate available for each single frame. In Fig. 5 it can be observed that the subjectively perceived quality first decreases, then remains constant at low level. The test persons could not notice any difference in quality for 15 and 25 frames per second. This means that there is a lower bound slightly above 40 qp, which is reached at approximately 15 frames per second.

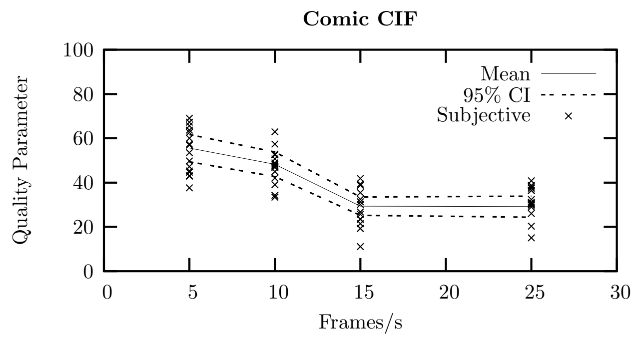

Fig. 6. Comic (CIF, 64 kbit/s)

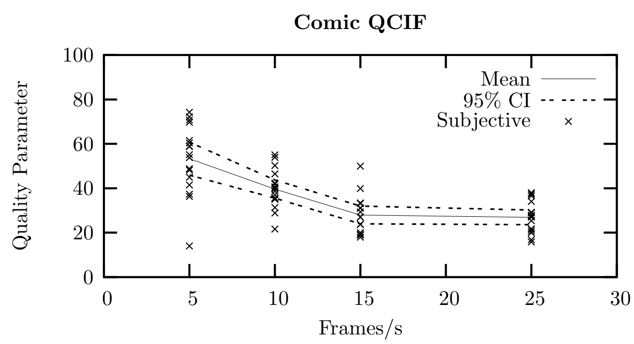

Fig. 7. Comic (QCIF, 32 kbit/s)

The results for the sequence *Comic* are shown in Fig. 6 and Fig. 7. The sequence contains little TA but more SA than *Conference*. In Table 4 it has been shown that the subjectively perceived quality of *Comic* is more sensible to the variation of the frame rate than *Conference*. This leads to the assumption that a lower frame rate must be selected for clips containing more spatial activity because more encoding bitrate is needed for each frame to produce the same picture quality as for videos with lower spatial activity. A frame rate increase implies a decrease of encoding bitrate available for each frame, and therefore the overall consequence is a more significant quality decrease of the video clip. In Fig. 6 and Fig. 7 it can be noticed that the quality decrease for the frame rates 5 to 15 frames per second is intensified with respect to *Conference*. Similar to *Conference*, a lower bound (30 qp) of subjectively perceived quality of *Comic*, reached at 15 frames per second can be observed.

The results for the sequence *Advertising* are shown in Fig. 8 and Fig. 9. This sequence contains some moderate spatial and temporal activity. Smoothing motion by raising a very low frame rate from 5 to 10 frames per second for sequences with medium temporal activity increases the subjective quality perception, but a further increase has negative effects because the decreasing picture quality becomes the dominant factor for the subjectively perceived video quality. In Fig. 8 and Fig. 9 it can be observed that the subjectively perceived quality increases from 5 to 10 frames per second, but decreases from 10 to 25 frames per second. This observation can be made for both CIF and QCIF resolutions. However, this effect is more accentuated in QCIF.

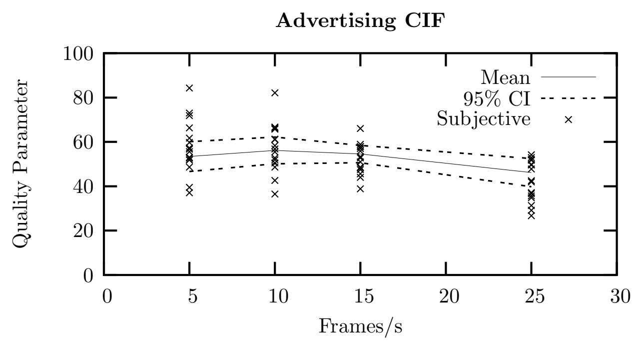

Fig. 8. Advertising (CIF, 128 kbit/s)

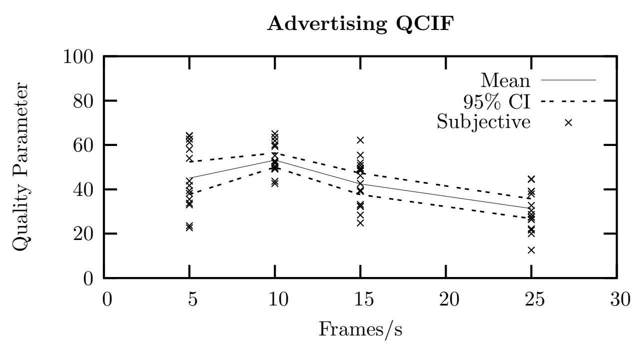

Fig. 9. Advertising (QCIF, 48 kbit/s)

The results for the sequence *Soccer* are shown in Fig. 10 and Fig. 11. This sequence has higher spatial activity and much higher temporal activity than the other clips. It has been used in the experiments with two different encoding bitrates. At a very high bitrate the curve shown in Fig. 10 is comparable to the result obtained for sequence *Advertising* in CIF resolution (Fig. 8). The quality increase is extended to 15 frames per second and the overall ratings are much higher. In this case the picture quality is high enough to ensure a high rating that can be improved by smoothing the motion. From 15 to 25 frames per second the decrease of the picture quality seems to become the dominant factor and the perceived quality decreases again. In the case of low bitrate-encoded *Soccer* sequence presented in Fig. 11 the same behavior as for *Advertising* in the QCIF resolution can be observed. The picture quality becomes dominant at 10 frames per second.

In general it can be said that for test sequences with *high* TA a much higher bitrate had to be used to obtain the same quality level. The increase of the SA had no effect on the needed

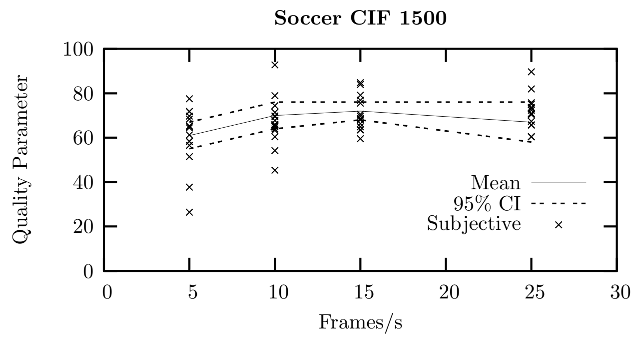

Fig. 10. Soccer (CIF, 1500 kbit/s)

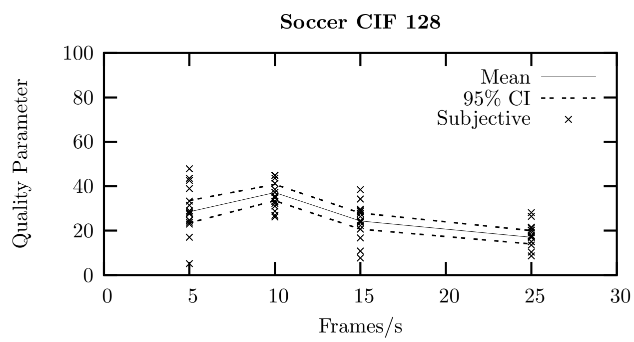

Fig. 11. Soccer (CIF, 128 kbit/s)

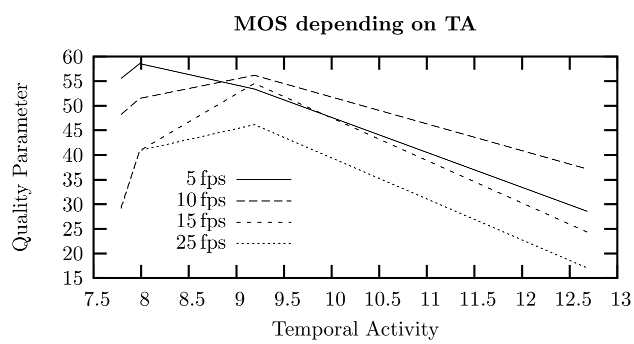

Fig. 12. MOS versus Temporal Activity.

bitrate. This leads to the question how the MOS depends on temporal activity. Fig. 12 gives an overview of our results. It shows the behavior of the MOS in relation to TA of each chosen frame rate. The MOS first reacts with an increase and then a decrease in all cases. For the TA interval from 7.7 to 8.7 TA a frame rate of 5 frames per second returns the highest MOS, while in the TA interval from 8.85 to 12.7 a choice of 10 frames per second delivered the best result. In both intervals a frame rate of 25 frames per second yields the lowest result.

4 Analytical Models

In order to estimate similarities between the observed MOS ratings for different videos we created analytical models for the MOS curves and compared their model parameters. The behaviour of the obtained quality ratings changes at a frame rate of 15 frames per second, therefore different models were created for the two intervals 5 to 15 frames per second and 15 to 25 frames per second. The major differences in quality perception were observed in the first interval while the quality parameter remained constant or decreased in the second interval. This means that there are three interpolation points available for an approximation of the subjective ratings in the first interval, resulting in a quadratic model. In Fig. 13 and Fig. 14 an approximation $\hat{qp}(x; ta, r, bps)$ of the subjective rating curves, defined by equations (2) to (8), is proposed. The parameters are defined as shown in Table 5.

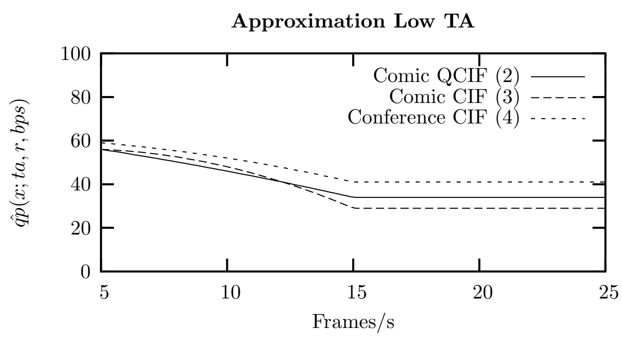

Fig. 13. Approximation of clips containing low TA.

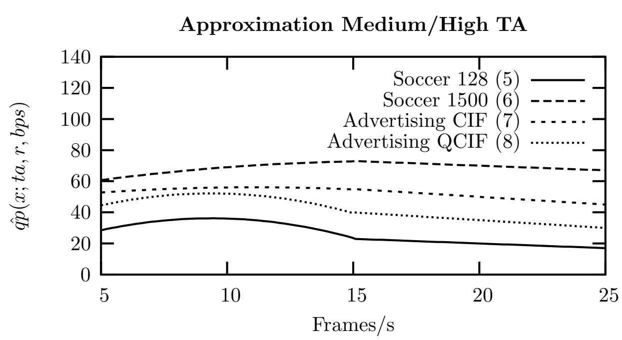

Fig. 14. Approximation of clips containing medium/high TA.

Table 5. Parameters of the approximation function $\hat{qp}$.

Parameter	Name	Range of values
x	frame rate	$\{5, 10, 15, 25\}$
ta	temporal activity	$\{\text{low } (l), \text{medium } (m), \text{high } (h)\}$
r	resolution	$\{\text{QCIF } (q), \text{CIF } (c)\}$
bps	encoding bitrate	32 – 1500

The subjective quality perception of the first interval was approximated by a polynomial function of degree two, and to the second interval by a linear function. The logistic function proposed by the ITU [3] was not used because its shape is not suitable for the collected results obtained in this work. The polynomials for the low-TA case are defined as follows:

$$\hat{qp}(x; l, q, 32) = \begin{cases} -0.04x^2 - 1.4x + 64 & \text{if } x \leq 15 \quad \text{(Comic QCIF)} \\ 34 & \text{else} \end{cases} \tag{2}$$

$$\hat{qp}(x; l, c, 64) = \begin{cases} -0.22x^2 + 1.7x + 53 & \text{if } x \leq 15 \quad \text{(Comic CIF)} \\ 29 & \text{else} \end{cases} \tag{3}$$

$$\hat{qp}(x; l, c, 64) = \begin{cases} -0.08x^2 - 0.2x + 62 & \text{if } x \leq 15 \quad \text{(Conference CIF)} \\ 41 & \text{else} \end{cases} \tag{4}$$

The polynomials for the medium/high-TA case are defined as follows:

$$\hat{qp}(x; h, c, 128) = \begin{cases} -0.4x^2 + 7.5x + 1 & \text{if } x \leq 15 \quad \text{(Soccer 128)} \\ -0.6x + 32 & \text{else} \end{cases} \tag{5}$$

$$\hat{qp}(x; h, c, 1500) = \begin{cases} -0.09x^2 + 3x + 48 & \text{if } x \leq 15 \quad \text{(Soccer 1500)} \\ -0.6x + 82 & \text{else} \end{cases} \tag{6}$$

$$\hat{qp}(x; m, c, 128) = \begin{cases} -0.09x^2 + 2x + 45 & \text{if } x \leq 15 \quad \text{(Advertising CIF)} \\ -x + 70 & \text{else} \end{cases} \tag{7}$$

$$\hat{qp}(x; m, q, 48) = \begin{cases} -0.4x^2 + 7.5x + 17 & \text{if } x \leq 15 \quad \text{(Advertising QCIF)} \\ -x + 55 & \text{else} \end{cases} \tag{8}$$

Some equations show obvious similarities with respect to their polynomial coefficients. For instance, concerning $x \leq 15$, (2), (3), and (4) have offsets which are comparable in size. Also, there is a striking similarity between (5) and (8) and between (6) and (7) for $x \leq 15$. Since the respective pairs show completely different content (Soccer and Advertising), the similar models indicate a more general subjective principle behind the measurements.

On the other hand, for $x > 15$, the approximations show obvious similarities concerning the coefficients of x between (5) and (6) and between (7) and (8). Here the similarity can be explained by the same content that is shown in the respective pairs.

5 Conclusion

As already shown by the analysis of variance, variations of the parameter frame rate have a significant influence on the subjectively perceived video quality. The amount and direction of this influence depend on the parameters TA and SA of the video sequence. The performed investigation leads to the following conclusions:

- The value of the frame rate corresponding to a maximum quality level depends on the encoding bitrate. For low encoding bitrates (128 kbit/s) the maximum is reached at about 10 frames per second and for a very high bitrate (1500 kbit/s) the maximum is reached at a value around 15 frames per second.
- If the frame rate is increased, the subjectively perceived quality decreases for sequences with low temporal activity until a lower bound is reached. Thus the minimum frame rate of 5 frames per second is recommended in this case.

- The previously described decrease is intensified for sequences with high spatial activity. Therefore the optimal choice for the parameter frame rate remains unchanged.
- If the parameter frame rate is increased, sequences containing high temporal activity first react with an increase of the subjectively perceived quality, followed by a decrease.
- Piecewise polynomial models for low-TA videos on the one hand, and for medium/high-TA videos on the other hand show striking similarities, indicating a general tendency of subjective user ratings for varying frame rate.

We conclude that our findings are consistent with the findings of other authors, showing that for human observers a minimum image quality is most important and must be ensured even at the expense of a dramatic drop of frame rate. But unlike previous works, in our findings the optimum frame rate may even be as low as 10 or even 5 frames per second.

References

1. ITU-T H.261 (1993), *Video codec for audiovisual services at px 64 kbit/s*, ITU-T H.261.
2. ANSI_T1.801.03 (1996), *Digital Transport of One-Way Video Signals-Parameters for Objective Performance Assessment*, ANSI_T1.801.03-1996.
3. ITU-R BT.500-11 (2002), *Methodology for the subjective assessment of the quality of television pictures*, ITU-R BT.500-11.
4. ITU-T_P.910 (1999), *Subjective video quality assessment methods for multimedia applications*, ITU-T P.910.
5. R.T. Apteker and J.A. Fisher and V.S. Kisimov and H. Neishlos (1995), *Video Acceptability and Frame Rate*, IEEE Multimedia, Vol. 3, pp. 32-40.
6. G. Ghinea and J.P. Thomas (1998), *QoS Impact on User Perception and Understanding of Multimedia Video Clips*, ACM Press New York, pp. 49-54.
7. K. Fukuda and N. Wakamiya and M. Murata and H. Miyahara (1997), *QoS Mapping between Users Preference and Bandwidth Control for Video Transport*, Proceedings of Fifth IFIP International Workshop on Quality of Service 97, pp. 291-302.
8. M. McCarthy and A. Sasse and D. Miras (2004), *Sharp or Smooth? Comparing the Effects of Quantization vs. Frame Rate for Streamed Video*, Proceedings of ACM Conference on Human Factors in Computing Systems, ACM Press New York.
9. G. Yadavalli and M.A. Masry and S.S. Hemami (2003), *Frame rate preferences in low bit rate video*, Proceedings of the IEEE Intl. Conf. on Image Processing.
10. F. Pereira and T.F. Ebrahimi (2002), *The MPEG 4 Book*, Prentice Hall.
11. J. Hartung and B. Elpelt and K.-H. Klösener, *Lehr- und Handbuch der angewandten Statistik*, Oldenbourg München, Wien.
12. S. Winkler (1999), *Issues in Vision Modelling for Perceptual Video Quality Assessment*, Signal Processing, nr. 2, pp. 4-6.
13. A. Wörner (1999), *Digital Quality Analyzer DVQ*, News from Rohde & Schwarz, nr. 163, Vol. 78, pp. 231-252.
14. John Verzani (2004), *Using R for Introductory Statistics*, Chapman & Hall/CRC.