QUALITY ASSESSMENT OF VIDEO STREAMING IN THE BROADBAND ERA

Jan Janssen, Toon Coppens and Danny De Vleeschauwer

Alcatel Bell, Network Strategy Group, Francis Wellesplein 1, B-2018 Antwerp, Belgium
{jan.janssen, toon.coppens, danny.de_vleeschauwer}@alcatel.be

ABSTRACT

Since the transport capacity in the access networks to the Internet and in the Internet core itself has increased substantially, PC applications that make use of streaming video of good quality are nowadays emerging. The well-known MPEG-1 video compression standard is an adequate candidate to be used in such kind of applications. In this paper, we determine the bit rate needed to attain good quality for streaming video. We use the well-known Mean Square Error (MSE) and the Just Noticeable Difference (JND) implemented in the commercially available JNDmetrixIQ software, which models the characteristics of the human visual system, as objective quality measures. We correlate both these measures with the Mean Subjective Score (MSS) resulting from a small-scale subjective experiment and show that the JND correlates better than the MSE in the region of interest (i.e., where the quality is not too bad). In this process we also calibrate the JND scale, i.e., we determine the JND value that corresponds to "good" quality. Finally, we show that the bit rate needed to attain good quality depends on the video content but (for MPEG-1) a bit rate of 1.5 Mbit/s is sufficient for any kind of video.

1. INTRODUCTION

During the last years, the transport capacity of both the Internet core and the access networks to it, has increased dramatically. In particular, broadband access technologies (e.g. ADSL, cable, …) provide typical downstream link rates in the order of Mbit/s, allowing more bandwidth-intensive services than a traditional dial-up connection.

Streaming video, i.e., video that needs to be displayed almost immediately (i.e., some small amount of buffering time) after its request, is an example of such a bandwidth-demanding service. On-demand streaming is initiated by the end user who requests certain content (e.g. news, weather forecast, …) to be streamed from the Internet. Live streams, on the other hand, can be captured by any end user and are typically used for special occasions (e.g. music shows, sport events, …). Also broadcasting video over the Internet can be considered as a streaming service.

In uncompressed form, video sequences require a huge amount of bit rate. In principle, for every frame pixel, 3 bytes have to be reserved for its RGB color values. However, as the human eye is known to be more sensitive to luminance than chrominance, a subsampling technique may be used when using the YC_bC_r color space, with Y standing for luminance and C_b and C_r for chrominance difference values. In the latter color space, it suffices to store only a C_b and a C_r value for every 2 pixels, often referred to as 4:2:2 subsampling. For Standard Intermediate Format (SIF) sequences of 352 × 288 pixels and 25 frames/s, this leads to a bit rate of 352 × 288 × 25 × 16 bit/s = 40.55 Mbit/s. To reduce this bit rate, video compression techniques can be used. The most wide-spread video codecs nowadays were developed by the Moving Picture Experts Group (MPEG) and are referred to as MPEG-{1, 2} [2], [3], [5]. A segment-oriented successor of these codecs (MPEG-4 [1]) has already been introduced, but this is outside the scope of this paper. The basic idea of the MPEG-{1,2} codecs is quite similar, i.e., they both divide the successive frames in blocks and perform Discrete Cosine Transform (DCT) and motion-compensated prediction techniques on them.

Obviously, the quality of MPEG-encoded video streams is strongly related to the bit rate reserved for it. Streaming video services require the effective bit rate of the video stream (= the bit rate of the encoded video + some packetization overhead) to be such that it fits within the access capacity of the user. Otherwise, information will be lost and the expected video quality will not be reached.

In this paper, we try to determine the bit rate needed to encode SIF sequences at decent quality levels using the Constant Bit Rate (CBR) mode of MPEG-1.

In order to do so, the quality of the encoded video streams will be assessed with 2 objective measures, i.e., Sarnoff's Just Noticeable Difference (JND) [9] and the Mean Square Error (MSE). Also, a small subjective quality assessment experiment was performed. As an important side-result of this paper, we also compare and discuss the results of these 3 evaluation methods.

This paper is organized as follows. In Section 2, we describe the original, high-quality video sequences that will be under test in this paper. Section 3 starts with a description of both objective quality measures (JND and MSE). Afterwards, the outcome of both measures are reported, discussed and correlated. The results of our subjective quality experiment are described in Section 4. Also, some correlation with the objective results is given. Finally, the conclusions are reported in Section 5.

2. REFERENCE AND DEGRADED VIDEO SEQUENCES

For the high-quality reference sequences used in this paper, we started from PAL sequences publicly available at the website of the Visual Quality Experts Group (VQEG) [10]. These sequences contain 220 interlaced frames at a rate of 25 frames/s. The resolution of each frame is 720 × 576, and the used color space is YC_bC_r, with 4:2:2 subsampling.
In order to reduce the vertical resolution from 576 to 288, the even lines (i.e., the 2nd field of an interlaced frame) are removed. For the reduction in horizontal resolution from 720 to 352 pixels, we used an anti-aliasing filter for the luminance values. As the chrominance difference values C_b and C_r were subsampled in the original files, they could be copied.
Five different reference video sequences were constructed in this way, i.e., "barcelona", "fries", "F1 car", "mobile & calendar" and "moving graphic". The main characteristics of these sequences are summarized in Table 1.

Sequence name	Characteristics
Barcelona	Details, saturated colors, masking effect
Fries	Movement, fast panning, scene cut
F1 car	Movement, text, scene cut
Mobile & calendar	Details, saturated colors
Moving graphic	Synthetic, details, moving text

Table 1: Characteristics of reference video sequences

From every reference video sequence, we made some degraded versions, obtained by MPEG-1 compression using the state-of-the-art software LSX-MPG encoder from Ligos [7]. We encoded in CBR mode at 9 different bit rates (250, 500, 750, 1000, 1250, 1500, 1750, 2000 and 2500 Kbit/s) and adhered to the standard settings of the encoder (implying amongst others a Group Of Pictures (GOP) structure equal to IBBPBBPBBPBB).

3. OBJECTIVE QUALITY ASSESSMENT

3.1. Quality metrics

Below, we shortly describe the two objective quality metrics used in this paper. Observe that they are both double-ended methods, i.e., they calculate a quality score based on the reference *and* the degraded video sequences.

3.1.1. JND

The JNDmetrixIQ [8] is an objective image/video quality analysis tool, based on Sarnoff's proprietary JND model of the human visual system [9] and calibrated against psychophysical data.

The tool requires a reference and degraded video sequence as input, and performs measurements quantifying the Just Noticeable Differences (JNDs) between these input (reference and degraded) sequences. These measurements can be made at several levels of detail (i.e., per pixel, per frame or per sequence), and for luminance and chrominance separately or combined. In particular, the JNDmetrixIQ tool outputs so-called JND maps, with JND values per pixel not indicating only the magnitude of the differences but also the location. These pixel JND values form the basis for the so-called frame JND values, which are in turn averaged in some way leading to the so-called sequence JND value.
The JND measure is defined in such a way that a JND value of 1 corresponds to a 75% probability that an average user looking at two stimuli (e.g. frames) will notice a difference. JND values above 1 are defined incrementally in [4]. The translation from JND values to discrimination probabilities is reported in Table 2.

JND value	1	2	3	4	5
Discrimination probability	0.75	0.938	0.984	0.996	0.999

Table 2: Discrimination probabilities for JNDs from 1 to 5

Although the probability of an average viewer seeing a difference between two stimuli converges fast to 1 for increasing JND values, it does not imply that the corresponding quality is necessarily bad. Some quality descriptions (expressed in terms of the differences between the two stimuli) corresponding to JND values up to 5 are summarized in Table 3. These descriptions will be endorsed by our subjective experiment in Section 4.

JND value	Quality description
1	Essentially undistinguishable
3	Not obviously different = the same except under detailed examination
5	Differences are readily apparant

Table 3: JND values and quality descriptions

The results of the JNDmetrixIQ have shown to be highly correlated with subjective results for a large number of (detection, discrimination and rating) tasks. We refer to [8], [9] for more detailed information (amongst others on the latter topic).

3.1.2. Mean Square Error

A more simple, but very often used objective measure to evaluate image/video quality is the Mean Square Error (MSE). It is usually defined in terms of luminance pixel values, i.e.,

$$MSE = \frac{1}{I\,J}\sum_{i=1}^{I}\sum_{j=1}^{J}\left(Y_{i,j} - \tilde{Y}_{i,j}\right)^2 \quad , \qquad (1)$$

where $Y_{i,j}$ and $!_{i,j}$ denote the luminance value of pixel (i,j) of the reference and degraded $I \times J$ frame under consideration, respectively. Similar, but less often used MSE variants can be defined for the chrominance difference values C_b and C_r. Yet, as this metric is based on the exact differences between reference and degraded material and does not incorporate a model of the human visual system, it is well known to be often incorrect in predicting subjective quality judgments.

3.2. Results

Our aim is to determine the bit rates required to encode SIF video sequences at decent quality using the (JND and MSE) objective quality measures described above. As these measures compare frames on an individual basis, they do not take temporal artifacts into account[1]. However, the frame rate considered in this paper (of 25 frames/s) is generally considered to be high enough for smooth video display, resulting in a negligible influence of jerky effects on the overall video quality. Remark that there is a difference between frame rate and screen refresh rate. A frame rate of 25 Hz is enough to smoothly capture most kind of motions. The refresh rate of a Cathode Ray Tube (CRT) display needs to be 70 Hz to avoid flickering.

In Figure 1, the sequence JND value is plotted versus the video bit rate for the 5 fragments under consideration. As mentioned in the previous section, the sequence JND value is obtained by some averaging function applied to the total (= combined luminance and chrominance) JND values of all frames.

As expected, the JND values decrease (and, thus, the quality increases) for increasing bit rates, i.e., there is a trade-off to be made between bit rate and quality. According to Table 3, JND values of interest lie in the range of 3 to 5. The corresponding range of bit rates largely depends on the specific nature of the video sequence. In Table 4, we report the bit rates (linearly interpolated between the considered ones) required to achieve an average quality level corresponding to a JND value of 4.

Name sequence	Bit rate (Kbit/s) for JND = 4
Barcelona	909
Fries	1149
F1 car	1493
Mobile & calendar	1216
Moving graphic	751

Table 4: Bit rate (in Kbit/s) required for a sequence JND value of 4

[1] The JNDmetrixIQ has the capability of incorporating temporal artifacts in its calculations, but this option was disabled.

Seemingly, the range of bit rates to achieve this specific quality level varies between about 750 and 1500 Kbit/s. The synthetic "moving graphic" video required the smallest bit rate, followed by "barcelona" that claimed to contain (spatial) masking effects. A movie-like sequence as "fries" required slightly more than 1 Mbit/s, while high-detailed sequences (like "mobile & calendar") and especially high-motion, sport material (e.g. "F1 car") need an even larger bit rate.

The previous results are all expressed in terms of total JND values. As the higher sensitivity for luminance (compared to chrominance) of the human visual system is modeled in the JNDmetrixIQ, similar results would have been obtained if only luminance artifacts were taken into account. This is illustrated in Figure 3, where we depict the luminance, chrominance and total JND values for the "F1 car" sequence encoded at 750 Kbit/s.

Next, we consider the trade-off between bit rate and MSE, using the latter as quality measure. For the reason indicated above, it suffices to consider the MSE based solely on luminance values as it is defined in eq. (1).

Figure 2 contains the sequence MSE value for the 5 video sequences encoded at different bit rates, obtained by averaging the MSE values over all frames.

For increasing bit rates, the sequence MSE decreases, or, equivalently, the quality increases. Yet, as there is no generally agreed-upon, content-independent MSE value to aim at, it is much harder to derive any concrete results from this figure. For illustration purposes, the bit rates needed corresponding to sequence MSE values of 50 and 100, respectively, are reported in Table 5.

Name sequence	Bit rate (Kbit/s) for MSE = 50	Bit rate (Kbit/s) for MSE = 100
Barcelona	2197	1103
Fries	733	494
F1 car	2297	1094
Mobile & calendar	> 2500	1639
Moving graphic	488	387

Table 5: Bit rate (in Kbit/s) required for a sequence MSE value of 50 and 100, respectively

These MSE-based bit rates are clearly different from the ones reported in Table 4, both with respect to magnitude as relative ranking.

In the previous reasoning, we used the sequence (read: averaged) JND and average MSE as quality measures of entire video sequences. To this respect, it is worthwhile noting that MPEG encoders (e.g. Ligos) in CBR mode do not imply constant quality along the different frames. Based on the selected bit rate, the encoder assigns a specific number of bits for encoding I, P and B frames. These amounts are fixed and do not take into account the complexity of the frame. As such, every frame is encoded with a different

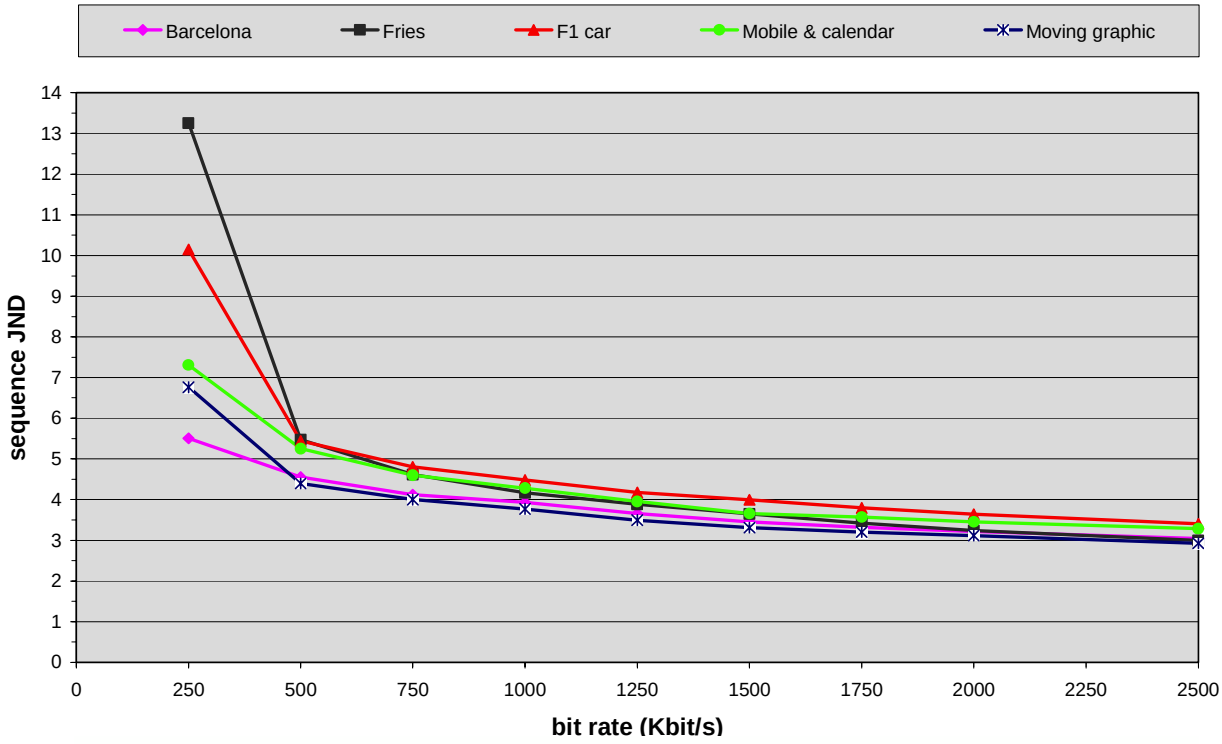

Figure 1: Sequence JND values vs bit rate

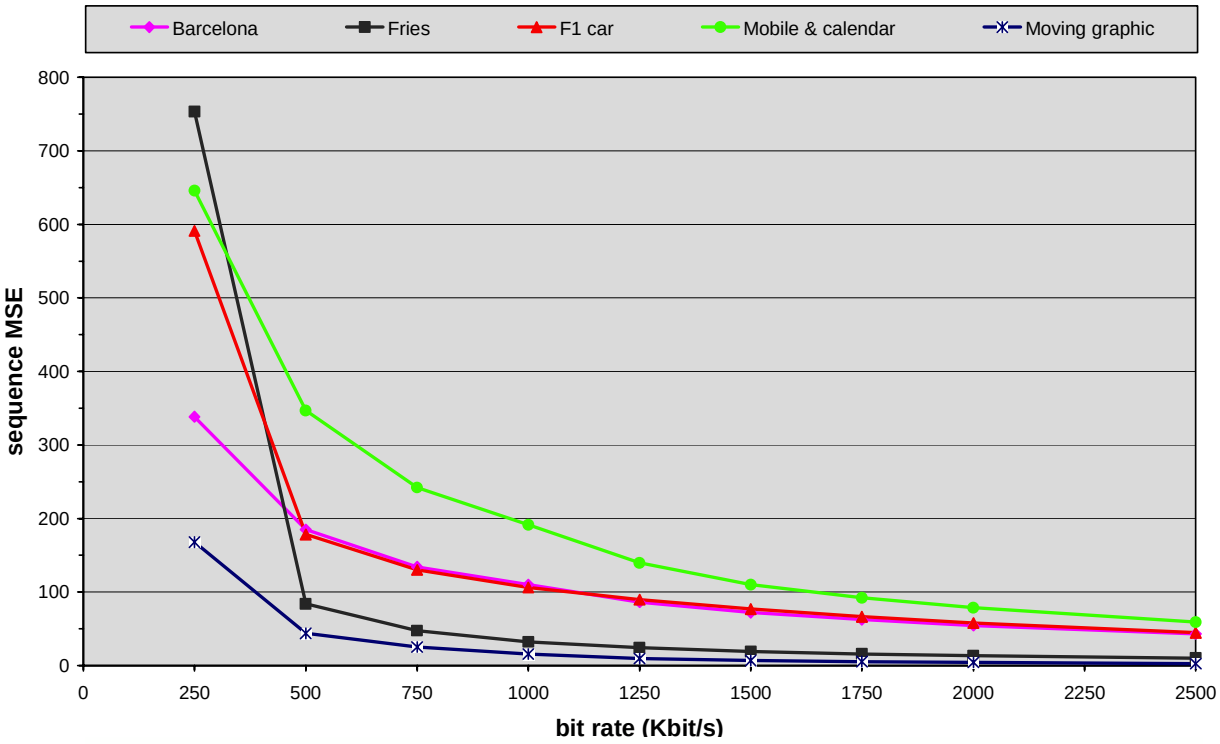

Figure 2: Sequence MSE values vs bit rate

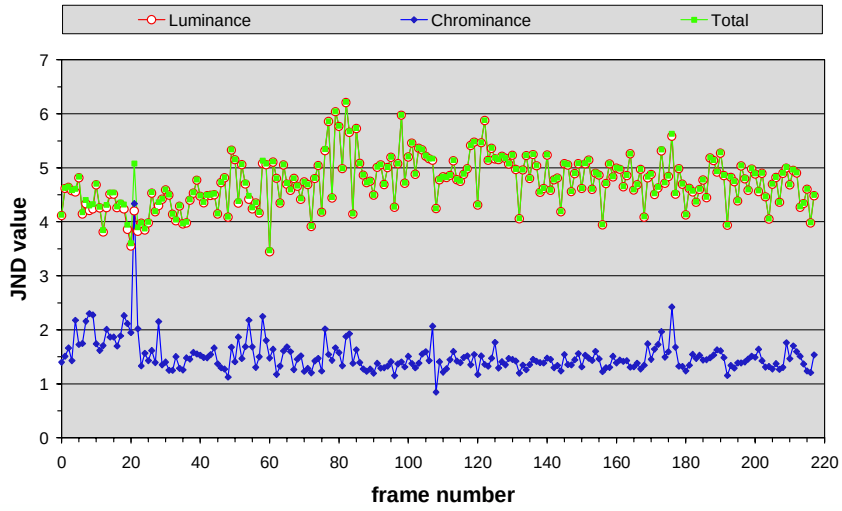

Figure 3: Luminance, chrominance and total JND values for “F1 car” encoded at 750 Kbit/s

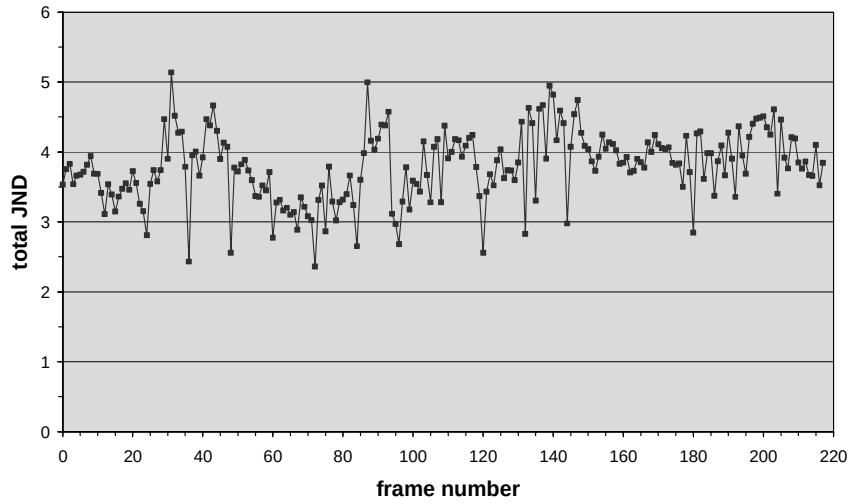

Figure 4: Frame JND values vs frame number for the “fries” sequence encoded at 1250 Kbit/s

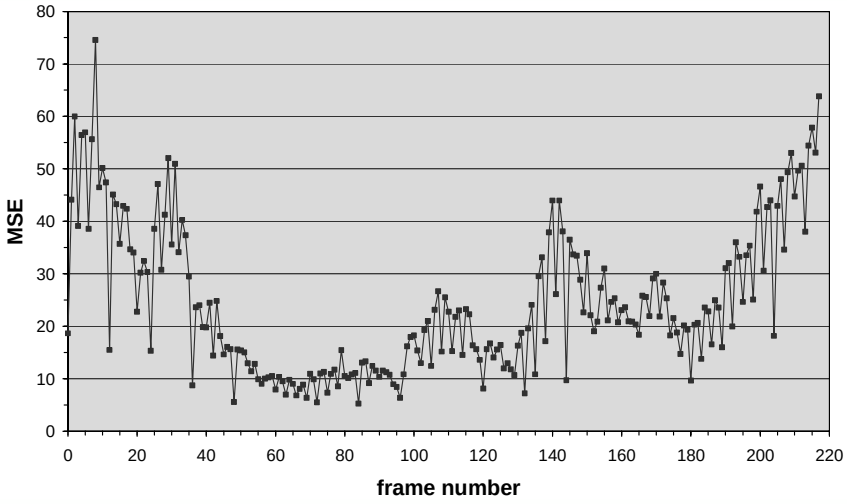

Figure 5: Frame MSE values vs frame number for the “fries” sequence encoded at 1250 Kbit/s

level of detail (i.e. using different quantization scales), and, thus, at a different quality. An illustration of the variability of the JND and MSE frame values for the "fries" video sequence encoded at 1250 Kbit/s is given in Figure 4 and Figure 5, respectively. The JND values are spread closely around their mean value, while the MSE varies within a wider range of values, depending specifically on the content of the specific frames. More precisely, detailed textures (boiling fat and a bricked wall) are gradually moving out and into the camera viewpoint in the beginning (before frame 40) and at the end (after frame 180) of the sequence, respectively, clearly leading to a decrease and increase of the MSE values. Before the scene cut (at frame 99), the amount of texture remains more or less constant, and between frame 120 and 180, portions of texture (a striped shirt and the fast-food restaurant's price list) first gets in, and then out of view. The corresponding JND values, on the other hand, suggest that the human visual system is not that sensitive to texture differences.

4. SUBJECTIVE QUALITY ASSESSMENT

In addition to the former objective quality evaluation, we also performed a small subjective experiment in which 24 colleagues (18 males and 6 females) with an average age of 31 years were asked to subjectively evaluate some of the degraded video sequences.

We tried to adhere to the relevant ITU-T Recommendation P.910 (on subjective video quality assessment for multimedia applications) [6] where possible. That is, we opted for the Absolute Category Rating (ACR) or single-stimulus method, in which the degraded test sequences are presented one at a time. The 11-point category scale, shown in Table 6, was used for rating purposes. Note that the endpoints on this scale are anchoring points that are not assignable. After a training phase, all sequences were shown twice in a random order on a 19" high-quality computer monitor with a resolution of 1024 × 768 pixels and a refresh rate of 70 Hz. The background color was set to 50% grey (R = G = B = 128).

10	NO FURTHER IMPROVEMENT POSSIBLE
9	Excellent
8	
7	Good
6	
5	Fair
4	
3	Poor
2	
1	Bad
0	A WORSE QUALITY CANNOT BE IMAGINED

Table 6: 11-grade numerical quality scale

The Mean Subjective Scores (MSSs), averaged over all participants, are plotted versus the bit rate in Figure 6. The subjective quality is (as expected) proportional to the bit rate, except for the "barcelona" sequence, for which the 1250 and 1750 Kbit/s sequences were rated nearly the same quality. Two explanations can be brought up against this discrepancy. First, statistical fluctuations may occur in our subjective test with only a limited number of participants of a very specific type (nearly all young telecom engineers). In particular, the standard deviation on the subjective scores of the 1250 and 1750 Kbit/s "barcelona" sequences equal 1.11 and 1.32, respectively. Second, several participants expressed their concerns about their ratings for "barcelona". Even after the complete test (and, thus, after several repetitions of this video fragment), several participants could not describe its content.

In Table 7, the bit rates corresponding to "good" quality (a mean subjective score of 7) are reported for the different sequences. When we compare these bit rate numbers with the ones of Table 4 and Table 5, one immediately sees that the trend is more similar to the one of the former table. Furthermore, as mentioned earlier, subjective results are always subject to some statistical fluctuations, and are expensive and time-consuming to organize. Therefore, we will use the JND model (and the results derived from these numbers, e.g., the bit rate values of Table 4) in the context of (objective) video quality measurements.

Name sequence	Bit rate (Kbit/s) for "good quality"
Barcelona	1225
Fries	1076
F1 car	1185
Mobile & calendar	1610
Moving graphic	1197

Table 7: Bit rate (in Kbit/s) required for good subjective quality

To illustrate the latter observation in more detail, the couples of associated sequence JND and MSS values for all subjectively tested bit rates and sequences are depicted in Figure 7. Figure 8 is a similar figure in terms of sequence MSE values.

We observe that for the lowest-quality (250 Kbit/s) sequences, the sequence JND values are widely spread (between 5.5 and 13.25), although the mean subjective scores are nearly similar. This implies that the JND measure is not that accurate for such low quality levels. Yet, for the quality levels of interest (i.e., the dashed rectangle), the correlation between the sequence JND values and mean subjective scores (over all sequences) is quite high, i.e., -0.86. For the same sequences, the correlation between the sequence MSE values and mean subjective scores is substantially lower, i.e., -0.66. Interpreting these results on the figures can be

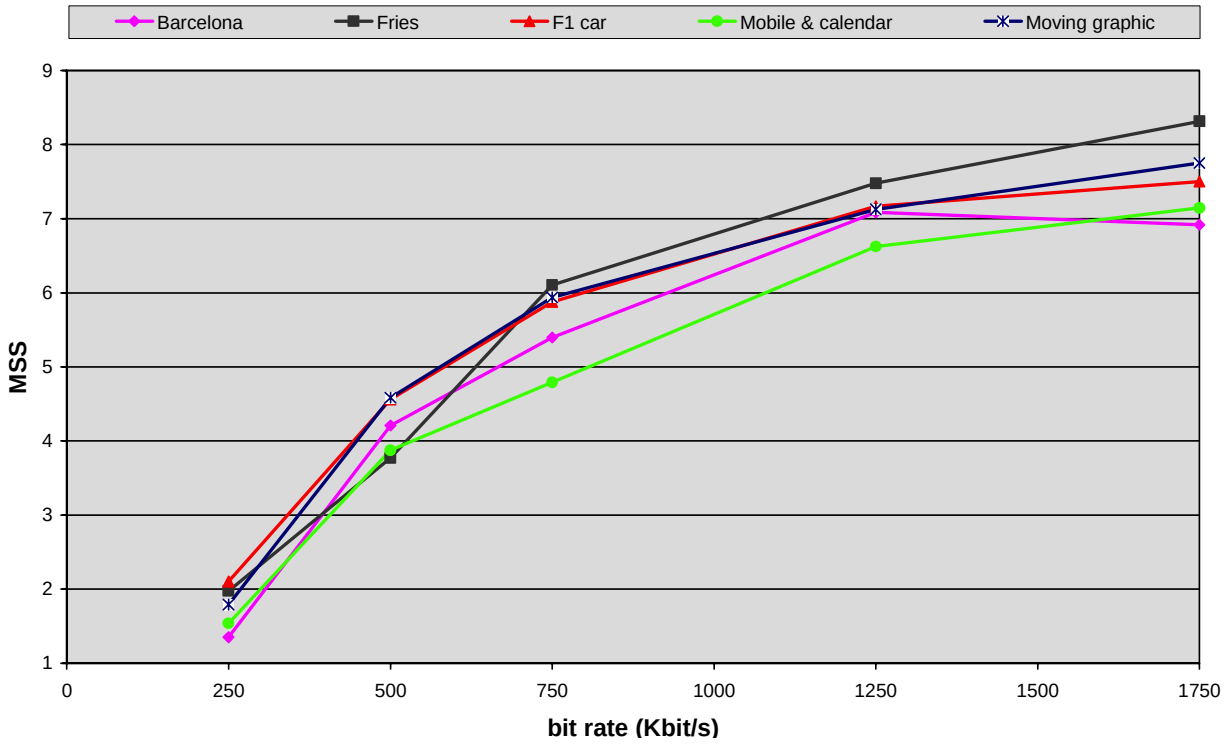

Figure 6: Mean subjective scores vs bit rate

done by drawing a trendline on the couples of interest. The average distance of these couples to this trendline will be largest in the MSE case of Figure 8.

As an immediate consequence of this high correlation between the objectively obtained JND values and mean subjective scores, it is also relatively easy to relate both quality measures to another. "Good" subjective quality indeed corresponds to a JND value around of 4, i.e., the 7 on the subjective scale of Figure 7 translates to about 4 on the JND scale.

5. CONCLUSIONS

In this paper we have investigated which bit rate a video sequence (typically incorporated in a PC application) requires in order to attain a good enough quality. We considered 5 video sequences each containing a different type of content (sports, movie, computer graphics, …). To determine the quality we considered two objective measures: the well-known Mean Square Error (MSE) and the Just Noticeable Difference (JND) calculated in the commercially available JNDMetrixIQ, which mimics the human visual system. We correlate the results obtained with both methods with the results of a small subjective experiment in which an audience of 24 people were asked to rate the sequences encoded at different rates giving a Mean Subjective Score (MSS).

It turned out that in the region of interest (i.e. not too low a quality or, equivalently, not too low a bit rate) the JND correlates far better with the MSS than the MSE. In that process we also calibrate the JND scale, i.e., we determine that JND value that corresponds to what the audience found to be "good" quality. Since large-scale subjective experiments are quite cumbersome and time consuming, the JND is a useful practical quality measure for the purpose of testing PC-based video streaming applications described in this paper.

Finally, using this JND measure, we showed that the bit rate required to obtained good quality depends on the type of content and ranges from as low as 750 Kbit/s for computer graphics, over 1 Mbit/s for movie-like sequences up to 1.5 Mbit/s for fast moving scenes with lots of detail (e.g. sports).

6. REFERENCES

[1] R. Koenen, "MPEG-4 overview (V.18 - Singapore Version)", ISO/IECJTC1/SC29/WG11 N4030, March 2001, http://mpeg.telecomitalialab.com/standards/mpeg-4/

[2] L. Chiariglione, "Short MPEG-1 description", ISO/IECJTC1/SC29/WG11 MPEG96, June 1996, http://mpeg.telecomitalialab.com/standards/mpeg-1/

[3] L. Chiariglione, "Short MPEG-2 description", ISO/IECJTC1/SC29/WG11 MPEG00, October 2000, http://mpeg.telecomitalialab.com/standards/mpeg-2/
[4] IEEE Draft Standard P1486, "Draft standard for the subjective measurement of viusal impairments in digital video using a Just Noticeable Difference scale", January 2002, http://grouper.ieee.org/groups/videocomp/private/P1486-D06.pdf
[5] ITU-T Recommendation H.262, "Generic coding of moving pictures ans associated audo information: Video", February 2000.
[6] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications", September 1999.
[7] Ligos Corporation, http://www.ligos.com/
[8] Sarnoff's JNDmetrix, http://www.jndmetrix.com/
[9] Sarnoff Corporation white paper, "JND: a human vision system model for objective picture quality measurements", June 2001, http://www.sarnoff.com/common/pdf/int/JND_whitepaper_0106.pdf
[10] Visual Quality Experts Group, http://www.vqeg.org/

ACKNOWLEDGEMENTS

This work was carried out within the framework of the project CoDiNet sponsored by the Flemish Institute for the promotion of Scientific and Technological Research in the Industry (IWT).

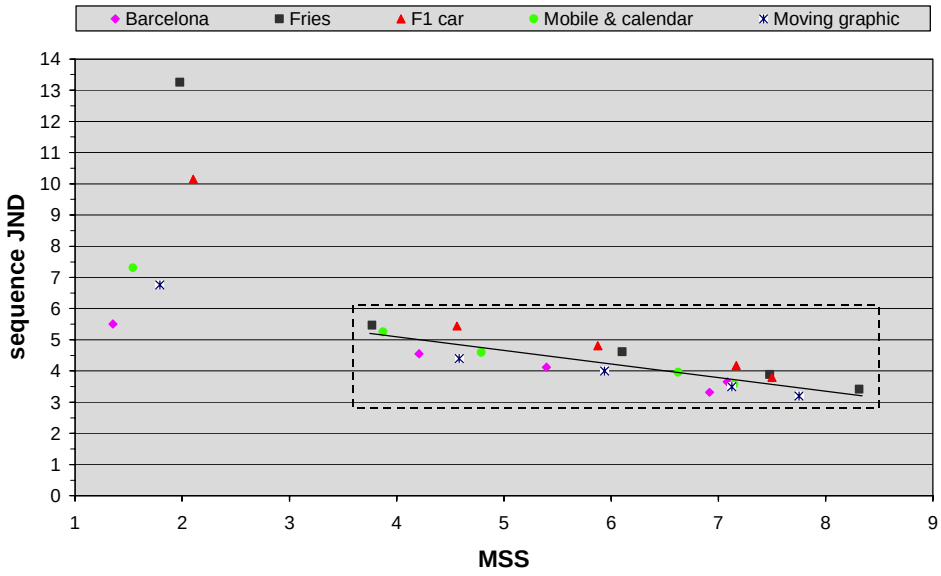

Figure 7: Sequence JND values vs mean subjective scores

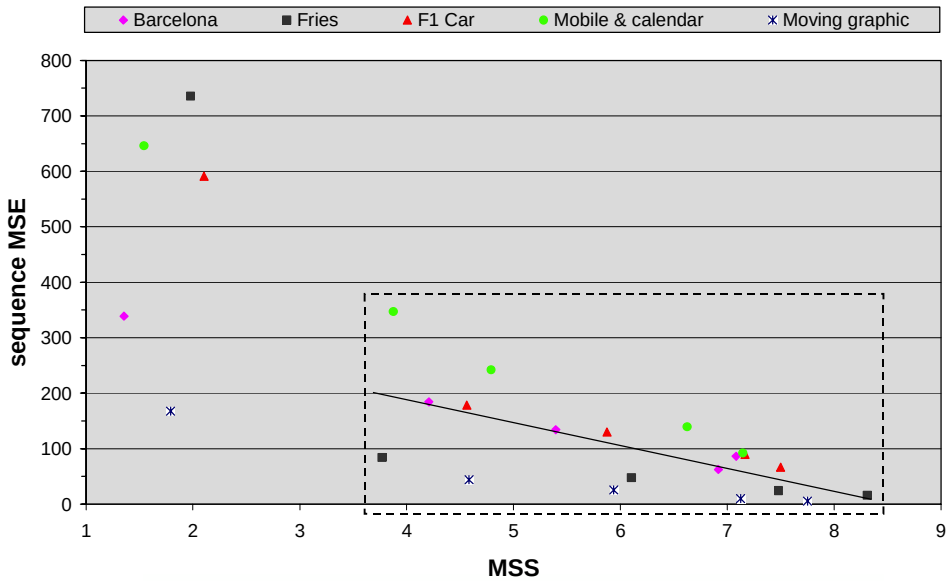

Figure 8: Sequence MSE values vs mean subjective scores