Perceived Quality for Transported Video

Nele Van den Ende
Reinder Haakma
Maddy Janse
Peter van der Stok
Philips Research, Eindhoven, the Netherlands
High Tech Campus 31, NL-5656AA Eindhoven
nele.van.den.ende@philips.com

Abstract

Wireless networks typically suffer from packet loss and variable bit-rates due to bandwidth variations when, for example, video is transported between different devices in the home. Users who are watching video sent over wireless networks will therefore often experience jerkiness and blocking in the image. To intercept the effects caused by those problems, adaptation methods such as I-Frame Delay and Signal-to-Noise Ratio scalability were devised. To assess the perceived quality of MPEG-2 video optimized by these adaptation methods, subjective tests were carried out. These tests demonstrate, amongst others, that scene content and duration of the quality loss affects the perceived quality. Furthermore, we observed a saturation effect.

Key words: perceived quality, streaming video, adaptation methods, wireless networks

1. Introduction

Video traffic is often sent between devices such as DVD players and televisions over networks. The current trend here is that the networks go from being only wired to both wired and wireless. Obviously, the positive feature of wireless networks is that users no longer have to deal with wires in their surroundings. At present, wireless networks often suffer from problems caused by interference from other networks or consumer electronic devices such as microwaves. Those network problems include bandwidth variations, which can cause packet loss, which in turn can lead to jitter. Packet loss can lead to visual artifacts such as blocking, whereas high jitter creates jerkiness. Bandwidth variations can arise from interference through consumer electronic devices (e.g. microwaves), and usually results in lowering the available bandwidth. When the available bandwidth is lower than the bit-rate from the video stream, video quality is impaired. Depending on the size of the network fluctuations and decoder implementations, observers can see a black screen, blocking artifacts, or a frozen image which seems to flicker.

1.1 Adaptation Methods for Streaming Video

Different adaptation methods can be applied to remedy problems that arise from using wireless networks. These adaptation methods, however, lead to very different effects and artifacts for observers. The adaptation methods under consideration were devised for MPEG (which stands for Moving Picture Experts Group) encoding and especially address the effects of wireless network properties in relation to the human visual system. MPEG encoding is designed to compress a video, and to allow playback of the compressed video on different types of platforms. It works by introducing three different kinds of frames in

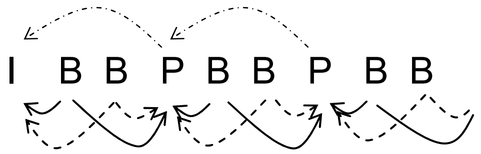

Figure 1: Example of a Group of Pictures (GOP). Arrows show how frames depend on each other. Note that frames are not always transmitted in this order, and have to be reordered by the MPEG decoder.

the video stream, as can be seen in Figure 1. First, the intra-coded or I-frames, which are independent and do not contain references to other frames. I-frames should occur about twice every second in the video stream. Predictive-coded or P-frames are dependent on the previous I or P-frame in the stream, while bi-directionally predictive coded or B-frames depend on both the previous I or P-frame and the subsequent I or P-frame. In MPEG encoding, frames are grouped together in so-called ’Groups of Pictures’ (GOP), with the usual order being something along the lines of IBBPBBPBBPBB. Consequently, GOPs are encoded and sent in a different order than they are played. The two adaptation methods we considered are the I-Frame Delay approach and the Signal-to-Noise Ratio scalability approach.

The principle of the I-Frame Delay (IFD) approach is based on dropping the least important frames in a video stream. That is, when network congestion occurs, and the buffer containing the remaining frames flows over, the IFD method drops the least important frames (B-frames), to assure that the frames which contain the most important information are received. The decoder then repeats the last frame, which usually results in users perceiving ‘jerky movement’ in the movie. Blocking artefacts due to partial frame loss do not occur for the user to perceive. Furthermore, two parts should be added to the infrastructure: a tagger to identify packets as they come along and a dropper to decide which packets can be dropped and which are allowed to move on. The buffer, in this case, should be large enough to contain two I-frames. Packets currently sent are the most important ones, whereas waiting packets are the ones that could be dropped if necessary. Another category is other, which could contain audio. Audio, though, is only partially incorporated in the model and will not be considered here. The technical advantage of the IFD approach is that only the sender of the video has to be modified for the bit-rate adaptation. The introduction of IFD adaptation for a video stream reduces the available bandwidth for video somewhat because it introduces processor overhead costs. However, without IFD a wireless transported video stream suffers from artefacts, loss of audio and ‘hiccoughs’ (i.e. the movie stops and starts randomly) [Kozlov, van der Stok & Lukkien, 2005].

The Signal-to-Noise Ratio scalability (SNR scalability) method [Jarnikov, van der Stok & Lukkien, 2005] deals with variable throughput problems by dividing the original video in several layers: one base layer and (if possible) several enhancement layers. When there is interference in the network and the available bit-rate drops abruptly, one or more enhancement layers can be dropped. This assures that, though the image quality of the video is not always perfect, users always see a moving image on their screen. The advantage of this method is that it each layer can be decoded by standard non-scalable MPEG-2 decoders. The base layer provides the basic video quality and the enhancement layers increase the video quality. In addition, the base layer is independent of the enhancement layer, and the frames within the enhancement layers are independent of each other. A base layer and its enhancement layers can be stored separately and transmitted in a single combined stream or separate streams. When there are bandwidth variations over a short term, they can be dealt with on a frame-by-frame basis because frames within enhancement layers are independent of each other. The independence of the frames is ensured when the layer consists of B-frames with an I-frame at the beginning of the sequence and a P-frame at the end. Overhead costs are also present, and depend on the size of the base-layer, the overall bit-rate of the video and the number of the enhancement layers introduced. To conclude, the SNR scalability approach uses more bit-rate than video that has not been scaled to reach the same level of video quality, but it can overcome network problems that non-scalable video cannot.

Hence, both the IFD and the SNR scalability method have advantages and disadvantages. Not applying any adaptation method is not an option though, because end-users will be subjected to unacceptable video quality. To understand how both adaptation methods reflect on the user experience, we set up two experiments. In the first experiment, the IFD method is compared to normal DVD video quality. The second experiment compares the SNR scalability approach to normal DVD video quality. Both are designed to see how each method compares to the perceived quality of video sent through a wired network.

1.2 Previous research

Earlier research for the case of low network bandwidth (<1 Mbps) has shown that sustained blockiness or bluriness in a video has a negative impact on the perceived video quality [Masry, Hemami, Osberger & Rohaly, 2001]. Masry et al. also discovered that the optimal frame rate, according to observers depends on the type of motion displayed in the video. Video sequences that already show jerky motion are perceived as better quality when the spatial quality is enhanced for lower frame rates. The perceived video quality of video sequences portraying smoother motion seems rather unaffected when the frame rate is changed. Results from experiments reported by Wang, Speranza, Vincent, Martin & Blanchfield [2003] show that artifacts from quantization errors (e.g. blocking, false countouring) are rated lower on perceived video quality than either jerkiness or image blurring. McCarthy, Sasse & Miras [2004] showed that their observers appeared more sensible to reductions in frame quality than in frame rate for videos shown on small screens (325x288 and 176x144 pixels). Note that these experiments were done with small displays and a CIF format (325x288), so it is difficult to say how easy they are transferred to larger video formats and screens.

Further research done by Zink, Kunzel, Schmitt & Steinmetz [2004] shows that less and longer quality drops are better received by observers than frequent but short-lived and more severe quality drops. Of course, this might only apply for larger screens, since earlier research also shows that observers prefer a lower frame rate for smaller screens [McCarthy, Sasse & Miras, 2004]. Hands and Avonds [2001] found that the lowest perceived video quality (maximum distortion) was one of the most determining effects for the viewers' perceived video quality rating. Looking at the presented results, it is difficult to generalize towards one clear conclusion on which one can base additional experiments. However, keeping all the possibilities in mind, and in conjunction with research from Haakma, Jarnikov & van der Stok [2005], our two experiments were set up.

2. Perception Experiments

2.1 Method

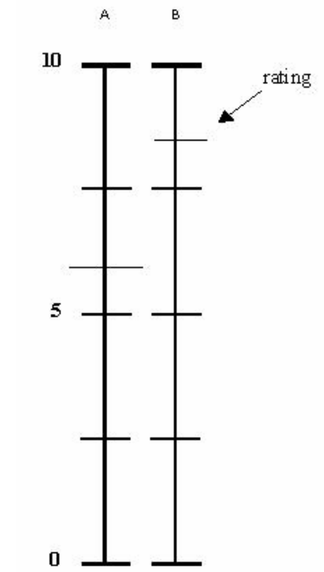

Figure 2: Example of the rating scale for the DSCQS used in both experiments.

When it is not possible to provide the full range of the video quality in the test conditions, it is generally better to use the double-stimulus continuous quality-scale (DSCQS). Observers are shown a pair of videos, but one is an unimpaired reference and the other one is shown through the process or system under evaluation (here either IFD or SNR scalability). Note that assessors are not told which video is the reference, and they have to indicate their opinion of both videos on continuous quality scales. The scales are printed in pairs (as can be seen in Figure 2), because of the double video presentation. Keep in mind that results from this test have to be interpreted as difference scores, and not as absolute scores. Accordingly, it is invalid to associate the collected scores with a single quality description.
In addition, it is claimed that the DSCQS is less sensitive for context, i.e. that participants' ratings are less influenced by the severity and ordering of the modified video sequences throughout the test session. Chosen video sequences are 10 seconds long, in accordance with the ITU-R BT.500-11 DSCQS's recommendations. Variant II was used, thus observers were shown the sequences consecutively per pair. Double presentation of all pairs took a total of 48 minutes. Each pair consisted of a reference and adapted video material.

Both experiments took approximately 1 hour per observer, since a break every 15 minutes is recommended to keep them from tiring. The position of the reference was changed in a pseudo-random fashion throughout the whole test. The observers were asked to assess the overall quality by inserting a mark on a graphical continuous quality scale, which are printed in pairs for the double presentation of the material. In total there were 24 conditions, but the stimulus set was shown twice in random order. Hence, observers had to compare a minimum 48 video pairs.

2.2 Experimental Design

Both experimental designs elaborate on previous work of Haakma et al (2005). To keep consistent with guidelines to test network traffic we used a high (6 Mb/s) and a low (3 Mb/s) bit-rate. The video material consisted of two action scenes: a fight scene from *Matrix: Reloaded* and a dance scene from *Feet of Flames*. They were shown on a 42" Philips plasma screen.

To cover IFD video quality evaluation, 16 observers participated in the first experiment. We look at what happens when every seventh and every fourteenth B- frame is left out for a certain period of time: 2, 4 and 8 seconds in the middle of the video material. In total, 24 different sequences were included in the stimulus set and each condition was judged twice. Bit-rate loss ranged from 7.45 Kb/s to 33 Kb/s, which is a rather small loss given the 3 and

6 Mb/s[1]. All shown video sequences are coded in one layer. This was a within subject design.

For the second experiment, 14 observers were asked to assess the SNR scalability adaptation method. Here, the shown video sequences are encoded in two layers: a base layer and an enhancement layer, whose size (in Mb/s) are both manipulated. The base-layer is either 1/3 or 2/3 of the available bit-rate (1, 2 or 4 Mb/s, also called base-layer ratio), and the EL is dropped during 2, 4 or 8 seconds[2]. Modification takes place in the middle of the video sequence, which is dropping the EL for this set-up. Again a within subject design is used.

2.2 I-Frame Delay

Data from this experiment were analysed with ANOVA (repeated measures). The variables used were repetition, duration of the quality loss, scene and B-frame loss[3]. Our main questions we wanted to answer are:

- Is leaving out more B-frames (compared to leaving out less B-frames) perceived as worse quality?
- Does the duration of a quality drop influence perceived video quality?
- Do observers notice the difference between the shown bit-rates?

From the results of the statistical analysis, it is not possible to conclude that higher B-frame loss is always rated lower. When the bit-rate is of lower quality (3 Mb/s), it depends on the scene content whether or not the modified video sequence really scored worse than the reference video sequence. Figure 3 illustrates this. When a bit-rate of 6 Mb/s is employed, duration plays a role. When B-frame loss lasts longer, it becomes easier to see. Especially 1/7 B-frame loss over 8 seconds was rated low by observers. Note: bit-rate changes were very small, and at most 7.45 kb/s was lost. This is reflected in the observers' scale usage and in the relatively small effect sizes.

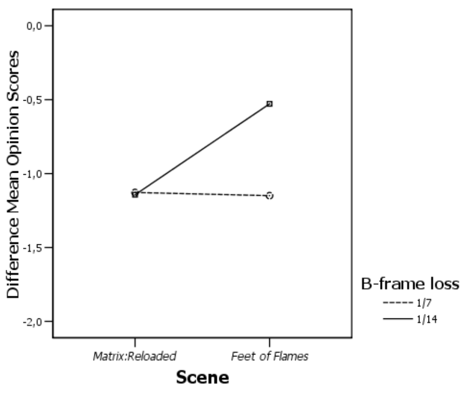

Figure 3: Interaction between the variables B-frame loss and scene for IFD (ANOVA, 3 Mb/s).

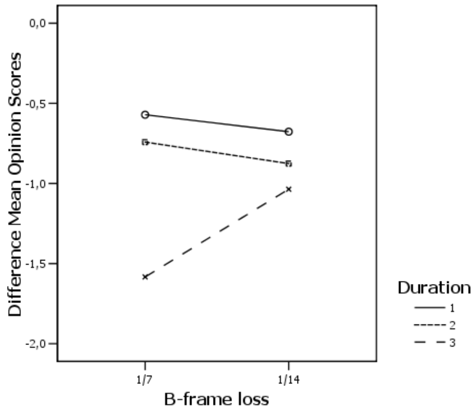

Figure 4: Interaction between B-frame loss and duration for IFD (ANOVA, 6 Mb/s).

---

[1] There were two references: when a modified video sequence of 3 Mb/s was shown, it was accompanied by a reference of 3 Mb/s, with a similar course of action pursued for the 6 Mb/s conditions.

[2] Only one reference of 8 Mb/s was used.

[3] Because the references differed per bit-rate for the IFD adaptation method evaluation, we analysed the conditions with 3 Mb/s separately from those with 6 Mb/s.

Compared to Haakma et al (2005) however, observers noted a small difference between the amounts of frames thrown away. Although the perceived video quality still decreases, the decline is smaller and more varied than when 25%, 33% and 40% of all available frames are skipped. Masking might play a role here: if B-frame loss happens right after a scene change, it is less likely that observers will pick up on it. Duration only played a role in the 6 Mb/s conditions and then only when B-frames were dropped with a 1/7 ratio (see Figure 4). No differences between shown bit-rates were found, but a repetition effect for 3 Mb/s illustrates the presence of a possible carry-over (or learning) effect. When participants see the same image sequences more often they begin to learn where to look for the effects. It seemed that in the second presentation they rate modified sequences in the same direction, but harsher.

2.3 Signal-to-Noise Scalability

Again, data were analysed with ANOVA (repeated measures). The variables this time were repetition, duration of the quality loss, scene, bit-rate and base-layer ratio. For this experiment our questions were:

- Do observers notice the difference between the shown bit-rates?
- Does the duration of a quality drop influence perceived video quality?
- Does a quality saturation effect really show? Do observers really not notice when video quality is enhanced once a certain level is reached?

Results from the analysis confirm that observers notice the difference of the shown BLs, although it depends on scene content, and sometimes on the duration of the EL drop as well. Figure 5 illustrates this in 4 different graphs. Hands & Avonds [2001] concluded that duration did not influence perceived video quality as much as the lowest quality drop. However, results show that a drop of 8 seconds under some circumstances is perceived as lower quality than one of 2 or 4 seconds (figure 5A)[4]. Also note that Hands & Avonds [2001] did not repeat their tests as often as was the case here. It is possible that the carry-over effect is mostly responsible for this slight duration effect, and that observers, after a while, incorporate the length in their quality judgement. Further research into this area is necessary to either confirm or deny these options.

Lastly, there was the quality saturation effect to consider. It seems likely that such an effect really exists. It is possible that when image quality reaches a point where it is too impaired or too blurry, observers do not notice whether it becomes worse or not. Likewise, if the image quality becomes very sharp and very clear, a quality increase will go unnoticed.

---

[4] These results only cover the effect of the duration over all quality variations, not the interaction between the amount of quality loss (EL lowering) and the duration.

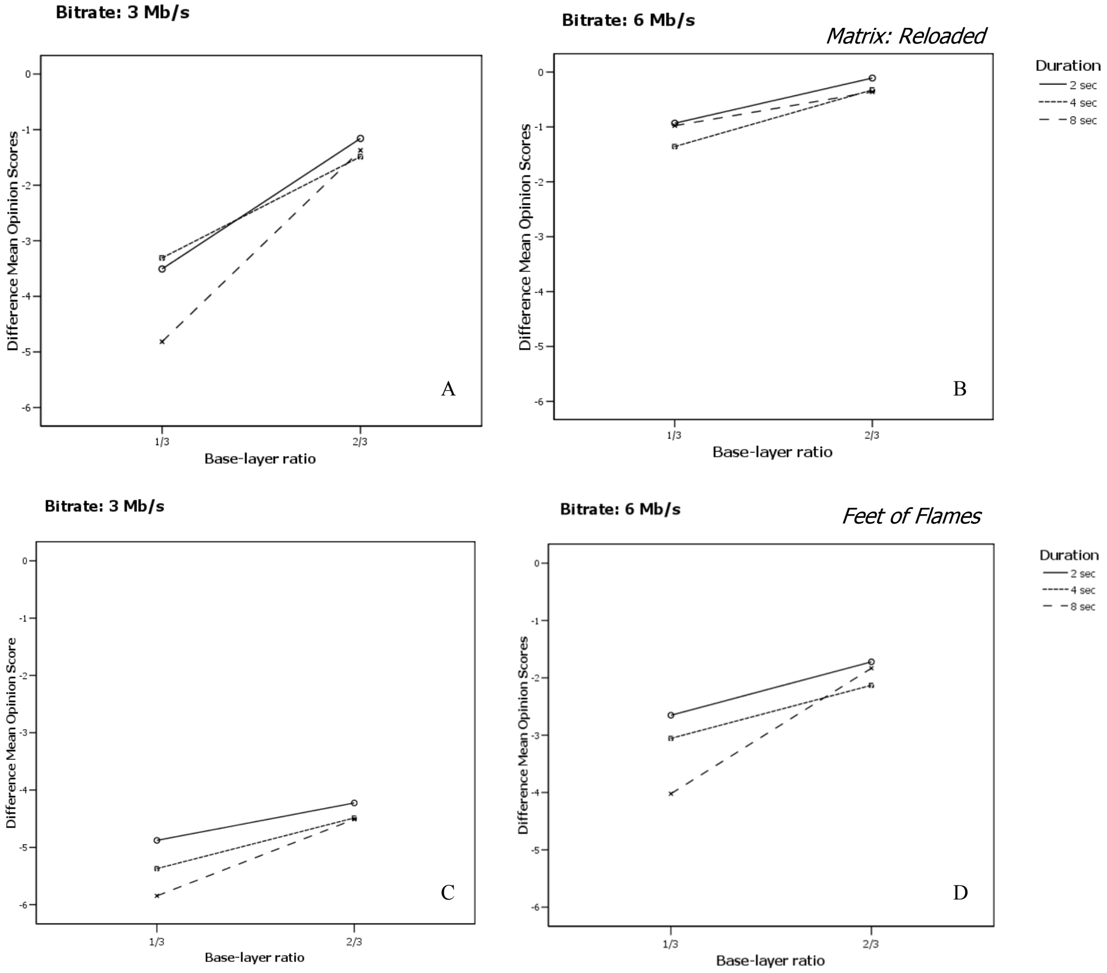

Figure 5: Graphs of the interaction between bit-rate, base-layer ratio and duration (SNR scalability ANOVA).

3. Discussion and Conclusion

We found that the influence of IFD on the perceived video quality was rather small. These results confirmed the results from Haakma et al. [2005], in which the B-frame loss was manipulated from 1/2 to 1/4. In our experiment we focused on a B-frame loss of 1/7 and 1/14 to extract a clearer view at the development of the effect of B-frame loss on perceived video quality. These effects were small, because the B-frame loss was small. The influence of duration and amount of quality loss on perceived video quality depended on both scene content and bit-rate. In general, when B-frame loss is more pronounced and lasts longer, differences between the reference and the modified video sequences become easier to detect.

The results from studying the influence of SNR scalability on perceived video quality show that observers perceived the different bit-rates of the base-layers. However, the extent to which this happens again depends on scene content and on the duration of the quality drop. Additionally, a saturation effect could be observed: when perceived quality is already low, observers do not notice further lowering of the bit-rate. Similarly for high-perceived quality: when a bit-rate is sufficiently high, observers do not notice further increases. These upper and lower limits depend on the scene content, though. We also compared these data on the perceived video quality with the commonly used objective measure peak signal-to-noise ratio (PSNR). This comparison shows that PSNR can function as predictor for the same video content, but not for comparison between video content [Van den Ende, Meesters & Haakma, accepted].

In conclusion, we explored the physical effects of the adaptation methods for streaming video on the perceived video quality. Physical here means that the chosen video clips were not analysed for content, only for the type of motion and camera-movement. Content is a very important variable here and it is possible that the type of content, i.e. the semantics of the content, influence the perceived video quality. For example, when the content is predictable (e.g. a familiar movie, advertising), observers might be less disturbed by a loss of video quality. For live sports or news items, people might be less forgiving because they cannot predict what will happen. Additionally, control over the disturbances or the type of disturbances can also be an issue, not to mention the trade-off between audio and video. Currently, audio is not addressed in either the IFD or SNR scalability method but it will have to be considered because it affects the overall user experience. Future research will have to address the aforementioned issues.

References

Haakma, R., Jarnikov, D., & van der Stok, P. (2005). Perceived quality of wirelessly transported videos. In: *Dynamic and robust streaming in and between connected consumer-electronic devices.* Ed: Peter van der Stok. Springer, pp 213-239.

Hands, D.S. & Avonds, S.E. (2001). Recency and duration neglect in subjective assessment of television picture quality. *Applied Cognitive Psychology, 15*, 639-657.

Jarnikov, D., van der Stok, P., & Lukkien, J. (2005). Wireless streaming based on a scalability scheme using legacy MPEG2 decoders. *Ninth IASTED, International Conference on Internet & Multimedia Systems & Applications.*

Kozlov, S., van der Stok, P., & Lukkien, J. (2005). Adaptive scheduling of MPEG video frames during real-time wireless video streaming. *IEEE, International Symposium on a World of Wireless, Mobile and Multimedia Networks*.

Masry, M., Hemami, S. S., Osberger, W., & Rohaly, A.-M. (2001) Subjective quality evaluation of low bit rate video. *SPIE Conf. on Human Vision and Electronic Imaging, vol. 4299*, San Jose, CA.

McCarthy, J.D., Sasse, M.A., & Miras, D. (2004). Sharp or smooth? Comparing the effects of quantization vs. frame rate for streamed video. *CHI 2004*, Vienna, Austria.

Van den Ende, N., Meesters, L. & Haakma, R. (2006). Effect of MPEG-2 Compression Parameters on Perceived Video Image Quality. Accepted for the Society for Information Display 2006 International Symposium, Seminar and Exhibition.

Wang, D.; Speranza, F.; Vincent, A.; Martin, T;, & Blanchfield, P. (2003). Towards optimal rate control: a study of the impact of spatial resolution, frame rate, and quantization on subjectuve video quality and bit rate. *SPIE, Visual Communications and Image Processing, vol 5150*, 198-209.

Zink, M.; Kunzel, O.; Schmitt, J.; & Steinmetz, R. (2003). Subjective Impression of Variations in Layer Encoded Videos. *Eleventh Workshop on Quality of Service*.