Evaluation of Subjective Video Quality of Mobile Devices

Satu Jumisko-Pyykkö
Institute of Software Systems
Tampere University of Technology
Korkeakoulunkatu 10, P.O. Box 553,
33101 Tampere, Finland
+358503610038
satu.jumisko-pyykko@tut.fi

Jukka Häkkinen
Department of Psychology, University of Helsinki,
P.O.Box 9, 00014 University of Helsinki, Finland
Multimedia Technologies Laboratory, Nokia Research
Center, P.O.Box 407, 00045 Nokia Group, Finland
+358504839483
jukka.hakkinen@helsinki.fi

ABSTRACT
Subjectively perceived video quality is a critical factor when adopting new mobile video applications. When video is used in mobile networks the most important requirements are related to low bitrates, framerates and the screen size of mobile device. In two tests we investigated the effects of codecs and combinations of audio and video streams with low bitrates and different contents on the perceived video quality of mobile devices. The first test showed that the codec H.264 produced the most satisfying video quality, but the quality was not high enough for the presentation of textual information. In the second test, the audio-video ratio 32/128kbps was found to be the most pleasant, but there were content dependent variations.

Categories and Subject Descriptors
H.5.1 Multimedia Information Systems: Evaluation/ methodology.

General Terms: Experimentation, Human Factors

Keywords: Subjective evaluation, audiovisual quality, framerate, bitrate, picture ratio, video, quality, mobile device

1. INTRODUCTION
Mobile video and television are emerging services. Mobile television has already been launched in some countries, but the service but only the first steps in its adoption are been taken. Multimedia messaging is also expected to increase in popularity. Although these services look appealing, the end-users' subjectively perceived quality of them is a critical factor for their success. Consequently, subjective quality evaluation tests during product development are necessary if the service is to reach an acceptable quality level [10].

Subjective audiovisual parameters related to mobile television have not been widely studied. There are only a few published studied about subjective evaluation of low framerates, bitrates and modern codecs used with devices with small displays [17][12]. The relation between the audiovisual contents has, until recently, also been a relatively unexplored area [18][8].

MM'05, November 6–11, 2005, Singapore.

This study focuses on two main themes: Firstly, we investigated the role that different codecs play in video quality. Secondly, we studied the relation of audio and video streams on audiovisual quality.

2. PERCEIVING AND PRODUCING QUALITY
Video quality is a combination of perceived quality and produced quality. It is an integrated set of perceptions of overall excellence of audiovisual material that vary from noticeable unacceptable and acceptable to pleasing levels and maximum perceived quality. In mobile video, the produced quality is usually between an acceptable and a low level of quality.

2.1 Quality perception
Quality perception varies according to the sensory channel. If the quality experience is considered at the sensorial level, it can be reduced to sensory thresholds and modulation transfer functions (MTF). However, human quality interpretation depends on the relationship between sensory processing and higher level processing that includes emotions, knowledge, expectations and situational schemas.

In audiovisual perception, visual and auditory information are integrated into a unified perceptual experience in a more complicated way than just adding two perceptual channels. For example, in the McGurk effect mismatched visual and acoustics material are integrated into a single audiovisual experience [13]. Also correct synchronization is necessary for unified perceptual experience to bind the image and sound together [15].

Recent studies show that good audio quality can enhance visual quality [14] and visa versa [3]. The importance of auditory and visual information depends on content [8][18]. For example, Winkler & Faller found that when the complexity of an audiovisual scene increases at very low bitrates, the importance of the auditory channel increases. Changing bitrates out of the accepted ranges of the different modalities (audio 16-24 kbps; video 32-40 kbps) lowered quality. In Hands' [8] multimedia quality model, the video quality is weighted more towards high-motion content and talking head content as both modalities have approximately the same affect the quality judgment.

2.2 Quality production
In mobile video production, limited bandwidth and the limitations of the devices (e.g. display size, processing capabilities) are the main issues. Spatial resolution, framerate and frame quality are the factors that affect video quality [7] and with 3G mobile networks, questions of low bitrates, low framerates and small screen sizes are the most crucial.

Bitrate and codec reveal the frame quality of video [1]. The latest video compression standard, H.264, has excellent coding efficiency and network adaptation thus enabling a significant reduction in bitrate compared to other released standards such as H.263 and MPEG-4. For example H.264 has features for decreasing visible coding errors (e.g. small block sizes and de-blocking filters) [16].

Objective framerate as a temporal resolution of video appears to the perceiver as natural motion at high level [7] and as distinct snapshots [1] at low level. The relation between the framerate and perceived quality is not linear and is dependent on the content used [2]. In a recent study sports fans tolerated framerates as low as 6fps as long as single frame quality was sufficiently high and the content was personally significant. These sports videos were taken on handheld devices (6-24fps/QCIF) [12]. At low bitrates (24-48 kbps), a frame rate of 8fps gives better video quality than a framerate of 15kbps at the same bitrate from a PC with QCIF frame size [18].

Modern audio encoding e.g. MP3 and AAC is based on efficient perceptual coding [4]. The main parameters affecting quality are the spatial parameters relevant to the monophonic or stereophonic sound and temporal parameters relevant to the sampling rate [2]. With spatial parameters stereophonic sound is better, but at low bitrates mono is more pleasing [18]. The most common impairment appears as unpleasant distortions, preceding noise and the sound of double recording typical at low bitrates and with the use of headphones [4].

This study concentrates on comparing video codecs and audio-video bitrate ratios used in mobile devices.

3. RESEARCH METHODS

The research method was based on ITU-T recommendations for subjective evaluation of video and audiovisual quality [10] and it was used in two different studies. First, the video study investigated the video quality factors and second, the audio-video study explored the combination of video and audio.

3.1 Participants

75 subjects participated in the video study and 60 in the audio-video study. They were stratified according to age (18-65 years) and sex and the number of professional evaluators was restricted to 20%.

3.2 Test procedure

The participants were screened for visual acuity (20/40), color vision and hearing [9][J.H.3]. They were also surveyed for demographics and briefed about the test procedure.

Before the test, participants were shown the highest and the lowest quality samples as an example of the quality scale and they use for the test. In the test, the material was shown using the single stimulus method where the clips are viewed one by one and rated independently [10]. Participants marked the quality score of a clip on an answer sheet using a discrete scale from 0 to 10. The test was followed by an interview concerning content recognition, interest in the content and an open interview on the participants' evaluation criteria during the test.

3.3 Selection of Test Material

Resolution and popularity were the main criteria in the selection of the content. The test materials were from popular TV-programmes from the Finnish broadcasting network and were chosen according to Finnish TV-broadcast ratings [6]. The selected materials were also suitable for mobile TV broadcasting.

The richness of temporal and spatial resolution was the criteria for clip selection for the content of each category selected . The contents are presented in Table 1 (Tele-text was used only in video test).

Table 1. The content selection for the test

Genre	Content	Spatial details	Temporal motion	Audio
News	Evening news	Low	Low	Speech
Sport	Ice hockey	Medium	High	Speech
Series	CSI	Medium	Medium	Speech
Animation	Simpsons	Low	High	Speech
Tele-text	Newsfeed	High	Low	-
Music video	Rock music	High	High	Music

3.4 Test Material Production Process

The original material for clips was sourced from DVB MPEG-2 and midi DV-tapes and converted to PAL format AVI frames (InterVideo WinProducer (3.0B001.111C2A). AVI frames were used as the input to produce the sample clips. The original audio samples (stereo, 32kHz) were normalized and converted to 16kHz sampling rate mono. The parameters for 10 second-long sample clips are presented in Table 2 and the encoding tools in Table 3.

Table 2. The parameters for the both tests

VIDEO TEST						
Device	Nokia 6600			S-E P800		
Video codec	H.263	Real Video8	H.264	XviD		
Picture ratio	QCIF	QCIF	QCIF	QCIF	QCIF	SIF-SP
Bitrate	80	80	80	80	128	128
Framerate	12,5					
Contents	6					
AUDIO-VIDEO TEST						
Device	Nokia 7700		Nokia 7700 | S-E P800			
Codecs	H.263-AAC		H.263-AAC | XviD -MP3			
Total bitrate	100		160			
Audio bitrate	24	16	64	32	16	
Video bitrate	76	84	96	128	144	
Framerate	6	12,5	12,5	12,5	12,5	
Contents	5					
Picture ratio	QCIF					

3.5 Presentation of Test Materials

The headphones supplied with the devices were used for audio playback. The loudness of audio signal from the headphones to

ear was adjusted to 75dB by using a human ear simulator. The devices were attached to a stand and the viewing distance was set to 440mm. General viewing conditions for the laboratory were set according to ITU recommendations [10][11]. Two devices were used in both studies and the starting devices were randomly selected. All clips were played from the device memory . The Nokia 6600 and 7700 used a RealOne Player and the Sony-Ericsson a P800 SmartMovie.

Table 3. The codecs and programs used in encoding

Codec	Program
H.263, AAC	Nokia Multimedia Converter Pro 2.0
RealVideo 8.0	RealProducer 8.0.0.45
H.264	Customized software
XviD, SmartMovie, XviD Codec, Lame MP3 encoder	SmartMovie 2.20

4. RESULTS

4.1 Video test: Codecs, picture ratio and bitrates in visual quality

Codec comparisons were made between codecs H.263, H.264 and RealVideo (the XviD was left out because it only reached the target bitrates occasionally). Figure 1 shows that the best quality scores were acquired in the following order, first H.264, second codec, RealVideo8 and third H.263 (ANOVA $F(2,1347)=143.0$ $p<0.001$). Post-hoc tests indicated that all the pairwise differences between the codecs were significantly different ($p<0.001$). When the contents were taken into account, the results remained the same (ANOVA $p<0.05$; Post-hoc comparison of codecs $p<0.001$ in every comparison) with the exception of tele-text content ($p>0.196$).

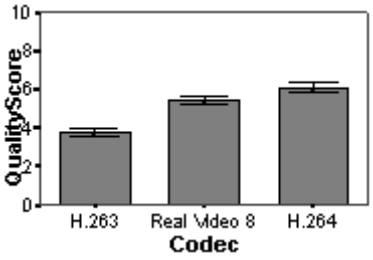

Figure 1. Quality evaluation scores for codecs (The error bars show 95% confidence interval of mean).

4.2 The audio-video test: The most pleasing combinations of audio and video streams

The audio-video study compared different audio-video bitrate ratios. Results are only given for parameters with total bitrate of 160kpbs due to comparability. The ratio had significant effect on evaluations, (ANOVA $F(4,1347)=11.2$ $p<0.001$) and for all content the audio-video ratio 32/128kbps was the most pleasant and it was significantly different ($p<0.05$) from all other ratios with the exception of 64/96kbps ($p>0.276$) (Figure 2).

Content evaluated at different audio-video ratios affected the evaluations significantly. The most pleasant animation test clips were played at a ratio of 32/128kbps which differed significantly from all other ratios ($p<0.01$). The high-motion ice hockey clip with in the middle level of visual detail and speech, as audio, was strongly dependent on the quality of the visual information. The ratio 16/144kbps produced the most pleasant ice hockey clip which differed significantly from all other ratios ($p<0.05$). Music video clips appeared the best when played using ratios of 32/128kbps and 64/96kbps ($p<0.05$) and difference compared to 16/144 was significant ($p<0.001$). Even though the music video clips had high motion and lots of spatial details, the music clearly set the requirements for audio input. The news clip appeared the best when it was played using an audio-video ratio of 64/96kbps followed by a ratio of 32/128kbps ($p>0.05$). A ratio of 16/144kbps was regarded as insufficient for news because of low audio bitrates ($p<0.001$). The subtitled TV series clip was considered to appear the most pleasant when the video had a high enough bitrate. The audio-video ratios 32/128kbps and 16/144kbps gave the most pleasant results and the difference in quality between these and all other ratios was significant ($p<0.05$).

5. DISCUSSION

This study focused on the subjective video quality of mobile devices. In the video study, we examined how codec affected the perceived quality and in the audio-video study we wanted to find the optimal audio-video bitrate ratio combination from three different combinations.

5.1 Codecs and picture ratio

The codec H.264 produced the most satisfying video quality followed by RealVideo8 and H.263. The result was the same with different content except the tele-text. The higher quality of H.264 was expected as it represents the state-of-the-art video coding technology. Also Winkler & Faller's [18] study supports this finding. They compared codecs at extremely low bitrates and concluded that the quality of content viewed using H.264 was perceived significantly better than content viewed using H.263 and MPEG-4.

5.2 Audio-visual bitrate ratio

In the audio-video study, the bitrate ratio 32/128kbps appeared to offer the best results with all contents, although the difference when compares to 64/96kbps was small. Bitrate preferences are summarized in Table 4.

Table 4. The most pleasant audio-video bitrate ratios

Content	Animation	Ice hockey	Music video	News	TV series
AV-ratio (kbps)	32/128	16/144	32/128, 64/96	64/96	32/128, 16/144

The role of audio was emphasized in the music video clip which was visually demanding due to high motion and spatial details. Although the visual aspects of music videos are generally artistic and take a narrative approach with the enjoyment being enhanced with visual effects such as lights, scene cutting, and camera effects [5], our results still suggest that participants relied more on the music in their evaluations. In general, the music video clip collected the highest ratings, which suggests that the limitations of mobile devices do not reduce the enjoyment of the content as much as with other contents. Consequently, music video might be one the first contents that become popular with mobile device users.

Ice hockey and TV series, as high motion and detailed information heavy contents were demanding for video bitrates. For both contents, perceiving meaningful visual details e.g.

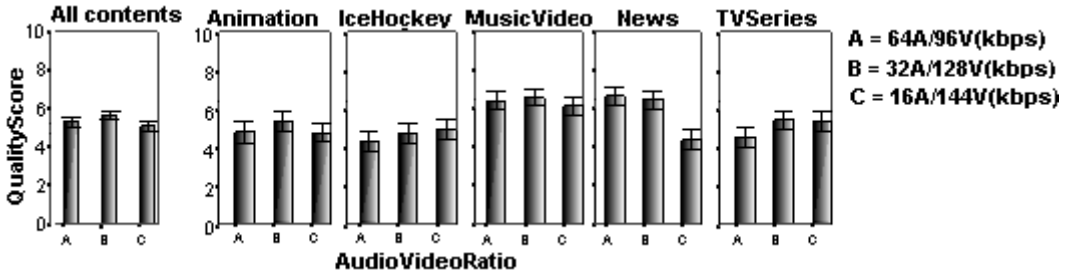

Figure 2. The quality evaluation scores with different audiovisual ratios (The error bars show 95% confidence interval of mean).

subtitles, hockey puck and players is necessary. This result is in agreement with Hands' [8] earlier findings in which video was considered more important than audio in high-motion sports contents. Animation represents the medium level content where both audio and video are important. From the view of parameters tested the result was fairly expected.

5.3 Conclusions

The state-of-art codec H.264 produced the most satisfying quality for presenting video on mobile devices but the quality was not enough for presenting details of tele-text. Demands for presenting different contents clearly require different audio-video bitrate ratios at relatively low bitrates levels. These content dependant requirements are necessary for bandwidth optimization to deliver acceptable quality in new products such as mobile televisions.

6. ACKNOWLEDGEMENTS

This study was funded by Radio- ja televisotekniikan tutkimus Oy (RTT). RTT is a non-profit research company specialized in digital television data casting and rich media development. We also wish to thank prof. Kaisa Väänänen-Vainio-Mattila, Ville Ilvonen, Päivi Aho and Mika Filander for their work in the project and Nick Zacharov in guidance with audio.

7. REFERENCES

[1] ANSI T1.801.02, Digital Transport of Video Teleconferencing /Video Telephony Signals – Performance Terms, Definitions, and Examples. New York: ANSI 1996.

[2] Apteker, R.T., Fisher, J.A.,Kisimov, V.S., Neishlos, H. Video Acceptability and Frame rate. *IEEE Multimedia*, 3(3):32—40, 1995.

[3] Beerends, J.G. & de Caluwe, F.E. The influence of video quality on perceived audio quality and vice versa. *Journal of the Audio Engineering Society*, 47 (5), 355-362, 1999.

[4] Brandenburg K, MP3 and AAC explained, AES 17th *International Conference on High Quality Audio Coding.* Italy: September 1999.

[5] Casey, B., Casey, N., Calvert, B., French, L., Justin, L. *Television Studies – The Key Concepts.* London: Routledge, 2002.

[6] Finnpanel. Television audience measurements. http://www.finnpanel.fi. (visited 05/2004).

[7] Fukuda K. Integrated QoS Control Mechanisms for Real-Time Multimedia Systems in Reservation-Based Networks, PhD Thesis. Osaka University; 2000..

[8] Hands, D. S. A Basic Multimedia Quality Model. *IEEE Transactions on Multimedia*, Vol. 6, No. 6, December (2004). pp. 806-816.

[9] ISO *Standards Handbook 35. Acoustics.* 1st ed. p.386. Switzerland; International Standardization Organization, 1990.

[10] ITU-R BT.500-11 *Methodology for the subjective assessment of the quality of television pictures*, International Telecommunications Union – Radiocommunication sector, 2002.

[11] ITU-T P.910 *Subjective video quality assessment methods for multimedia applications*, International Telecommunications Union – Telecommunication sector, 1999.

[12] McCarthy, J. D., Sasse M. A. and Miras D. "Sharp or Smooth?: Comparing the Effect of Quantization vs. Framerate for Streamed Video", 2004, *Proc. of the 2004 conference on Human factors in computing systems*. Vienna: 2004. p. 535-542.

[13] McGurk, H. and MacDonald, J. Hearing lips and seeing voices. *Nature*, 264, 746-748, 1976.

[14] Reeves, B. & Nass, C.. *The media equation*: How people treat computers, television, and new media like real people and places. Cambridge University Press. 1996

[15] Van Wassenhove V., Grant K.W., Poeppel D. Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America*, 102(4),1181-6, 2005.

[16] Wiegand T., Sullivan G. J., Bjøntegaard G., Luthra A., Overview of the H.264/ AVC Video Coding Standard, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.13,No.7(2003). pp. 1-19.

[17] Winkler, S. & Faller, C. Audiovisual quality evaluation of low-bitrate video. *Proc. SPIE/IS&T Human Vision and Electronic Imaging*, vol. 5666. San Jose, United States of America: January 2005. pp. 139-148

[18] Winkler, S. & Faller, C. Maximizing audiovisual quality at low bitrates. *Workshop on Video Processing and Quality Metrics for Consumer Electronic*s. Scottsdale, United States of America: January 2005.