Effect of TV Content in Subjective Assessment of Video Quality on Mobile Devices

Satu H. Jumisko*, Ville P. Ilvonen, Kaisa A. Väänänen-Vainio-Mattila
Tampere University of Technology, Korkeakoulunkatu 10, 33720 Tampere

ABSTRACT

Selection of test materials in subjective assessment methodology recommendations is based mainly on technical parameters. Materials should test the ability of the codec to cope with spatial and temporal redundancy. However consumers watch TV for a reason - one of the main criteria is the interesting content. In this study we examined whether the content recognition and subjects' personal interests have an effect on quality assessment. We also studied subjective assessment criteria for video quality. The study was done using small resolution and low bit rate video in mobile phones in a laboratory environment. Altogether 135 subjects, aged 18-65 years, participated in the tests. The test started with a subjective assessment of video quality using well-known TV content. Afterwards a survey was done to measure content recognition and level of interests in the content. The test session ended up with a qualitative interview about evaluation criteria. Our studies showed that there is a connection between interest in content and given quality score with TV content. Therefore we raise a concern on content selection and recommend measuring the evaluator's interest in content in subjective assessment studies. The study on subjective evaluation criteria revealed that subjects pay attention on content and quality impairments especially in regions of interest.

Keywords: Subjective evaluation, perception, mobile TV, quality, video, audiovisual

1. INTRODUCTION

People have been consuming video on TV for decades. Motivations for TV consumption vary from purely entertaining to communicational and educating. Consumption of TV has been integrated into people's lives on a daily basis, to such an extent that people are dependent on – even addicted to – this medium. However, it is not the medium alone but the content that creates this attraction. Furthermore, linking TV content to other media such as radio, magazines and more recently, the Internet creates the cross-media environment which can increase the addiction. The total quality of different media combinations – together with the high-quality content – forms the basis for the user experience that the consumers' decision to consume is based on.

Consuming visual media on a mobile device is a relatively new phenomenon. The first steps of using mobile video have been taken with multimedia messaging service (MMS). This service was launched in Europe, for example in Finland in the year 2000 and now allows sending of video clips together with other media items such as text and audio messages to other MMS-capable mobile terminals. MMS is expected to become as wide-spread as text messaging [11]. Other usages of mobile video – including viewing streaming video on mobile terminals – are still on the experimental level. Mobile TV with a video-on-demand feature is predicted to be one of the "killer applications" of mobile terminals [21].

TV image quality has developed from the black-and-white picture of the 50's to the color TV of the 70's to the emerging wide-angle and digital TV qualities. Similar developments have taken place with audio standards; first from mono to stereo sound, and in late 90's, to Dolby surround audio systems which are now being adopted beyond early adopters by the mainstream consumers. Quality of TV image is improving even though the acceptable quality has been achieved a long time ago. Consumers become more demanding when they are presented with new solutions. With mobile TV the development still concerns investigations of what is the acceptable quality level.

The media field is under continuous change. Adoption of new media is a relatively slow process in which consumers' needs must be understood from the broadest issues of social context to the smallest details of technical implementation. The broadest questions deal with the total cross-media field and understanding users' motivations to consume these me-

* satu.jumisko@tut.fi; phone +35850-361-0038

Multimedia on Mobile Devices, edited by Reiner Creutzburg,
Jarmo H. Takala, Proc. of SPIE-IS&T Electronic Imaging,
Vol. 5684 © 2005 SPIE and IS&T · 0277-786X/05/$15

dia in different contexts of use. On the detailed level, the features and qualities of the mobile terminal have a central effect on the attractiveness of mobile video applications. Usability of the device, the quality and size of the display play a role in making the viewing experience worthwhile. Subjective experience of the video quality is one of the central factors affecting the user experience and eventually, consumer acceptance.

In order to optimize the usage of the limited data bandwidth available in mobile terminals, different parameter combinations of video (and audio) need to be tested for users' experience of the quality. Different contents need to be tested to get valid results regarding the best parameter combinations. The current recommendations [2,16] for subjective video quality assessment set aside the criteria for content selection based on users' interest. This paper presents results of the study of a total of 135 subjects on the effect of content in subjective assessment of video on mobile devices.

Chapter two discusses the user needs for TV contents in general and specifically for mobile TV. Chapter three presents the issues of human perception that are behind the subjective video quality evaluation. Chapter four describes the test method and setup, and chapter five presents the results on the effect of the TV content in subjective assessment of mobile video. The subjective evaluation criteria are also discussed. Chapter six concludes the paper.

2. TV CONTENT AND USER NEEDS FOR MOBILE TV

TV is an entertainer and an informer. It has a central role in people's everyday life and is integrated into the daily routines of single people, couples, families and even children. It has psychological, sociological, economical and political aspects [20]. Psychologically television can be a disturber or a comforter. TV is a culture creator via films and stories by spreading information around the world. Economically TV is a part of media industry with enormous markets. Politically TV influences people via its programmes. [3] When moving towards mobile TV, TV on-demand or interactive TV, TV will develop new roles. As these routines develop, there is potential for TV to become integrated into people's lives even in a more holistic way.

The modern television is mixture of many genres. Traditionally, genres organize and classify TV programmes and they can be used as a way to combine types of contents with target audiences and the related industries. Genres are tools for the TV industry in scheduling, targeting and maintaining popularity. Commonly mentioned genres are news, soaps, drama, comedy, sports and science fiction. Different genres have their special characteristics. News as a significant genre represents an official and mostly objective information source. Sports give excitement and possibilities for escaping the everyday existence. Drama like crime and hospital series represents reality in a fictive view. Soap operas are also entertainment drama with special emphasis on family and personal relationships. Boarders between genres are narrow and nowadays they are often mixed. [3]

Whereas traditional TV is mostly consumed at home, mobile TV is expected to be used in a wide variety of contexts and situations. Potential places are, for example, airports, vehicles, waiting halls, hotels and summer cottages [21]. Mobile TV can be also seen as an entertainer for children while the family is travelling or for adults while waiting for something, i.e. "killing time". Naturally, a central target group of mobile TV are people who travel a lot and are heavy media users.

In a mobile TV field study of 81 users in Finland in 2002-2003 [21] it was established that the users of mobile TV will choose news, sports and other entertaining contents such as TV series. The study also showed that people see mobile TV as TV, not as an interactive media service. People expected to get the same programs on their mobile TV terminal as on normal TV. Their motivation to consume mobile TV was seen as a substitute to the evening newspaper: killing time, fighting loneliness, keeping up-to-date and browsing the content. Furthermore, "TV anytime" (video-on-demand) was an attractive usage concept that gives consumers the freedom to watch TV on their mobile terminal when ever they want. Children were also attracted to the mobile TV usage; they enjoyed watching different children's programs and cartoons.

As was discussed in the introduction, quality factors affecting the user experience of mobile TV consist of issues varying from detailed usability aspects to contextual offerings of media combinations. In order to design the mobile TV quality that is subjectively acceptable and attractive to the consumers, we need to understand what happens in the hu-

man mind while watching TV. The next chapter will focus on the viewpoint of human perception and how it affects subjective evaluation.

3. TOWARDS UNDERSTANDING HUMAN PERCEPTION IN QUALITY EVALUATIONS

In subjective evaluation test methods the test participants assess video clips after watching them [2]. The outcome of the evaluation is determined by the quality of the processed image and by the user's previous experiences and interests. The former relates to the artefacts produced by image processing, which interact with the low level sensory processing of the human visual system. If the artefacts are so noticeable that the human visual processing is able to detect them, the judged quality might be impaired. For example, motion, brightness, sharpness and colours are features that affect the quality judgement at this level. The latter processing level relates to the high-level cognitive processing of the images. This processing is affected by previous knowledge, expectations, attitudes and emotions [12,13]. From this viewpoint the user is interpreting the meaning of the video clips while rating them and thus she can be seen as an active information constructor.

The majority of image quality models and objective metrics are based on the low level sensory processing. For example the image quality model of Peter Barten [1] is based on modelling the modulation transfer function of the human visual system. There are also other models which have tried to capture the essence of image quality judgements with low level sensory processing features of the human visual system [10,17,25]. However, there are examples which indicate that low level models are not sufficient in characterizing the image quality judgements.

For example, a recent study showed that extremely low frame rates are acceptable if there is sufficient interest in the content material [14]. In the study, participants were recruited according to their interest in soccer. The results showed that soccer fans found the quality of only 6 frames per second video acceptable 80% of the time. McCarthy et. al concluded from this result that the visual system can tolerate relatively gross interruptions and that participants can accept surprisingly bad quality video clips if there is sufficient interest in the topic. The result indicates that cognitive influences can override sensory level processing and that more research on cognitive factors is needed.

Some steps towards the active human approach have already been taken. Regions of interest (ROI) have been studied with eye-tracking techniques [14,23]. This information can be used in video coding because some regions are more interesting and necessary for understanding the contents than others. Wang et al. [24] have developed video coding algorithms exploiting these characteristics of the human visual system. When video quality is to be compromised it can be done in the areas that are not in the centre of attention.

Recently there has also been multimedia quality theories that integrate sensory and cognitive properties of the user. One such model is quality of perception (QoP) which combines the different levels of subjective evaluation of quality [5]. The quality of perception combines different approaches in subjective evaluation of multimedia quality. Gulliver et al. [6] divided QoP further into sections of information assimilation and satisfaction. Satisfaction has two dimensions, enjoyment and level of objective quality, which can be evaluated separately. They have shown that clip contents affect the perception of quality in the information transfer more significantly than the frame rate, device type and mobility. They also point out the users' ability to distinguish subjective enjoyment and objective level of quality. According to Ghinea and Chen [4]the quality of service parameters like frame rate and colours do not impact user's quality of perception. Instead the selected multimedia content and dynamics affected the user's understanding and enjoyment.

Understanding and applying the knowledge of human perception set new requirements to the codec development and quality evaluation methods. TV genres have roles as both an entertainer and information source. Central questions rising from the levels of perception and TV genres related to the widely used subjective evaluation methods [2] are: Can we assess overall video quality independently of the TV content? Are the familiar or interesting TV contents attaining higher ratings in subjective evaluation tests? How do the evaluators experience the quality in their in own words? In this paper we explore the answers to these questions.

4. METHOD –TEST SETUP

The test setup was an application of ITU-T recommendations for subjective evaluation of video and audio quality [2]. There were two studies, a video study and an audio-video study. The video study investigated the image quality factors and the audio-video study explored the combination of video and audio. These studies were part of a larger audio-visual quality research project whose results are beyond the scope of this article. Both studies were conducted at Tampere University of Technology between July and October 2004.

4.1 Participants

All together 135 participants participated in the two studies, 75 in the video study and 60 in the audio-video study. None of the participants participated in both of the studies. The number of professional evaluators (i.e. people who worked on developing information technology) was restricted to 20%. In both studies participants were stratified by age and sex. Participants were from 18 to 65 years old and they were screened for normal or corrected to normal visual acuity and color blindness. In addition they were screened for normal hearing threshold in the audio-video study [11]. Correction to required hearing threshold level (Q=0.1) was done according to ISO standard [8], due to recruitment difficulties with elderly people.

4.2 Test Configuration

Both studies were conducted in the laboratory presented in Figure 1. In the beginning of the test session the participant was screened for vision and hearing (Area 1 in Figure 1). Participants were briefed about the test procedure and they were shown all the test material using Single Stimulus method [2]. Participants marked their quality score of a sample on an answer sheet. The scale was from 1 to 10. In the final part participants were asked to name the content program and give their level of interests in the content (Area 2) on a scale from 1 to 5. Recognition of content was done by asking participants to name the program in the content. After the test session there was an interview to study the criteria participants had used when evaluating the quality (Area 3).

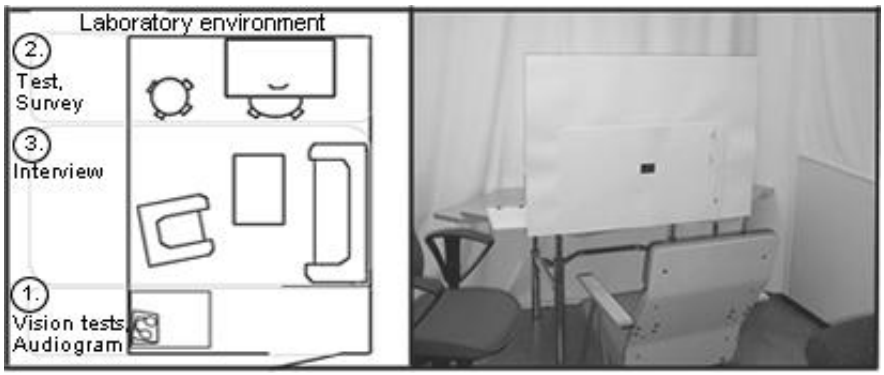

Figure 1 - Laboratory setup

4.3 Test Material based on popular TV contents

The test materials were TV-contents from Finnish broadcast network. Materials are presented in Table 1 and Figure 2.

Genre	Content
News	Evening news
Sports	Ice Hockey – WC 2004 Finland – Russia
Series	CSI (Crime Scene Investigation)
Cartoon	Simpsons
TeleText	Newsfeed
Music video	Sessions, musicvideo (rock)

Table 1 Genres and contents

Figure 2 Screen shots of the study contents

There were two criteria for content selection; popularity and richness of spatial and temporal resolution. The content in each genre was chosen according to Finnish TV-broadcast ratings measured by Finnpanel [4]. Programs were chosen by the highest ratings in the category in order to ensure that all the contents could be recognized by the major part of the test participants. Ratings used were from the year 2003. All contents were used in the video study; the audio-video study was done without Text-TV content because it has no audio. Two of the contents, CSI and Newsfeed, included textual information.

Resolution was the criteria for clip selection within each content. Clips were categorized by the amount of spatial and temporal resolution as presented in Table 2.

Temporal – Motion / Spatial - Detail	**Low**	**Normal**	**High**
Low	News		Animation
Normal		Series	Sports
High	Text-TV		Music Video

Table 2 Temporal and spatial resolution in the selected contents

The original material for clips was sourced from DVB MPEG-2 and DV-tapes to PAL format AVI frames. AVI frames were used as input to produce the sample clips. Sample clips for the video test were produced with H.263, RealVideo8 and XviD codecs using QCIF and SIF-SP picture ratio and bit rates 80kbps and 128kbps. Frame rate was 12.5 fps. Sample clips for the audio-video study were produced with H.263/AAC and XviD/MP3 codec combinations. Total bit rates were 100kbps and 160kbps. Picture ratio was QCIF and frame rate varied from 6 to 12.5 fps. Altogether 36 test samples were produced for video study and 40 for audio-video study. All the samples were presented three times to each test participant in random order.

4.4 Presentation of test samples

Mobile devices were set on a special stand for presentation. The stand made it possible to adjust the viewing distance and angle equal for all participants. Viewing distance was set to 400mm according to preferred viewing distance (PVD) test made for 15 subjects at Tampere University of Technology. In this PVD test, subjects were shown one video clip three times and asked to adjust the distance to their preference. PVD was measured from forehead to device screen.

General viewing conditions for the laboratory environment were set according to BT.500-11 [2]. In the audio-video study both devices were hidden to neutralize the effect of the device. All clips were played from the memory of the device using manual selection from play lists by trained laboratory assistants. Two devices were used in both studies. The order of devices was changed in every other test. The device was changed in the middle of the test.

Video study was carried out with two different devices: Nokia 6000 and Sony-Ericsson P800. RealMedia Player was used for playback with Nokia 6600 and SmartMovie with Sony-Ericsson P800. The audio-video study was also carried out with two different devices: Nokia 7700 and Sony-Ericsson P800. Playback was done with RealMedia Player with Nokia 7700.

4.5 Data analysis methods

4.5.1 The effect of TV content

In this quantitative part of the research, data were analysed with the SPSS for Windows version 12.0.1. One and two factor analyses of variances (ANOVA) and t-test were used for testing significant differences in variable categories [7]. The quality rating is a dependent variable. Recognition of the content and the level of interests in the content were independent variables. Significance level of $p<0.05$ was adopted in this study.

4.5.2 Subjective evaluation criteria

Qualitative data consisted of transcriptions of the interviews. All the participants were interviewed after the subjective evaluation test. In the interview, participants were asked the leading question:

"What were the factors you paid attention to while evaluating the video quality?"

The leading question was followed by supporting questions such as "What do you mean with x? Could you be a little more specific?" All the interviews were audio recorded and transcribed. The data was analysed using inductive analysis [12]. Transcriptions were searched for concerns that participants expressed during the interview. The concerns were reduced to simplified expressions and further classified under themes. Themes were interpreted as evaluation criteria of participants. All together 75 transcribed interviews were analyzed in the video study and 60 in the audio-video study.

5. RESULTS

The results are presented in two parts for each of the two test set-ups. The first part presents the results on the effect of TV contents. The second section concentrates on the subjective evaluation criteria.

5.1 Effect of TV content in the video-audio study

The unrecognised content got higher evaluations compared to the recognised content. The findings are shown in Figure 3a. The quality evaluations were significantly influenced by the recognition of the contents [$t(2398)=9.089$ $p<0.001$]. Ice hockey [$t(58) = 2,68$ $p=0.010$] and music video [$t(58) = 2,68$ $p=0.010$] contents established to have significant effect. Recognition of news, animation or TV series had no effect on the evaluation.

Interesting or attractive content were rated higher than less interesting. Figure 3b presents the effect of interesting content. Also the interest in the content affected significantly the quality evaluation [$F(4,2395)=11.22$ $p<0.001$]. The music video established a significant influence on the evaluation [$F(4,55)=4.17$ $p=0.005$]. The interest had no effect on other contents.

The content recognition and interests in the content also had a significant interaction in the quality evaluation [$F(9,2390)=19.80$ $p< 0.001$]. Music video [$F(7,52)=4.40$ $p=0.001$] and news [$F(7,52)=2.19$ $p=0.050$] demonstrate the effect. Findings are illustrated in Figure 3c. No significant effects were found in the animation, ice hockey or TV series.

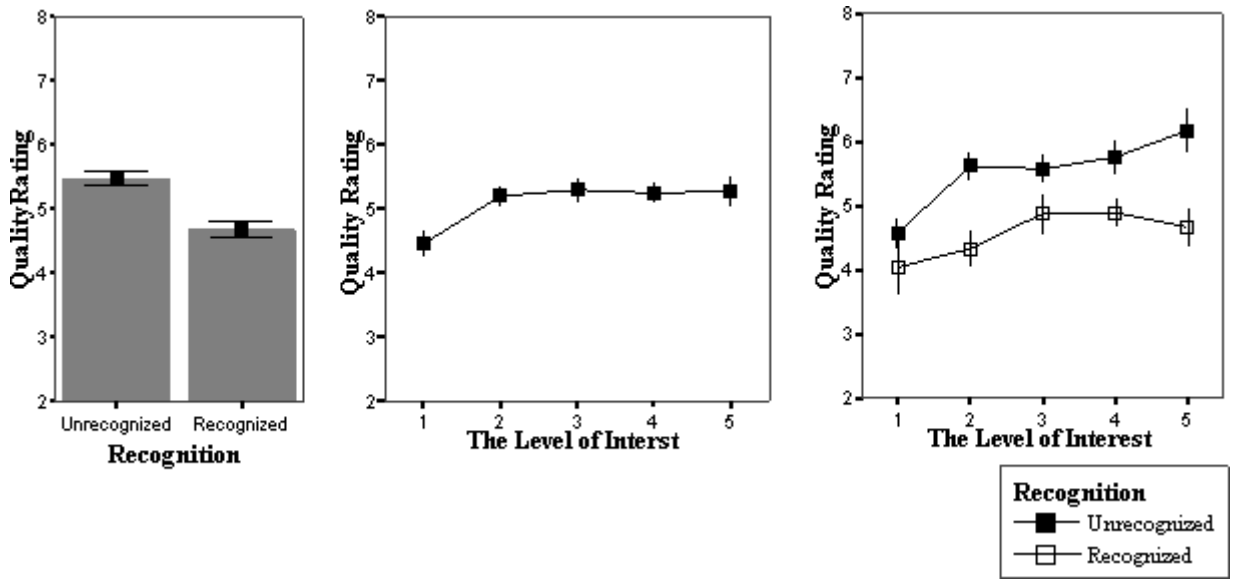

Figure 3 a) Quality ratings with recognition, b) Quality ratings with level of interest and c) Quality ratings with level of interest and recognition. 95% confidence interval is used in the analyses.

5.2 Subjective evaluation criteria in the audio-video study

Altogether 366 concerns were found from the data. Table 3 presents the main themes to which 5% or more of subjects' concerns could be allocated.

Theme	N of mentions	%
Audio	92	25
Accuracy	57	16
Regions of interest	56	15
Content	38	10
Colors	26	7
Block distortion	22	6
Jerkiness	19	5

Table 3 Subjective evaluation criteria themes in the interviews of audio-video study

Participants paid most of their attention to the audio; in particular to the impairments on speech and the informational meaning of the audio. Other important factor with audio was that it was connected to the content. Speech was important in news and music in music video, and audio compensated impairments in the visual part of the video.

"Then the audio quality. How clear it is and those background noises. I mean that even a slight impairment on the background disturbed me very much. So it was more that you can recognize the speech." (Woman, 24)

"And in the end…the [music] playing can be dusky because the situation in the video was dusky or something like that. It was amazing that it fitted to it... surprisingly well. But of course it depends on the song." (Man, 45)

Other themes, beside the content, fall under different impairments in the visual part of the video. Those impairments are already well defined [22] and recognized as factors [16] of total quality. The content is not commonly regarded as an important part of the subjective evaluation of video quality. In this study, the effect of content was expressed explicitly. The content had a dualistic nature in the comments of the participants. Either it had an entertaining or informational value. The common denominator is the interest for content.

"Well it matters what kind of content it is. It affects that ice-hockey is fun to watch and news feels important and you rate them better for that. So the content has an effect and not only plain quality in such an evaluation." (Man, 31)

Also comments on the video quality were connected to the contents through the interviews. Certain impairments identified with examples from certain content. Comments on impairments on regions of interest were reflected this way.

"That can you see the faces well and, in people, numbers of the icehockey players, faces in that soap opera [CSI]. So it was the genuineness of the people. That was it natural. In that TV reporter it could be seen pretty well." (Woman, 54)

5.3 Effect of TV content in the video study

The recognition of the content had an effect on the overall quality evaluations [$t(2248)=2.115$ $p=0.035$]. Recognized contents were given lower ratings compared to unrecognised clips. Results are show in Figure 4a. The effect did not significantly appear individually in any of the contents ($p>0.05$).

The interesting contents collected higher ratings compared to uninteresting ones in the video test [$F(5,2244)=6.874$ $p<0.001$]. The results are provided in Figure 4b. There is a significant difference ($p<0.01$) between evaluations given for uninteresting content rated as 1 and highly interesting contents rated as 4 and 5. The interests in the content did not rise up when screening content by content ($p>0.05$).

The recognition and interests in the content did not have interaction on evaluations [$F(5,2238)=4,896$ $p=0.134$]. Neither was there a noticeable interaction in any of the contents ($p>0.05$).

In both tests unrecognised contents were assessed with better scores compared to recognized contents. Ratings for interesting contents were given with slightly higher ratings compared to uninteresting contents. Even if the levels of explanation were relatively low, these results give direction to taking the content recognition and interest as factors in evaluation.

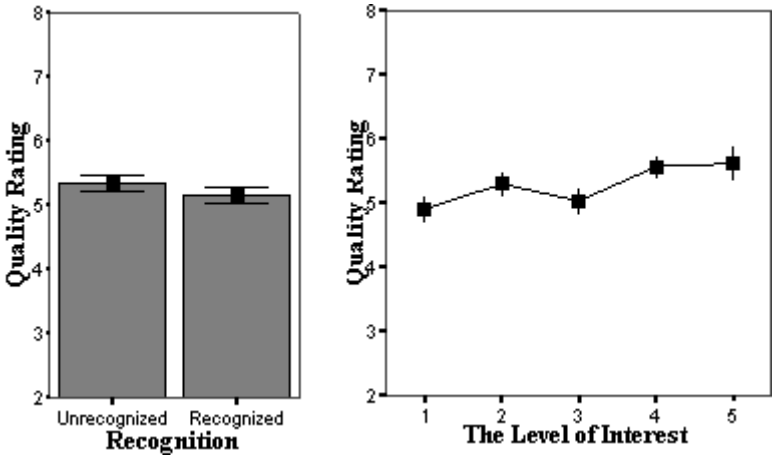

Figure 4 a) Quality ratings with recognition and b) Quality ratings with level of interest. 95% Confidence interval is used in the figures.

5.4 Subjective evaluation criteria in the video study

Altogether 513 expressions of concern were found from the data. Table 4 presents the main themes found in the video study interviews.

Theme	N of mentions	%
Accuracy	86	17
Regions of interest	68	13
Picture ratio	63	12
Text	59	12
Colors	45	9
Block distortion	43	8
Content	42	8
Difference between devices	30	6

Table 4 Subjective evaluation criteria themes from the interviews of audio-video study

The themes are quite similar to the ones in the audio-video study. Some of the differences in themes can be explained with the test setup. Picture ratio became an important theme because participants found watching the larger picture more enjoyable despite the more visible impairments. This refers to the difference between the sizes of the device screens.

"Would I watch it, I mean would it really be worth watching from such a small screen when the quality of the video is lousy you would not watch it at all. At least I'm not ready to pay for it.
Interviewer: What were the other factors you paid attention to when evaluating video quality?
Hmm... one major point was that the picture was larger in that eric [S-E terminal]. So I didn't care if the quality was a little worse as the larger picture in a manner of speaking forgives the bad quality" (Woman, 26)

Regions of interest and text were strongly connected to the content in the comments. When there are visible blocks or other impairments, they are meaningful on the regions of interest.

"And then that you can see the interesting details, for example can you see the puck in the ice-hockey and well of course can you read the text in the newsfeed or can you see the faces of people." (Man, 24)

This result about the importance of regions of interest is also supported by the study in which eye tracking methods has been used to collect information about gaze distribution [14]. Gaze distribution is fixed on a relatively small area compared to the total size of the image.

6. DISCUSSION AND CONCLUSIONS

6.1 Audio-video test

Familiar TV contents collected lower ratings than unfamiliar in the audio-video test. This effect was especially evident with the lively music video and ice hockey contents. It seems that evaluators with previous knowledge about the genre are more demanding for the acceptability of quality.

In the audio-video test music video as an interesting content was evaluated with positive quality scores. One reason for this can be nature of the genre. Music videos are generally experienced as artistic and commercial type of genres. The narrative is approaching high enjoyment in short time made with special lights, cuttings, camera effects, etc. [3]. Music video clip was also the only content having music played instead of speech.

Unrecognised but interesting contents also collected high ratings. In the audio-video test music video and news represented results in this category. In the view of technical parameters the music video clip represents both high temporal and spatial resolution. It represents demandable parameters. In the view of human interests and content recognition have

effect on quality evaluation – especially in the case of music video. It seems that in a test situation some clips are experienced as being interesting even tough their genre is unfamiliar. Lower ratings are again given for known and uninteresting content.

In the audio-video test, quality evaluation criteria were focused on the audio. Audio compensated the impairments in the visual part of the video and impairments in speech were found to be very distracting. These results can be applied to content in very low bit rate mobile video. Audio should be clear enough because it contains the relevant information and the visual part only supports the audio if needed at all. This questions the traditional TV concept, visual radio, the other way around if visual details are hard to discern. An unrecognizable talking head in news gives only a little additional information to the speech. Music is the interesting part in the music video and visuals only support it with artistic impressions.

These results raise interesting questions. What is the effect of content in subjective evaluation of video quality if the user has no interest in the content and the content is not familiar to the user? The meaning has an important role in the perceptual process. Therefore it is not possible to bypass it selecting meaningless contents for studies on subjective evaluation of video quality. However, further studies to identify the regions of interest within the visual part of the content are needed. That information can be used in video coding to clarify the meaningful visual parts in the video to improve the quality.

6.2 Video test

The effects of the interesting and recognized content were light in the video test. Overall findings were done in recognized and interesting content but not in their combined effect. In a video test situation participants were evaluating only visual quality. However, considering everyday TV consumption with audio this evaluation situation differs meaningfully. The evaluation of soundless contents might mask the powerful effect of recognised and interesting content in particular content type.

Video evaluation criteria focused on impairments, especially in regions of interest. Content was not one of the main themes in quality analysis. However, comments on impairments were reflected through the contents. Different regions are interesting in different contents. This result can be utilized in further studies. In combination with eye tracking methods these areas could be identified in different contents. With this information modest quality mobile video could be improved significantly by applying lower compression to these regions. Recent studies [14,19] have shown that users are not impacted easily by variation in frame rate. Therefore efforts on better quality can be focused on spatial level quality rather than temporal level.

In the video test, connections could be noticed between participants' qualitative comments and quantitative ratings. In quantitative evaluations interesting and recognized content had light effects on evaluation. Similarly in the evaluation criteria, content was mentioned related to regions of interests, but it was not the main criteria. In both results the phenomena related to the content stayed more in the background.

6.3 Both studies

Results in two different test set-ups can be compared critically keeping in mind that the used parameters differed. When summarizing the results on high level it seems that stronger effect of interests and recognition in the content can be found in the video-audio test. The fact that this test resembled more the "normal" TV watching might be a reason for this. On the other hand, lack of audio gives the evaluator better chance to focus on impairments of the visual part. However, in both tests the effects of recognizing the content and interests had same directions. Familiar and uninteresting contents were evaluated with lower ratings.

As we have stated in this paper, selection of content in subjective evaluation should be paid more attention to as part of the test process. Recommendations pay attention to selecting the materials based on wide range of resolution. For selection of test materials, the recommendations acknowledge that television programmes could be used in content selection [2]. When using the TV content in test clips, overall quality evaluation may be involved with evaluator's previous knowledge, attitudes, emotions etc. These aspects belong to the mixed top-down and bottom-up processes of the human perception [12][13]. In our study, connections between recognition and interests in TV contents to quality evaluation

were found. Both technical and human factors are necessary to be understood in the content selection. More research is still needed to find the systematic way of selecting contents in different evaluation methods.

When developing video applications for mobile devices, subjective evaluation is an important method of finding acceptable media quality. At the moment, quality is still very modest. At the same time the consumers are experiencing development of extremely high quality TVs. This contrast brings forward many interesting evaluation-related problems to further research. What other human factors should be taken into account in laboratory evaluation? Which factors have a role when moving outside the lab to the real context of use? What are the demands of different consumer segments? While selecting the test materials it seems to be necessary to take into account also the top-down nature of the perceptual processes. In the large scale this means moving from the viewpoint of people as passive information processors towards a viewpoint of people as active humans. Thus, further steps need to be taken in understanding the user experience on a larger scale – the preferred ways to consume content on the mobile TV.

ACKNOWLEDGEMENTS

This study is funded by Radio- ja televisotekniikan tutkimus Oy (RTT). RTT is a non-profit research company spesialized in digital television datacasting and rich media development. For more about RTT, please visit www.rtt.tv. We also wish to thank Päivi Aho and Mika Filander for their invaluable help in recruiting a large number of participants and conducting the tests. We are also indebted to Jukka Häkkinen at Nokia Research Center for his help and comments in theory and analysis.

REFERENCES

1. Barten, P.G. J. *Contrast Sensitivity of the Human Eye and Its Effects on Image Quality*. SPIE Press. Washington. 1999.
2. BT.500-11. *Methodology for the subjective assessment of the quality of television pictures*. ITU-R. 2002
3. Casey, B., Casey, N., Calvert, B., French, L., Justin, L. *Television Studies – The Key Concepts*. p.291. Routledge. London, 2002.
4. Ghinea, G & Chen, S. Y. "The impact of cognitive styles on perceptual distributed multimedia quality". British Journal of Educational Technology. **Vol 34 No. 4**. p.393-406. 2003
5. Ghinea, G. & Thomas, J. P. "QoS impact user perception and understanding of multimedia video clips. Proceedings of ACM Multimedia '98" p.49-54. Bristol. 1998
6. Gulliver, S.R. ,Serif, T. and Ghinea, G. "Pervasive and Standalone Computing: the Perceptual Effects of Variable Multimedia Quality." International Journal of Human Computer Studies. **Vol 60 No. 5-6**. p.640-665. 2004
7. Hinton, P. R., Bronlow, C., MaMurray, I.,Conzens, B. *SPSS Explained*. Routledge. New York. 2004
8. ISO 7029 SFS-EN Acoustics. Statistical distribution of hearing threshold as a function of age. 2000
9. ISO Standards Handbook 35. *Acoustics*. 1st ed. p.386. Switzerland. 1990
10. Keelan, B., W. *Handbook of Image Quality, Characterization and Prediction.* Marcel Dekker, Inc. New York. 2002
11. Koskinen,I., Kurvinen, E., Lehtonen, T-K. *Mobile Image*. IT-Press. p.160. Helsinki. 2002.
12. Matlin, M., W. Cognition. International Thomson Publishing. 2002.
13. Neisser, U., *Cognition and Reality, Principles and Implications of Cognitive Psychology*, W.H. Freeman and Company, San Francisco, 1976.
14. McCarthy, J. D., Sasse M. A. and Miras D. *"Sharp or Smooth?: Comparing the Effect of Quantization vs. Frame Rate for Streamed Video"*, 2004, *Proc. of the 2004 conference on Human factors in computing systems*. p. 535-542.Vienna. 2004
15. Finnpanel. *Television audience measurements*. http://www.finnpanel.fi. visited 05/2004.
16. P.910. *Subjective video quality assessment methods for multimedia applications*. ITU-T. 1999
17. Pappas, T. N. & Safranek, R. J. *Perceptual Quality Assessment of Compressed Image and Video*. p. 669–684 in *Handbook of Image and Video Processing*, A. C. Bovik, ed. Academic Press. New York. 2000
18. Patton, M. Q. *Qualitative Evaluation Methods*. p. 306-313. Sage Publications. London. 1986
19. Serif, T., Gulliver, S.R., Ghinea, G. "*Infotainment Across Access Devices: the Perceptual Impact of Multimedia QoS"*. Proceedings of the 2004 ACM Symposium on Applied Computing. Nicosia. 2004
20. Silverstone, R. *Television and every life*. Routledge, New York. 1994.

21. Södergård, C. (ed.). *"Mobile television, technology and user experiences. Report on the mobile TV project."* VTT Publications 506. p.298. Espoo. 2003
22. T1.801.02, *"Digital Transport of Video Teleconferencing/Video Telephony Signals – Performance Terms, Definitions, and Examples"*. ANSI.1996
23. Vuori, T., Olkkonen M., Pölönen M., Siren A., Häkkinen J., *"Can eye movements be quantitatively applied to image quality studies?"*. *Proceedings of the third Nordic conference on Human-computer interaction*. p.335-338. Tampere. 2004
24. Wang Z., Lu L., Bovik, A.C. "*Foveation scalable video coding with automatic fixation selection"* IEEE Transactions on Image Processing. **Vol 12 No. 2**. p.243-254. 2003
25. Winkler, S. *"Issues in Vision Modelling for Perceptual Video Quality Assessment"*. Signal Processing **vol. 78, No. 2.** pp. 231-252. 1999