Experienced Quality Factors - Qualitative Evaluation Approach to Audiovisual Quality

Satu Jumisko-Pyykkö [1]*, Jukka Häkkinen [2], Göte Nyman [2]
[1] Institute of Human-centered Technology, Tampere University of Technology,
PO Box 553, 33101 Tampere, Finland
[2] Department of Psychology, University of Helsinki, PO Box 9, 00014 Helsinki, Finland

ABSTRACT

Subjective evaluation is used to identify impairment factors of multimedia quality. The final quality is often formulated via quantitative experiments, but this approach has its constraints, as subject's quality interpretations, experiences and quality evaluation criteria are disregarded. To identify these quality evaluation factors, this study examined qualitatively the criteria participants used to evaluate audiovisual video quality. A semi-structured interview was conducted with 60 participants after a subjective audiovisual quality evaluation experiment. The assessment compared several, relatively low audio-video bitrate ratios with five different television contents on mobile device. In the analysis, methodological triangulation (grounded theory, Bayesian networks and correspondence analysis) was applied to approach the qualitative quality. The results showed that the most important evaluation criteria were the factors of visual quality, contents, factors of audio quality, usefulness - followability and audiovisual interaction. Several relations between the quality factors and the similarities between the contents were identified. As a research methodological recommendation, the focus on content and usage related factors need to be further examined to improve the quality evaluation experiments.

Keywords: Subjective evaluation, audiovisual quality, audio, video, qualitative methods

1. INTRODUCTION

New mobile multimedia services are increasing in popularity. In the production of these services, huge amount of data, mobility, and the properties of the devices set special requirements, which can cause noticeable impairments on mobile video or television presentation. To provide the benefits of the media services the produced quality has to be set according to end user's quality criteria. This is a complicated task in the multimedia environment, as the quality perception is determined by interaction of audio and video variables. Currently, the multimodal effects on quality are relatively unexplored. The existing studies provide some indications that the importance of visual and audio parameters varies according to the content, context and task [10,13,21,25,42].

Subjective quality is usually formulated quantitatively based either on methodological recommendations of International Telecommunication Union among engineering society [16,17] or measures of acceptability or quality of perception measurements representing more user-oriented approaches [10,26]. Qualitative research methods, which include gathering the reasons for quality judgments, interpretations, and experiences of quality are used less often. In this paper, our purpose is to use qualitative methods to investigate subjective audiovisual quality in the context of compressed television material for mobile devices. We gathered subject's quality evaluation criteria with a semi-structured interview after subjective quality evaluation experiment. The results of the study can be used to understand the characteristics of multimodal quality, to improve the quality evaluation measures of new mobile multimedia applications, and to define the experienced quality in this context.

The organization of the paper is following: An overview to produced and perceived quality is given in chapter 2. Issues related to audiovisual perception are presented in chapter 3. Chapter 4 summarizes the research methods. The results drawn from three different methods of analysis are summarized in the chapter 5. Chapter 6 presents the discussion and concludes the study.

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

2. COMPROMISING PRODUCED AND PERCEIVED QUALITY

Multimedia quality is a compromised combination between produced and perceived quality. The main goal in this optimization is to produce quality under tight technical constrains, e.g. high data compression with as little negative perceptual effects as possible. The optimization is typically done separately for different media and multimodal e.g. audiovisual perceptual challenges are not taken into account.

1. Quality production

In mobile multimedia production, special requirements are set because of huge amount of data, limited bandwidth and constrains of devices. High level of compression or errors in vulnerable transmission channel can cause visible quality degradation. For example, quality perceptions can be affected by *spatial resolution* of video (the number of pixels in each frame and the dimensions of a frame in height and width e.g. [25]), *frame rate* (the number of frames per second (fps) e.g. [12,26] and the frame quality factors like *bitrate (*the number of bits used to code a particular piece of data (bps) e.g. [25,26], *codec* (compression standards for compression/decompression e.g. [21,22,42]. In the audio compression the relevant factors are for example *spatial factors* (the monophonic or stereophonic sound) and temporal parameters like the *sampling rate* (the number of samples per unit of time, Hz, [1]). Also the combinations of compression factors like bitrate allocation have been under investigations [25,33,42]. Error rates, like packet loss [38] or error rates in mobile television transmission channel (DVB-H) are the topics in the studies of transmission quality [19,20].

2. Quality perception

Quality perception is an active process. Each sensory modality has its special characteristics that depend on the physical dimensions of stimuli. The purpose of early sensory processing is to extract relevant visual features from the incoming sensory information. In vision, sensory processing uses brightness, form, color, stereoscopic and motion information in creating the early perceptual experience whereas pitch, loudness, timbre and location are the attributes of auditory processing [7,11,37]. However, the final quality judgment is always a combination of low-level sensorial and high-level cognitive processing. In cognitive processing stimuli are interpreted and their personal meaning and relevance to intentions and goals are determined. For example individual emotions, knowledge, expectations and schemas representing reality affect the weight that each sensory attribute is given and these factors enable human contextual behavior and active quality interpretation [29,30].

The unified multimodal experience of audiovisual material is created when the information from audio and visual channels are combined. Different modalities can complement and modify the perceptual experience created by other perceptual channels and therefore multimodal experience is more than the simple sum of two different perceptual channels [13]. The McGurk effect is a classical example of audiovisual integration in speech perception in which the mismatched visual and acoustical materials are integrated into unified experience which differs from both presented material [27]. The detailed integration process of audiovisual material itself is complex and still relatively unknown. Traditionally audiovisual perception is examined as isolated processes of different channels (also studied in separate research societies) and multimodality is a combinatory mechanism on the top. However, the multimodality has been also approached from the behavioral and early combination point of view [5]. Even though the processing is not known in depth, synthesis of auditory and visual material is requirement for unified perception. The proper temporal synchronization between the sources is needed in synthesis. Inadequate synchronization reduces the clarity of message and distracts the viewer form intended content [31].

The studies of multimodal quality emphasize the characteristics of audiovisual perception. Good audio quality is known to affect visual quality and vice versa [2,31]. On the other hand the studies of comparing relative importance of auditory and visual data have resulted in the notion of “content dependent”. One media seems to be more important than other in different content, context and tasks [10,13,20,21,22]. Content-based multimedia model of Hands [13] describes the role of different modalities for sport and head and shoulders data in the form of objective quality metric. In the high motion sport content, video quality has relatively more weight than audio. Both modalities affect the quality judgment approximately equally, although the audio quality is havinga slightly more significant role in talking head and shoulder content.

The conclusions of content dependency have also been reported in the context of audiovisual quality in mobile television. According to Jumisko-Pyykkö & Häkkinen [21], the relative significance of audio bitrates was emphasized in the head and shoulder news contents and music video, but video bitrates were more important in the ice hockey and TV-series contents. Winkler & Faller [42] have summarized content dependency in more general level: The more complex audiovisual scene in the low bitrates, the more important is quality in auditory channel. In the study of instantaneous unacceptability comparing the effects of transmission errors, the content dependency is affected by variation of noticeable error frequency and length of the errors [20]. For example, audio errors are more unacceptable in non-artificial contents, like news and sport. These studies indicate that the content type has a significant effect when audio and video content interact in producing the quality perception.

Synchronization has also been under investigations in audiovisual quality estimations. For example, Kitawaki et al.'s [23] model of perceptual quality for audiovisual communication services e.g. interactive real-time audiovisual services emphasizes that the overall quality is the combination of audiovisual quality, delay and their interaction. Knoche et al. [24] have studied the dependency between frame rate and audiovisual skew. In the low frame rates (10fps) the audio lag (at least +40ms) is important for understandable speech perception in the delivery of audiovisual material.

3. QUANTITATIVE, QUALITATIVE AND MIXED METHODS OF QUALITY EVALUATION

Multimodal quality assessments are typically studied with quantitative tests methods based on Telecommunication Union (ITU) research methodological recommendations. The recommendations offer guidelines to design and run the experiments and they are relatively widespread among engineering society. To judge the overall multimedia quality International Telecommunication Union's the recommendation P.911 provides the research methodology for three different retrospective research methods [17]. In mostly used method called Absolute Category Rating (ACR), also known as single stimulus method, the short test stimuli (<10s) are presented one by one, rated independently and retrospectively. This method is especially applicable for system or performance evaluation with wide quality variation and its usage has been reported in several audiovisual quality studies with relatively low qualities for mobile presentation purposes [22,33,42]. Other methods of recommendations are powerful in testing small differences between the test materials, because the evaluations are given in relation to other materials (pairwise comparison and degradation quality methods).

The second recently reported quantitative experimental method approaches the multimodal quality based on Fechner's psychophysical method of limit [26]. The threshold of acceptance is achieved by gradually decreasing or increasing the intensity of the stimulus in discrete steps in every 30 seconds. While watching, participants evaluate quality continuously and say the point of acceptable/ unacceptable quality. In the analysis, binary acceptance ratings are transformed to a ratio calculating the proportion of time during each 30-second period that quality was rated as acceptable. The results are expressed as acceptance % of time. This method is applied in several recently reported studies [e.g. 25].

There is also quantitative user-oriented approach to evaluate the Quality of Perception (QoP) [10,12]. Multidimensional QoP is a combination of information assimilation and satisfaction formulated from dimensions of enjoyment and subjective, but content independent objective quality (e.g. sharpness). Information assimilation is measured with questions of different media contents and both of the satisfaction factors are assessed with a scale 0-5. Final QoP is sum of information assimilation and satisfaction that sets the stimuli into preference order. In addition to these quantitative majority methods, there have also been attempts to objectively quantify the quality experience by applying physiological measurements. For example, user stress, commonly measured by blood volume pulse and heart rate has been used [40,41].

In contrast to wide use of quantitative experimental multimedia quality evaluations, only few studies have used qualitative research methods to tackle the quality. McCarthy et al. [26] have reported the participant's reasons for unacceptable quality after the quality evaluation task. Frame rates (6-24fps) and frame quality described with quantization factor (2-24) were varied with soccer content and evaluated by soccer fans in the experiment. The main factors were: recognizing the players was impossible, problems to follow the ball, the close-up shots were fine but the long/ distant camera shots were poor and the jerky movement. The similar procedure has later reported by Knoche et al. [25] when the evaluation task compared image resolutions, video and audio encoding bitrates were with four different contents. The mostly reported unacceptability reasons across the content types were lack of general details (20%), insufficient image size (18%), eye fatigue (10%) and effort (8%). The frequency of other reasons (text detail, object detail, shot types, facial detail, jerky pictures, audio fidelity, color & contrast) varied content dependently.

Watson & Sasse [39] have reported qualitative results from study of effects of quality degradations on internet speech. The different levels of packet loss, echo, loudness and bad microphone conditions were compared in the study. During the quality evaluation experiment, after scoring each stimulus quantitatively (0-100) the participants were asked to describe the reasons for each rating. The descriptions of quality on the low level packet loss (5%) were fuzzy, buzzy, metallic, electronic, and robotic. The words robotic, metallic, digital, electronic, broken up and cutting out were the mostly described in the case of high packet loss (20%). Bad microphone condition gave impressions of distant, far away, muffled, being on telephone, walkie-talkie or in a box. In all of these studies, qualitative approach has been used to support the quantitative results and the actual data-collection method and method of analysis are reported weakly.

In the field of image quality research, the Interpretation-based Quality (IBQ) approach combining the quantitative and qualitative quality issues have been introduced recently. Nyman et al's. [29] IBQ combines the subjectively interpreted quality with objective produced image quality factors. Their measurement procedure starts with qualitative classification task in which the participants group the image materials into self-defined categories. The quality and subjective liking of the image groups are then estimated. Qualitative part is followed by the quantitative psychometric evaluation. The IBQ hybrid model is a user-oriented approach to connect the produced and perceived quality.

Taken together, most of recent audiovisual quality studies has focused on to organize predefined objective variables into perceptually quantitative preference order. However, these methods typically do not open up the studied concepts from the subjects' point of view. As Reiter & Köhler [32] have pointed out *"the term overall quality is a fuzzy one which is often individually interpreted by test subjects and its meaning depends on subjects personal background"*. Reiter & Kohler see the term of overall quality as a problem that should be solved by defining the number of different quality attributes for the subjects. In contrast to their approach, we want to see this human behavior as a challenge and therefore the term overall quality should be opened up to understand the audiovisual quality phenomena in multimedia applications. It is important to see if the predefined variables are actually the one people judge in the quality evaluations or if there are some significant hidden issues behind.

2. RESEARCH METHOD

4.1 Participants

The study was conducted with 60 participants in the laboratory environment at Tampere University of Technology during the summer of year 2004. The participants were equally stratified by age (18-65 years) and gender. Maximum number of people categorized as innovators and early adopters according to their attitudes toward technology and professional evaluators defined as people studying, working or otherwise engaged in information technology or multimedia processing/presenting were restricted to 20%.

4.2 Test procedure

The test procedure contained three parts [Figure 1]. In the beginning of test session, sensorial test and demographic data-collection took a place. Visual acuity (20/40), color vision and normal hearing threshold with correction according to the age were measured [15].

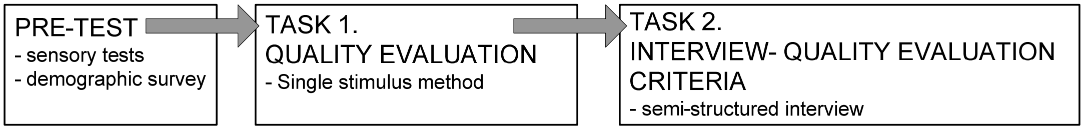
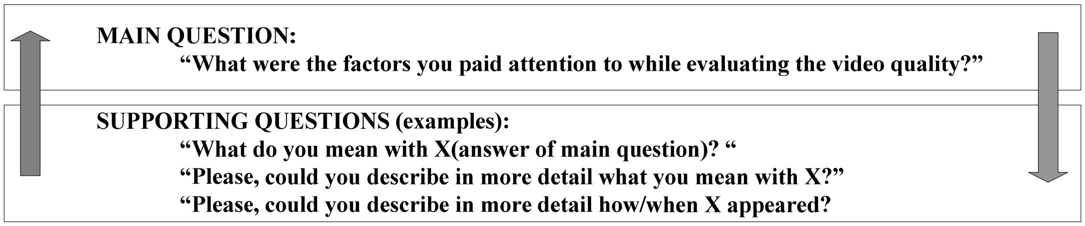


Figure 1 The different tasks of test procedure

The quantitative quality evaluation task was the second part of the test procedure. We applied the single stimulus method in which stimuli lasting 10 seconds are shown one by one and rated independently [16,17] The evaluation task started with anchoring that introduced the quality extremes and contents of stimuli materials and with a short training period. The quality evaluations were given on an unlabelled scale from 0-10. At the end of task, the participants' interests in the content and content recognition were studied.

The final part of test session was the semi-structured qualitative interview about subject's quality evaluation criteria. The semi-structured interview is beneficial when the research field is relatively unexplored and prior expectations are not set [4,6]. The semi-structured interview enabled the data-collection by two assistants. The effects of interviewer are reported to be smaller in semi-structured interview compared to open interview [4]. The interview contained main and supporting questions [Figure 2]. The main question with slight variations was presented several times during the interview and the supporting questions clarified the answers of main questions.



MAIN QUESTION:

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

SUPPORTING QUESTIONS (examples):

"What do you mean with X(answer of main question)? "

"Please, could you describe in more detail what you mean with X?"

"Please, could you describe in more detail how/when X appeared?

Figure 2 The main and supporting questions in semi-structured interview

4.3 Stimulus Material

The stimuli were selected according to their audiovisual characteristics, popularity and potential to mobile television presentation. In five stimulus clips, the levels of spatial details and temporal motions varied in the visual domain and speech and music in audio domain [Figure 3]. The variation of audiovisual characteristics is needed to test the performance of codecs.

Potential genres for mobile television presentation (news, sport and series) were chosen [36] and the content in different genres was chosen according to Finnish TV-broadcast ratings [8]. The programs with the highest rating in each category during the year 2003 were selected to the experiment because of their popularity. The selection was done in order to provide the potential contents for mobile TV broadcasting. The contents in each genre are presented in Table 1.

Table 1 Genre and contents of stimuli

Genre **(Content)**	**Image**
News (Evening news)	
Sports (Ice Hockey)	
Series (CSI)	
Cartoon (Simpsons)	
Music video (Sessions, Rock)	

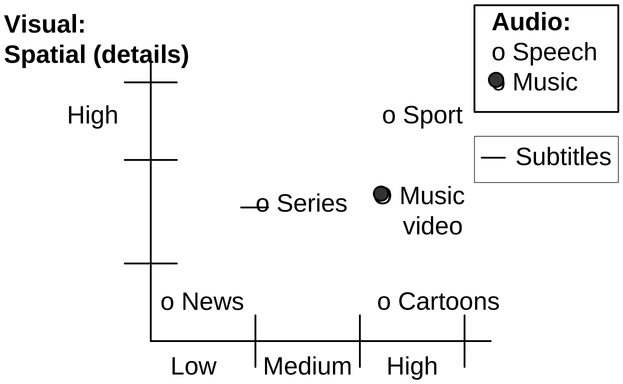


Figure 3 Audiovisual characteristics of stimuli material

Table 2 Encoding parameters

AUDIOVISUAL VIDEO QUALITY					
Device	Nokia 7700		Nokia 7700 / Sony-Ericsson P800		
Codecs	H.263-AAC		H.263-AAC / XviD – MP3		
Total bitrate (kbps)	100		160		
Audio bitrate (kbps)	24	16	64	32	16
Video bitrate (kbps)	76	84	96	128	144
Framerate (fps)	6	12,5	12,5	12,5	12,5
Picture ratio	QCIF				

4.3 Stimulus production process

The original material for stimuli 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 encoding parameters for 10 second-long stimuli clips are shown in Table 2. XviD was encoded with SmartMovie 2.20. and H.263 and AAC used Nokia Multimedia Converter Pro 2.0.

4.4 Presentation of stimuli materials

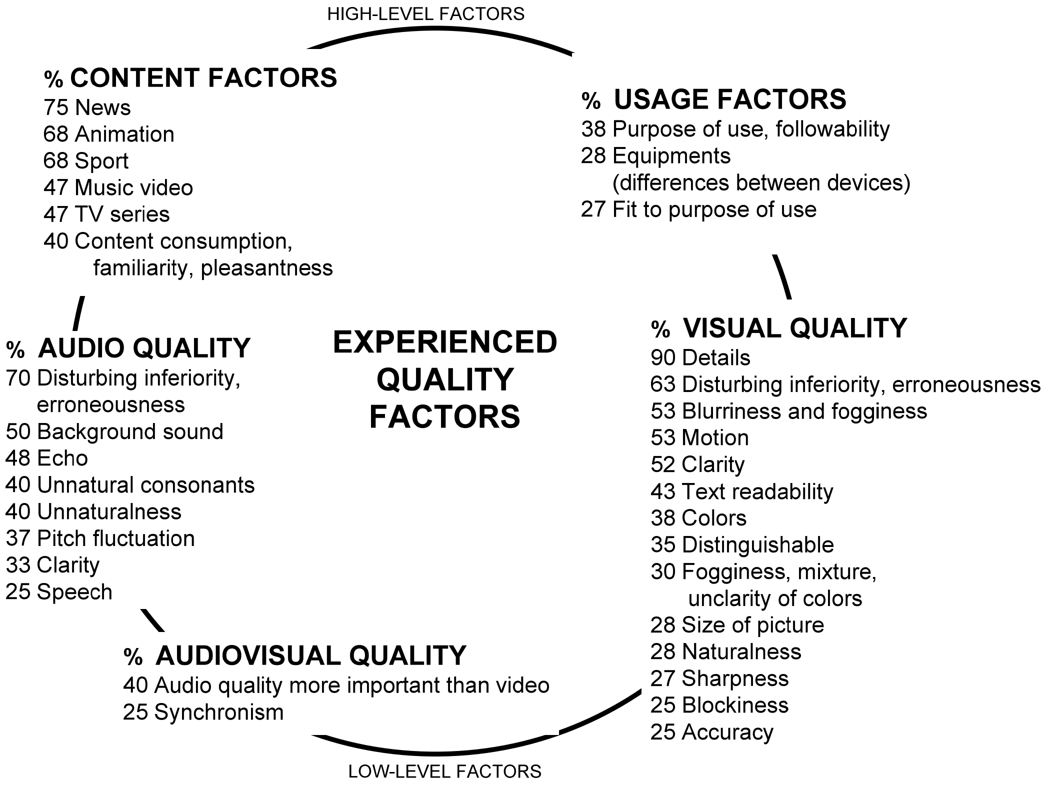
General viewing conditions for the laboratory environment were set according to ITU recommendations evaluators [16,18]. Headphones were used for audio playback and they were the default headphones provided with Nokia 7700 and Sony-Ericsson P800. The audio signal strength from the headphones to the ear was adjusted to 75dB. The devices were set on a special stand for a presentation that enabled to adjust the viewing distance (440mm, [22]), an angle equal (less than 30 degrees from the normal level) and to hide the effect of different devices. All clips were played from the memory of the device using a manual selection from the play lists by the trained laboratory assistants. Two devices were used in both studies and the starting device was randomly selected. The stimuli clips presented with the device Nokia 7700 used a RealOne Player and the Sony-Ericsson a P800 SmartMovie. All stimuli materials were presented three times in randomized order.

5. RESULTS

5.1 Experienced quality factors in one dimension

The qualitative analysis was based on grounded theory presented by Strauss & Gorbin [35]. Grounded theory approach can be applied in the research areas of which there is little a priori knowledge, such as in the case of multimodal quality experience, and when the research aims at understanding the meaning or nature of a person’s experiences. The theory or its building blocks are derived from data with systematical steps of analysis. At the beginning of analysis, all data was read through. Due to the large sample size of qualitative data, detailed open coding was only done for randomly selected 10/60 written interviews to discover the concepts and their properties. All concepts were organized into 64 categories and all data put in order according to these categories. In the categorization, several mentions of the same topic were recoded only once.

The most often mentioned experienced quality aspect or factor categories are presented in Figure 4. At least 25% of participants had at least one description in each of these categories. The results are presented with the aid of five different major categories called content, usage, audio, audiovisual, and visual quality aspects/factors.



% = percent of all participants have mentioned

Figure 4 The experienced quality factors. At least 25% of participants have mentioned the presented categories.

Audio, audiovisual and visual quality factors reflect the characteristics derived from the presented stimulus material and therefore they are grouped into low-level factors. Some of the low level factors, like motion and colors seem to directly refer to the early stages of perceptual processes. Visual detail was the most typically mentioned category. In both audio and visual quality factors, poor quality descriptions about disturbing inferiority and erroneousness were among the most mentioned classes:

> *"… speech clarity… it is just important that there is not additional echo sound. It really started to distract me"* (female, 24 years)
>
> *" actually in the ice-hockey video I was only able to see the puck couple of times. It annoys quite a lot in sport type of contents"* (male, 22 years)

Both classes had also descriptions with negative meaning, like audio unnaturalness or visual blurriness or fogginess and with neutral meaning like audio and visual clarity. In audiovisual quality factors, it was most mentioned that audio quality was more important than video quality.

> *" … for me the audio is more important* (than video) *because I need to watch it from such a small screen"* (male, 50 years)
>
> *" …like in news, the most of information comes from audio and therefore it should be good enough"* (female, 54 years)

In contrast to material-driven low-level factors, content and usage factors represent the more abstract constructions which we call high-level factors. They illustrate more high-level perceptual processes taking into account human goal-oriented actions, like evaluations of fitness to the purpose of use and knowledge, like names of contents. The contents were among the most mentioned factors in study. Especially news, animation and sport contents were named. In the usage factors, purpose of use – followability was the most referred category:

> *"from such a small image, so it is not that accurately possible to follow --- the players"* (female, 53 years)

5.2 Connections between experienced quality factors

Several descriptions of quality factors were typically connected to each others. For example the audio, visual or audiovisual factors were portrayed with examples of stimuli contents. These dependencies were modeled with Bayesian modeling and correspondence analysis.

5.2.1 Bayesian modeling

Bayesian modeling analysis was applied in order to examine the dependencies between the experienced quality factors. Bayesian methods fit well to the context in which the variables are nominal and non-normal in nature and to small sample sizes as they are typically the presumptions for the sophisticated statistical methods [28]. In Bayesian modeling, the model is inductively constructed from data and the statistical hypothesis testing cannot be done. Uncertainty is formulated from conditional probabilities to evaluate the Bayesian inference taking into account probabilities of unknown things in the given data and background information. The b-course modeling tool was used in the analysis [3].

Bayesian classification modeling was conducted in order to figure out the best predictors for each content type. There are two reasons to do content by content examination. The contents constructed one of the main categories in grounded theory analysis. It also seems that experience of quality is not averaged over the contents but it is rather content dependant as previous studies have concluded [13,20,21,25]. The strongest variables are presented content by content in Table 3. The classification results show that some contents (animation, sport and TV series) have more visually related experience factors than other contents which are rather described with both audio and visual factors (music video, news). The estimated accuracy of the model was more than 90% (M=90.1%, SD=2.2%) and the number of estimated models approximately 39000.

In addition to the contents, the Bayesian classification was conducted to the category Disturbing inferiority, erroneousness and Naturalness of visual and audio quality as well as to the categories Fit and Not fit to purpose of use. In all of the cases over 39000 models were evaluated. Disturbing inferiority, erroneousness of audio and visual material, as they among the most represented categories gave the most meaningful classification results. The results of both are presented in Table 4. The accuracy of audio model was 90.4% and visual 88.7%. In the all other classification factors, the prediction power of the classification performance was relatively weak (mostly <0.34%).

The dependency modeling was also used for all data set to find the dependencies between the all variables of data without any classifying factor. Altogether 1970108 model candidates were evaluated. The strongest causal influences between the audio experience factors were: Speech was described with Clarity*, Disturbing inferiority and erroneousness with Unnaturalness, Unnatural consonants and Background sound. The connected visual factors were: Motion is illustrated with Jerkiness, Details with Sport content and Distinguishable. The strength is announced as the final model is one millionth (* one billionth) times probable if the arc is removed between the factors.

Table 3 Experienced quality factors for each content type. The percentage illustrates the decreased prediction power in the classification performance if the variable is left out of the model.

CONTENT	EXPERIENCED FACTORS [Predictive acc.]
Animation	(V): Graininess [0,84]
	(V): Borderlines of colors [0,64]
	(V): Clarity of colors [0,51]
	(V): Disturbing inferioty, erroneousness [0,34]
Sport	(V): Distiguishable [1,36]
	(U): Not fit to purpose of use [1,01]
	(V): Shooting distance and angle [1,01]
	(AV): Video quality more important than audio [0,67]
	(V): Colors [0,34]
	(V): Graininess [0,34]
TV Series	(V): Naturalness [0,84]
	(V): Text readability [0,51]
	(C): Content consumption, familiarity and pleasantness [0,51]
	(U): Purpose of use, followability [0,34]
	(V): Accuracy [0,34]
	(V): Disturbing inferioty, erroneousness [0,34]
	(V): Motion [0,34]

CONTENT	EXPERIENCED FACTORS [Predictive acc.]
Music Video	(U): Fit to purpose of use [0,67]
	(V): Naturalness [0,67]
	(A): Music [0,67]
	(A): Undisturbed naturalness [0,34]
	(V): Lightness [0,34]
News	(A): Unnatural consonants [1,52]
	(A): Speech [1,35]
	(AV): Audio quality more important than video [1,18]
	(A): Echo [1,18]
	(V): Clarity [0,51]
	(V): Details [0,51]
	(U): Purpose of use, followability [0,34]
	(V): Distiguishable [0,34]
	(V): Blockiness [0,34]
	(V): Text readability [0,34]
	(A): Undisturbed naturalness [0,34]

(A) = Audio, (V)= Visual, (AV)= Audiovisual, (U)= Usage, (C)= Content

Table 4 Experienced factors of disturbing inferiority and erroneousness. The percentage illustrates the decreased prediction power in the classification performance if the variable is left out of the model.

DISTURBING INFERIORITY, ERRONEOUSNESS	
EXPERIENCED AUDIO QUALITY FACTORS	**EXPERIENCED VISUAL QUALITY FACTORS**
Unnaturalness [2,02]	Fogginess, mixture, unclarity of colors [1,85]
Background sound [1,18]	Blockiness [1,69]
Echo [1,01]	Eye fatigue [1,52]
Foggy, mellowness, muddling [0,51]	Picture size [1,35]
Speech [0,51]	Accuracy [1,35]
Technical parameters [0,34]	Details [1,18]
	Colors [1,18]

5.2.2 Correspondence analysis

The aim of correspondence analysis is to show the relationships between the factors presented with the rows and columns of a correspondence table. Nyman et al. [13] and Radun et al. [30] have for example applied this method to combine qualitative categories to quantitative evaluations in the image quality studies. In our analysis, we are restricted to qualitative categories and we conducted several trials in correspondence analysis (e.g. taking into account only the most mentioned categories presented earlier, all categories and examining audio, visual and other factors in own groups.) However, the requirement, the certain level of inertia was only filled with some content categories but not with descriptive categories. Because of this unstructured data, the reliable results cannot be drawn from this data with correspondence method of analysis.

6. DISCUSSION AND CONCLUSIONS

The aim of this study was to examine the factors that contribute to the subjective audiovisual quality experience of compressed television materials on mobile devices. These experience factors were gathered with the qualitative semi-structured interview that was conducted after the quality evaluation experiment. Audio-video bitrate ratio and video framerate were varied in the experiment. One-dimensional holistic representation of experience factors was derived based on the use of grounded theory analysis and connections between the factors were analyzed with Bayesian network modeling.

6.1 Experience of audiovisual quality and its definition

In our study, the experience of audiovisual quality was based on five main groups each consisting of different subjective quality variables.. Three of the groups, i.e. those concerning audio, audiovisual and visual quality represented the low-level data-driven aspects of perception. They also illustrated that the experienced quality is mediated via different senses each with their own quality characteristics. Combination of the variable groups was balanced between importance and timing of two senses. Our results are in line with previous studies of modeling audiovisual properties [13,14,23] as well as the descriptions of quality features [25,26].

Two of the quality contributing groups, usage and content described the quality from higher level abstraction point of view. They illustrated rather more interpretation and knowledge based approach on perception of quality than directly reflecting the presented material. The usage group that included the aspects of purpose of use, followability demonstrates the classical J. J. Gibson's ecological approach to perception [9]. According to him we perceive objects as affordances showing possibilities for acting in the environment. As a conclusion, the experience of audiovisual quality cannot be separated from these potential action-related properties.

The experienced quality was described with respect to the audiovisual contents. The importance of one sensory channel over another was demonstrated in two ways. In sport content, the importance of visual over audio appeared as one aspect and in news content vice versa which is in line with earlier studies [13,21]. In addition, visually demanding sport [21] and TV-series [21] and animation as audiovisual [21] stimuli contents were mostly connected to visual quality aspects whereas audio-dominated music video [21] and news [13,21] to both audio and visual aspects of quality. This difference between the sensory channels indicates a genuine content dependent, relative importance or weight of one sensory channel over another.

Our results can be summarized in the form of definition for the experienced audiovisual quality. Experience is typically used, but not a well-defined term in HCI studies. *The experienced audiovisual quality is an integrated set of audio, visual and their interaction specialized audiovisual perceptual aspects which characteristics dependents on presented material, and excellence or distraction of aspects is evaluated according to the goals of usage.* The definition is applicable among the population of non-professional evaluators in passive viewing task and in a specific usage context.

6.2 Research methods of quality

The relevance of viewer's goals in the viewing situation should be considered when quality optimization experiments are designed. Typically, user studies are roughly done at the level of quality optimization (e.g. bitrate allocation including this work) and usability measurements [32]. The usability studies might have a higher level of application readiness and more focused methods in user group, tasks and contexts selection than the quality optimization studies. In the optimization studies, the user evaluates quality of the presented material in an artificial laboratory setting with a limited set of contents. In our results, the quality was represented by the usage-related descriptions. These results lead to the discussion if quality evaluations, conducted for certain application field, should be given more from the usage point of view taking into account the user's goals and context. For example, could the followability or watchability represent the quality in mobile video optimization?

The qualitative research method presented in this paper needs some further development. First of all, content based interviews with stimulus materials are needed to improve the understanding of content dependence. Secondly, several test materials from the same content are necessary to improve the external validity of content or clip-type dependent conclusions. In addition, the effect of positive/negative quality factors might be valuable to gather in more detail and/or with mixed methodologies to model the intensity of subjective quality. These structural improvements can be expected to give more organized data for using sophisticated methods of analysis and to raise the reliability of modeling the connections between different quality aspects. In long term, this would increase the understanding between produced and experienced quality.

To conclude, this study holistically describes experience factors of audiovisual quality and presents the definition for experienced audiovisual quality in the context of video presented in mobile devices. These quantitative results were clearly supported by earlier studies. This study also highlighted the significance of usage in the terms of followability (watchability) to be considered in the quality optimization studies. In the further work, the content and usage related factors need to be examined more in detail to improve the quality evaluation experiments to create a broader view to the experienced quality factors.

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, Mika Filander and Ville Ilvonen for their invaluable work in the project, and Tuomas Leisti and Jenni Radun for their help in the correspondence analysis. Satu Jumisko-Pyykkö's work is supported by the Graduate School in User-Centered Information Technology and this publication preparation work by HPY' Research Foundation.

REFERENCES

1. Apteker, R.T., Fisher, J.A., Kisimov,V.S., Neishlos, H."Video Acceptability and Frame rate,"*IEEE Multimedia,* 3(3):32-40, 1995.
2. 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).
3. B-Course – Bayesian Data Analysis and Modeling: http://b-course.hiit.fi/
4. Clark-Carter, D. *Quantitative Psychological research*. New York Psychology: Press, 2002.
5. Coen, M., "Multimodal Integration - A Biological View," *In Proceedings of IJCAI'01*, Seattla, WA, 2001.
6. Coolican, H., 2004. *Research methods and statistics in psychology* ,4th ed, London: J. W. Arrowsmith Ltd, 2004.
7. Evans, F.F., "Auditory processing of complex sounds: An overview*,*" *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, Volume 336, Issue 1278, pp. 295-306 (1992).
8. Finnpanel. *Television audience measurements*. http://www.finnpanel.fi. visited 05/2004
9. Gibson, J.J., *The Ecological Approach to Visual Perception*, Houghton Mifflin, Boston. Lawrence Eribaum, 1979.
10. Ghinea, G. & Thomas, J. P. "QoS impact user perception and understanding of multimedia video clips, " *Proc. of ACM Multimedia '98"*. Bristol, pp. 49-54 (1998).
11. Grill-Spector, K. & Malach, R.,"The human visual cortex,"*Annual Review of Neuroscience*, Vol 27, 649-677(2004).
12. Gulliver, S.R., Serif, T. and Ghinea, G. "Pervasive and Standalone Computing: the Perceptual Effects of VariableMultimedia Quality," *International Journal of Human Computer Studies*. Vol 60 No. 5-6. p.640-665(2004).
13. Hands, D. S., "A Basic Multimedia Quality Model," *IEEE Transactions on Multimedia*,Vol.6,No.6, 806-816(2004).
14. Hollier, M. P., Rimell, A. N., Hands, D. S. and Voelcker, R. M. "Multi-modal perception," *BT Technology Journal* Volume 17, Number 1, 35-46 (1999).
15. ISO Standards Handbook 35, *Acoustics*, 1st ed. p.386. Switzerland. 1990.
16. ITU-R BT.500-11 *Methodology for the subjective assessment of the quality of television pictures*, International Telecommunications Union – Radiocommunication sector, 2002.
17. ITU-T P.911 Recommendation P.911, *Subjective audiovisual quality assessment methods for multimedia application*, International Telecommunication Union – Telecommunication sector, 1998.
18. ITU-T P.920 *Interactive test methods for audiovisual communications*, International Telecommunications Union – Telecommunication sector, 2000.
19. Jumisko-Pyykkö, S., Vinod Kumar M. V., Liinasuo, M, Hannuksela, M. "Acceptance of Audiovisual Quality in Erroneous Television Sequences over a DVB-H Channel, " *Proc. of Second International Workshop in Video Processing and Quality Metrics for Consumer Electronics*, Scottsdale, USA, January 2006.

20. Jumisko-Pyykkö, S., Vinod Kumar, M. V., Korhonen, J., “Unacceptability of Instantaneous Errors in Mobile Television: From Annoying Audio to Video,” Proc. 8th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI 2006) in Espoo, Finland, September pp. 1-8 (2006).
21. Jumisko-Pyykkö, S. and Häkkinen, J., “Evaluation of Subjective Video Quality on Mobile Devices”. *Proc. of ACM Multimedia 2005*, Singapore, pp.535 – 538 (2005)
22. Jumisko-Pyykkö, S., Häkkinen, J. *“I would like see the face and at least hear the voice”: Effects of Screen Size and Audio-video Ratio to Perception of Quality in Mobile Television*. Proc. of 4th European Interactive TV Conference Athens, Greece, May 18-19, 2006.
23. Kitawaki, N., Arayama, Y., Yamada, T. “Multimedia opinion model based on media interaction of audiovisual communications,” *Proc. 4th International Conference of MESAQIN 2005*, pp.5-10 (2005)
24. Knoche, H., de Meer, H., Kirsh, D. “Compensating for Low Frame Rates,”. *Proc. of CHI 2005*, 4-7 April 2005, Portland, OR, USA (2005).
25. Knoche, H., McCarthy, J. D., Sasse, M. A. “Can Small Be Beautiful? Assessing Image Size Requirements for Mobile TV,” In Proceedings of ACM Multimedia 2005, 561-, 6-12 November, Singapore (2005).
26. McCarthy, J. D., Sasse M. A. and Miras D. “Sharp or Smooth?: Comparing the Effect of Quantization vs. Frame Rate for Streamed Video”, *Proc. of the 2004 conference on Human factors in computing systems*. Vienna. p. 535-542(2004).
27. McGurk, H. & MacDonald, J. “Hearing lips and seeing voices”, Nature 264, 746-748 (1976).
28. Myllymäki, P., Silander, T., Tirri, H., Uronen,P., “B-Course: A Web-Based Tool for Bayesian and Causal Data Analysis,” *International Journal on Artificial Intelligence Tools*, Vol 11, No. 3, pp. 369-387(2002).
29. Nyman, G., Radun, J., Leisti, T., Oja, J., Ojanen, H., Olives, J-L., Vuori, T., Häkkinen, J. “What do users really perceive: probing the subjective image quality,” *Proceedings of SPIE*, Vol. SPIE-6059, pp. 13-19 (2006).
30. Radun, J., Virtanen, T.,. Nyman, G., Olives, J-L., “Explaining multivariate image quality - Interpretation-Based Quality Approach,” Proc. of ICIS 06, Rochester, NY, USA, 119-121 (2006).
31. Reeves, B. & Nass, C., *The media equation: How people treat computers, television, and new media like real people and places.* Cambridge University Press. 1996.
32. Reiter, U. & Kohler, T. “Criteria for the subjective assessment of bimodal perception in interactive AV application systems,” *Proc. 9th International Symposium on Consumer Electronics*, 14-16 June 2005 pp. 186- 192 (2005).
33. Ries, M. Puglia, R., Tebaldi, T., Memethova, O., Rupp, M. “Audiovisual Quality Estimations For Mobile Streaming Services,” Proc. of the 2nd International Symposium on Wireless Communication Systems, Siena, Italy (2005).
34. Rogers E. M. *Diffusion of Innovations*, 5th ed., New York: Free Press, 2003. p. 560.
35. Strauss, A., and Corbin, J., *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.),Thousand Oaks, CA: Sage, 1998.
36. Södergård, C. (ed.). *”Mobile television, technology and user experiences. Report on the mobile TV project.”* VTT Publications 506. Espoo. 2003.
37. Ts’o, D.Y. & Roe A. W., “Functional compartments in visual cortex: segregation and interactions”. In *The Cognitive Neurosciences* (Gazzaniga MS, ed.). M.I.T. Press, Cambridge, MA, pp. 325-337 (1995).
38. Watson, A. Sasse, M.A., "Multimedia conferencing via multicast: Determining the Quality of Service required by the end-user". *Proc. International Workshop on Audio-Visual Services over Packet Networks*, Aberdeen, 1997.
39. Watson, A. &Sasse, M. A., “The Good, the Bad, and the Muffled: The Impact of Different Degradations on Internet Speech,” *Procs 8th ACM International Conference on Multimedia*, Marina Del Rey, CA; pp.269-302(2000).
40. Wilson, G. & Sasse, M. A., “Do Users Always Know What's Good For Them? Utilising Physiological Responses to Assess Media Quality,” *Proc. of HCI 2000* (September 5th-8th, Sunderland, UK), pp.327-339(2000)
41. Wilson, G. & Sasse, M. A., “Investigating the Impact of Audio Degradations on Users: Subjective vs. Objective Assessment Methods,” *Procs OZCHI'2000*, Sydney, Dec. pp. 135-142 (2000)
42. 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. pp. 139-148(2005)