"Sharp, bright, three-dimensional" - Open Profiling of Quality for mobile 3DTV coding methods

Dominik Strohmeier[a] and Gerhard Tech[b]

[a]Ilmenau University of Technology, P.O. Box 100565, 98693 Ilmenau, Germany;
[b]Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute
Einsteinufer 37, 10587 Berlin, Germany

ABSTRACT

The choice of the right coding method is a critical factor in the development process of mobile 3D television and video. Several coding methods are available and each of these is based on a different approach. These differences lead to method specific artefacts - content and bit rate are as well important parameters for the performance. In our study, we evaluated Simulcast, Multi View Coding, Mixed Resolution Stereo Coding and Video + Depth Coding. Therefore each method has been optimized at a high and a low bit rate using parameters typical for mobile devices. The goal of the study was to get knowledge about the optimum codign method for mobile 3DTV, but also to get knowledge about the underlying rationale of quality perception. We used Open Profiling of Quality (OPQ) for comparison. OPQ combines quantitative rating and sensory profiling of the content. This allowed us to get a preference order of the coding methods and additional individual quality factors that were formed into a quality model. The results show that MVC and V+D outperform the other two approaches, but content itself is still an important factor.

Keywords: stereo video coding, sensory profiling, visual quality evaluation, subjective evaluation, 3DTV

1. INTRODUCTION

After the first boom in the 1950s, today 3D cinema and video is supposed to be the next big thing in entertainment. 3DTV is one of the emerging technologies in the sector of entertainment technologies. It is expected to come to our living rooms soon. People will be watching 3D content on large screens, either with the help of glasses or on autostereoscopic displays. However, studies on user needs and expectations[1] have shown that 3DTV can also be an appealing technology for mobile devices. The results show that although the technologies of mobile TV and 3DTV originally differ very much related to technology (e.g. screen size) or usage context (e.g. living room vs. mobility), prospective users see potential use for a converging mobile3DTV system.

Currently, mobile 3D television and research is in the focus of research projects. One of the fields of research is the development of the core technologies for an optimized mobile 3DTV delivery over DVB-H system.[2] Currently, different coding methods have been adapted for mobile 3D content to efficiently compress content for transmission over DVB-H. This process needs to consider limited bandwidth of the channel and calculation power of the mobile receiver device. According to user requirements for mobile 3D stereo and video, mobile3DTV is expected to be mainly private viewing. So only two encoded views are taken into account. From the users' point of view the developed coding methods need to provide good enough overall quality to satisfy his needs and expectations. Only end users' quality acceptance will guarantee success of the system under development. So, in parallel to the technological developments, large-scale user studies are conducted to optimize the critical components of the system following a user-oriented quality evaluation methodology.[3] Overall quality of the system is studied with respect to the end-product, prospective users, and their user requirements.

The aim of this paper is twofold. Firstly, it targets the question of the most applicable coding method for mobile 3D television and video. We present the results of a large scale study with 47 test participants who

Further author information: (Send correspondence to Dominik Strohmeier)
Dominik Strohmeier: E-mail: dominik.strohmeier@tu-ilmenau.de, Telephone: +49 (0)3677 69–2671
Gerhard Tech: E-mail: gerhard.tech@hhi.fraunhofer.de, Telephone: +49 (0)30 31002–416

evaluated four different coding methods. Secondly, we applied a new research method called Open Profiling of Quality (OPQ). In contrast to common quality evaluation methods, OPQ extends the quantitative quality evaluation with a sensory profiling of the items under test. The goal of OPQ is to elicit individual quality factors and to connect them to users' quality preferences. User-centered quality evaluation aims to understand the rationale that leads to a certain quality preference of the items under test.

The paper is organized as follows. In section 2, we give an overview of existing quality evaluation methods and their application in 3DTV research. Section 3 introduces the different coding methods and explains the basic algorithms. Chapter 4 presents the used research method and explains the different steps of Open Profiling of Quality. The results of the study are presented in section 5. Section 6 finally discusses and concludes the paper.

2. TOWARDS USER-ORIENTED QUALITY EVALUATION FOR MOBILE 3DTV

2.1 Related psychoperceptual quality evaluation studies

For a long time, video quality has been analyzed following a psychoperceptual approach. Psychoperceptual methods have been standardized by several standardization bodies. The most applied methods for (stereoscopic) video quality assessment can be found in the ITU recommendations.[4,5] All psychoperceptual methods try to quantify test participants' quality sensation. Therefore, test participants judge the overall quality of a set of short video clips on a numerical scale. The methods have been widely applied to evaluate the quality of critical components in the context of mobileTV and 3DTV. In the context of mobile television, studies have focused on evaluating the impact of coding errors and quality degradation due to low bit rates. Additionally, the impact of changes in image size or framerates has been evaluated.[6–8] With respect to stereo-coding, there are studies on stereo-coding errors in images and videos.[9] An overview about psychoperceptual evaluations in the context of 3D television can be found in Meesters et al.[10]

However, stereoscopic quality evaluations have shown that the experienced overall quality is more than pure image or video quality. Depth impression has always been claimed to provide an additional added value of 3D. Studies verified this added value provided by depth.[11] However, the theory failed when distorted stimuli were used in the quality assessment. So, Seuntiens[12] concludes in his 3D quality experience model that the added value of 3D is a combination of depth and video quality.

What is mentioned but not included in Seuntiens' model is that also other factors seem to have impact on 3D quality perception. Beside evaluation of video quality user studies in the field of 3DTV have focused on the impact of different quality factors on presence, the users feeling of being there. Simulator Sickness or visual discomfort[13] has been evaluated for mobile 3D presentation by Häkkinen et al.[14] Its impact is shown, but the measurement is still in progress. Recently, Lambooij et al.[15] studied different ways to measure the impact of visual fatigue and its influencing factors associated with stereoscopic displays.

2.2 Mixed method approaches in audiovisual quality evaluation

An additional challenge in current quality evaluations is "to arrive at a better understanding of the attributes underlying a multidimensional concept like image quality".[10] Mixed method approaches combine quantitative and qualitative research methods to elicit individual quality factors to combine these with quality preferences. In audiovisual quality evaluation, two general approaches have been applied. While Experienced Quality Factors[16] and Interpretation-Based Quality (IBQ)[17] use interviews as qualitative tools, our Open Profiling of Quality (OPQ) uses methods from Sensory profiling to elicit individual quality factors. Coming from the food sciences, sensory evaluation is "a scientific discipline used to evoke, measure, analyse and interpret reactions to those characteristics of food and materials as they are perceived by senses of light, smell, taste, touch and hearing."[18] Interview-based approaches and sensory profiling have been applied successfully in mobile TV and 3DTV research.[16,19,20]

3. CODING METHODS FOR STEREOSCOPIC VIDEOS

Four coding methods for stereoscopic video are evaluated in this paper. These methods are Simulcast, Multi View Coding (MVC), Video + Depth Coding (V+D) and Mixed Resolution Stereo Coding (MRSC). The latter three methods target advanced technologies for savings of bit rate. Nevertheless, these savings come with new video artifacts as well as with additional processing steps at the sender and receiver side. A brief overview is given in table 1. A detailed description is presented below.

Table 1. Overview of different stereo coding approaches

	Simulcast	**MVC**	**V+D**	**MRSC**
Advanced bit rate saving by	-	exploration of view similarities	high compressibility of depth	reduced resolution of one view
Expected Artifacts	coding	coding	coding, rendering, interpolation	coding, blurr
Additional processing at sender	-	interview prediction	depth estimation	down-sampling
Additional processing at receiver	-	interview prediction	view synthesis	up-sampling

3.1 H.264/AVC Simulcast

The Simulcast approach is straight forward. Both views of a stereo sequence are coded and transmitted independently. In this scope coding has been carried out using H.264/AVC, the latest video standard of ISO/IEC MPEG and ITU-T VCEG.[21] In contrast to other methods pre- or post-processing is not needed before coding and after decoding. The complexity of video processing on the receiver is only determined by the H.264/AVC decoder. A drawback of the Simulcast approach is that redundancy between the two similar channels is not exploited. Thus, the data rate is doubled compared to 2D coding.

3.2 H.264/MVC

H.264/MPEG-4 Multi View Coding[22] extents H.264/AVC with interview prediction. For coding of stereo video a frame of the second view can not only be predicted from its preceding frames, but also from the corresponding frame of the first view. The similarity of the views can be exploited and leads to further rate reductions. To keep backward compatibility with single view decoders, the interview prediction is only applied in one direction. Hence it is possible to decode the first view independently from the second view. As for simulcast coding no pre- or post processing is required. Nevertheless, the complexity of the H.264/MVC decoder increases due to the additional features needed for interview prediction, as for example an increased size of the decoded picture buffer.

3.3 Video plus Depth Coding using MPEG-C Part 3 and H.264/AVC

The views of a stereo sequence are very similar. A mapping of the samples from one view to the other view is given by the geometry of the depicted scene and the camera setup. For rectified stereo views the mapping can be described by scalar values representing the disparity between corresponding samples of the first and the second view. With known camera parameters depth can directly be derived from the disparity. If a view and its depth are given, a second virtual view can be synthesized by shifting the samples of the view by the disparity derived from its depth. Therefore a stereo sequence can be converted to the V+D format.

The transmission of V+D data can be carried out using the ISO/IEC 23002-3 ("MPEG-C Part 3") standard.[23] This standard allows an independent coding of video and depth, so an optimized bit rate allocation for both signals is possible. The required depth estimation can be done offline on the sender side. At the receiver side view synthesis has to be performed. Main features of view synthesis algorithms are pixel shifting and the interpolation in disoccluded regions.

3.4 Mixed Resolution Stereo Coding using H.264/AVC

The perception of stereo video is influenced by binocular effects. Distortions that exist independently in the two stereo views can amplify or suppress each other. If the presented stereo views have different resolutions, the suppression theory states that perceived image quality is dominated by the view with higher spatial resolution. In case of stereo views with different amounts of blocking artifacts, the perceived quality is given by their mean quality.[9] The Mixed Resolution approach utilizes these attributes of human perception. Bit rate is saved by down-sampling one view and can be spend for an increased coding quality. Hence a tradeoff between spatial sub-sampling and quantization is carried out, that should result in a higher overall quality. Down-sampling can be done at the sender side offline. Although up-sampling and interpolation has to be done at the receiver side, the total computational complexity does not increase. Decoding of the sub-sampled right view needs much less operations[24] than up-sampling. Beyond the simple Mixed Resolution approach evaluated in this scope more enhanced Mixed Resolution concepts are under investigation.[25] These approaches utilize interview prediction, as well as optimized down-sampling and reconstruction methods.

4. RESEARCH METHOD

4.1 Participants

47 participants (age: 16-37, mean: 24) equally stratified in gender took part in this study. All participants were recruited according to the user requirements for mobile 3D television and system. A parents' consent was collected for all underaged participants before the study. All participants were screened for normal or corrected to normal visual acuity (myopia and hyperopia, Snellen index: 20/30), color vision using Ishihara test, and stereo vision using Randot Stereo Test ($\leq$ 60 arcsec). The sample consisted of naive participants who didn't have any previous experience in quality assessments.[26] They were no professionals in the field of multimedia technology. 15 test participants were chosen randomly to take part in the additional sensory profiling part of the study. No participant reported Simulator Sickness symptoms in the test. Simulator Sickness was measured using Simulator Sickness Questionnaire.

4.2 Test Procedure

Open Profiling of Quality (OPQ) approach was chosen for the quality evaluation. OPQ combines standardized quantitative evaluation methods and sensory quality profiling method. The quantitative quality evaluation consisted of pre-test, evaluation, and a post-test task. Pre- and post-task tests included demographic data collection, screening and Simulator Sickness measurement. Following, a training and anchoring was conducted in which test participants watched a subset of test items. They represented all contents and extreme values of the study. Participants familiarized with the evaluation task and the usage of quality evaluation scales. In the evaluation, Absolute Category Rating (ACR) according to ITU-R P.910[5] was chosen. In ACR, stimuli are presented consecutively and rated independently after each test item. Test participants rated general overall quality acceptance on a binary yes-no scale[27] and the overall quality satisfaction on an 11-point unlabeled scale. The rating was done on paper to avoid that an additional rating screen acts as an indirect monoscopic quality reference. Each test item was evaluated twice. The order of the stimuli were randomized to avoid bias effects. The quantitative session took 90 minutes in total.

The sensory profiling part of OPQ adapts Free Choice Profiling in which test participants develop their individual quality attributes. In the *attribute elicitation task*, test participants were asked to use their own words to evaluate perceived quality. While watching a subset of 24 randomly chosen test items, test participants wrote down their quality attributes (preferably adjectives) that described their individual quality sensation. In the following *attribute refinement task*, test participants selected a maximum of 15 attributes. These attributes must be unique, i.e. describe one specific quality aspect. Additionally, test participants must be able to define the attributes precisely. The selected attributes were written on a score card attached to a 10cm long line. This line is labeled from 'min' to 'max' at the ends. In the evaluation task, the participants then rated overall quality on these attributes for each test item independently one after another. Therefore they marked the sensation intensity of each attribute on the line, 'min' refers to no sensation of the attribute, 'max' to maximal sensation of the attribute for the respective item under assessment. The sensory profiling session took 75 minutes in total.

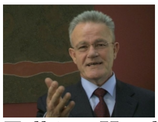
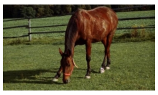
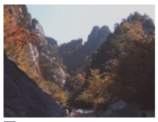
4.3 Test Material and Apparatus

The tests were conducted in the Listening Lab at Ilmenau University of Technology. This laboratory offers a controlled test environment. The laboratory settings for the study were set according to ITU-R BT.500.[4] A NEC autostereoscopic 3.5 display with a resolution of 428px x 240px was used to present the videos. This prototype of a mobile 3D display offers equal resolution for monoscopic and autostereoscopic presentation. It is based on lenticular sheet technology. The display was connected to a Dell XPS 1330 laptop via DVI. The laptop served as a playback device and control monitor during the study. The randomized orders of the test items had been stored as different playlists for each part of the study before.

4.3.1 Selected test content

The selected contents are in line with the user requirements for mobile 3D television and video. They represent different genres like documentary, sports, or animation. Additionally, the videos represent a variety of spatial details, temporal resolution, and amount of depth.[3] Six different videos with a length of 10 seconds were chosen for the test. A screenshot of each content can be seen in Table 2.

Table 2. Test contents and their characteristics.

Details	**Bullinger**	**Butterfly**	**Car**	**Horse**	**Mountain**	**Soccer2**
	 Talking Head	 Animation	 Feature Film	 Documentary	 Documentary	 Sports
Spatial	medium	high	high	medium	high	medium
Temporal	low	medium	high	low	low	high
Depth	low	medium	medium	high	high	high

4.3.2 Preparation of the test sequences and parameters under assessment

Codec and Codec settings Four coding methods were selected for this study – Simulcast, Mixed Resolution, Multiview Video, and Video+Depth Coding. The different approaches are described in section 3. Coding has been carried out using two codecs. For the Simulcast, Mixed Resolution and V+D approach the H.264/AVC Reference Software JM 14.2 has been used. The MVC stimuli have been coded using the H.264/MVC reference Software JMVC 5.0.5. Codec settings have been chosen to match the requirements for application in mobile devices. To further decrease coding complexity, the baseline profile was used for encoding. This includes a coding structure of IPPP and the use of CAVLC (Context Adaptive Variable Length Coding). To provide frequent random access point to the transmitted bit stream, the period of intra frames was set to 16. Hence the period of I-frames is a little more than one per second and enables fast reentry after burst errors. Due to the low resolution of the test sequences, the search range for motion-compensated prediction of the encoder was set to 48.

Quality levels The coding approaches have been evaluated at a high and low quality level. Due to a variable compressibility of different sequences it is not useful to set these quality levels to fixed bit rates. A rate sufficient for a high quality for one sequence might produce a low quality for other sequences. To guarantee comparable low and high quality for all sequences, individual bit rate points had to be determined for each sequence. For all sequences of the test set the quantization parameters (QP) of the encoder for Simulcast Coding was set to 30 for the high quality and 37 for the low quality. This results in a low and high bit rate for each sequence of the coding test set. Resulting bit rates are shown in Table 3 and have been used as target rates for the other three approaches.

Table 3. Bit rates of the test sequences for low and high bit rate level

Bit rate level	Bullinger	Butterfly	Car	Horse	Mountain	Soccer2
Low (in kbit/s)	74	143	130	160	104	159
High (in kbit/s)	160	318	378	450	367	452

Simulcast and MVC settings The generation of test stimuli using the Simulcast and the MVC approach is straight forward. To achieve the target bit rates shown in Table 3 the quantization parameters for the left and the right were changed jointly. Thus left and right view are of same quality. For the Video plus Depth and the Mixed Resolution approach a optimization was carried out to find an optimal bit rate distribution between full and down-sampled view or video and depth, respectively.

V+D settings For the V+D approach depth has been estimated using a Hybrid Recursive Matching algorithm.[28] For view synthesis the simple algorithm presented by Yongzhe et al.[29] has been used. The PSNR was calculated based on the average MSE of the left and the rendered right view. The rendered right view from uncoded data has been taken as reference for the rendered right view from coded data. Rendering artifacts already existing in the uncoded data are neglected with this approach. Hence the PSNR calculated this way only evaluates the coding quality and not the overall quality. The optimization has been carried out using QPs from 18 to 44 and a step size of two for the left view. For depth QPs from 8 to 44 or 18 to 44 depending on the sequence have been used. The step size has been two. Sequences coded with QP combinations for video and depth that result in the highest PSNR and the bit rates defined in Table 3 have been taken as test stimuli. If necessary, coding with intermediate QP-combinations has been done to match the target bit rates more precisely. The optimization results in a bit rate for depth of approximatly 10% to 30% of the total rate depending on the complexity of depth.

MRSC settings For generation of Mixed Resolution sequences, the right view was decimated by a factor of about two in horizontal and vertical direction. For up- and down-sampling tools provided with the JSVM reference software for Scalable Video Coding have been utilized. The approach suggested by Brust et al.[24] was used to determine optimal bit rate distribution between the full and the down-sampled view. Thus, the PSNR was calculated from the average MSE of the full and the up-sampled low resolution view. The down- and up-sampled original view was taken as reference for the coded and up-sampled low resolution view. This approach takes the binocular suppression theory into account since distortions introduced down-sampling are neglected. Only the coding quality is evaluated but not the overall quality. The PSNR calculated this way can be utilized for optimization of MR coding, but not for a objective comparison of the coding methods. The optimization has been performed with a QP range from 18 to 44 with a step size of 2 for the left view as well as the down-sampled right view. Sequences matching the bit rates defined in Table 3 and coded QP combinations for video and depth that lead to a maximum PSNR have been taken as test stimuli. Therefore also coding with intermediate QP-combinations has been done if necessary. Findings of the optimization are bit rates for the down-sampled view of approx 30% to 45% of the total rate depending on the test sequence.

4.4 Method of Analysis

4.4.1 Psychoperceptual data analysis

The psychoperceptual data was analyzed as follows. As no normal distribution was given for the dependent variables (Kolmogorov-Smirnov: $P < .05$), non-parametric tests were applied. Acceptance ratings were analyzed using Cochran's Q and McNemar-Test. Cochran's Q is applicable to study differences between several and McNemars test is applied to measure differences between two related, categorial data sets. Comparably, to analyze overall quality ratings a combination of Wilcoxon test and Friedman test was applied. Friedman test is applicable to measure differences between several related samples. Wilcoxon then measures differences between two related, ordinal samples. SPSS 15.0. was used for analyzing psychoperceptual data.

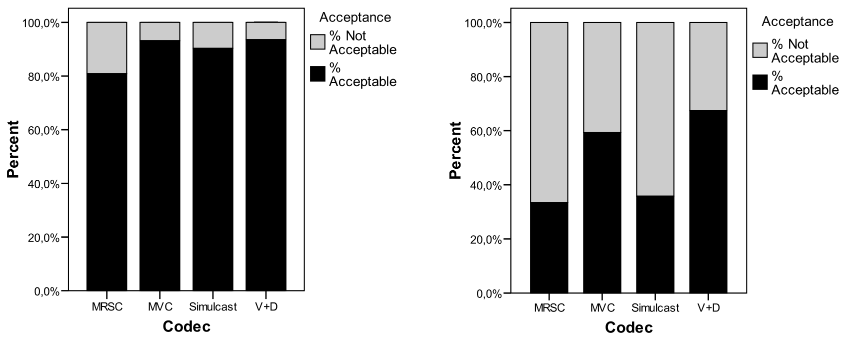
4.4.2 Sensory data analysis and External Preference Mapping

To be able to analyze the participants' individual attributes and related ratings, i.e. configurations, they must be matched according to a consensus configuration. Generalized Procrustes Analysis (GPA)[30] rotates and translates the individual configurations by minimizing the residual distance between the configurations and their consensus. The scaled data sets are analyzed using Principal Component Analysis (PCA). From the high-dimensional input matrix, PCA creates a low-dimensional model of the perceptual space. The relation between the participants' attributes and the PCA model can be plotted as correlation plots which show the correlation of each individual attributes with the principle components of the low-dimensional model. Finally, the preference data of the psychoperceptual evaluation and the data of sensory profiling can be combined in an External Preference Mapping (EPM).[31] Sensory analysis was run using XLSTAT 2009.5.01.

5. RESULTS

5.1 Acceptance of Overall Quality

Figure 1(a) and 1(b) show the acceptance scores of each coding method at high and low quality level. All coding methods provide highly acceptable quality at the high quality level. They all get an acceptance score of 80% and higher. At low quality level, MVC and V+D still provide acceptable quality at an acceptance score of 60% and more. At high quality all differences are significant (all comparisons: $P < .05$) except MVC vs. V+D ($P > .05$). At low quality all differences are highly significant (all comparisons: $P < .001$) except MRSC vs. Simulcast ($P > .05$).



(a) High quality level (b) Low quality level

Figure 1. Quality acceptance scores for all coding methods.

5.2 Satisfaction with Overall Quality

The analysis of the overall quality ratings shows that Multiview Video Coding (MVC) and Video+Depth (V+D) are outperforming the other methods. MVC and V+D perform on a comparable level in a general view. Content was a determining factor in comparing the overall quality that is provided by different coding methods (Friedman: all comparisons $P < .001$). Figures 2 and 3 show the results averaged over content (All) and content by content. The mean satisfation scores differ significantly for high and low quality level (Friedman test: all comparisons $P < .001$). In both bitrate settings, coding methods had a high significant effect on the overall quality (Friedman test: all comparisons $P < .001$), except for sequence Bullinger in high quality case (Friedman: $F_R = 2.942, df = 3, P > .05, ns$). Comparing general performance of the coding methods the results show that MVC and V+D perform on the same level (High quality: $Z = -.828, P = .407, ns$, Low quality: $Z = -.316, P = .752, ns$). They significantly outperfom MRSC and Simulcast (all comparisons: $P < .001$).

5.2.1 High Quality level

The results for the high quality level can be found in Figure 2. The results show that MVC and V+D get the best mean satisfaction scores ($Z = -.828, P > .05, ns$). They significantly outperform MRSC and Simulcast (all comparisons: $P < .001$). MRSC gets the worst mean satisfaction score (MRSC vs. Simulcast: $Z = -7.233, P < .001$). For more detailed evaluation, results are shown per content in Figure 2. Video+Depth significantly gets the best mean satisfaction scores at content Car, Mountain, and Soccer2 (all comparisons: $P < .01$). At content Butterfly, MVC outperforms the other three coding methods (MVC vs. Simulcast: $Z = -3.006, P < .01$). MRSC always gets the worst mean satisfaction scores. Interestingly, there are no significant differences between the coding methods at content Bullinger (Friedman: $F_R = 2.942, df = 3, P > .05, ns$).

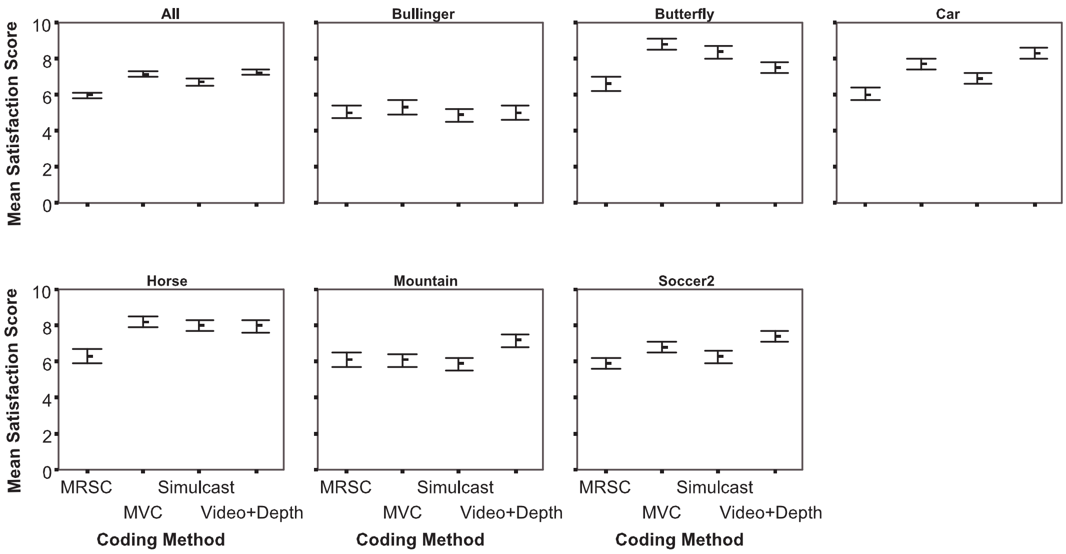


Figure 2. Mean Satisfaction scores averaged over all contents and content by content for high quality level.

5.2.2 Low Quality level

As it can be seen in Figure 3 MRSC and Simulcast perform similar at low quality Level (MRSC vs. Simulcast: $Z = -.316, P > .05, ns$). While for high quality level MVC and V+D outperformed each other at different contents, they now perform similar for all contents (all comparisons: $P > .05$), except Butterfly ($Z = -7.078, P < .001$) and Car ($Z = -6.662, P < .01$). The differences at Bullinger are very small compared to the other content.

5.3 Sensory Evaluation of Overall Quality

5.3.1 Interpretation of the GPA model

Quantitative evaluation has shown that there is a clear preference for MVC and Video+Depth method. These two methods provide the best perceived quality for low and high quality level. However, the quantitative data does not give information about a rationale for this preference ranking. 13 assessors developed a total of 102 individual quality attributes in the sensory profiling tasks. By applying GPA, the individual data has been scaled and transformed in a low-dimensional model. This model can help to explain the identify a common structure of the data set. The first two components of the GPA model explain 88,36% of the variance of the data with two components. The scores of each attribute are depicted in Figure 4. The correlation plot in Figure 5 illustrates the correlation of each individual attribute with the two components of the GPA model. The higher the correlation of an attribute with at least one component, the more important this attribute is to explain differences between the items. Attributes between the inner and the outer circle explain between 50% and 100% of the variance. These attributes are considered to be more important and so are more emphasized in the following interpretation.

To understand the meaning of the model, the first goal is to identify the resulting components. From the correlation plot in Figure 5 one can see that component 1 correlates on negative polarity with attributes like

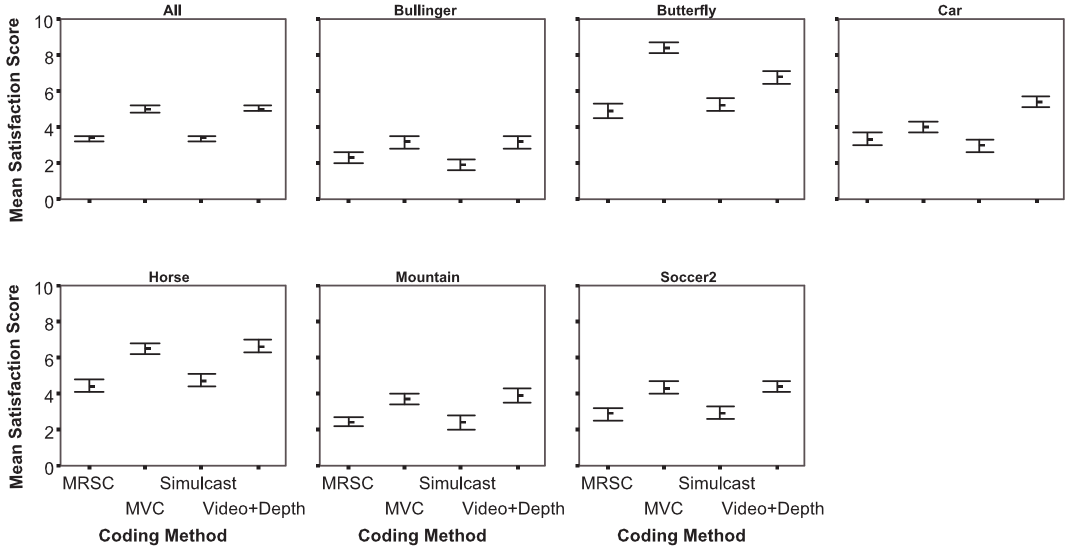


Figure 3. Mean Satisfaction scores averaged over all contents and content by content for low quality level.

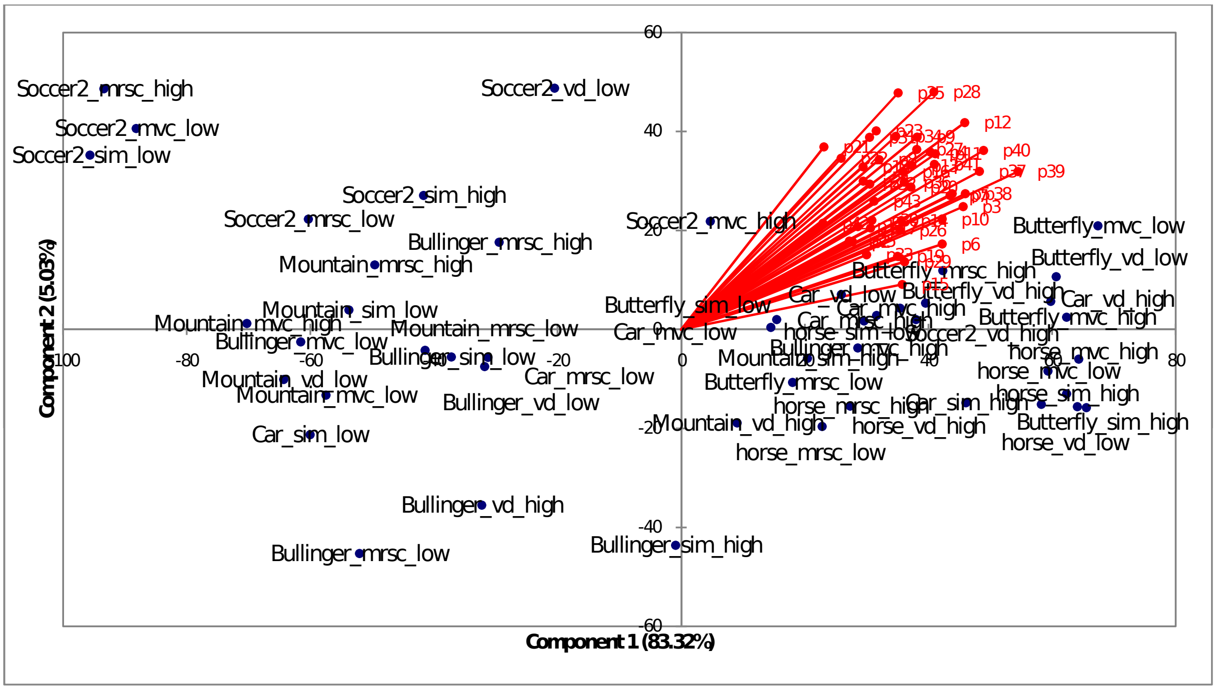


Figure 4. The score plot of the Generalized Procrustes Analysis. It shows the items in the resulting GPA model. The arrows mark the users' preferences obtained by using External Preference Mapping.

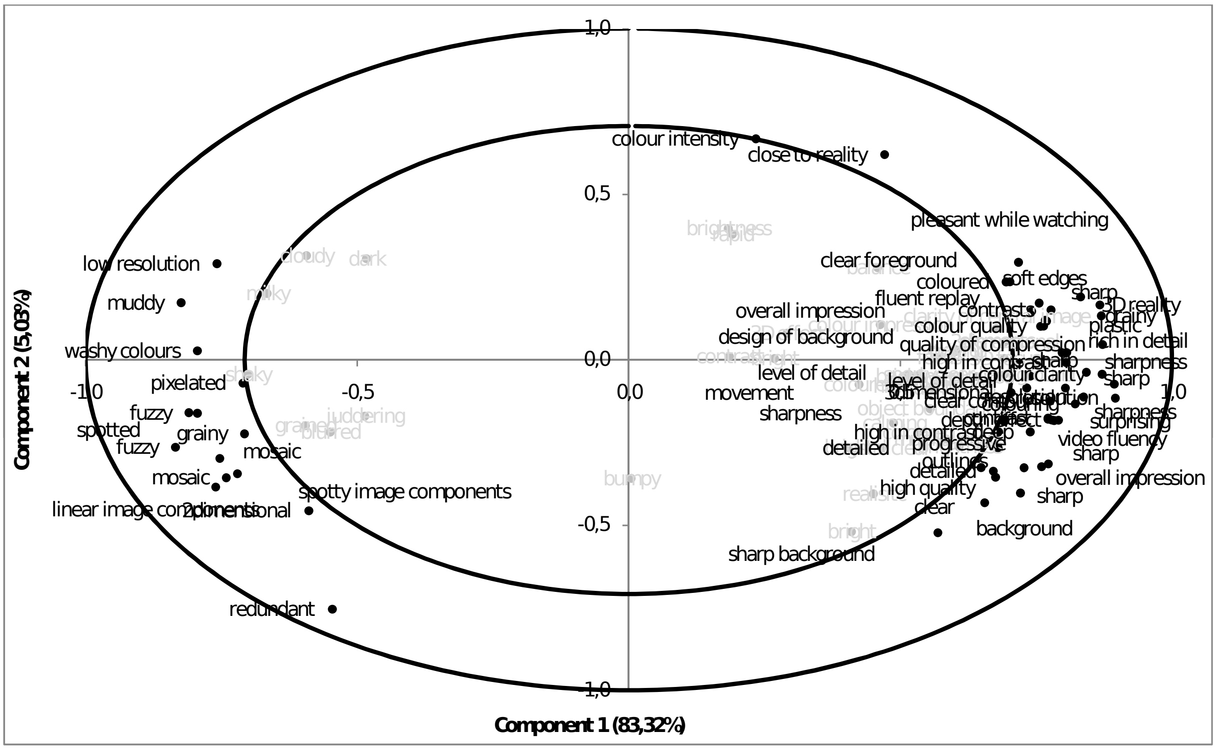


Figure 5. The correlation plot of the Generalized Procrustes Analysis showing the correlation of each individual attribute with the first two components of the model. The inner ring marks 50%, the outer ring marks 100% explained variance.

'blurry', 'blocky', or 'grainy'. In contrast, its positive polarity is described by attributes like 'sharp', 'detailed', and 'resolution'. This component seems to describe the kind of artifacts that people perceive. The second dimension can be identified from the item distribution in Figure 4. Content with a high amount of movement (Car, Soccer2, and Butterfly) is located on positive polarity, more static content (Bullinger, Mountain, Horse) is located on the negative one. Concluding, perceivable artifacts and movement, or temporal changes in the content, are the most important components to explain the overall quality rationale.

Interestingly, there is no model component that describes depth perception. Figure 5 shows that 3D-related attributes like 'spacious', '3D reality', or 'background depth' correlate with the positive polarity of component 1. No depth-related attributes can be found on its negative polarity. This finding goes along with results of other studies.[20] Depth impression only seems to contribute to quality if artifacts are low. If the video quality is low due to coding artifacts, then this quality degradation will exceed the additional value provided by the stereoscopic video presentation. At last, it must be mentioned that sensory profile does not differentiate different coding structures. Perceived quality of different coding methods seems to be dependent on a combination of the content, its characteristics, and the existing artifacts. The final step of analysis lets us combine users' quality preference and the sensory profile using EPM. The arrows in Figure 4 mark the assessors' preferences. Expectedly , it shows a clear preference structure for artefact-free content. The best-rated sequences (c.f. Figure 2 and Figure 3) highly correlate with component 1. Least preferred items are all Bullinger clips at the opposite side of the arrows.

6. DISCUSSION AND CONCLUSION

In our study we evaluated coding methods for mobile 3D television and video. The first goal of the study was to find an optimized coding method. Four different methods were chosen for evaluation. The results show that Multiview Coding and Video+Depth provide the best overall quality. The two methods represent contrary methods in the coding of 3D video. While MVC uses inter- and intraview dependences of the two video streams (left and right eye), the Video+Depth approach renders virtual videos from a given view and its depth map. Our results show that the performance of the coding methods strongly depends on the content and its characteristics. In overall, there is no siginificant difference. A second goal of the study was the application of sensory profiling. The use of OPQ method allowed us to collect individual quality factors. With the help of the profiles we were able to understand rationale used to evaluate experienced quality. The results show that for 3D video artifacts are still the determining quality factor. Expected added value through depth perception was rarely mentioned by the test participants. When it was mentioned then the profile shows that it is connected to artifact-free videos. These results are inline with previous studies.[9, 20] Depth perception and artifacts are both determining 3D quality perception.[12] But in contrast to Seuntiens' model,[12] our profiles show that it is a hierarchical dependency. Only if artifacts are at a low level, then depth perception seems to contribute to the added value of 3D video and television.

The results show that 3D quality and its perception depends on several factors that interact with each other. Importance of these factors (video quality, depth) seems to change depending on the provided video quality. The expected added value provided by the depth impression only enhances users' quality perception when artifacts are low in the presented material. Sensory Profiling techniques such as the presented Open Profiling of Quality approach are a promising way to be able to detect these quality models holistically by giving the test participants full freedom in describing their quality sensation.

ACKNOWLEDGMENTS

Dominik Strohmeier would like to thank Katja Eulenberg, Klemens Göbel, Sebastian Kull, and Sören Nagel for their support in conducting the tests. Thanks goes also to Satu Jumisko-Pyykkö and Karsten Müller for their fruitful advices in test design. The authors would like to thank KUK Filmproduktion (Horse, Car: http://www.kuk-film.de/) and Electronics and Telecommunications Research Institute (ETRI) (Mountain, Soccer2: http://www.etri.re.kr/) for providing stereoscopic content. The Butterfly sequence was rendered from 3D models provided on www.bigbuckbunny.org (copyright by Blender Foundation). MOBILE3DTV project has received funding from the European Communitys ICT programme in the context of the Seventh Framework Programme (FP7/2007-2011) under grant agreement n^o 216503. The text reflects only the authors views and the European Community or other project partners are not liable for any use that may be made of the information contained herein.

REFERENCES

[1] Jumisko-Pyykkö, S., Weitzel, M., and Strohmeier, D., "Designing for user experience: what to expect from mobile 3d tv and video?," in [*UXTV '08: Proceedings of the 1st international conference on Designing interactive user experiences for TV and video*], 183–192, ACM, New York, NY, USA (2008).

[2] Gotchev, A., Smolic, A., Jumisko-Pyykkö, S., Strohmeier, D., Akar, G. B., Merkle, P., and Daskalov, N., "Mobile 3d television: development of core technological elements and user-centered evaluation methods toward an optimized system," *Multimedia on Mobile Devices 2009* **7256**(1), 72560J (2009).

[3] Jumisko-Pyykkö, S. and Strohmeier, D., "Report on research methodologies for the experiments," *Technical Report Mobile3DTV* (2009).

[4] Recommendation ITU-R BT.500-11, *Methodology for the Subjective Assessment of the Quality of TelevisionPictures, Recommendation ITU-R BT.500-11.* ITU Telecom. Standardization Sector of ITU (2002).

[5] Recommendation ITU-T P.910, *Subjective video quality assessment methods for multimedia applications, Recommendation ITU-T P.910.* ITU Telecom. Standardization Sector of ITU (September 1999).

[6] Winkler, S. and Faller, C., "Audiovisual quality evaluation of low-bitrate video," *Human Vision and Electronic Imaging X* **5666**(1), 139–148, SPIE (2005).

[7] Jumisko-Pyykkö, S. and Häkkinen, J., "Evaluation of subjective video quality of mobile devices," in [*MULTIMEDIA 05: Proceedings of the 13th annual ACM international conference on Multimedia*], 535–538, ACM, New York, NY, USA (2005).

[8] Jumisko-Pyykkö, S., "'I would like to see the subtitles and the face or at least hear the voice': Effects of picture ratio and audio—video bitrate ratio on perception of quality in mobile television," *Multimedia Tools Appl.* **36**(1-2), 167–184 (2008).

[9] Stelmach, L., Tam, W. J., Meegan, D., and Vincent, A., "Stereo image quality: Effects of mixed spatio-temporal resolution," *IEEE Transactions on Circuits and Systems for Video Technology* **10**, 188–193 (2000).

[10] Meesters, L., IJsselsteijn, W., and Seuntiens, P., "A survey of perceptual evaluations and requirements of three-dimensional tv," *Circuits and Systems for Video Technology, IEEE Transactions on* **14**, 381–391 (March 2004).

[11] IJsselsteijn, W., de Ridder, H., and Vliegen, J., "Subjective evaluation of stereoscopic images: effects of camera parameters and display duration," *Circuits and Systems for Video Technology, IEEE Transactions on* **10**, 225–233 (Mar 2000).

[12] Seuntiens, P., "Visual experience of 3d tv," *PhD Thesis* (2006).

[13] Lambooij, M., IJsselsteijn, W., Fortuin, M., and Heynderickx, I., "Visual discomfort and visual fatigue of stereoscopic displays: A review," *Journal of Imaging Science and Technology* **53**(3), 030201 (2009).

[14] Häkkinen, J., Pölönen, M., Takatalo, J., and Nyman, G., "Simulator sickness in virtual display gaming: a comparison of stereoscopic and non-stereoscopic situations," *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services* , 227–230 (2006).

[15] Lambooij, M., Fortuin, M., Ijsselsteijn, W. A., and Heynderickx, I., "Measuring visual discomfort associated with 3d displays," *Stereoscopic Displays and Applications XX* **7237**(1), 72370K, SPIE (2009).

[16] Jumisko-Pyykkö, S., Häkkinen, J., and Nyman, G., "Experienced Quality Factors Qualitative Evaluation Approach to Audiovisual Quality," *Proceedings of the IS&T/SPIE 19th Annual Symposium of Electronic Imaging* , Convention Paper 6507-21 (2007).

[17] Radun, J., Leisti, T., Häkkinen, J., Ojanen, H., Olives, J.-L., Vuori, T., and Nyman, G., "Content and quality: Interpretation-based estimation of image quality," *ACM Trans. Appl. Percept.* **4**(4), 1–15 (2008).

[18] Stone, H. and Sidel, J. L., [*Sensory evaluation practices: Food science and technology*], vol. 4th edition, Elsevier Academic Press., Boston, USA (2004).

[19] Häkkinen, J., Kawai, T., Takatalo, J., Leisti, T., Radun, J., Hirsaho, A., and Nyman, G., "Measuring stereoscopic image quality experience with interpretation based quality methodology," *Image Quality and System Performance V* **6808**(1), 68081B, SPIE (2008).

[20] Strohmeier, D., Jumisko-Pyykkö, S., and Kunze, K., "New, lively, and exciting or just artificial, straining, and distracting – a sensory profiling approach to understand mobile 3d audiovisual quality," *Proceedings of the Fifth International Workshop on Video Processing and Quality Metrics for Consumer Electronics - VPQM 2010* (2010).

[21] ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG-4 AVC), ITU-T and ISO/IEC JTC 1, *Advanced Video Coding for Generic Audiovisual Services* (November 2007).

[22] ISO/IEC JTC1/SC29/WG11, *Text of ISO/IEC 14496-10:200X/FDAM 1 Multiview Video Coding.* Doc. N9978, Hannover, Germany (July 2008).

[23] ISO/IEC JTC1/SC29/WG11, *ISO/IEC CD 23002-3: Representation of auxiliary video and supplemental information.* Doc. N8259, Klagenfurt, Austria (July 2007).

[24] Brust, H., Smolic, A., Müller, K., Tech, G., and Wiegand, T., "Mixed resolution coding of stereoscopic video for mobile devices," *3DTV Conference* (2009).

[25] Fehn, C., Kauff, P., Cho, S., Kwon, H., Hur, N., and Kim, J., "Asymmetric coding of stereoscopic video for transmission over t-dmb," *3DTV Conference* (May 2007).

[26] Bech, S. and Zacharov, N., [*Perceptual Audio Evaluation – Theory, Method and Application*], Wiley, Chichester, England (2006).

[27] Jumisko-Pyykkö, S., Kumar Malamal Vadakital, V., and Hannuksela, M., "Acceptance threshold: Bidimensional research method for user-oriented quality evaluation studies.," *International Journal of Digital Multimedia Broadcasting* **2008**, Hindawi Publishing Corporation (2008).

[28] Atzpadin, N., Kauff, P., and Schreer, O., "Stereo analysis by hybrid recursive matching for real-time immersive video conferencing," *IEEE Trans. Circuits Syst. Video Techn.* **14**, 321–334 (March 2004).

[29] Merkle, P., Wang, Y., Müller, K., Smolic, A., and Wiegand, T., "Video plus depth compression for mobile 3D services," *3DTV Conference* (2009).

[30] Gower, J., "Generalized procrustes analysis," *Psychometrika* **40**(1), 33–51 (1975).

[31] Schlich, P., "Preference mapping: Relating consumer preferences to sensory or instrumental measurements," *Bioflavour'95* , INRA Editions (2005).