USER-CENTERED QUALITY OF EXPERIENCE: IS MOBILE 3D VIDEO GOOD ENOUGH IN THE ACTUAL CONTEXT OF USE?

Satu Jumisko-Pyykkö, Timo Utriainen
Tampere University of Technology

ABSTRACT

To become successful, 3D video on mobile devices needs to provide higher constructed quality than the existing systems and function properly in the heterogeneous mobile contexts of use. The goal of the paper is to explore how video quality is experienced under the different presentation modes (2D/3D). Our data-collection procedure applies a hybrid User-Centered Quality of Experience (UC-QoE) evaluation approach by combining 1) quantitative preference ratings, 2) qualitative descriptions of quality, 3) situational audio/video data-collection, and 4) conducting a study in a laboratory and dynamic field settings. The experiments are conducted on a portable device using parallax barrier display technology and simulcast stereo video encoding with relatively low total bitrates relevant for mobile broadcasting. We vary presentation mode (2D/3D), video, and audio bitrates. The results showed the superiority of 2D visual presentation mode over the 3D mode in all used assessment tasks. To improve constructed 3D video quality on mobile devices, further work needs to target on the factors of ease of viewing, viewing comfort and spatial excellence on top of added depth perception to provide good viewing experiences in the heterogeneous contexts of use.

1. INTRODUCTION

3D video is on the way to mobile devices. We refer to 3D as a system which presents video using a stereoscopic presentation mode providing a different image for each eye. To become successful, it does not only need to fill the user's requirements and needs, but also to make sure that the critical system components, such as constructed quality, is high enough to outperform the existing quality level. To date, there are only a limited number of studies published comparing the subjective visual quality between different presentation modes (2D and 3D) on the mobile screen sizes [6], [15]. There is no previous work done in exploring how quality of 3D video is experienced on mobile devices in the real-life-like settings. Even the work on contextual 2D quality on mobile devices is very limited [8], [13].

The goal of the paper is to explore how video quality is experienced under the different presentation modes (2D/3D). We use a hybrid User-Centered Quality of Experience (UC-QoE) evaluation approach. The experiments are conducted on a portable device using parallax barrier display technology and simulcast stereo video encoding with relatively low total bitrates relevant for mobile broadcasting. We vary the presentation mode (2D/3D), video and audio bitrates, and conduct the experiments in both controlled and field settings (home-like, travel-bus, waiting-station).

2. RESEARCH METHOD

The study contained two parts, referred as Experiment1=E1 and Experiment2=E2.

Participants: A total of 60 participants (aged 18-45, 50% male/50% female) participated to the study. The participants were divided equally between the experiments.

Procedure: The pre-test session contained sensorial tests (visual (20/40) and stereovision acuity (.6), color vision), demographic data collection and combined anchoring and training where the extremes of quality samples and all contents were presented.

The bidimensional research method of acceptance was used [11]. The stimuli were presented one by one and rated independently and retrospectively [7]. After each clip, participants marked retrospectively the overall quality satisfaction score using a discrete unlabeled scale from 0 to 10 and the acceptance of quality for viewing mobile 3DTV on a binary (yes/no) scale. The evaluation task was repeated in the different evaluation contexts (E1 contained two and E2 three contexts).

The post-test session contained an interview of quality experiences. In E1, we also used a stimuli-assisted description task in which the participant was shown a set of stimuli one by one in random order and after each clip they verbally described the factors they paid attention to.

Test material: The stimuli represented the characteristics of potential contents for mobile 3D TV

[12] and contained a proportion of audiovisual characteristics chosen based on six expert evaluations (moderate inter-rater reliability: *Cohen's kappa* = 0.62 [11]; Table 1). The length of stimuli is approx. 35s.

Table 1 Stimuli content descriptions and audiovisual characteristics.

Screenshot	Genre, content description and audiovisual characteristics
V_{SD}=visual spatial details, V_{TD}=temporal motion, V_D=amount of depth, V_{DD}=depth dynamism, V_{SC}=amount of scene cuts, A=audio characteristics	
	**Animation** - Knight's Quest 4D (37s) The knight walks into a trap while trying to rescue another knight and falls down a shaft, but manages to get to the other side safely using a grappling rope. V_{SD}: high, V_{TD}: high, V_D: med, V_{DD}: high, V_{SC}: high, A: music, effects
	**User-created** - Liberation of Plzen (37s) A commemorative parade of the liberation of Plzen. Scenes with marching troops and army vehicles with ambient noise and music played by the marching band. V_{SD}: med, V_{TD}: low, V_D: med, V_{DD}: high, V_{SC}: low, A: music
	**Documentary** - Upper Middle Rhine Valley (34s) The clip starts with a panning camera over a river valley. Scenes of a vineyard with people gathering grapes and a man pouring gathered grapes into a container. V_{SD}: low, V_{TD}: med, V_D: high, V_{DD}: low, V_{SC}: low, A: music
	**Series** - Virtual Visit to Suomenlinna (32s) Scenes from the Suomenlinna fortress. A boy and a woman walk in a scene with a tree in the foreground. In the last scene they run up a grassy hill. V_{SD}: low, V_{TD}: low, V_D: med, V_{DD}: med, V_{SC}: low, A: speech, music

Parameters: The chosen parameter combinations vary presentation mode, video and audio bitrates. They are listed in a form of video bitrate/framerate/audio bitrate – presentation mode:

E1: 160kbps/15fps/48kbps – 2D and 3D modes
320kbps/15fps/48kbps – 2D and 3D modes
768kbps/15fps/48kbps – 3D mode
E2: 320kbps/15fps/18kbps – 2D and 3D modes
320kbps/15fps/48kbps – 2D and 3D modes

The parameters were chosen according to previous studies (e.g. [5], [4], [13]) conducted on standard mobile TV where video bitrates varied between 100kbps and 500kbps and framerates between 6 and 25 frames per second. 15 fps is commonly used in present day mobile television networks [1]. 48kbps audio for AAC-HEv2 has been deemed sufficient [3] and 18kbps was chosen for the low bandwidth scenario with moderately easily detectable encoding artifacts. 320kbps was chosen for the base video bitrate since it's commonly used in DVB-H networks [1] and is also tied to basic-speed 386kbps 3G mobile phone networks [2]. 160kbps was selected to be a clearly low bandwidth scenario with moderately easily detectable encoding artifacts with most of the chosen contents. All video bitrates are for the entire video, i.e. including both video channels with 3D content to compare quality of conventional 2D mobile television broadcasts to 3D stereovideo simulcast broadcasts. 2D was encoded so that it received the same total bitrate as the two 3D channels combined, as this is the way conventional 2D broadcasts would be done. Video bitrate of 768kbps was selected for 3D to provide the comparable bandwidth per video channel to the 2D mode.

Encoding procedure: The video codec, x264 'Skystrife' b1077 using H.264/AVC baseline profile and the audio codec, Nero AAC 1.3.3.0 with AAC-HEv2 in stereo mode using normalized volume were used for encoding. The clips were produced using meGUI 0.3.1.1010 via Avisynth 2.5.7 scripting to avoid intermediate steps that could hinder produced quality. The clips were encoded into a widescreen 16:9 letterbox resolution of 640px x 360px. For 3D the clips were squeezed horizontally to half their width and placed side-by-side for encoding. The 2D clips used the left video channel. Audio was adjusted to 75 dBA (+10 dBA for peaks) and presented using in-ear-type headphones Sony (MDR-E818LP). The stimuli material was presented centered on the screen covering a 3.3 inch area in diameter and was played using VLC 0.8.6 video player in full-screen mode. The video clips were presented on a Telson's mobile 3D device with a 4.3 inch 800px x 480px touch-enabled transmissive autostereoscopic parallax barrier screen.

Context of use: Context of use is defined as: *'Context represents the circumstances under which the activity (--) takes place'* [10]. According to the model of Context of Use in Human-Mobile-Computer Interaction (CoU-MHCI) [10], the context of use contains five context components: 1) physical, 2) temporal, 3) task, 4) social, and 5) technical and information context, their subcomponents and properties: 1) magnitude, 2) dynamism, 3) patterns and 4) typical combinations. The controlled laboratory evaluation (set up according to [7]) was used in both studies parallel to the evaluations in the context of use. The selection of contexts was based on the most common cases of user requirements for mobile (3D) TV [12]. The central components relevant to this study in each context of use are described in Table 2 following the CoU-MHCI model.

The presentation order of all stimuli was fully randomized. The order of evaluation contexts was fully randomized in E1 and partly in E2 (starting either at the station or in the lab).

Table 2 Central characteristics of evaluation contexts of use. The presentation follows model CoU-MHCI [10], in the table [D] denotes a dynamic property of context.

Components/ properties	Lab	Home	Bus	Station
Physical context				
Functional place	Laboratory conditions [7]	Simulated home	Local bus	Café, railway station
Sensed attributes (Audio, Visual)	A: quiet V: calm, indoor	A: quiet V: calm, indoor	A: noisy [D] V: noisy, light [D]	A: noisy [D] V: noisy, light [D]
Movements (Movement, Position)	M: * P: straight	M: none P: lean	M: bus [D] P: lean [D]	M: none P: lean [D]
Artifacts (other than answer sheet)	*	Accessories (e.g. pillows)	Bag	Bag, refreshments
Temporal context				
Duration	30min	30min, macro	25min, macro	25min, macro
Time of day	Vary	Vary	Vary	Vary
Actions-time	*	Extra time	Extra time, pressure at end	Extra time, pressure at end
Task context				
Multitask 1	Quality evaluation	Quality evaluation	Quality evaluation	Quality evaluation
Multitask 2	*	Relax	Search bus-stop (visual)	Check time for train (audio, visual)
Interruptions	*	None	Physical and social [D]	Physical and social [D]
Task type	*	Entertain	Entertain	Entertain
Social context				
Persons present	Moderator	None	Bystanders, moderator [D]	Bystanders, people near, moderator [D]
Interpersonal actions	*	*	Possible	Possible
Technical and informational context				
Other systems	*	*	Mobile usability lab	Mobile usability lab
Properties				
Level of dynamism	Static	Static	Dynamic (physical, task, social)	Dynamic (physical, task, social)
Other related factors				
Motivations	*	Entertain, pass time, relax	Entertain, pass time	Entertain, pass time
Viewing distance	Fixed (40cm, 10x picture height [13])	Freedom to adjust	Freedom to adjust	Freedom to adjust
Device volume	Fixed (75dBA [7])	Freedom to adjust	Freedom to adjust	Freedom to adjust

3. RESULTS

1. Preference ratings of acceptance and satisfaction – The studied parameter combinations had a significant influence on acceptance and satisfaction ratings in both experiments. The 2D parameter combinations were the only to provide acceptable quality (E1: 92.1%, E2: 83.1% of acceptance) while all 3D parameter combinations were below the 50% acceptance threshold (E1: 40.4%, E2: 46.1% of acceptance). The most satisfying quality was presented with the 2D presentation mode in all contents and contexts of use (Wilcoxon pairwise comparisons p<.001; Figure 1). The result also show that the significant increase in video bitrate (E1: 768kbps) in 3D does not improve the quality to reach the level of 2D (p<.001). The bus context had 15 participants, while the rest had 30.

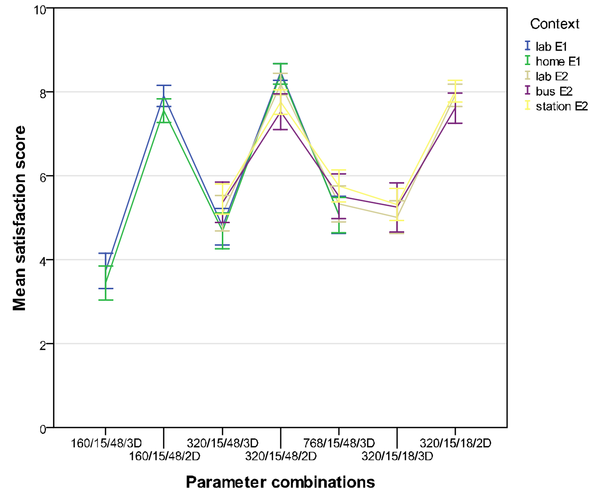

Figure 1 Mean satisfaction scores for 2D and 3D parameter combinations in all studied contexts of use averaged over all contents. Error bars show 95% CI.

2. Descriptive experienced quality (E1) – Analysis of transcribed interviews was conducted following the principles of Grounded Theory framework through systematical steps of open coding, concept development and categorizing [16].

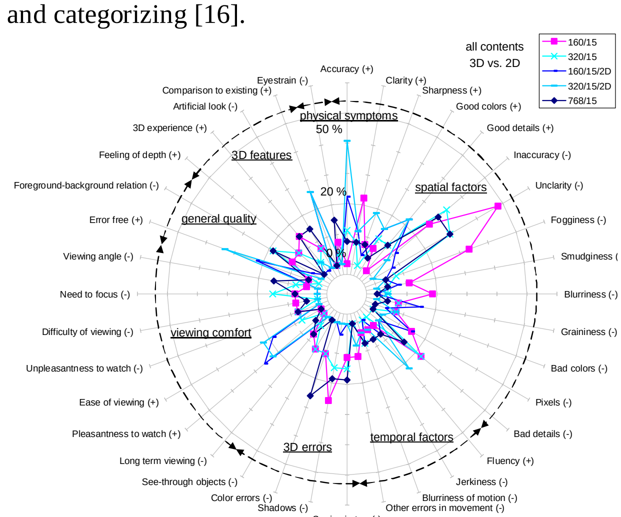

Figure 2 Descriptions for experienced quality in the stimuli-assisted description task with the percent of participants mentioning the category. Figure lists only categories mentioned by at least 10% of the participants. (+) denotes a positive category, (-) a negative.

Subcategories were further categorized into main categories: spatial factors, temporal factors, 3D errors, viewing comfort, general quality, 3D features and physical symptoms (Figure 2). The descriptions for the four different stimuli contents were combined.

Videos shown in 2D presentation mode were most commonly associated to the experience of positive quality (Figure 2). They were described with *accuracy, good colors, good details, pleasantness to watch, ease of viewing* and *error-freeness*. In contrast, videos shown in 3D were described with both positive mentions of *feeling of depth, 3D experience, clarity* and negative mentions of *inaccuracy, unclarity, fogginess, bad details, shadows, seeing in two, need to focus* and *foreground/background relation.*

The participants descriptions illustrates these key factors:

-- *reflections [bothered the most], or whatever you could call them, when you saw objects in two or three, it really started to put a strain on your eyes (male, 30)*

-- *[I liked when] the picture was sharp, clear, with good colors. So that the picture was free of errors (male, 39)*

-- *[I didn't like] when the picture was unclear, or when it moved the picture went blurry (female, 44)*

3. Adjustments of viewing conditions (E2) – User's movements in relation to the mobile device were analyzed based on data of the mobile usability laboratory containing a portable multi-video camera system [13] from a total of 12 participants. We coded the video material with an accuracy of 1s to identify e.g. gaze position, user's movements and actions, and changes in physical context. The adjustments of viewing conditions were coded as actions where the movement of head and/or device were done in parallel to viewing the stimuli.

Participants moved less to adjust the viewing conditions when 2D presentation mode was used compared to the 3D presentation mode (averaged over contents and contexts, Wilcoxon: Z=-4.01 $p<.001$: 2D mean=2.54 SD=2.07; 3D mean=4.35 SD=3.30). For both presentation modes, a significantly higher number of adjustments appeared in the real life contexts compared to the laboratory conditions (lab vs. others $p<.05$; bus vs. station $p>.05$; Figure 3).

4. Summary of results – Table 3 summarizes the results gained from the different methods for 2D and 3D presentation modes for both experiments.

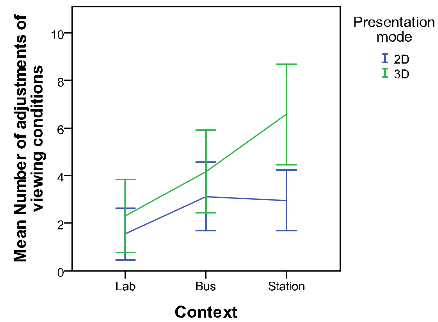

Figure 3 Number of adjustments of viewing conditions per video in different contexts for 2D and 3D presentation modes. Error bars show 95% CI.

Table 3 Summary of results for 2D and 3D presentation modes for both experiments

Results	2D presentation mode	3D presentation mode
Quantitative preference ratings of satisfaction and acceptance (E1, E2)		
	- Quality satisfaction high - Acceptable level of quality - Video: 320kbps better than 160kbps	- Quality satisfaction: low - Unacceptable quality level - Audio: no influence of bitrates
Qualitative descriptive experienced quality (E1)		
	- Mostly positive descriptions - Accuracy, good colors, good details; pleasantness to watch, ease of viewing; error free (+)	- Positive / negative descriptions - Feeling of depth, 3D experience, clarity (+) - Inaccuracy, unclarity, fogginess, bad details; shadows, seeing in two; need to focus; fore-/background relation (-)
Adjustments of viewing conditions (E2)		
	- Few	- Several, especially in the field

4. DISCUSSION AND CONCLUSIONS

The goal of the paper was to investigate how the video quality is experienced under the different visual presentation modes. The superiority of 2D mode was shown in all three measures we used. Firstly, the 2D mode provided the most pleasant and highly acceptable quality while the evaluations of 3D were below the 50% acceptance threshold. Strohmeier et al. [17] have also shown the superiority of 2D on the small display sizes. In their study, the 3D mode reached 70% of acceptance of quality showing a significantly higher quality level compared to our results. The difference might be explained by the different display technologies used (lenticular sheet vs. parallax barrier). If the display technology is a critical factor of the end-to-end chain of system, it can be suggested that comparisons of system factors (from capture to encoding and transmission) should employ at least two display technologies to be able to draw strong display independent conclusions.

The superiority of 2D also appeared in the qualitative descriptive experiences of quality such as good details, accuracy, pleasantness to watch, ease to view, error freeness. 3D was attached positively to *feeling of depth*, *3D experience* and *clarity*, and negatively to erroneousness such as *spatial factors*, *3D impairments* and *hardness of viewing*. Our descriptive results are in line with recent studies [6], [15], [17]. For example, Strohmeier et al. [17] described that 2D quality is described as *pleasant*, *focusable image*, *sharpness*, *naturalness*. In contrast, impaired 3D was described as *blurred*, *unstable*, *stressful* and *restless*. In the case of successful 3D, the impressions are described as *brilliant*, *layered*, *spatial*, *depth and its sharpness*. Taken together, challenges in improving 3D quality in relation to 2D on small displays needs to target on spatial excellence of 3D, viewing comfort and ease of viewing.

Our situational data, collected in three contexts of use, revealed that viewing using 3D presentation mode required a significantly higher number of adjustments for viewing conditions compared to 2D. In the real-life-like contexts, more adjustments were needed compared to the controlled settings. Face down viewing position, variable light, vibration circumstances and actively divided attention between viewing and surrounding are a central part on the context of viewing for mobile devices in natural circumstances [9]. This result indicates that ease of viewing and gaining of the sweet-spot for viewing are the challenges for the system in the actual contexts of use.

To conclude, the constructed 3D video quality on mobile devices needs to guarantee ease of viewing, viewing comfort, spatial excellence (accuracy, error-freeness) on top of the added depth experience to provide good viewing experiences in the heterogeneous contexts of use.

5. ACKNOWLEDGEMENTS

This work is supported by the European Commission within the ICT program of FP7 under Grant 216503 with the acronym MOBILE3DTV(http://www.mobile3dtv.eu/). The authors would like to thank Cinovent (Upper Middle Rhine Valley), Red Star Studio (Knight's Quest 4D), Stereoscape (Virtual Visit to Suomenlinna) and Centre of Computer Graphics and Visualization from the University of West Bohemia (Liberation of Plzen) for their contribution as stereoscopic content providers.

6. REFERENCES

[1] DVB-H.org. http://www.dvb-h.org/, accessed 12/2008.

[2] European Broadcasting Union (EBU), "Digital Video Broadcasting (DVB): DVB-H Implementation Guidelines". European Telecommunications Standards Institute. ETSI TR 102 377 V1.2.1 (2005-11), 2005.

[3] European Broadcasting Union (EBU). "Subjective listening tests on low-bitrate audio codecs". *Tech 3296*, June 2003.

[4] Faria, G., Henriksson, J., Stare, E., and Talmola, P. "DVB-H: Digital Broadcast Services to Handheld Devices". *Proc of the IEEE*, Vol. 94, No. 1, January 2006.

[5] Gulliver, S., and Ghinea, G. "Defining User Perception of Distributed Multimedia Quality". *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*. Volume 2, Issue 4 (November 2006), pp. 241-257, 2006

[6] 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". *Proc. of the IS&T/SPIE's Electronic Imaging 2008*, Vol.6808,pp.68081B-68081B-12. doi:10.1117/12.760935

[7] ITU-T P.911 Recommendation, "Subjective audiovisual quality assessment methods for multimedia applications". International Telecommunication Union – Telecommunication Standardization Sector, 1998.

[8] Jumisko-Pyykkö, S., and Hannuksela, M.M. "Does Context Matter in Quality Evaluation of Mobile Television?". *Proceedings of MobileHCI '08*.

[9] Jumisko-Pyykkö, S., and Utriainen, T. "D4.4 v2.0 Results of the user-centered quality evaluation experiments". *Technical report*, November 2009.

[10] Jumisko-Pyykkö, S., and Vainio, T. "Framing the Context of Use for Mobile HCI", *International Journal of Mobile-Human-Computer-Interaction (IJMHCI)*, in press.

[11] Jumisko-Pyykkö, S., Kumar Malamal Vadakital, V., and Hannuksela, M.M. "Acceptance Threshold: Bidimensional Research Method for User-Oriented Quality Evaluation Studies". *International Journal of Digital Multimedia Broadcasting*, 2008.

[12] Jumisko-Pyykkö, S., Weitzel, M., and Strohmeier, D. "Designing for User Experience: What to Expect from Mobile 3D TV and Video?". Proc of 1st UXTV conference 2008

[13] Knoche, H., Sasse, M. A. (2008) The sweet spot: How people trade off size and definition on mobile devices. In *Proceedings of ACM Multimedia 2008*

[14] Oulasvirta, A., and Nyyssönen, T. "Flexible Hardware Configurations for Studying Mobile Usability", *Journal of Usability Studies*, Volume 4, Issue 2, Feb 2009, pp. 93-105

[15] Shibata, T., Kurihara, S., Kawai, T., Takahashi, T., Shimizu, T., Kawada, R., Ito, A., Häkkinen, J., Takatalo, J., and Nyman, G. "Evaluation of stereoscopic image quality for mobile devices using interpretation based quality methodology". *Proc. SPIE 7237*, *72371E* (2009), doi:10.1117/12.807080

[16] Strauss, A., and Corbin, J. *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.), Thousand Oaks, CA: Sage, 1998.

[17] 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". *Proc of VPQM 2010*.