EXPERIENCED AUDIOVISUAL QUALITY FOR MOBILE 3D TELEVISION

Timo Utriainen, Satu Jumisko-Pyykkö

Tampere University of Technology

ABSTRACT

For mobile 3D television, substantial optimization of system resources is needed to provide adequate quality to the viewers while sparing valuable and limited system resources. In the end, the experienced quality of 3D needs to outperform the quality of existing services. The goal of this paper is to explore the influence of audiovisual encoding parameters on viewers' experienced quality on mobile 3D television. We conducted two extensive subjective quality evaluation experiments where presentation modes (2D/3D), framerates, and video and audio bitrates with several content types were varied. The experiments were carried out on a portable autostereoscopic device using parallax barrier display technology utilizing simulcast stereo video encoding with relatively low total bitrates relevant for broadcasting onto mobile devices. The results showed the superiority of 2D presentation mode, importance of visual over audio quality and that a significant increase in bitrate-framerate resources did not improve the visual quality of 3D. Further work needs to address several display techniques as a part of quality evaluation studies to provide reliable comparisons of critical system factors.

Index Terms — Audiovisual quality, experienced quality, 3D, autostereoscopic, three-dimensional displays, mobile television

1. INTRODUCTION

3D video is making its breakthrough from big movie screens onto mobile devices. We refer to a 3D system as a system which presents video using a stereoscopic presentation mode providing a different image for each eye. To become successful, the novel system 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. Subjective quality evaluation experiments are conducted for optimizing critical system components during the process of system development. In mobile 3D television, huge amounts of audiovisual 3D data, limited bandwidth, vulnerable transmission channel, and constraints of the receiving devices (such as screen size, computational power and battery-life) set tight boundaries for optimization of system resources. The ultimate goal of the optimization is to ensure the system under development fills the viewers' requirements without wasting valuable and limited system resources.

There are few previous studies exploring the influence of depth on user' experienced quality on small mobile screens. Shibata et al. [1] carried out the comparison between 2D and 3D presentation modes on a mobile device and 3D on a large screen with image contents. The results showed the superiority of 3D over 2D on small displays while these experiences were relatively modest compared to the large screens. In addition, 3D accompanied to the presentation on a small display can contribute to visual fatigue [1]. Later, the excellence of 3D over 2D has also been shown with video in the case where the content is free from visually detectable artifacts [2]. These studies indicate that 3D on mobile screens can enhance the user's experienced quality.

To optimize the critical system components for mobile 3D television and video, different video encoding methods have been compared by Strohmeier & Tech [3]. In the current datarates of mobile television, the MVC and video+depth slightly outperformed simulcast while MRSC provided clearly the lowest subjective quality. The authors also concluded that visible stereoscopic artifacts have a strong negative influence on experienced quality and they can override the added value of depth.

Although the recent studies have concentrated on visual quality of 3D on mobile devices; services, such as mobile television, are expected to have two modalities – audio and video. It is known that audiovisual quality is more than the simple sum of quality derived from one channel [4] and visual quality can influence on audio quality and vice versa [5]. The previous subjective audiovisual quality studies carried out for mobile television have shown the need for the content dependent optimal share of audio and video resources, e.g. bitrates [6], [7]. For example, relatively higher need for visual quality is needed for high motion sport contents compared to audibly important news content [7], [4]. Strohmeier et al. [2] have studied the influence of audio (mono/stereo) and visual presentation modes (2D/3D) on perceived quality. In overall, their results showed the strong dominance of visual presentation over audio presentation modes and no interaction between the modalities. However, the study was conducted with extremely high video bitrates (tens of megabytes), outside the zone of applicable resources intended for broadcasting onto mobile devices. The previous studies to explore audiovisual quality have been mainly done with 2D video on mobile devices while the work on audiovisual 3D quality is limited.

Taken together, the focus of previous work has been mainly on visual quality of 3D video and comparisons of presentation modes for video while less attention has been paid on quality under the realistic technical resources and possibilities for multimodal optimization for mobile video or television. The aim of this paper is to explore the influence of audiovisual encoding parameters on experienced quality on mobile 3D television. We conduct two experiments on a portable device using parallax barrier display technology and simulcast stereo video encoding with relatively low total bitrates relevant for mobile broadcasting. We varied presentation mode (2D/3D), framerates, and video and audio bitrates, and conducted the experiments in the controlled laboratory circumstances.

978-1-4244-6378-7/10/$26.00 ©2010 IEEE

2. RESEARCH METHOD

The study contained two parts, referred to as Experiment 1 (E1) and Experiment 2 (E2).

Participants: 60 participants (aged 18-45, 50% male/50% female) took part in the study. They were divided equally between the two experiments.

Procedure: Sensorial tests (visual (20/40) and stereovision acuity (≤ 60 arcsec), color vision), demographic data collection and combined anchoring and training, where the extremes of quality samples and all contents were presented, were included in the pre-test session.

The bidimensional research method of acceptance was used [8]. The stimuli were presented one by one and rated independently and retrospectively [9]. 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 3D TV on a binary (yes/no) scale. Evaluation task was repeated in different evaluation contexts while this paper presents the results from the controlled laboratory settings.

Test material: The stimuli represented the characteristics of potential contents for mobile 3D TV [10] and contained a chosen proportion of audiovisual characteristics based on six expert evaluations (moderate inter-rater reliability: *Cohen's kappa* = 0.62 [8]; Table 1). The length of stimuli was approx. 35s each.

Table 1. Stimuli content 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 modes, framerates, and video and audio bitrates (Table 2). The parameters were chosen according to previous studies (e.g. [11], [12], [13]) conducted on conventional mobile TV where video bitrates varied between 100kbps and 500kbps and frame rates between 6 and 25 frames per second. 10 and 15 fps are commonly used in present day mobile television networks [14]. AAC-HEv2 with 48kbps audio bitrate has been deemed excellent [15] and 18kbps was chosen for the low bandwidth scenario with moderately easily detectable encoding artifacts. 320kbps was chosen for the base video bitrate for its common use in present day DVB-H networks [14], additionally it refers to basic-speed 386kbps 3G mobile phone networks [16]. For a low bandwidth scenario 160kbps was selected with moderately easily detectable encoding artifacts with most of the chosen contents. All video bitrates include the entire video stream, 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. To provide comparability however, additional higher video bitrates (768kbps and 1536kbps) were selected for 3D to provide at least the comparable bandwidth per video channel than what was used with 2D mode. The highest video bitrate (1536kbps) was selected to see how high-bandwidth on-demand services could benefit from higher available bitrates.

Table 2. Parameter combinations for experiments 1 (E1) and 2 (E2)

Parameter combinations	**E1**	**E2**
video bitrate/framerate/audio bitrate - presentation mode		
160kbps/10fps/48kbps - 3D	√	
160kbps/15fps/48kbps - 2D and 3D	√	
320kbps/10fps/48kbps - 3D	√	
320kbps/15fps/18kbps - 2D and 3D		√
320kbps/15fps/48kbps - 2D and 3D	√	√
768kbps/10fps/48kbps - 3D	√	
768kbps/15fps/18kbps - 3D		√
768kbps/15fps/48kbps - 3D	√	√
1536kbps/24fps/48kbps - 3D	√	

Encoding procedure: Video was encoded using x264 'Skystrife' b1077 utilizing the H.264/AVC baseline profile and audio was encoded using Nero AAC 1.3.3.0 with AAC-HEv2 in stereo mode using normalized volume prior to 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. The stimuli material was presented centered on the screen using VLC 0.8.6 video player in full-screen mode covering a 3.3 inch area in diameter on the screen. The video clips were presented on a Telson mobile 3D device with a 4.3 inch 800px x 480px touch-enabled transmissive autostereoscopic parallax barrier screen interleaved at pixel level.

Context of use: The tests were conducted in a controlled laboratory environment set up according to [9]. The presentation order of all stimuli was fully randomized.

3. RESULTS – EXPERIMENT 1

Acceptance of Quality – In overall, the stimuli shown using 2D presentation mode reached the 80% acceptance level in all contents while 3D stimuli were below the threshold of 50%. As an exception, user-created content provided in 3D mode reached the 50% acceptance threshold in four parameter combinations (320kbps/10fps; 768kbps/10fps and 15fps; 1536kbps/24fps). As the satisfaction ratings are used for the detailed analysis of quality preferences, we identified the acceptance threshold on the satisfaction scale [8]. Acceptable quality was connected to scores between 5.8 and 9.1 (Mean=7.4, SD=1.68) on satisfaction scale and unacceptable quality to scores between 1.5-5.6 (Mean=3.6, SD=2.05) with a significant difference between their distributions (χ^2 (10)=278.1, p<.001).

Quality satisfaction – Among all tested parameter combinations, the most satisfying quality was presented with 2D presentation mode when averaged across the contents (difference to others p<.001). The higher 2D bitrate (320kbps) outperformed the lower bitrate (160kbps) (p<.001). In the content by content examination, 2D presentation mode provided more satisfying quality than 3D presentation mode (p<.001) and the higher 2D bitrate (320kbps) outperformed the lower bitrate (160kbps) in all other contents (p<.001), except the series content (p<.05).

The bitrates had a significant impact on satisfaction scores (F_R=43.90, df=3, p<.001; Figure 1) when averaged over the content. The two highest bitrates (758kbps and 320kbps) provided equally the most satisfying quality (p>.05) over the lowest bitrate (160kbps; p<.001). This result appeared independently on used framerate. Content by content analysis follows this main tendency with the exception of documentary content where all bitrates were equally rated (p>.05).

The framerates had significant impact on satisfaction scores (F_R=49.45, df=2, p<.001). The framerates were equally evaluated in the high bitrates (p>.05). In the lowest bitrate, (160kbps) the 10fps was preferred over 15fps (p<.05).

The comparison to the maximum studied bitrate-framerate parameter combination reveals that the increase of these resources would not increase quality satisfaction. The maximum case was equally rated with 768kbps/10fps, 15fps and 320kbps/10fps (p<.05) parameter combinations and outperformed the other 3D parameter combinations when averaged over the contents (p<.05). Content by content analysis replicated this main result.

4. RESULTS – EXPERIMENT 2

Acceptance of Quality – Similarly to the experiment 1, the stimuli shown in 2D presentation mode reached the 80% acceptance level while the level of acceptance for 3D mode was mainly below the 50% threshold. The 3D parameters which made an exception were shown with animation (video: 768kbps, audio: 48kbps) and user-created contents (video: 320kbps and 768kbps, audio: 18kbps). The acceptance threshold was identified on the satisfaction scale. The satisfaction scores between 6.2 and 9.1 (Mean=7.6, SD=1.46) were experienced as acceptable while unacceptable quality was attached to scores between 2.5-5.8 (Mean=4.2, SD=1.65). The distributions between retrospectively rated satisfaction and acceptance differed significantly (χ^2 (10)=647.0, p<.001).

Quality Satisfaction – 2D presentation mode was experienced giving the most satisfying quality (p<.001) and this result is independent on the content. The combination 320kbps with 48kbps audio gave the most satisfying audiovisual experience (difference to 320kbps, 18kbps, 2D p<.05).

The studied video bitrates impacted on quality satisfaction scores when averaged across the contents (F_R=7.34, df=1, p<.01) while detailed pair-wise comparisons did not reveal any significant differences. Exceptionally, 768kbps was rated significantly (p<.05) higher than 320kbps when used with 18kbps audio for the series content.

The studied audio bitrates did not have a significant impact on quality satisfaction scores in 3D presentation mode (F_R=0.24, df=1, p>.05). Higher audio bitrate gave the most satisfying quality in the 2D presentation mode (p<.05). With documentary, 48kbps audio bitrate was rated significantly higher than 18kbps when used with 2D mode. For series content, 320kbps video bitrate with 48kbps audio bitrate was rated significantly (p<.05) higher than 18kbps when presenting with 3D mode.

5. DISCUSSION

The goal of this paper is to explore the influence of audiovisual encoding parameters on viewers' experienced quality on mobile 3D television. The results showed that with the used display technology (parallax barrier) the 2D presentation mode provided the more pleasant and highly acceptable quality compared to the 3D presentation mode. In overall, the provided quality of 3D was experienced as unacceptable. In the recent study of Strohmeier et al. [2] conducted using lenticular sheet display technology, the 3D presentation mode reached 70% of acceptance of quality showing a significantly higher quality level compared to our results. The difference might be explained by the excellence of different display technologies used (lenticular sheet vs. parallax barrier). Our previous work has shown also that with parallax barrier display technology it is challenging to find the optimal viewing conditions for 3D and the display can cause a higher level of visual fatigue compared to other display types [18], [20].

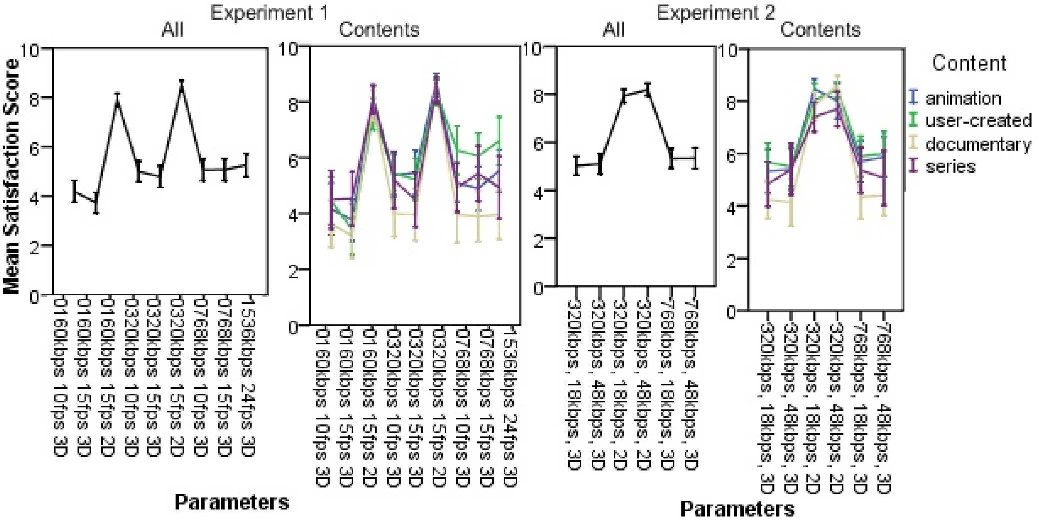

Figure 1. Influence of different parameter combinations on quality satisfaction in the experiment 1 and 2. The bars show 95% CI of mean.

Based on the results of the qualitative descriptive study for the same parameters, several perceivable stereoscopic impairments (e.g. seeing in two, color errors, see-through objects) were associated to 3D presentation independently on the used parameters [17]. This indicates that the presence of visible impairments can only not override the added value of depth, as suggested by previous studies [3], [19], but can also outweigh the differences in the levels of produced quality. For future work, it would be valuable to conduct the comparisons of system factors (from capture to encoding and transmission) using at least two display technologies to be able to draw strong display independent conclusions of the quality of system factors over the whole broadcasting chain.

Our results showed also the dominance of visual over audio quality under the tested parameters. There are three possible explanations to this result. Firstly, the used audio codec (AAC-HEv2) is especially efficient in maintaining adequate audio quality even at very low bitrates [15], which might have resulted in very small differences between the stimuli. Secondly, visual quality might act as the most changing variable collecting the greatest attention as suggested by the peak-end theory [21]. Thirdly, for the naive participants the differences in audio can be hard to detect under the parallel existence of visual video, as suggested in [2], [22]. Further work needs to understand the nature of bimodal interaction under relatively low quality circumstances, and the audio and visual factors need to be varied in spatial, temporal and depth domains. In addition, it would be worth of exploring the cross-modal influence when 3D on mobile devices is accompanied with very high-level audio (e.g. rendering multi-channel 5.1 audio set-up for the headphones) to make detectable enhancements to audio as well.

6. REFERENCES

[1] T. Shibata, S. Kurihara, T. Kawai, T. Takahashi, T. Shimizu, R. Kawada, A. Ito, J. Häkkinen, J. Takatalo, and G. Nyman, ”Evaluation of stereoscopic image quality for mobile devices using interpretation based quality methodology,” in *Proc. SPIE 7237, 72371E, Stereoscopic Displays and Applications XX*, San Jose, CA, USA, 2009.

[2] D. Strohmeier, S. Jumisko-Pyykkö, and K. Kunze, “New, lively, and exciting or just artificial, straining, and distracting? A sensory profiling approach to understand mobile 3D audiovisual quality,” in *Proc. 4th Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics VPQM*, Scottsdale, USA, Jan. 2010.

[3] D. Strohmeier, and G. Tech, “Sharp, bright, three-dimensional: open profiling of quality for mobile 3DTV coding methods,” in *Proc. Multimedia on Mobile Devices at Electronic Imaging 2010*, San Jose, California, USA, Jan. 2010.

[4] D. S. Hands, “A Basic Multimedia Quality Model,” *IEEE Trans. Multimedia*, vol. 6, no. 6, pp. 806-816, Dec. 2004.

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

[6] S. Winkler, and C. Faller, “Audiovisual quality evaluation of low-bitrate video,” in *Proc. SPIE/IS&T Human Vision and Electronic Imaging*, vol. 5666, San Jose, USA, pp. 139-148, 2005.

[7] S. Jumisko-Pyykkö, “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 and Applications*, vol. 36, no. 1-2, pp. 167–184, 2008.

[8] S. Jumisko-Pyykkö, V. K. Malamal Vadakital, and M. M. Hannuksela, “Acceptance threshold: bidimensional research method for user-oriented quality evaluation studies,” *Int. J. Digital Multimedia Broadcasting*, Volume 2008 (2008), Article ID 712380, 20 pages.

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

[10] S. Jumisko-Pyykkö, M. Weitzel, and D. Strohmeier, “Designing for user experience: what to expect from mobile 3D TV and video?,” in *Proc. 1st Int. Conf. Designing Interactive User Experiences For TV and Video*, Silicon Valley, California, USA, Oct. 2008, UXTV '08, vol. 291, ACM, New York, NY, pp. 183-192.

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

[12] G. Faria, J. Henriksson, E. Stare, and P. Talmola, “DVB-H: Digital Broadcast Services to Handheld Devices,” in *Proc. IEEE*, vol. 94, no. 1, Jan. 2006.

[13] H. Knoche, J. McCarthy, and A. Sasse, “Can small be beautiful? Assessing image resolution requirements for mobile TV,” in *Proc. 13th Annual ACM Int. Conf. Multimedia*, pp. 829-838, 2005.

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

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

[16] 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.

[17] S. Jumisko-Pyykkö, and T. Utriainen, ”D4.4 v2.0 Results of the user-centred quality evaluation experiments,” *MOBILE3DTV Technical report*, Nov. 2009. http://sp.cs.tut.fi/mobile3dtv/results/tech/D4.4_Mobile3DTV_v2.0.pdf

[18] S. Jumisko-Pyykkö, and T. Utriainen, “User-centered quality of experience of mobile 3DTV: How to evaluate quality in the context of use?,” in *Proc. Multimedia on Mobile Devices at Electronic Imaging 2010*, San Jose, California, USA, Jan. 2010.

[19] P.J.H. Seuntiëns, “Visual Experience of 3D TV,” *PhD thesis*, Eindhoven: Technische Universiteit Eindhoven, 2006.

[20] S. Jumisko-Pyykkö, T. Utriainen, D. Strohmeier, A. Boev, and K. Kunze, “Simulator sickness – Five experiments using autostereoscopic mid-sized or small mobile screens,” in *Proc. 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, submitted.

[21] B.L. Fredrickson, “Extracting meaning from past affective experiences: The importance of peaks, ends and specific emotions,” *Cognition and Emotion*, 14(4), pp. 577-606, 2000.

[22] W. R. Neuman, A. N. Cringler, and V. M. Bove, “Television sound and viewer perceptions,” *9th Int. Conference: Television Sound Today and Tomorrow*, Feb. 1991.