NEW, LIVELY, AND EXCITING OR JUST ARTIFICIAL, STRAINING, AND DISTRACTING
A Sensory profiling approach to understand mobile 3D audiovisual quality

Dominik Strohmeier¹, Satu Jumisko-Pyykkö², Kristina Kunze¹

¹ Ilmenau University of Technology, Germany
² Tampere University of Technology, Finland

ABSTRACT

Quantitative studies are commonly used to evaluate subjective quality of audiovisual multimedia systems. As a result these methods produce preference ratings for the stimuli under test. In this study, we examined influences of mono and stereo audio and visual presentation format on perceived overall quality and perceived depth. We used a mixed method approach combining psychoperceptual evaluation and qualitative attribute elicitation to get a more holistic understanding of 3D audiovisual quality.

1. INTRODUCTION

Mobile 3D television and video has gained in importance in the development of new mobile applications and services. The goal of the European research project MOBILE3DTV [1] is the development of core technologies to launch autostereoscopic television on mobile devices streamed over DVB-H. While mobile TV has been of interest in user-centered quality evaluation for a long time, only little knowledge exists about user requirements for mobile3DTV at the moment [6]. According to the results in [6], the added value of 3D seems to be a crucial requirement of prospective mobile 3DTV users. The key question is about positive perceived quality changes when switching from 2D to 3D. This added value of 3D has been targeted in several studies. As a basic assumption, stereoscopic 3D experience is seen as a combination of video quality, depth perception and visual comfort while watching [10]. Especially depth perception is assumed to be responsible to measure the added value of 3D. However, studies into stereoscopic video quality failed to show the added value of 3D induced by the additional depth perception. On the one hand, Ijsselsteijn et al. [11] showed that depth adds added value for uncompressed 3D images. On the other hand, Stelmach et al. [12] found that perceived overall quality is mainly depending on artifacts, not on depth perception, when artifacts are present in the stimulus material. This dilemma has lead into a discussion on how to measure the added value of 3D. Presence and Naturalness have been proposed Deeper evaluating this dilemma of added value and its measurement, Seuntiens concludes in his 3D experience model [10] that 3D visual experience can be measured best with the concept of Naturalness.

However, Seuntiens’ stereoscopic image quality model [10] excludes the impact of visual comfort from the 3D quality experience. As confirmed by several studies, viewing comfort or visual fatigue are crucial factors that need to be included into 3D experience [14]. Recent studies focus on the evaluation of impacting factors on visual comfort for stereoscopic perception [13]. Only one study by Häkkinen et al. [3] targeted comparison of stereoscopic and non-stereoscopic presentation on mobile screens focusing on simulator sickness symptoms in both cases.

Perceived quality has been evaluated mainly following standardized methods of the psychoperceptual approach [7]. These methods describe the quality in the terms of preference order but they do not try to explain the subjective interpretation of quality. The interpreted quality can help to identify the underlying structures, give explanations for quantitative ratings, and finally help to understand the phenomenon. These descriptive methods are especially useful when studying multimodal and heterogeneous stimuli [4].

Mixed Method research combines quantitative and qualitative data evaluation to compensate the weaknesses of one methodology with the strength of the other one. In our Open Profiling of Quality approach (OPQ), we combine psychoperceptual quality evaluation in line with the existing recommendations and sensory profiling methods that have been adapted mainly from food industry.

2. RESEARCH METHOD

2.1. Participants

45 test participants (13 female, 32 male; aged 15-30 - mean 24) took part in the psychoperceptual evaluation task. All test participants passed a screening for visual acuity, color and 3D vision, and hearing acuity. Among
sample, we further randomly selected 15 participants to conduct the sensory profiling task.

2.2. Test material and apparatus

Six different contents were used to create the stimuli under test (Table 1). The videos were selected according to criteria of spatial details, temporal resolution, amount of depth, and the user requirements for mobile 3D television and video [6]. Targeting different depth perception in auditory and visual channel, the videos were varied in video (monoscopic or stereoscopic) and audio (mono or stereo) resulting in 24 videos under test. Audio mono and stereo tracks were exported from Adobe Premiere and normalized. Monoscopic and stereoscopic video tracks were generated in Shake and afterwards exported together with the audio tracks. The videos were coded with mp4v codec using Simulcast at 25fps for the video track and 16bit at a sample rate of 48kHz for the audio track. The presentation order of the stimuli was randomized. A NEC autostereoscopic 3.5” display was used to present 2D monoscopic and stereoscopic videos with the same auditory and visual channel, the videos were varied in depth, and the user requirements for mobile 3D television and video (monoscopic or stereoscopic) resulting in 24 videos under test. Audio was presented by using AKG K-450 headphones.

Table 1: Snapshots of the six contents under assessment (V = visual spatial details, V = temporal motion, V = amount of depth, V = amount of scene cuts, A = audio characteristics)

<table>
<thead>
<tr>
<th>Screenshot</th>
<th>Genre and their audiovisual characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animation - Knight’s Quest 4D (18s)</td>
<td>V = high, V = high, V = med, V = high, A: music, effects Video bitrate: 10 Mbit/s</td>
</tr>
<tr>
<td>Documentary - Cave (18s)</td>
<td>V = high, V = med, V = high, V = low, V = low, A: orchestral music Video bitrate: 22 Mbit/s</td>
</tr>
<tr>
<td>Videoconference – Bullinger (23s)</td>
<td>V = med, V = low, V = med, V = low, V = low, A: male voice Video bitrate: 10,5 Mbit/s</td>
</tr>
<tr>
<td>User-created Content – Oldtimers (16s)</td>
<td>V = high, V = high, V = high, V = med, V = low, A: train sound Video bitrate: 20 Mbit/s</td>
</tr>
<tr>
<td>Documentary – Upper Rhine Valley (18s)</td>
<td>V = high, V = med, V = high, V = high, V = med, A: ambient music Video bitrate: 21 Mbit/s</td>
</tr>
</tbody>
</table>

2.3. Procedure

The evaluation was done in two phases – a psychoperceptual evaluation and sensory profiling task. Absolute Category Rating (ACR) according to ITU-R P.910 [7] was chosen for the psychoperceptual task. In ACR, stimuli under test are presented consecutively and rated independently. Test participants rated general acceptance of quality on a binary scale [5], the perceived overall quality and 3D impression (perceived depth) on an 11-point unlabeled scale. Sensory profiling is defined as methods “to evoke, measure, analyze and interpret reactions to those characteristics of foods and materials as they are perceived by the senses of sight, smell, taste, touch and hearing.” [8]. OPQ as audiovisual sensory evaluation approach adapts Free Choice Profiling methodology in which test participants develop their individual quality attributes. In an attribute elicitation phase the test participants were asked to write down individual quality attributes. Thereby, the participants could openly choose attributes according to their individual perception. The attribute refinement phase included then defining the attributes in their own words. Following, each attribute was attached with a 10cm long line labeled with ‘min’ and ‘max’ at the ends. In the evaluation task, the participants then rated overall quality of the test set on these attributes again independently one after another.

3. RESULTS

3.1. Psychoperceptual evaluation

Acceptance of quality – In overall all presented stimuli provided highly acceptable quality level. On average, 2D presentation mode reached the acceptance level of 90% and all stimuli reached at least acceptance of 88%. For 3D visual presentation mode, the average acceptance level of quality was 79% while none of stimuli went below 63% of acceptance.

Overall satisfaction - Parameter combinations influenced on overall quality satisfaction when averaged over the content (Fr=92.2, df=3, p<.001; Figure 1). The most satisfying quality was provided by 2D visual presentation mode over the 3D mode (p<.001). In the both visual presentation modes, mono and stereo audio were equally evaluated (p>.05). The results of content by content analysis follow this main tendency. The analysis of content called ‘cave’ is an exception. Although there is not an overall effect of parameter combinations on satisfaction (Fr=4.46, df=3, p=.215, ns) in this content, detailed pairwise comparisons show that 3D presentation mode provides higher quality under equal audio conditions (3D vs 2D – mono: Z=-2.53, p<.001; 3D vs 2D – stereo: Z=-3.12, p<.001). However, 2D accompanied
Figure 1: Influence of parameter combinations (visual and audio presentation mode) on overall quality satisfaction and 3D impression. The bars show 95% CI of mean.

with stereo audio reaches the equal quality level to 3D with mono audio presentation (Z=−1.61, p=.108, ns)

3D impression - The parameter combinations influenced on perception when averaged over the contents (Fr=596.4, df=3, p<.001 Figure 1) The highest level of depth perception was provided by stimuli with 3D presentation mode (p<.001). Under the 3D mode, the used audio presentation mode did not influence on depth perception (Z=−1.45, p=.14, ns) while stereo mode slightly outperformed mono when 2D mode was used (Z=−2.91, p<.01).

3.2. Sensory Profiling

Participants developed 130 individual quality attributes (mean 8.7, min 3, max 14) in the FCP task. Generalized Procrustes Analysis was used to transform this input into a low-dimensional space. The analysis results in a two-dimensional model that can be found in Figure 2. The resulting two-dimensional space (81.17% cumulated explained variance) shows that video quality is still the determining quality factor. As can be seen in Figure 2 the stimuli group along dimension 1 (68.21% explained variance) in clusters of monoscopic and stereoscopic presentation. A second output of the analysis is the word plot which is depicted in Figure 3. It shows the correlation of each individual quality attribute with the two components of the low-dimensional model. The word plot in Figure 3 shows that all monoscopic videos correlate with attributes like ‘sharp’, ‘flat’, or ‘stress-free’, while stereoscopic videos are mainly described with artifact-related (negative) attributes as ‘blurred’, ‘unstable’, or ‘stressful’. As exception, stimulus Cave is correlating with e.g. ‘brilliant’, ‘layered’, and ‘spatial’.

4. DISCUSSION

The goal of our study was to explore the influence of audio and visual presentation modes on perceived quality for mobile devices. Mono/stereo presentation modes were varied for audio and 2D/3D for video. Our data-collection procedure combined both quantitative preference ratings and descriptive Open Profiling of quality. The use of OPQ allowed eliciting individual quality attributes and connecting these attributes to the preference ratings of the stimuli under test. The results showed that the provided quality level was good being clearly above 60% of acceptance threshold. This indicates that quality of 3D would be sufficient for consumer products and being higher than previous study carried out with different display technology and highly compressed video [15].

All used quality measures (audiovisual quality satisfaction, depth impression and quality descriptions) are strongly influenced by visual, but neither by audio presentation mode nor their interaction. Similar to our results, non-significant influences of audio on audiovisual quality has been concluded in previous studies in the context of large displays and surround sound systems in a good quality level [16, 17]. Neuman et al. [16] concluded that under the audiovisual task, untrained participants have difficulties in detecting between mono or stereo audio under the video viewing task. Lessiter & Freeman [17] underlined that feeling of presence is not enhanced by audio mode. It is also possible that the visual variable acted as the most changing variable in the experiment and captured the greatest attention as suggested by peak-end theory [18].
Figure 2: Item plot of GPA results. The plot shows that there is a cluster of monoscopic and one of stereoscopic stimuli along Dimension 1. Also to be seen is the separate cluster of stereoscopic (3D) Cave cluster.

Figure 3: Word plot of the GPA results. The plot shows the correlation of the individual attributes with the dimensions of the GPA model. Identified main cluster information is shown along the dimension.
The results showed also a controversial impact of 3D presentation mode on overall quality and depth impression. While use of 3D mode increased the depth impression it decreased the overall satisfaction. The descriptive GPA results gave further explanations to these results by underlining the inferiority (spatial, stressfulness, flickering, eye-strain) in the case of 3D. However, our results also showed that in artifact-free case, 3D can reach higher perceived quality compared to 2D. In that case the perceived depth and the exciting 3D sensation make the stereoscopic videos subjectively better. This result indicates that the added value induced by the depth perception in stereoscopic presentation is only valid when level of visible artifacts is low giving the further support for previous studies [10, 11, 12]. Further work needs to address the most annoying artifacts to improve the 3D presentation in the sufficient level of technical resources for portable devices.

ACKNOWLEDGMENT

MOBILE3DTV project has received funding from the European Community’s ICT programme in the context of the Seventh Framework Programme (FP7/2007-2011) under grant agreement n° 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