PROFILING EXPERIENCED QUALITY FACTORS OF AUDIOVISUAL 3D PERCEPTION

Dominik Strohmeier[a], *Satu Jumisko-Pyykkö*[b], *Ulrich Reiter*[c]

[a] Institute for Media Technology, Ilmenau University of Technology, Ilmenau, Germany
[b] Unit of Human-Centered Technology, Tampere University of Technology, Tampere, Finland
[c] Centre for Quantifiable Quality of Service in Communication Systems (Q2S), Norwegian University of Science and Technology (NTNU), Trondheim, Norway[1]

ABSTRACT

Experienced multimodal quality is often assessed using quantitative experiments disregarding participants' experienced quality factors. The goal of this paper is to explore multimodal quality by using both quantitative methods as well as qualitative methods. Monoscopic and stereoscopic visual presentation modes and related room acoustic simulations for small and large spaces were varied. The experiments were conducted using a 15" autostereoscopic display and a 4.0 loudspeaker setup. Experienced quality was measured using both quantitative evaluation and Free-Choice Profiling with 20 naive participants. The results show that participants were qualitatively able to differentiate variables, although no quantitative excellence of stimuli was identified. The results also show that individual sensorial preferences are part of any multimodal quality evaluation. Therefore, we recommend using mixed methods to help in understanding heterogeneous multimodal quality. Individual sensorial preferences need to be further examined to improve multimodal quality evaluation experiments.

Index Terms — Audiovisual, quality, multimodality, experienced quality, quality of experience, 3D

1. INTRODUCTION

Since the classic paper by Beerends and De Caluwe [15], we know that audio and video in an audiovisual presentation influence each other with respect to perceived quality. Although perceived audiovisual quality has been a topic of research for more than a decade, the exact mechanisms of that mutual influence are still unknown. Part of the reason for this can be found in the difficulties in experimental design and data analysis for audiovisual assessments. Whereas more traditional engineering approaches try to quantify perceived overall quality by using some kind of rating scale often resulting in a *mean opinion score* (MOS), they mostly fail to explain the reasons behind a subject's rating.

Yet, audiovisual quality is a key factor for the success of novel multimedia applications, services, and devices. For the optimization of these, knowledge of the salient attributes that contribute to a quality impression is indispensable. Therefore, developing experimental methods that allow the identification of such attributes is a crucial factor in the research of quality perception.

Looking at perceived audiovisual quality from a more technical point of view, most work has been done on evaluating audiovisual quality in so-called home cinema and multimedia systems. These are systems that usually comprise mid- to larger-size video screens and multichannel loudspeaker setups. An ITU recommendation for audiovisual quality testing of such equipment exists [16]. For interactive applications using very large screens and multichannel sound, Reiter has published on perceived audiovisual quality [17]. For small screens with the use of headphones, work has been done by Ries et al. [24] and Korhonen et al. [25]. This has been extended to stereoscopic display by Strohmeier's work on quality perception of small 3D screens [13][19]. In this paper, we focus on perceived quality of audiovisual content presented on a mid-size auto-stereoscopic display, comparing 2D and 3D visualization, and multichannel loudspeaker reproduction.

The purpose of the paper is twofold. Firstly, we want to present the results of our Free-Choice Profiling approach to understand audiovisual quality. Secondly, we want to show that the quality model can also be used to identify differences among assessors related to different perceptual styles. The paper is organized as follows.

In section 2 we give a short introduction to applications of sensory profiling for the evaluation of perceived audio and video quality. Section 3 presents the applied research method. Section 4 shortly introduces the method of analysis. Section 5 summarizes the results. In section 6 we discuss identified individual sensorial preferences. Section 7 discusses and concludes the paper.

2. DESCRIPTIVE QUALITY EVALUATION

Currently standardized quality assessment methods [5][6] lack the possibility to provide information to understand the reasons that led to certain quality ratings. Qualitative methods must be applied to be able to elicit these experienced quality factors. Different qualitative methods exist. Current approaches in the evaluation of still images and videos use semi-structured interviews [2][3][9]. Items under test are described in a semi-structured interview, targeting elicitation of experienced quality factors. The interview task extends in both methods standardized quality evaluation. The interview is either conducted prior to or after the quality evaluation.

In contrast to interviews, methods adapted from sensory profiling target direct elicitation of individual quality factors. Stone and Sidel define sensory evaluation as a "scientific research methodology used to evoke, measure, analyze and interpret reactions to those characteristics of food and materials as they are perceived by senses of light, smell, taste, touch and hearing" [7]. Existing approaches in quality evaluation task of audio [4] and video [13] allow test participants to develop their individual quality attributes. These attributes are then used to

[1] "Centre for Quantifiable Quality of Service in Communication Systems, Centre of Excellence" appointed by the Research Council of Norway, funded by the Research Council, NTNU and UNINETT.

evaluate experienced quality of the test items. Multivariate data analysis is then applied to identify a common quality derived from the individual attributes.

In the following, we present the results of a study to show how sensory profiling can be applied to evaluate audiovisual quality. We present the resulting quality model as a common quality rationale of all test participants. In extension to other approaches, we also show that the model can be used to identify differences in how the different assessors experienced audiovisual quality.

3. RESEARCH METHOD

3.1 Participants

A total of 20 participants (age: 18-27 years, mean age: 22) took part in this study, 8 female and 12 male. All subjects were tested for visual acuity (myopia and hyperopia: Snellen index: 20/40), color vision, and stereo vision (≤ 60 arcsec). Additionally, they passed a hearing threshold test in accordance with ISO 7029 [26]. None of the participants had been working in the field of 3D video. In addition, all subjects were inexperienced in the field of subjective quality evaluation.

3.2 Stimuli

Two independent variables were used in the study to be able to change the audiovisual depth perception:

- Video presentation mode: monoscopic and stereoscopic video presentation
- Room acoustics: large room acoustics and small room acoustics

The test material was rendered using the IAVAS I3D player. The large room showed a classroom and the sound source of a male speaker was represented by a manikin. A snapshot of the scene can be found in Figure 1(a). The small room was a student's living room. Here, the sound source of drum and bass music was represented by a laptop. A snapshot of this second room is shown in Figure 1(b). Users' movement through the room was automated and contained a straight approach towards the sound source and turns to the left and right. In total, 8 different test items of 15 seconds each were created.

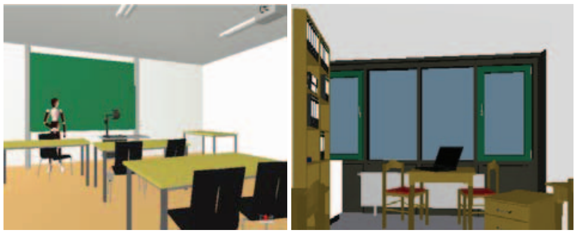

Figure 1 Screenshot of the two virtual rooms used in the assessment showing a) classroom on the left and b) living room on the right. Manikin and laptop represent the sound source in each room.

The rooms were designed using Maya software. They were exported into Binary Format for Scenes (BIFS). The audio was included using Advanced Audio BIFS. Each audio file was AAC encoded at a bit rate of 128kbps. The room acoustics were modeled using the Perceptual Approach as specified in the ISO/IEC standard 14496 (MPEG-4) [28]. For each room, a suitable model was applied, taking into account the different sizes and acoustical characteristics of the rooms. To vary depth in audio perception, the room models were exchanged between the rooms, too.

3.3 Stimulus presentation

The tests were conducted in the Listening Lab at Ilmenau University of Technology. The videos were presented on a 15" Sharp AL3DU stereoscopic display based on parallax barrier technology. The parallax barrier is built as a secondary LCD layer which can be switched on and off. This way it is possible to change from monoscopic to stereoscopic view. The viewing distance was set to 50cm in the beginning. Test participants were allowed to adjust their viewing distance, so that they could perceive the video three-dimensionally. The sound was played back on a four channel surround setup, with loudspeakers located at ±30 and ±110 degrees off the frontal direction, and at a distance of 1 meter from the assessor. The setup had been evaluated successfully before [8].

3.4 Test procedure

Open Profiling of Quality is a mixed method that combines evaluation of quality preferences and the elicitation of idiosyncratic experienced quality factors. It therefore uses quantitative psychoperceptual evaluation and, subsequently, an adaption of Free Choice Profiling [13][27]. In our 'open' approach, participants are free in the elicitation of overall quality factors, i.e. not limited to certain aspects of quality under investigation. They use their own attributes to evaluate each item under test. As an additional advantage, OPQ does not require extensive training as other descriptive methods [12]. Although it might be a hard task for test participants to develop their quality factors, the approach has been found to be applicable for naïve test participants.

Our test procedure was divided into three sessions. The first session started with the visual and auditory tests. Then, test participants filled out a demographic questionnaire. Following, the test participants had some time to practice to find their viewing position. Then an Absolute Category rating was applied to evaluate overall quality quantitatively [6]. In a first training task, participants watched a subset of the test items and practiced to evaluate perceived overall quality on an 11-point unlabeled scale. The training was followed by the evaluation and each item was assessed two times.

The second session was conducted after a short break of 15 minutes. The test participants were introduced to the Free-Choice Profiling task. They were told to think about quality factors that they used to evaluate overall quality in the first session. In the subsequent attribute elicitation task, each test item was shown three times. Participants had a blank sheet of paper and were asked to write down all their individual quality attributes. In the Open Profiling of Quality approach [13] we do not limit the test participants in what is quality. We want them to take into account every aspect that they see as important targeting a holistic understanding of the quality rationale. After the elicitation task a refinement task was conducted. In the refinement task, test participants define their final set of attributes by excluding from their list all quality factors that a) are not unique, and b) cannot be defined precisely. The final set of attributes was then written on a score card. Each one was attached to a 10cm long line. Each line labeled from 'min' to 'max' at the ends represents the quality sensation of the attribute. 'Min' means no sensation of the attribute, 'max' relates to a maximum sensation. An example of the score card can be found in Figure 2. Altogether, participants finished session 1 and 2 after 45 minutes.

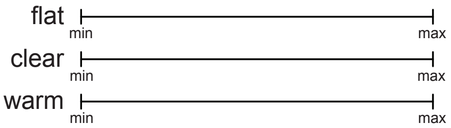

Figure 2 An example of a participant's score card

The third session was conducted within two days later. This session was divided into training and the evaluation task again. At first, each test participant was shortly reintroduced to FCP. His score card was presented to him. In a training task, test participants were shown a subset of the test items twice. Test participants evaluated each test items on their score card by ticking each line marking their sensation of each attribute. The evaluation task followed the training. Test participants evaluated each test item on a new score card. The third session took about 45 minutes.

4. METHOD OF ANALYSIS

The quantitative data was analyzed using non-parametric statistical analysis as no normal distribution was given for the test items (Kolmogorov-Smirnov: P<.05). Friedman test was applied to check if the independent variables impacted on the dependent one. Significant differences between two related items can then be measured using Wilcoxon test. Data was analyzed using SPSS 15.

In the Free-Choice profiling task, each test participant produces an individual configuration. A configuration is the judges per participant for his attributes among all test items. As a measure of sensation, the distance from the min-label to the participant's tick on the line is measured.

The individual configurations can be analyzed by applying Generalized Procrustes Analysis (GPA) [1]. GPA matches each configuration to a consensus, a common configuration of all participants. The idea is that individual configurations describe the same sensation although different terms are used. The results of GPA are comparable to those of a Principal Component Analysis [1][12]. MS Excel 2007 and XLSTAT 2010.2 were used for the data analysis.

5. RESULTS

5.1 Results of the quantitative evaluation

Visual presentation modes and room acoustic simulations did not have significant influence on overall quality perception (Friedman, χ^2 = 3.341, df = 7, p >.05, ns). All stimuli were equally rated (all pair wise comparisons p>.05).

5.2 Interpretation of the GPA model

The GPA model explained 81% of the data variation with three components (PC1: 36.2%, PC2: 25.06%, PC3:19.75%). The results of the data analysis are presented as GPA score plots in Figure 3 and as attribute correlation plots in Figure 4. Altogether a set of 289 individual quality attributes was developed. 212 of the attributes are located between the inner and outer circle of the correlation plot. These circles indicate 50% and 100% of explained variance, respectively. These attributes are emphasized in the following interpretation. With help of the GPA score plots in Figure 3, PC1 and PC2 (or their rotation) can be identified as 'content' and 'video quality', respectively. PC3 represents the 'audio quality'. Although the interpretation was done based on the test items or their related test parameters, we refer to the quality aspects of content (classroom, living room), video representation (2D, 3D), and room acoustics (large, small). This first finding confirms that test participants chose the quality factors according to the chosen test parameter.

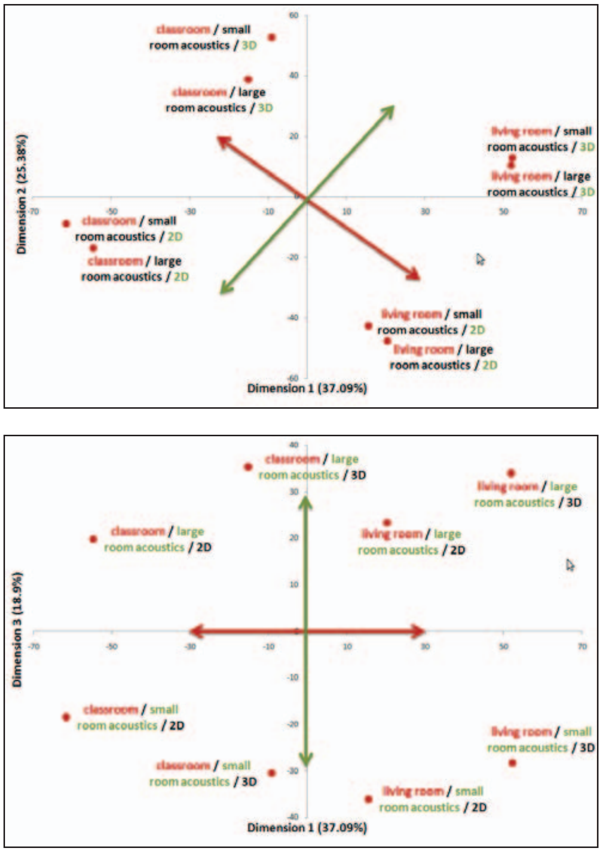

Figure 3 The GPA score plots showing the test items in the GPA model

5.3 Interpretation of the quality attributes

To be able to interpret the assessors' idiosyncratic attributes, the correlation of each attribute with the components of the model is calculated. The results are plotted as correlation plots in Figure 4. PC1 was identified as content dimension and is described with attributes like 'tidy' or 'comic like'. In addition, affective attributes like 'monotone' or 'sympathy' can be found.

The second dimension PC2 was identified as the video quality dimension. Its polarities describe the monoscopic (2D) and stereoscopic (3D) video mode, respectively (c.f. Figure 3). 3D correlates with quality factors like 'three-dimensional' or 'spacious'. It shows that the test participants were able to perceive depth. However, the other attributes don't show that 3D provided an added value for the test participants. Stereoscopic videos correlate with attributes like 'blurry', 'smeary', and 'interlaced lines'. Also words like 'exhausting' can be found having high correlation with the stereoscopic videos. In contrast, monoscopic videos are described with quality attributes like 'sharp', 'colorful', and 'clear'. It seems that the video quality is mainly determined by the video artifacts for the 3D case. Although depth is perceived, no attributes can be identified that describe added value through 3D, as for example presence or involvement.

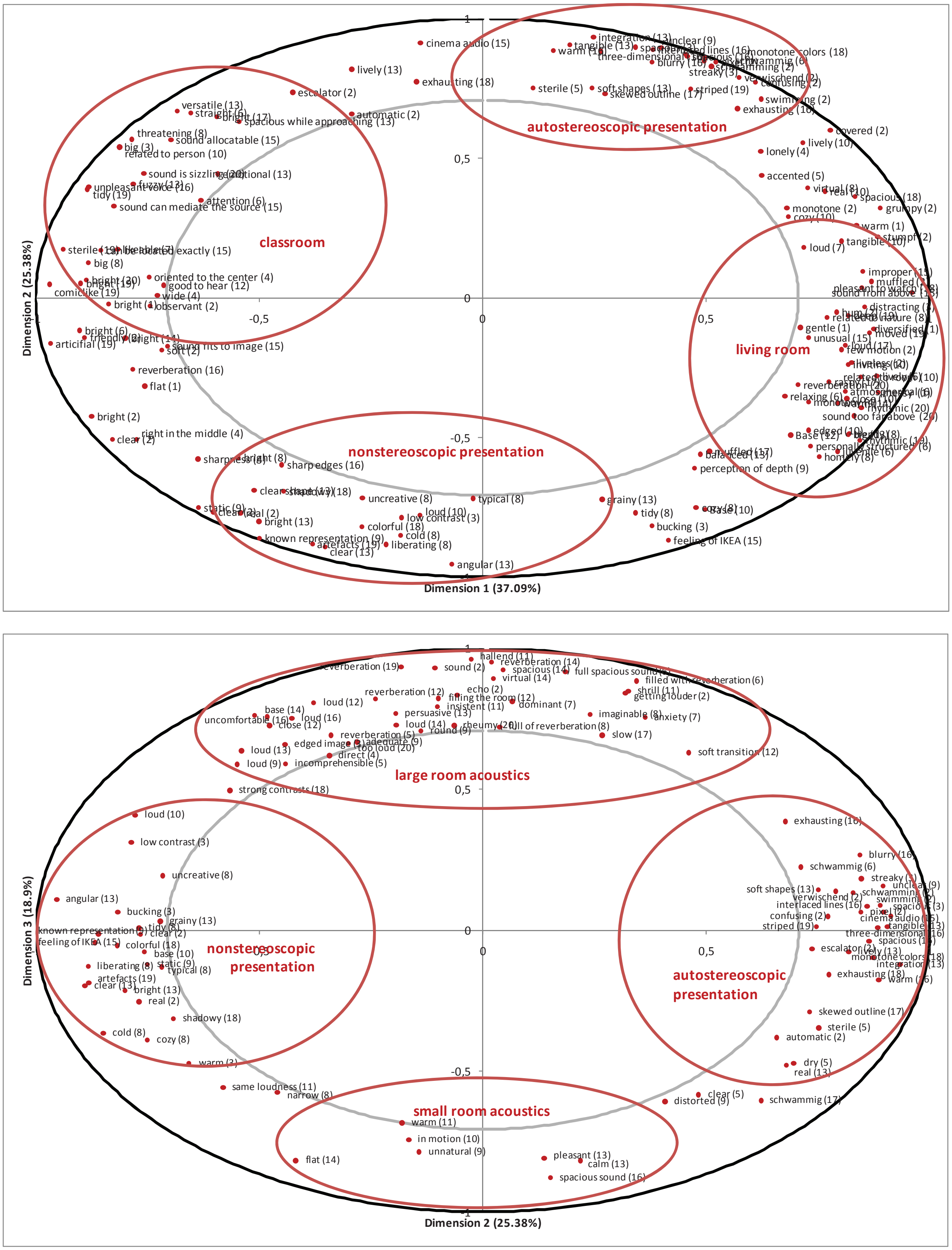

Figure 4 The correlation plots of individual quality attributes and GPA model. The plots show a) correlation with PC1 and PC2 and b) correlation with PC2 and PC3.

The third component PC3 is related to audio quality. Correlating attributes that describe the underlying parameter of changing room acoustics are ‘reverberation’, ‘echo’, or ‘becoming louder’[2]. They target the different room acoustics. Quality factors like ‘spacious’ and ‘spacious sound’ confirm the hypothesis that also audio contributed to a depth perception. However, PC3 is also described with attributes like ‘shrill’, ‘dominant’, or even ‘anxiety’. These affective quality factors show that perceived quality is more than just the perception of system quality.

Also attributes that correlate with more than one dimension can be interesting. Especially attributes that correlate with PC2 and PC3 as they describe audiovisual effects. Interdimensional attributes between audio and video dimension are rare as Figure 4(b) shows. Especially depth-related attributes that we expected to correlate with both dimensions correlate either with video (e.g. spacious (3)) or with audio dimension (e.g. spacious (14)). These results show ttha depth was perceived or rated independently either in auditory or visual perception. So, in the next section we will have a closer look at participants’ individual perceptual patterns.

6. INTERPRETATION OF PERCEPTUAL DIFFERENCES BETWEEN ASSESSORS

The goal of the Generalized Procrustes Analysis was to fit all assessors’ configurations to a common consensus. From these consensus configurations a model is calculated. The result targets to model a common quality rationale of all assessors. Individual differences among the test participants are not taken into account. However, it might be interesting to see how attributes from one configuration are spread in the GPA model after analysis. We attach the test participants ID to each of his or her quality factors. This allows us to identify each configuration’s attributes after the analysis.

Figure 5 is based on the GPA model. It shows again the correlation of assessors’ quality factors with the three components of the model. We removed the attribute names and only marked the attributes’ position on the correlation plot. Different marks belong to different assessors. In Figure 5, quality attributes of three different assessors are depicted. For the sake of clarity, only attributes between the 50% and the 100% explained variance circle are shown.

Figure 5(a) again shows the correlation of attributes with PC1 and PC2. It is remarkable that attributes from participant #1 (shown as triangles) only correlate with PC1 or the content dimension. In contrast, quality factors from participant #13 (squares) show high correlation with the dimension of video quality (PC2). Attributes from participant #14 (asterisk) are found rarely along PC1 and PC2. Their correlation is high with audio quality component as can be seen in Figure 5(b).

What can be seen from these plots is that participants use different parameters of the test items to derive their individual quality parameters from. An analogous analysis for other assessors showed that only few of them use two or even three parameters for deriving quality attributes

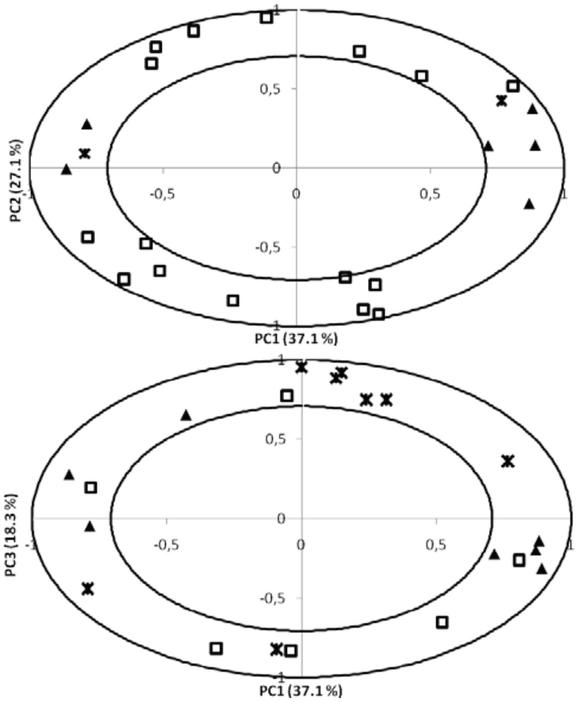

Figure 5 Correlation of attributes from participant #1 (triangles), participant #13 (squares), and participant #14 (asterisk) with the main components of the GPA model.

7. DISCUSSION AND CONCLUSIONS

The goal of this paper was to explore experienced multimodal quality. We varied monoscopic and stereoscopic visual presentation modes and related room acoustic simulations for small and large spaces. We used both quantitative evaluation and qualitative descriptive sensorial profiling methods in the data-collection procedure.

Our quantitative results did not show any differences in experienced quality between the variables. The results of sensorial profiling offered further explanations to this. Firstly, the non-significant difference was not caused by the non-detectable differences between stimuli, as participants qualitatively differentiated them. Secondly, perceived depth was underlined in both audio and visual modalities, thus contributing to the overall audiovisual perception. Thirdly, when visual 3D presentation mode was used, it was described as spacious and three-dimensional, but more importantly it was attached to several negative terms of inferiority. It is known that the added value induced by the depth perception in stereoscopic presentations is only valid when the level of visible artifacts is low, as has been shown in previous studies [19][20][21][22].

Our results also showed individual preferences towards quality of one modality. It is known that there are individual differences in human information processing styles towards different modalities. For example, the categorization into visual and verbal information processing styles is common [22]. Our results indicate that these different processing styles can also contribute to final multimodal quality judgments. There are two suggestions for further work. Firstly, the influence of different processing styles on multimodal quality perception under different quality levels and heterogeneous stimulus material needs to be addressed in detail to confirm the phenomenon. Secondly, for the practitioners of audiovisual quality, a well-validated tool is

[2] ‘Becoming louder’ is describing the effect of changes in the room acoustics when the animated movement was approaching the sound source.

needed for identifying the groups of different information processing styles and reporting these groups to characterize the sample.

8. ACKNOWLEDGEMENTS

Part of the work was done in MOBILE3DTV project which has received funding from the European Community's ICT programme in the context of the Seventh Framework Programme (FP7/2007-2011) under grant agreement no 216503.

9. REFERENCES

[1] Gower, J. 1975. Generalized procrustes analysis. Psychometrika 40, 1, 33-51.

[2] Jumisko-Pyykkö, S., Reiter, U., and Weigel, C. 2007. “Produced quality is not perceived quality - a qualitative approach to overall audiovisual quality”. In Proceedings of the 3DTV Conference.

[3] Jumisko-Pyykkö, S., Häkkinen, J., and Nyman, G. 2007. “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.

[4] Lorho, G. 2005. “Individual Vocabulary Profiling of Spatial Enhancement Systems for Stereo Headphone Reproduction”. Proceedings of Audio Engineering Society 119th Convention, New York (NY), USA, Convention Paper 6629.

[5] Recommendation ITU-R BT.500-11. 2002. Methodology for the Subjective Assessment of the Quality of Television Pictures, Recommendation ITU-R BT.500-11. ITU Telecom. Standardization Sector of ITU.

[6] Recommendation ITU-T P.910. 1999. Subjective video quality assessment methods for multimedia applications, Recommendation ITU-T P.910. ITU Telecom. Standardization Sector of ITU.

[7] Stone, H. and Sidel, J. L., Sensory evaluation practices, 3rd ed., Academic Press, San Diego, 2004.

[8] Strohmeier, D. and Jumisko-Pyykkö, S. 2008. “How does my 3d video sound like? – Impact of loudspeaker set-ups on audiovisual quality on mid-sized autostereoscopic display”. 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2008, 73-76.

[9] Radun, J., Leisti, T., Häkkinen, J., Ojanen, H., Olives, J.-L., Vuori, T., and Nyman, G. 2008. Content and quality: Interpretation-based estimation of image quality. ACM Trans. Appl. Percept. 4, 4, 1fi15.

[10] Reiter, U. and Jumisko-Pyykkö, S. “Watch, Press, and Catch Impact of Divided Attention on Requirements of Audiovisual Quality”. Proceedings of the 2007 HCI International Conference on Human Computer Interaction, Beijing, PR China. 2007 July 22-27.

[11] Williams, A. A., Langron, S. P. “Use of free-choice profiling for the evaluation of commercial ports”, Journal of the Science of Food and Agriculture. Vol. 35, pp. 558-68. May 1984

[12] Lawless, H. T., and Heymann, H. *Sensory evaluation of food: principles and practices*. Chapman & Hall, New York. 1999

[13] Strohmeier, D. and Tech, G. “Sharp, bright, three-dimensional: Open Profiling of Quality for mobile 3DTV coding methods”, Proceedings of the IS&T/SPIE Annual Symposium of Electronic Imaging, Convention Paper 7542-30. San Jose, CA, USA, 2010

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

[15] Beerends, J. G., and De Caluwe, F. E., “The influence of video quality on perceived audio quality and vice versa,” J. Audio Eng. Soc., vol. 47, no. 5, pp. 355-362, May 1999.

[16] Recommendation ITU-T P.911, Subjective audiovisual quality assessment methods for multimedia applications, International Telecommunication Union, Geneva, Switzerland, 1998, and Corrigendum 1, 1999.

[17] Reiter, Ulrich, “Bimodal Audiovisual Perception in Interactive Application Systems of Moderate Complexity”, PhD Thesis, TU Ilmenau, 2009

[18] Reiter, U., and Kühhirt, U.: “Object-Based A/V Application Systems: IAVAS I3D Status and Overview”, IEEE/ISCE'07, International Symposium on Consumer Electronics, Dallas, TX, USA, June 20-23, 2007.

[19] 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.”, in Proceedings of Fourth International Workshop on Video Processing and Quality Metrics for Consumer Electronics VPQM, Scottsdale, USA, Jan. 2010

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

[21] Ijsselsteijn, W., de Ridder, H., Vliegen, J. “Subjective evaluation of stereoscopic images: Effects of camera parameters and display duration”. IEEE Transactions on Circuits and Systems for Video Technology, 10:225–233, 2000

[22] Jumisko-Pyykkö, S. & Utriainen, T. “User-centered quality of experience: Is mobile 3D video good enough in the actual context of use?”, in Proceedings of Fourth International Workshop on Video Processing and Quality Metrics for Consumer Electronics VPQM, Scottsdale, USA, Jan. 2010

[23] Childers, T. L., Houston, M. J. and Heckler, S. E. (1985) “Measurement of Individual Differences in Visual Versus Verbal Information Processing”. In: Journal of Consumer Research, vol. 12, September, pp.125-134

[24] Ries, M., Puglia, R., Tebaldi, T., Nemethova, O., and Rupp, M. “Audiovisual Quality Estimation for Mobile Streaming Services”, 2nd Intl’ Symposium on Wireless Communication Systems, Siena, Italy, Sept. 2005.

[25] Korhonen, J., Reiter, U., Myakotnykh, E., ”On the relative importance of audio and video in the presence of packet losses”, 2nd Int. Workshop on Quality of Multimedia Experience QoMEX10, Trondheim, Norway, June 2010.

[26] International Organization for Standardization. ISO 7029 “Statistical distribution of hearing thresholds as a function of age”, Geneva, Switzerland, 2000

[27] Strohmeier, D., Jumisko-Pyykkö, S., Kunze, K. “Open Profiling of Quality – A mixed method approach to understanding multimodal quality perception,” submitted to Advances in Multimedia, Hindawi Publishing Corporation, 2010

[28] ISO/IEC 14496-1:2004, “Information technology - Coding of audio-visual objects - Part 1: Systems”, 3rd Ed., Geneva, Switzerland, 2004.