A UTILITY MODEL FOR SENSORY EXPERIENCE

Christian Timmerer, Benjamin Rainer, and Markus Waltl

Multimedia Communication (MMC) Research Group, Institute of Information Technology (ITEC),
Alpen-Adria-Universität Klagenfurt, Klagenfurt, Austria, *firstname.lastname*@itec.aau.at

ABSTRACT

Enriching multimedia with additional effects such as olfaction, light, wind, or vibration is gaining more and more momentum in both research and industry. Hence, there is the need to determine the influence of individual effects on the Quality of Experience (QoE). In this paper, we present a subjective quality assessment using the MPEG-V standard to annotate video sequences with individual sensory effects (i.e., wind, light, and vibration) and all combinations thereof. Based on the results we derive a utility model for sensory experience that accounts for the assessed sensory effects. Finally, we provide an example instantiation of the utility model and validate it against current and past results of our subjective quality assessments conducted so far.

Index Terms—Quality of Multimedia Experience, Sensory Effects, MPEG-V, Sensory Experience, Utility Model

1. INTRODUCTION

Recently, 3D video and Ultra-HD (4K) videos are gaining momentum and aiming at a more immersive viewing experience. Additionally, more and more researchers take up the challenge on enriching multimedia with additional effects such as light, wind, vibration, scent, etc. For example, [1] and [2] demonstrate the enhancement of multimedia content using additional light effects while others perform research in the area of olfaction together with multimedia content. The authors in [3] have shown that multimedia enriched with scent effects has a negative impact on the information recall of users. Others focus on synchronization issues between audio-visual content and olfactory data [4]. In [5], an olfactory virtual reality game is presented. This game allows the user to explore the relationship between olfaction and vision. The users are asked to distinguish between olfaction and vision following a memory game approach. In [6], the authors assessed the fairness of a networked olfactory game regarding the delay between the participating players. The participants have the task to judge whether the fruit matches the scent emitted. The results show that with an increase in delay the mean opinion score (MOS) of fairness is reduced.

In our previous work [7][8], we evaluated different sensory effects (i.e., wind, vibration, and light) and their impact on the perceived video quality. First, in [7], we have conducted a subjective quality assessment to investigate the impact of sensory effects on visible artifacts of video content. Second, in [8], we subjectively evaluated the influence of sensory effects on the emotional response. The results showed that sensory effects can mask visible artifacts and that the intensity of the emotions (e.g., for anger or fun) is increased in contrast to video sequences without sensory effects; thus, providing a more immersive viewing experience. Furthermore, the results showed that sensory effects enhance the Quality of Experience (QoE) for different genres (e.g., news, documentary, commercial, action, and sports).

The major contribution of this paper is a first utility model for sensory experience based on a subjective quality assessment which evaluates the impact of individual sensory effects and all combinations thereof. For the annotation of the sensory effects, we adopted Part 3 of the MPEG-V standard [9] which allows us to describe various effects for existing multimedia content. The results of this subjective evaluation provides insights regarding the influence of individual effects on the QoE and serves as basis for our utility model. In particular, they can be used to allow a processing engine to decide which effect should be effectively rendered based on the users' context.

Our utility model for sensory experience is defined complementary to existing approaches for predicting the QoE of audio-visual services. For example, these existing approaches aim to map QoS to QoE [10] or to predict the QoE [11][12] with a main focus on audio-visual services and do not take into account additional assets such as sensory effects. Other QoE models such as the one presented in [13] are based on perception, emotion, and sensation and mainly address adaptation and presentation issues without explicitly addressing sensory effects.

The remainder of this paper is organized as follows. Section 2 describes the subjective quality assessment and its results which evaluates the impact of individual sensory effects and combinations thereof. Our proposed utility model is described in Section 3 which is instantiated and validated in Section 4. In Section 5, we present a discussion of our results and conclude the paper with some future work in Section 6.

978-1-4799-0738-0/13/$31.00 ©2013 IEEE

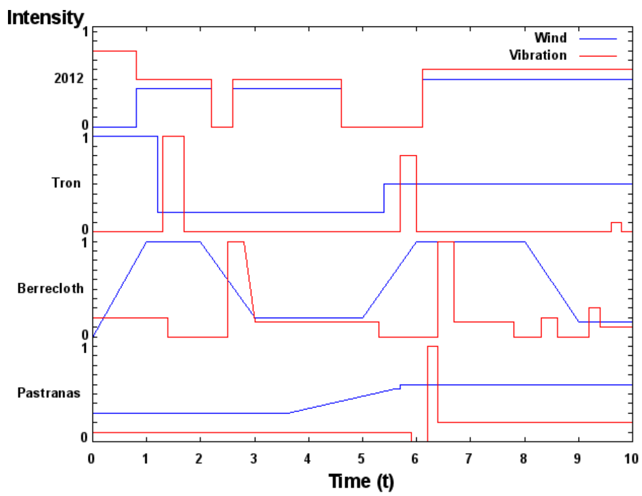

Figure 1. Overview of Video Sequences annotated with Sensory Effects.

2. QUALITY EVALUATION OF SENSORY EFFECTS

For the subjective quality assessment (henceforth, referred to as study) on the impact of individual sensory effects on the QoE and the combinations, thereof, we adopted the same test setup and environment as in our previous study [8]. Therefore, in this paper, we describe only the major differences and for detailed information concerning our setup the interested reader is kindly referred to [8]. In general, the study is based on ITU-R Rec. BT.500-13 [14].

2.1 Participants, Stimuli, and Setup

We invited 32 students (6 female and 26 male) aged between 20 and 47 with both technical and non-technical background. The results of our previous study revealed that there is no significant difference between male and female participants. The majority of 90.6% was aged between 20 and 31. Participants have been screened for handicaps (i.e., visual, audio, epilepsy) through a questionnaire and excluded if necessary.

As stimuli, we used four 720p and 10s long video sequences (taken from [15]); two sports (*Berrecloth* and *Pastranas*) and action (*2012* and *Tron*) video sequences, respectively. Figure 1 provides an overview of the video sequences showing its intensities and distribution of wind and vibration effects over time. The light effects are generated automatically by the Web browser plug-in as described in [8]. The SEM descriptions for the wind and vibration effects were manually annotated with the open source tool Sensory Effect Video Annotation (SEVino) [16].

The study was conducted in a controlled environment using the same hardware, software, and ambient conditions as described in [8]. The study itself was structured into three parts: an introduction (describing the purpose of the study and procedure) including a pre-questionnaire, the main evaluation, and a post-questionnaire. The overall duration for the study for each participant was around 15 minutes. The pre- and post-questionnaire were the same as in [8] and, thus, we focus only on the main evaluation which is described in the following sections.

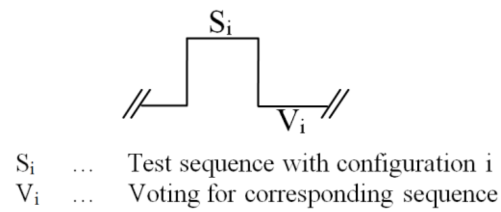

Figure 2. Single Stimulus Assessment Method.

Table 1. Video Sequence Effect Configurations.

Number	Configuration
1	No sensory effects
2	Light only
3	Wind only
4	Vibration only
5	Light and wind
6	Light and vibration
7	Wind and vibration
8	Light, wind, and vibration

Table 2. Rating Scale.

Interval	Label
80 – 100	Very high
60 – 80	High
40 – 60	Medium
20 – 40	Low
0 – 20	Very low

2.2 Assessment Method and Experimental Design

We used the Single Stimulus Continuous Quality Scale (SSCQS) method with a scale ranging from 0 to 100 as defined by ITU-R Rec. BT.500-13 [14]. We introduced a training phase to reduce a possible memory effect as suggested by [14]. Furthermore, this should eliminate or at least reduce the surprise effect for the participants in case they never experienced videos enriched with sensory effects. For the training phase, we selected the action sequence *Tron* as it provides high and low intensities of wind and vibration effects. Different configurations of the sensory effects attached to the video sequence were prepared as shown in Table 1 and presented to the participants in randomized order.

After the training phase, the remaining three video sequences (i.e., *2012*, *Berrecloth*, and *Pastranas*) were shown to the participants in the same manner as the training sequence. Figure 2 depicts the used single stimulus method with the voting period after each video sequence. Each video sequence and each configuration is randomly selected and presented to the participants (i.e., S_i with i=[1,8] according to Table 1). After each video sequence the

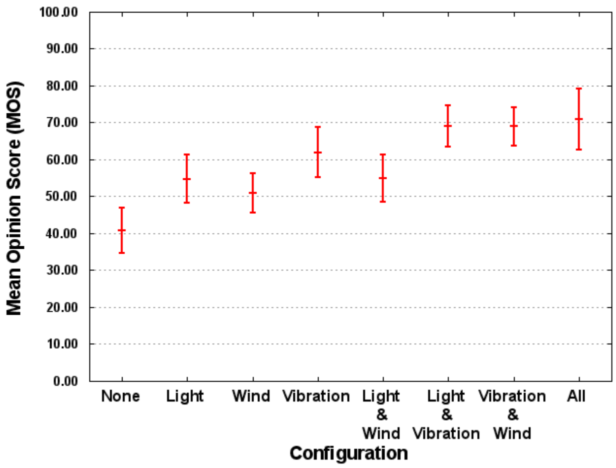

Figure 3. MOS and Confidence Intervals (95%) for *2012*.

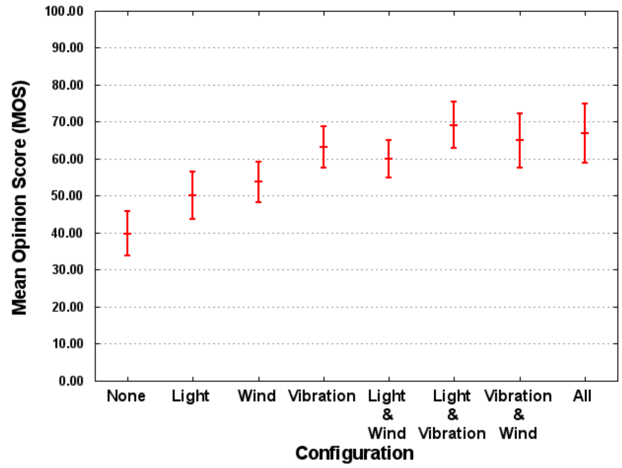

Figure 4. MOS and Confidence Intervals (95%) for *Pastranas*.

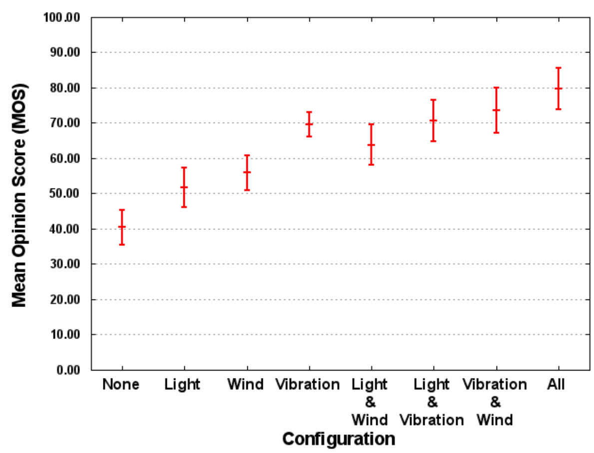

Figure 5. MOS and Confidence Intervals (95%) for *Berrecloth*.

participants had to rate the QoE of the video sequence with the corresponding effect configuration (i.e., V_i). The maximum voting time V_i was set to five seconds as suggested by [14].

The continuous rating scale of the QoE ranges from 0 to 100, whereas 100 represent a very high QoE and 0 indicates a very low QoE. The continuous rating scale was divided into five major levels as shown in Table 2 but only the "Very low" and "Very high" labels were presented to the participants during voting. The actual voting was performed by a slider as described in [8]. The other labels were only used during the introduction to explain the rating scale to the participants.

2.3 Results of the Sensory Effect Evaluation

The assessed data was analyzed according to [14]. The hypothesis that the data is normally distributed was checked by using the Q-Q plot and by calculating the kurtosis and skewness of the sample distribution. Outliers were screened using the β_2-Test as recommended by [14]. The screening of outliers identified one participant as outlier. For this outlier, we additionally checked if the ratings differ more than twice the standard deviation from the means which was the case. Thus, we excluded this participant from the results.

Figure 3 depicts the results for the video sequence *2012* from the *action* genre with all possible configurations (cf. Table 1). The results clearly indicate that without sensory effects the MOS for the QoE is around 40 which is between the categories "Low" and "Medium" according to Table 2. Adding sensory effects will increase the MOS. For example, adding light effects to the video sequence increases the MOS from around 40 up to about 55 which corresponds to the category "Medium". The MOS slightly decreases when exchanging the light effect with a wind effect. The highest impact of a single sensory effect on the QoE is achieved by using the vibration effect which increases the MOS up to approximately 62. One can see that using only vibration effects has a bigger impact on the QoE than using only light or wind effects. Surprisingly, the combination of light and wind effects does not result in a higher rating than the vibration effect on its own. Moreover, it shows a similar rating as with only light or wind effects, at least for this sequence. Any other effect combined with vibration increases the QoE. Finally, the highest MOS is achieved by combining all three sensory effects.

Figure 4 and Figure 5 depict the MOS for the two other video sequences *Pastranas* and *Berrecloth,* respectively, both from the *sports* genre. Figure 4 and Figure 5 show the same tendencies with respect to the MOS as indicated in Figure 3. Interestingly, it can be seen that in both cases the video sequences without sensory effects have almost the same and lowest MOS (approximately 40 MOS points). Furthermore, Figure 4 and Figure 5 nearly draw the same picture as Figure 3 regarding the video sequences presented with a single sensory effect only. Again, the vibration effect has the highest MOS for configurations with a single effect. The configurations combining two effects show an additional increase in MOS with a minor exception for

Pastranas (cf. Figure 4) where the combination "Vibration & Wind" is slightly lower than "Light & Vibration". This can be explained as in *Pastranas,* wind is rendered continuously throughout the entire sequence (cf. Figure 1). Additionally, the combination of all effects is slightly lower than with "Light & Vibration" (i.e., around two MOS points) which may indicate that the wind effect was not accurately used to accentuate specific scenes.

Summarizing the results, all three video sequences draw nearly the same picture. Thus, we can conclude that single sensory effects and combinations thereof increase the QoE, specifically vibration effects seem to have a higher influence than others which will be further investigated in the next sections. In particular, compared to our previous studies in this field [8], we had the indication that sensory effects do increase the QoE but we had no evidence on how strong single effects and their combinations do influence the QoE.

3. PROPOSED UTILTIY MODEL FOR SENSORY EXPERIENCE

As indicated in Section 1, the available models that try to map QoS to QoE or to estimate QoE from different parameters do not take into account additional assets such as sensory effects. The results from the study presented and discussed in Section 2 led us to the hypothesis that there exists a linear relationship between the number of effects and the actual QoE. Thus, we introduce a linear utility model for sensory experience. The aim of this model is to enable an estimation of the QoE of multimedia content with sensory effects (QoE_w) from the QoE of multimedia content without sensory effects (QoE_{wo}). Equation (1) shows our proposed utility model for sensory experience.

$$QoE_w = QoE_{wo} * (\delta + \sum w_i b_i) \tag{1}$$

In our utility model, w_i represents the weighting factor for a sensory effect of type i, e.g., in our setup $i \in \{\text{light(l)}, \text{wind(w)}, \text{vibration(v)}\}$. Please note that further sensory effect types (e.g., scent) may be incorporated easily, e.g., as soon as appropriate devices become available. The binary variables b_i ($b_i \in \{0,1\}$) are used to identify whether effect i is present for a given setup. Finally, δ is used for fine-tuning. The QoE_{wo} may be assessed through any existing model such as those given in [13] or by an appropriate QoS to QoE mapping [17]. Please note that the QoE_{wo} is not within the scope of this paper and, thus, we did not further investigate into this direction.

The results of the conducted study (cf. Section 2) request for a model that deals with all types of sensory effects separately. Therefore, we introduce the model illustrated by Equation (1) with weighting factors and binary variables for each type of sensory effect. An instantiation and validation of the proposed utility model is given in Section 4.

4. UTILITY MODEL INSTANTIATION AND VALIDATION

For the validation of our proposed utility model, we instantiate the model according to the study described in Section 2 and our previous study [8]. The instantiated utility model is shown in Equation (2).

$$QoE_w = QoE_{wo} * (\delta + w_l b_l + w_w b_w + w_v b_v) \tag{2}$$

In this instance of our model, the index *l* represents light, *w* stands for wind, and *v* denotes vibration. For validating our model, we employ Multiple Linear Regression (MLR) with the least square estimator [18]. For the response, we took the means of the MOS of each configuration (cf. Table 1). We estimate the weights δ and w_i for each sensory effect type and obtain the following values for the utility model as shown in Equation (3):

$$QoE_w = QoE_{wo} * (1.1 + 0.16 b_l + 0.17 b_w + 0.44 b_v) \tag{3}$$

The estimated variance with the used data is 10.9. Furthermore, we conducted an Analysis of Variance (ANOVA) on the obtained model [18]. The null hypothesis for the ANOVA was formulated as follows: "The variables (b_i) are zero ($b_i = 0$)". For the three variables, we obtained the following F-Statistics and p-values (F/p): b_l (7.73/0.04977), b_w (8.92/0.04), and b_v (57.49/0.0016). The F-Statistics and p-values clearly state that the null hypothesis can be rejected for a significance level of 5%.

Another very important measure is how well the regression model reflects the variability of the response, indicated by the square sample correlation r^2 [19], which is the ratio of the sum of squares to the total sum of squares of a variable. The highest variability is represented by b_v (73.57%) whereas b_l and b_w represent 10% of the response, respectively. The remaining percent points are the estimation error. These results reflect the tendencies that are observed by the results presented in Section 2.3. That is, the vibration effect has the highest impact on the estimated QoE. Figure 6 illustrates the used data for the MLR and the estimated response of the proposed utility model. Each estimated response according to our utility model is almost within the confidence interval of the response from the subjective quality assessment.

Figure 7 depicts the Mean Absolut Error (MAE) between the estimated response for all effects by our utility model and the data obtained by the subjective quality assessment described in Section 2 denoted by UTY. The MAE shows how much the estimated response differs on average from the assessed values. Additionally, Figure 7 shows the MAE compared to our previous studies in [8] for the genres action and sports. In [8], we presented three studies conducted at three different locations (Alpen-Adria-Universität Klagenfurt – AAU, University of Wollongong –

UOW, and Royal Melbourne Institute of Technology – RMIT). The MAE is relatively low (i.e., ±18 points on average) for all of the conducted studies and, thus, reflects the accuracy of the estimated response through our instantiated utility model.

5. DISCUSSION

In general, the results confirm our previous studies in [7][8], i.e., sensory effects increase the QoE. An interesting finding is that among the test sequences, the individual sensory effect types are perceived likewise in terms of QoE. Furthermore, all three video sequences show the same trend with respect to MOS. That is, the MOS almost always increases if more effects (and combinations thereof) for both genres sports and action are present. Our previous study in [20] suggests that sensory effects increase the QoE depending on the genre with action, documentary, and sports benefiting the most. However, note that the goal of the current study was different. In particular, in this study, we investigated the impact on individual effects and combinations thereof versus the impact on the genre in [20].

Looking at the results with one additional sensory effect type, the vibration effect achieved the highest MOS among all three video sequences. That is, vibration has the greatest impact on the QoE which is also confirmed in Section 4. Video sequences with two sensory effect types show the same tendency as video sequences with a single sensory effect type. Any configuration with light and wind is in all cases lower than configurations including vibration effects. Nevertheless, the highest MOS is achieved when all three sensory effect types are presented to the user.

The conducted subjective quality assessment led us to a first utility model for sensory experience. Other models do not take into account sensory effects and have their main focus on mapping QoS to QoE for audio/visual content [10][11][13]. The proposed utility model shows that there exists a linear relationship between the QoE without sensory effects and the QoE with sensory effects. Nevertheless, the QoE model without sensory effects may still follow a non-linear function which allows applying our model on top of it. Furthermore, we have defined a general utility model for sensory experience and provided an instantiation for light, wind, and vibration effects. The general utility model may accommodate further sensory effect types such as scent but would require another instantiation taking these additional sensory effect types into account. For example, when adding scent, it remains to be seen whether vibration still has the greatest impact on the QoE among other effects.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we presented our subjective quality assessment on the influence of different combinations of sensory effects on the QoE. Based on the results, we derived a first utility model for sensory experience. The results of this evaluation show how single sensory effects and combinations thereof increase the QoE. In particular, the results showed that vibration effects have the biggest influence on the QoE, whereas wind and light have a lower impact. Additionally, the results confirmed our previous results and, therefore, it is recommended to use sensory effects if suitable devices are available. Moreover, we introduced a general utility model for estimating the QoE with sensory effects from the QoE without sensory effects and appropriate weights. We instantiated and validated the proposed utility model and we showed that the instantiation provides reliable estimates of the MOS for different configurations of sensory effects. The analysis of the instantiated utility model showed that the highest variability is provided by the variable denoting whether vibration is available or not. The proposed utility model is the first model for sensory experience which can be used to estimate the QoE for a video sequence if sensory effects can be

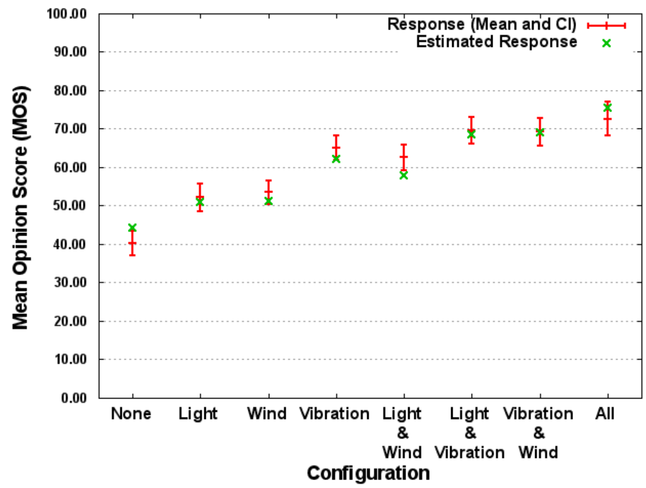

Figure 6. Response used for the MLR and the estimated response.

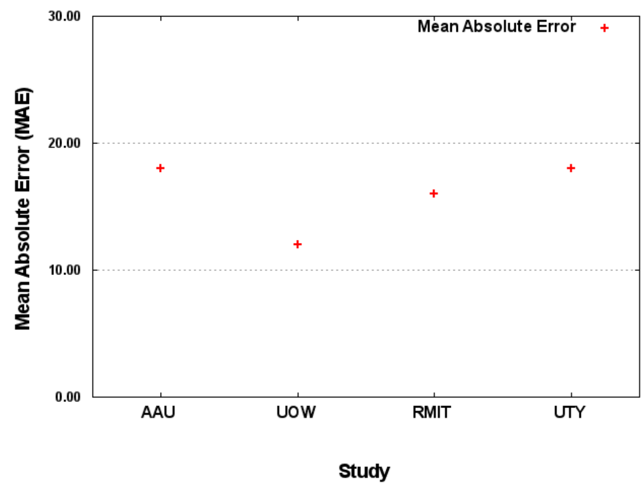

Figure 7. MAE of the estimated response and the actual assessed data of previous user studies.

rendered. Using the estimations provided by the utility model offers a first step towards an objective quality measurement method for sensory effects; thus, reducing the necessity for subjective quality assessments in this domain.

Future work topics in the area of sensory experience comprise the usage of additional effects (e.g., scent, fog) and the instantiation and validation of the proposed utility model with these additional sensory effects. Furthermore, the synchronization of sensory effects with multimedia content is also subject to future work. That is, how much earlier or later can sensory effects be rendered without being perceived as annoying? There is already some work on the synchronization of scent effects by [4], but nothing has been done yet with other effects. Another future work item is to conduct medical user studies with electrocardiographic (EEG) devices to investigate the impact of sensory effects on the indirect emotional response.

Acknowledgments: This work was supported in part by the EC in the context of the ALICANTE (FP7-ICT-248652), SocialSensor (FP7-ICT-287975), and QUALINET (COST IC 1003) projects and partly performed in the Lakeside Labs research cluster at AAU.

7. REFERENCES

[1] S. H. A. Begemann, "A Scientific Study on the Effects of Ambilight in Flat-Panel Displays", 2005. http://bit.ly/pxbZYA. Accessed 31.01.2013.

[2] P. Seuntiens, I. Vogels, A. van Keersop, "Visual Experience of 3D-TV with Pixelated Ambilight", *Proceedings of the 10th Annual International Workshop on Presence*, Barcelona, Spain, pp. 339-344, 2007.

[3] G. Ghinea, O. A. Ademoye, "Olfaction-Enhanced Multimedia: Bad for Information Recall?", *IEEE International Conference on Multimedia and Expo 2009 (ICME'09)*, pp. 970-973, 2009.

[4] Murray, N., Qiao, Y., Lee, B., A K, K., Muntean, G.-K, "Subjective Evaluation of Olfactory and Visual Media Synchronization", *Proceedings of the ACM Multimedia Systems Conference 2013 (MMSys)*, Oslo, Norway, 2013.

[5] A. Mochizuki, T. Amada, S. Sawa, T. Takeda, S. Motoyashiki, K. Kohyama, M. Imura, K. Chihara, "Fragra: A Visual-Olfactory VR Game", *Proceedings of the ACM SIGGRAPH'04 Sketches*, p. 123, 2004.

[6] Y. Ishibashi, S. Hoshino, Q. Zheng, N. Fukushima, S. Sugawara, "QoE Assessment of Fairness between Players in Networked Game with Olfaction", *Proceeding of the 11th Annual Workshop on Network and Systems Support for Games (NetGames),* pp. 1-2, 2012.

[7] M. Waltl, C. Timmerer, H. Hellwagner, "Improving the Quality of Multimedia Experience through Sensory Effects", *Proceedings of the 2nd Int'l. Workshop on Quality of Multimedia Experience (QoMEX'10)*, Trondheim, Norway, Jun. 2010.

[8] B. Rainer, M. Waltl, E. Cheng, M. Shujau, C. Timmerer, S. Davis, I. Burnett, C. Ritz, H. Hellwagner, "Investigating the Impact of Sensory Effects on the Quality of Experience and Emotional Response in Web Videos", *Proceedings of the 4th International Workshop on Quality of Multimedia Experience (QoMEX'12)*, Yarra Valley, Australia, pp. 278-283, July 2012.

[9] ISO/IEC 23005, "Information Technology - Media Context and Control", 2011.

[10] T. Zinner, O. Abboid, O. Hohlfeld, T. Hossfeld, P. Tran-Gia, "Towards QoE Management for Scalable Video Streaming", *Proceedings of the 21st ITC Specialist Seminar on Multimedia Application – Traffic, Performance and QoE*, pp. 64-69. 2010.

[11] T. Wang, A. Pervez, H. Zou, "VQM-based QoS/QoE Mapping for Streaming Video", *Proceedings of the 3rd IEEE International Conference on Broadband Network and Multimedia Technology (IC-BNMT)*, pp. 807-812, 2010.

[12] W. Cherif, A. Ksentini, D. Negru, M. Sidibe, "A_PSQA: Efficient real-time video streaming QoE tool in a future media internet context," *IEEE International Conference on Multimedia and Expo (ICME) 2011*, Barcelona, Spain, July 2011.

[13] F. Pereira, "A Triple User Characterization Model for Video Adaptation and Quality of Experience Evaluation", *Proceedings of the 7th Workshop on Multimedia Signal Processing (MMSP)*, Shanghai, China, pp. 1-4, 2005.

[14] ITU-R Rec. BT.500-13, "Methodology for the subjective assessment of the quality of television pictures", 2012.

[15] M. Waltl, C. Timmerer, B. Rainer, H. Hellwagner, "Sensory Effect Dataset and Test Setups," *Proceedings of the 4th International Workshop on Quality of Multimedia Experience (QoMEX'12)*, Yarra Valley, Australia, pp. 115-120, Jul. 2012.

[16] Sensory Experience Lab, http://selab.itec.aau.at. Accessed 30.01.2013.

[17] Fiedler, M., Hossfeld, T., Tran-Gia, P., "A generic quantitative relationship between quality of experience and quality of service", *IEEE Network*, vol.24, no.2, pp.36,41, March-April 2010.

[18] J. A. Rice, "Mathematical Statistics and Data Analysis", *Duxbury Advanced Series, Brooks/Cole*, ISBN 9780534399429, 2009.

[19] T. Keith, "Multiple Regression and Beyond", *Pearson/Allyn and Bacon*, ISBN 9780205326440, 2005.

[20] M. Waltl, C. Timmerer, H. Hellwagner, "Increasing the User Experience of Multimedia Presentations with Sensory Effects", *Proceedings of the 11th Int'l Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS'10)*, Desenzano del Garda, Italy, Apr. 2010.