INCREASING THE USER EXPERIENCE OF MULTIMEDIA PRESENTATIONS WITH SENSORY EFFECTS

Markus Waltl, Christian Timmerer, and Hermann Hellwagner

Multimedia Communication (MMC) Research Group, Institute of Information Technology (ITEC)
Klagenfurt University, Klagenfurt, Austria
E-mail: *firstname.lastname*@itec.uni-klu.ac.at

ABSTRACT

The term Universal Multimedia Experience (UME) has gained momentum and is well recognized within the research community. As this approach puts the user into the center stage, additional complexity is added to the overall quality assessment problem which calls for a scientific framework to capture, measure, quantify, judge, and explain the user experience. In previous work we have proposed the annotation of multimedia content with sensory effect metadata that can be used to stimulate also other senses than vision or audition. In this paper we report first results obtained from subjective tests in the area of sensory effects attached to traditional multimedia presentations such as movies that shall lead to an enhanced, unique, and worthwhile user experience.

1 INTRODUCTION

The Universal Multimedia Experience (UME) is nowadays well known within the research community and has been derived from Universal Multimedia Access (UMA) [1]. UME puts the user into center stage and in previous work we [2] – and also others [3] – have argued that the consumption of multimedia assets may stimulate also other senses than vision or audition, e.g., olfaction, mechanoreception, equilibrioception, or thermoception which shall lead to an enhanced, unique user experience. Therefore, the multimedia assets are enriched with additional metadata describing so-called sensory effects that are rendered on sensory devices like fans, vibration chairs, lamps, perfumer, etc.

However, research on ambient intelligence [4] concluded that there is a need for a scientific framework to capture, measure, quantify, judge, and explain the user experience. This paper contributes to this scientific framework with subjective tests based on our existing test-bed [2] (including extensions) and the conclusions drawn thereof. In anticipation of the results we can report that – depending on the genre – the user experience of multimedia presentations can be increased. As the sensory effects add additional dimensions to the overall quality assessment procedure we had to slightly extend traditional test procedures for our needs.

The remainder of this paper is organized as follows. Section 2 describes the Sensory Effect Media Player (SEMP) which has been developed in order to conduct the subjective tests. Section 3 provides the actual test environment including the extended test procedure and the evaluation results. Conclusions and future work items are presented in Section 4.

2 SENSORY EFFECT MEDIA PLAYER

The **S**ensory **E**ffect **M**edia **P**layer (SEMP) is a DirectShow-based media player with support for the Sensory Effect Description Language (SEDL) as defined in [5]. The amBX system [6] is used for rendering of the sensory effects. It consists of two fans, a wrist rumbler, two sound speakers and a subwoofer, two front lights and a wall washer. SESim as described in [2] only simulates these effects (among others) within a Graphical User Interface (GUI). SEMP, instead, presents these effects within the real world by utilizing the amBX equipment.

Figure 1 depicts the architecture of SEMP comprising a *data input layer* which is used for loading the Sensory Effect Metadata (SEM) and the corresponding audio/video resource. A SEM conforms to SEDL and describes various effects (e.g., light, vibration, wind). The *application core* consists of a SEM parser responsible for extracting the effects specified. The extracted effects are forwarded to the *effect manager* via the control component (*media player)*. The effect manager stores the effects and creates amBX commands for enabling/disabling the amBX devices with their corresponding parameters. For knowing when to activate an effect, the effect manager uses the timestamp received from *DirectShow*. DirectShow is used for handling the video. It renders the video within the video player and sends time and audio information to the (optional) *On-Screen-Display (OSD)*. Furthermore, DirectShow forwards the video to a so-called *SampleGrabber* which extracts a frame every 0.1 second which is used by the *average color calculator* to determine the average color to be used for the additional light effects. In particular, the frame is divided into three parts (left, center and right) and the calculated color for each part is send to the effect manager which steers the corresponding amBX lights. That is, the left third of the frame is used for the color of the light located on the left side of the screen, the

Table 1. New five-level enhancement scale.

5	Big enhancement
4	Little enhancement
3	Imperceptible
2	Annoying
1	Very annoying

middle third of the frame specifies the color for the wall washer behind the screen, and the right third of the frame defines the color of the light at the right side of the screen. The splitting of the frame was performed to allow for a more fine-grained color partitioning based on the content of the video.

Vibration and wind effects are not generated automatically. They are annotated manually by using SEVino [2].

3 TEST-ENVIRONMENT

The test procedure is based on the ITU-T Recommendation P.911 [7] and the setup of the environment is partly based on [8].

3.1 Location

All sessions of the experiment are conducted in an isolated room under the same ambient conditions. Before the session the following conditions should prevail:

- All nonessential electronic equipment is turned off.
- Telephones are unplugged.
- Windows are closed and covered with translucent blankets.
- All overhead lights are turned off.
- The entry door to the room is closed.
- A “Do not disturb” sign is placed on the outside of the door.
- The participant is asked to turn off any audible pagers, mobile phones, and/or watches.
- A ceil flooder is switched on to illuminate the room in a warm light.

3.2 Participants

For reliable results there should be between 10 and 20 participants. We have invited 25 participants. The participants were not directly involved in the work area and were not experienced assessors. Furthermore, there were a nearly equal number of male and female participants for getting a better range of results.

Before a session the participants were informed about the type of assessment, the opinion scale and the presentation of the test sequences. Questions were asked and answered only before the start of the session.

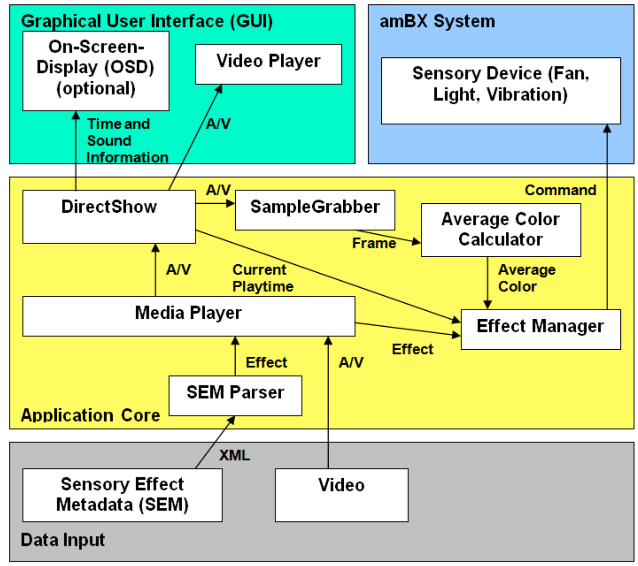

Figure 1. Architecture of the Sensory Effect Media Player.

3.3 Apparatus

The tests were conducted using the following hardware and software:

- Dell Precision 360: Pentium 4 3.2 GHz with 1 GB RAM and NVidia Quadro NVS (64 MB)
- amBX Premium Kit (Fan, Vibrator, Light, Sound)
- 26’’ Monitor with a resolution of 1680x1050
- Windows XP SP3
- Sensory Effect Media Player (SEMP)
- amBX Software (i.e., amBX System 1.1.3.2 and Philips amBX 1.04.0003)

3.4 Procedure for Evaluation

The test setup consists of a control station and the actual test computer. The control station was used to start the test sequences and for resolving possible playback issues of SEMP.

The participants sat down in a comfortable seat to get the best movie feeling and were placed within a distance of three times the height of the monitor.

The ITU-T Rec. P.911 defines the *Degradation Category Rating* (DCR) [7] which was used for the tests. This method takes a reference sequence and a modified sequence. In our tests the modified sequence is enriched with additional sensory effects (i.e., wind, vibration, and light). The ITU-T Rec. P.911 also defines a five-level impairment scale which we adopted for our needs as shown in Table 1. The table shows a new five-level enhancement scale because we would like to go beyond the traditional evaluation of the impairment of a video sequence. That is, in order to evaluate the experience of sensory effects we introduce an enhancement evaluation of a given video sequence.

Furthermore, the DCR defines sequences of around 10 seconds but the length was increased in order to allow for

Table 2. Video Sequences.

	Bit-rate (kbit/s)	Resolution	Duration (sec)	Genre	Nr. of Effects
Rambo 4	6484	1280 x 544	58.11	Action	10
ZIB Flash	8000.4	1024 x 576	83.04	News	1
Babylon A.D.	7141.4	1280 x 544	118.42	Action	28
Wo ist Klaus?	4327.3	1024 x 576	59.16	Commercial	16
Earth	6868.2	1280 x 720	66	Documentary	24
Formula 1	4886.2	1280 x 720	116.2	Sports	43

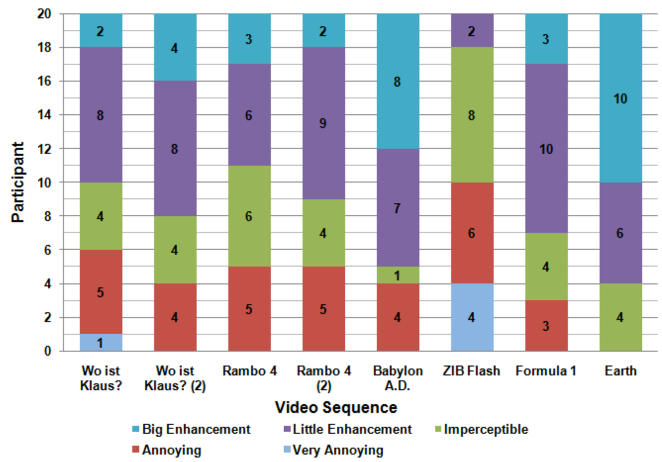

Figure 2. Evaluation results.

more sensory effects within one sequence. Table 2 shows the used video sequences with their bit-rates, resolution, duration and genre. It also depicts the number of effects for each sequence including wind and vibration. Note that light is not included because it is calculated automatically every 0.1 second which approximately leads to "video duration / 0.1" light effects. For the test we had in total eight video sequences. Two sequences (i.e., Rambo 4, Wo ist Klaus?) were presented twice but not directly one after the other. The reason for presenting them twice was to test the reliability of the participants. Additionally, the order of sequences was randomized for each participant.

As defined we presented first the reference sequence and second the same sequence enriched with sensory effects with a two second break in between. At the end of each paired presentation the participant should evaluate – within five seconds – the enhancement of the second sequence with respect to the reference sequence using the new five-level enhancement scale introduced above. This procedure conforms to the definition of DCR stated in [7]. It is important to note that the evaluation reflects the participants' overall opinion of the audio/video resource and sensory effect quality.

After all sequences were displayed the participant had to answer the post-experiment questions for which the participants had no time limit. The post experimentation part was used to receive more pieces of information about possible experimentation errors (e.g., too long sequences, too few sequences) or possible experimentation enhancements. The following questions were asked during the post-experiment part:

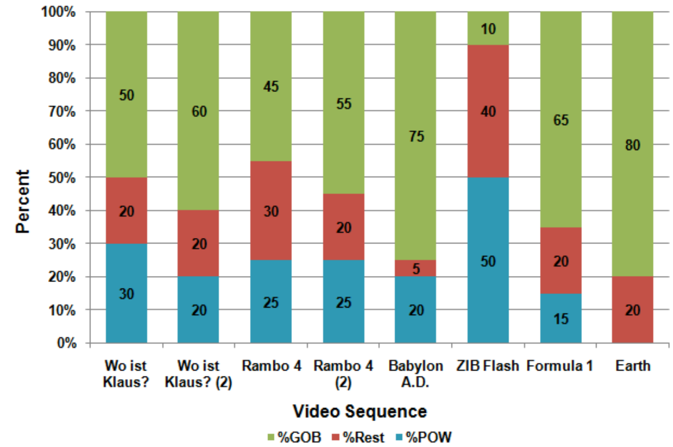

Figure 3. %GOB and %POW.

- How easy or difficult was it to determine the enhancement of the video?
- Would you have liked less or more time to hear/see the sequence with sensory effects?
- Did you direct your attention to any specific sensory effect when determining the quality of the experience?
- Where you ever mentally overloaded during any part of the experiment?
- Have you participated in an experiment similar to this one?
- Any other comments about what you liked or did not like, or things that should be changed during the course of this experiment?

The overall time of the experiment was around 30 minutes per participant.

3.5 Evaluation Results

25 students participated in the user study. There were 13 female and 12 male students between the age of 20 and 31 years. The students were not involved with the evaluation topic and had not been familiar with computer science or signal processing. The majority of students were coming from journalism and psychology. Furthermore, they have never been to a subjective test like this one. We eliminated five outliers according to the suggestion given in [9].

Figure 2 shows how many participants have rated which video in which category. The results show that effects have different influences at different genres. For example, the action, sports and documentary genre benefits from the additional effects. Only a few participants did not like the sensory effects for these genres. It is important to note that the enhancement strongly depends on the content. For example, *Rambo 4* is also an action sequence but the participants were not that keen of the effects. The commercial genre can also profit from the additional effects. Only the news genre will not profit from these effects due to the situation reports, e.g., still picture in background and only moderator talks about the event.

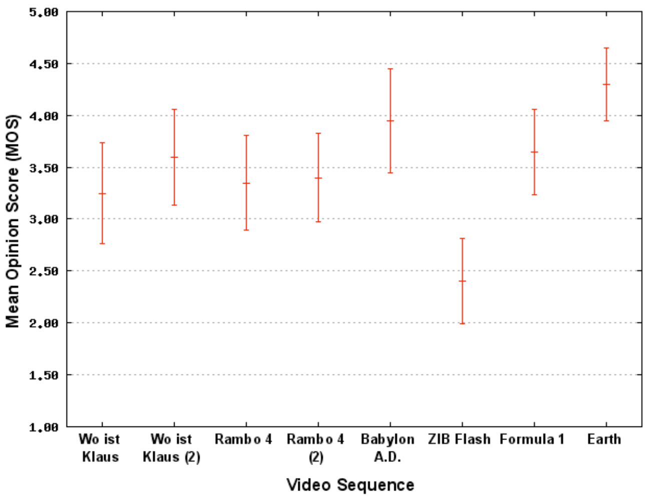

Figure 4. Mean opinion score and confidence interval.

Figure 3 depicts the results in percent of good or better (%GOB) by using the values from *Big Enhancement* and *Little Enhancement*. Furthermore, it also depicts the results in percent of poor or worse (%POW) by using the values from *Annoying* and *Very Annoying*. %Rest comprises the amount of values for *Imperceptible*. The diagram shows very good the preferences of the participants. For example, more than the half of the participants rated the news report as poor or imperceptible. Furthermore, one can see that the documentary and the action movies were rated very high.

In this paper we do not show results differentiated between female and male participants for two reasons. First of all, the number of participants was not big enough to be representative if split up. Second, the results vary not significantly between female and male participants.

Figure 4 presents the mean opinion score (MOS) and confidence interval for every video sequence. Here a 95% confidence interval is used. The figure clearly shows the low MOS for news and the high MOS for action and documentary. Furthermore, it depicts that the two videos presented twice differ in the results. This leads to the assumption that sensory effects will enhance the user experience the more often a video sequence with sensory effects is presented. However, this is also an indication that the test method may not function properly for this kind of content or evaluation. We will further study this in the future.

The evaluation of the post experiment questions gave the following results: first, the majority of the participants stated that they had *no* (15%) to *minor* (55%) difficulties to determine the enhancement of a video with sensory effects. 20% of the participants declared that they had *medium* difficulties. Only a small number of participants had *much* (5%) or *very much* (5%) difficulties to determine the enhancement. Second, they would prefer to have *much* (10%) or *little more* (40%) time to hear/see sequences with the sensory effects. 30% stated that the length was sufficient and the rest wanted *little less* (10%) or *less* (10%) time. The individual comments to the post-experiment questions can be summarized as the following: the vibration effects were annoying during the sequences and because of that most of the participants paid attention to the vibration effects. This leads to the assumption that vibration effects shall be used carefully in order to reduce the risk of producing an annoying impression to the user.

4 CONCLUSIONS AND FUTURE WORK

This paper presents a tool, the Sensory Effect Media Player, for testing sensory effects for an increased user experience. Furthermore, it describes one of the first subjective tests done in the area of sensory effects combined with the usage of SEDL. The results look promising and indicate that sensory effects increase the user experience for action, documentary, and sport movies and have little effect for news and commercials. Furthermore, the results depict that the evaluation method used for the tests is suitable for movies but not ideal for sensory effects evaluation.

Future work items include further subjective tests with different combinations of sensory effects but also video qualities. Additionally, we will work on standardizing a quality assessment method for sensory effects.

5 ACKNOWLEDGMENT

This work was supported in part by the European Commission in the context of the NoE INTERMEDIA (NoE 038419), P2P-Next project (FP7-ICT-216217), and ALICANTE project (FP7-ICT-248652).

6 REFERENCES

[1] F. Pereira, I. Burnett, “Universal Multimedia Experiences for Tomorrow,” *IEEE Signal Processing Magazine*, vol. 20, no. 2, Mar. 2003, pp. 63-73.

[2] M. Waltl, C. Timmerer, H. Hellwagner, "A Test-Bed for Quality of Multimedia Experience Evaluation of Sensory Effects", *Proc. 1st Int’l Workshop on Quality of Multimedia Experience (QoMEX’09)*, San Diego, USA, Jul. 2009.

[3] B.-S. Choi, S. Joo, H. Lee, “Sensory Effect Metadata for SMMD Media Service”, *Proc. 4th Int’l Conf. on Internet and Web Applications and Services (ICIW’09)*, Venice/Mestre, Italy, May 2009, pp. 649–654.

[4] E. Aarts, B. de Ruyter, “New research perspectives on Ambient Intelligence”, *Journal of Ambient Intelligence and Smart Environments, IOS Press*, vol. 1, no. 1, 2009, pp. 5–14.

[5] C. Timmerer, S. K. Kim, J. Ryu, (eds.) “Committee Draft of ISO/IEC 23005 Sensory Information,” *ISO/IEC JTC 1/SC 29/WG 11/N10785*, London, UK, Jul. 2009.

[6] amBX UK Ltd., http://www.ambx.com. (last accessed: Feb. 2010)

[7] ITU-T Rec. P.911, "Subjective audiovisual quality assessment methods for multimedia applications", Dec. 1998.

[8] R. L. Storms, M. J. Zyda, "Interactions in Perceived Quality of Auditory-Visual Displays", *Presence: Teleoperators and Virtual Environments*, vol. 9, no. 6, Dec. 2000, pp.557-580.

[9] ITU-R Rec. BT.500-11, "Methodology for the subjective assessment of the quality of television pictures", Jan. 2002.