Signal Processing: *Image Communication* 27 (2012) 909–916

Contents lists available at SciVerse ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

Assessing the quality of sensory experience for multimedia presentations ☆

Christian Timmerer*, Markus Waltl, Benjamin Rainer, Hermann Hellwagner

Alpen-Adria-Universität Klagenfurt (AAU), Institute of Information Technology (ITEC), Multimedia Communication (MMC), Universitätsstraße 65-67, 9020 Klagenfurt am Wörthersee, Austria

ARTICLE INFO

Available online 2 February 2012

Keywords:
Quality of Experience
Sensory experience
Subjective quality assessment
Experimental results
MPEG-V

ABSTRACT

This paper introduces the concept of sensory experience by utilizing sensory effects such as wind or lighting as another dimension which contributes to the quality of the user experience. In particular, we utilize a representation format for sensory effects that are attached to traditional multimedia resources such as audio, video, and image contents. Sensory effects (e.g., wind, lighting, explosion, heat, cold) are rendered on special devices (e.g., fans, ambient lights, motion chair, air condition) in synchronization with the traditional multimedia resources and shall stimulate other senses than audition and vision (e.g., mechanoreception, equilibrioception, thermoreception), with the intention to increase the users Quality of Experience (QoE). In particular, the paper provides a comprehensive introduction into the concept of sensory experience, its assessment in terms of the QoE, and related standardization and implementation efforts. Finally, we will highlight open issues and research challenges including future work.

1. Introduction and motivation

Multimedia content (i.e., combinations of text, graphics, images, audio, and video) has become omnipresent in our lives. Each day we consume dozens of multimedia assets when reading electronic newspaper, listening to podcasts or Internet radio, and watching digital television (TV). The quality of the multimedia content as perceived by the end user was and still is a challenging research topic, not only since the development of the E-model for audio. Recently, 3D content and technology (e.g., 3DTVs) have entered the consumer market opening another dimension of quality yet not being widely researched. In particular, the Video Quality Experts Group (VQEG) is working in the field of video quality assessment and is currently running a project among others in the area of 3DTV.

In our work we target yet another quality dimension addressing human senses that go beyond audition and vision. The consumption of multimedia assets may stimulate also other senses, such as olfaction, mechanoreception, equilibrioception, or thermoreception, opening a number of new issues with respect to the QoE that we find worth investigating. This work item was motivated by conclusions drawn from the research on ambient intelligence. That is, there is a need for a scientific framework to capture, measure, quantify, and judge the user experience [1]. In our approach the multimedia assets are annotated with sensory information describing sensory effects (e.g., additional ambient light effects, wind, vibration, scent, water spraying) which are synchronized with the actual multimedia assets and

☆ This work was supported in part by the European Commission in the context of the QUALINET (COST IC 1003) and ALICANTE (FP7-ICT-248652) projects.

* Corresponding author.
E-mail addresses: Christian.Timmerer@itec.uni-klu.ac.at (C. Timmerer), Markus.Waltl@itec.uni-klu.ac.at (M. Waltl), Benjamin.Rainer@itec.uni-klu.ac.at (B. Rainer), Hermann.Hellwagner@itec.uni-klu.ac.at (H. Hellwagner).

0923-5965/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.image.2012.01.016

rendered on appropriate devices (e.g., ambient lights, fans, motion chairs, scent vaporizer, water sprayer, etc.). The ultimate goal of this approach is that the user will also perceive these additional sensory effects giving her/him the sensation of being part of the particular multimedia asset and resulting in a worthwhile, informative user experience. In the context of this work, this kind of user experience is referred to as sensory experience.

Therefore, we have built a test-bed based on existing hardware devices [2] (incl. extensions) and conducted various subjective tests [3–5]. The aim of this paper is to provide a comprehensive introduction into the concept of sensory experience, its assessment in terms of the QoE, and related standardization and implementation efforts. Furthermore, we will highlight open issues and research challenges.

The remainder of this paper is organized as follows. Related work is discussed in Section 2. Section 3 describes the concept, system architecture, standardization and implementation aspect of sensory information and Section 4 provides the major findings from subjective quality assessments conducted so far. Section 5 is dedicated to open issues and research challenges. The paper is concluded with Section 6 including future work.

2. Related work

New research perspectives on ambient intelligence are presented in [1] which includes also sensory experiences calling for a scientific framework to capture, measure, quantify, judge, and explain the user experience. Thus, this paper is regarded as a major source of inspiration for our work which aims at contributing to this framework. In [6] the same authors report—based on user studies—that additional light effects are highly appreciated for both audio and visual contents.

In the context of the MPEG-V standardization some work has been published recently related to sensory experience and worth mentioning here, i.e., [7–9]. In [7] authors introduce a new generation of media service called Single Media Multiple Devices (SMMD) which is based on SEM as defined in MPEG-V. In particular, the SMMD media controller is described which maps sensory effects on appropriate sensory devices for the proper rendering thereof. The main focus of [7] is clearly implementation/engineering aspects whereas we concentrate on the QoE. Paper [8] can be regarded as an earlier version of [7] and additionally puts it in the context of UPnP, thus, focusing also on implementation/engineering aspects. In [9] authors present a framework for 4-D broadcasting based on MPEG-V, i.e., the main focus is on delivering SEM in the MPEG-2 Transport Stream and its decoding within the home network environment including the actual service discovery.

Note that sensory effects are not limited to installations, e.g., in home environments, there is already research to bring sensory effects to mobile devices [10]. Furthermore, Kim et al. [11] introduce—among others—new location-based mobile multimedia technology using ubiquitous sensor network-based five senses content. The temporal boundaries within which olfactory data can be used to enhance multimedia applications is investigated in [12] concluding that olfaction ahead of multimedia content is more tolerable than olfaction behind content.

Another area that is related to our work is multisensory research (e.g., [13]) which investigates how different senses interact and how their input is integrated to communicate with one another.

Finally, Grega et al. [14] provide a good overview of the state-of-the-art in QoE evaluation for multimedia services with a focus on subjective evaluation methods which leads us to related work in the area of QoE models such as [15–17]. Most of these models are restricted to a single modality (i.e., audio, image, or video only) or a simple combination of two modalities (i.e., audio and video). For the combination of audio and video content one may employ the basic quality model for multimedia as described in [15]. Another approach is known as the IQX hypothesis formulated as an exponential function [16]. In [17] a triple user characterization model for video adaptation and QoE evaluation is described that introduces at least three quality evaluation dimensions, namely sensorial (e.g., sharpness, brightness), perceptual (e.g., what/where is the content), and emotional (e.g., feeling, sensation) evaluation. Furthermore, it proposes adaptation techniques for the multimedia content and quality metrics associated with each of these layers. The focus is clearly on how an audio/visual resource is perceived, possibly taking into account certain user characteristics (e.g., handicaps) or natural environment conditions (e.g., illumination).

3. Sensory experience: concept, system architecture, standardization, and implementation

3.1. Concept and architecture overview

The concept and system architecture of receiving sensory effects in addition to audio/visual content is depicted in Fig. 1. The media and the corresponding Sensory Effect Metadata (SEM) may be obtained from a Digital Versatile Disc (DVD), Blu-ray Disc (BD), or any kind of online service (e.g., download/play or streaming portal). The media processing engine acts as the mediation device and is responsible for playing the actual media resource and accompanying sensory effects in a synchronized way based on the users setup in terms of both media and sensory effect rendering. Therefore, the media processing engine may adapt both the media resource and the SEM (and, consequently, the corresponding effects) according to the capabilities of the various rendering devices. The users digital living room is extended with additional rendering devices enabling the (increased) stimulation of senses other than audition and vision. For example, a motion chair, fan/ventilator, heater/cooler, etc. may be used to address the somatosensory (human sensory) sub-system, whereas scent vaporizer device stimulates the olfactory sub-system. The visual sub-system may be further stimulated using (additional) ambient light devices. Note that the term sub-system refers to the human sensory system comprising the sub-systems visual, auditory, somatosensory, gustatory, and olfactory.

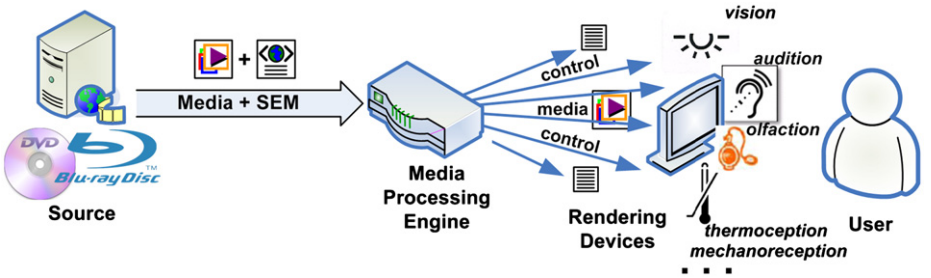

Fig. 1. Concept and system architecture of sensory experience [2].

3.2. Sensory Effect Description Language

The metadata format for describing such sensory effects is defined by ISO/MPEG in the context of MPEG-V Media Context and Control. In particular, Sensory Information (Part 3) [18] defines a Sensory Effect Description Language (SEDL), an XML Schema-based language, which enables one to describe sensory effects.

The actual sensory effects are not part of SEDL but defined within the Sensory Effect Vocabulary (SEV) for extensibility and flexibility, allowing each application domain to define its own sensory effects. A description conforming to SEDL is referred to as Sensory Effect Metadata (SEM) and may be associated with any kind of multimedia content (e.g., movies, music, Web sites, games). The SEM is used to steer sensory devices like fans, vibration chairs, lamps, etc. via an appropriate mediation device in order to enrich the experience of the user. That is, in addition to the audio–visual content of, e.g., a movie, the user will also perceive other effects such as the ones described above with the aim to improve the users' QoE.

The current syntax and semantics of SEDL are specified in [18]. However, in this paper we provide an EBNF (Extended BackusNaur Form)-like overview of SEDL due to the verbosity of XML. In the following the EBNF will be described.

```
SEM ::=timeScale [autoExtraction]
    [DescriptionMetadata] (Declarations|
    GroupOfEffects|Effect|ReferenceEffect)+
```

SEM is the root element which contains a *timeScale* attribute that defines the time scale used for the sensory effects within that description (i.e., the number of ticks per second). Furthermore, it contains an optional *autoExtraction* attribute and *DescriptionMetadata* element followed by choices of *Declarations*, *GroupOfEffects*, *Effect*, and *ReferenceEffect* elements. The *autoExtraction* attribute is used to signal whether automatic extraction of sensory effects from the media resource is preferable. The *DescriptionMetadata* provides information about the SEM itself (e.g., authoring information) and aliases for classification schemes (CS) used throughout the whole description. Therefore, appropriate MPEG-7 description schemes [19] and CS defined in [18] are used, which are not further detailed here.

```
Declarations ::=(GroupOfEffects|Effect|
    Parameter)+
```

The *Declarations* element is used to define a set of SEDL elements without instantiating them for later use in a SEM via an internal reference. In particular, the *Parameter* may be used to define common settings used by several sensory effects similar to variables in programming languages.

```
GroupOfEffects ::=
    timestamp [BaseAttributes]
    2*(EffectDefinition|ReferenceEffect)
    (EffectDefinition|ReferenceEffect)*
```

GroupOfEffects provides an author the possibility to reduce the size of a SEM description by grouping multiple effects sharing the same timestamp or *BaseAttributes* (cf. below). The timestamp provides information about the point in time when this group of effects should become available for the application. Depending on the application this information can be used for rendering and synchronization purposes with the associated media. The timestamp is provided as XML Streaming Instructions as defined in MPEG-21 Digital Item Adaptation [20]. Furthermore, a *GroupOfEffects* shall contain either at least two *EffectDefinition* or *ReferenceEffect*. The *EffectDefintion* comprises all information pertaining to a single sensory effect whereas the *ReferenceEffect* provides a reference to a previously declared *EffectDefinition*.

```
Effect ::=timestamp EffectDefinition
```

An *Effect* describes a single sensory effect (e.g., wind effect) with an associated *timestamp*.

```
EffectDefinition ::=
  [SupplementalInformation]
  [BaseAttributes]
```

An *EffectDefinition* may have a *SupplementalInformation* element for defining a reference region from which the effect information may be extracted in case *autoExtraction* is enabled. Furthermore, several optional attributes are defined which are called *BaseAttributes* and described in the following.

```
BaseAttributes ::=
  [activate][duration][intensity-value]
  [intensity-range][fade][priority]
  [location][alt][adaptability]
  [autoExtraction]
```

activate describes whether an effect shall be activated or deactivated; *duration* describes how long an effect shall be activated; *intensity-value* indicates the actual strength of the effect within a given *intensity-range* (note that the actual semantics and the scale/unit are defined for each

effect individually); *fade* provides means for fading an effect to the given *intensity-value*; *priority* defines the priority of an effect with respect to other effects within a group of effects; *location* describes the position from where the effect is expected to be perceived from the users perspective (i.e., a three-dimension space with the user in the center is defined in the standard); *alt* describes an alternative effect identifier by a URI (e.g., in case the original effect cannot be rendered); *adaptability* attributes enable the description of the preferred type of adaptation with a given upper and lower bound; *autoExtraction* with the same semantics as above but only for a certain effect.

3.3. Sensory Effect Vocabulary

The Sensory Effect Vocabulary (SEV) defines a set of sensory effects to be used within SEDL in an extensible and flexible way. That is, it can be easily extended with new effects or by derivation of existing effects thanks to the extensibility feature of XML Schema. The SEV is defined in a way that the effects are abstracted from the authors intention and be independent from the users device settings. This mapping is usually provided by the media processing engine and deliberately not defined in this standard, i.e., it is left open for industry competition. It is important to note that there is not necessarily a one-to-one mapping between elements or data types of the Sensory Effect Metadata and sensory device capabilities. For example, the effect of hot/cold wind may be rendered on a single device with two capabilities, i.e., a heater/air conditioner and a fan/ventilator. Currently, the standard defines the following sensory effects.

Light, *colored light*, and *flash light* for describing light effects with the intensity in terms of illumination expressed in lux. For color information there are three possibilities: First, color can be presented by using a classification scheme (CS) which is defined by the standard comprising a comprehensive list of common colors. Second, color information can be defined by the author via the hexadecimal color format known from HTML (e.g., #2A55FF). Third, color can be automatically extracted from the associated content (e.g., average color of a video frame). The flash light effect extends the basic light effect by adding the frequency of the flickering in times per second.

Temperature enables describing a temperature effect of heating/cooling with respect to the Celsius scale. *Wind* provides a wind effect where it is possible to define its strength with respect to the Beaufort scale. *Vibration* allows one to describe a vibration effect with its strength according to the Richter magnitude scale. For the *water sprayer*, *scent*, and *fog* effect the intensity is provided in terms of ml/h. Furthermore, the scent effect may use a set of pre-defined scent definitions via a corresponding CS.

Color correction provides means to define parameters that may be used to adjust the color information of a media resource to the capabilities of end user devices or impaired end users. For example, it is possible to adjust the color of the media resource to provide color blind users with a better experience than without the adjustment. Furthermore, the color correction allows the author to define regions of interest where it should be applied in case this is desirable (e.g., black/white movies with one additional color such as red).

Rigid body motion, *passive kinesthetic motion*, *passive kinesthetic force*, *active kinesthetic* and *tactile* describes a set of effects which may be used for kinesthetic and tactile devices. For example, the movement of a special pen is stored in a SEM description and after the user takes the pen it moves his/her hand to guide/demonstrate how a plan is drawn.

3.4. Usage example

In this section we provide snippets of SEM descriptions with an in-depth description how a *media processing engine* should handle this description to control sensory devices. Let us assume that we have a Web portal with different types of video like, for example, YouTube. In particular, one of the videos shows a scene of a boat on the open sea which may be annotated with the following sensory effects: *wind* and *temperature* based on the cold/warm breeze on the open sea, *rigid body motion* based on the boat movements, and *colored light* based on the color information within the video. As mentioned earlier the light effects could be calculated automatically from the content or defined manually. Listing 1 shows an excerpt for a colored light effect that is defined by the author. In this example blue lights will be presented at all light devices that are located in the center front of the user regardless if the light is above, below or directly in front of the user. The color is defined via the CS term for blue (i.e., #0000FF) but the hexadecimal value could also be used.

Listing 1. Example for a colored light effect.

```
<sedl:Effect xsi:type="sev:LightType"
color="urn:mpeg:mpeg-v:01-SI-ColorCS-NS:blue"
location="urn:...:center:front:*"
  si:pts="..."
.../>
```

The light breeze on the open sea could be defined by a wind effect accompanied by a temperature effect. Listing 2 presents the corresponding excerpt of a SEM description.

Listing 2. Example for a group of effects.

```
<sedl:GroupOfEffects si:pts="..."
duration="100"
location="urn:...:center:front:middle">
 <sedl:Effect
  xsi:type="sev:TemperatureType"
  intensity-value="0.393"
  intensity-range="0 1"/>
 <sedl:Effect xsi:type="sev:WindType"
  intensity-value="0.082"
  intensity-range="0 1"/>
</sedl:GroupOfEffects>
```

The group of effects comprises two effects that share the attributes defined within the *GroupOfEffects* element. This means that the enclosed effects start at the same timestamp as defined via the `si:pts` attribute. Furthermore, both effects have the same duration and the same `location`, i.e., the effects are perceived from the front with respect to the user which is indicated by `center`, `front`, and `middle`, respectively.

The first element within the group of effects describes a temperature effect indicated by `sev:TemperatureType`. This effect is responsible for rendering the temperature of the breeze. The effect defines a temperature of 0.393 on a range from 0 to 1. Note that this range is mapped by the media processing engine to the temperature scale supported by the device. Alternatively, one can also use a temperature range from [−30, +40] and an intensity value of about +20. The temperature effect can use, for example, an air-conditioner to provide the desired heating/cooling.

The second element, i.e., `sev:WindType`, is responsible to render the light breeze which is around 0.082 on a range from 0 to 1. Again, the media processing engine maps the capabilities of the actual devices rendering the effect. On the other hand, the author of the SEM description could have also stated the minimum and maximum range in terms of the Beaufort scale, i.e., [0, 13] and set the intensity of the effect to around 1. This effect can be rendered by fans (or ventilators) which are deployed around the user.

Finally, the movement of the boat may be handled by the rigid body motion effect as shown in Listing 3.

Listing 3. Example for a rigid body motion effect.

```
<sedl:Effect
 xsi:type="sev:RigidBodyMotionType"
 si:pts="...">
 <sev:Wave direction=":WAVE:left-right"
  startDirection=":WAVESTR:up"
  distance="10"/>
</sedl:Effect>
```

Assuming that the sea is very calm and the boat only moves slightly we can generate a movement of the boat that moves 10 cm up and down. The waves are simulated with a movement from left to right, starting with an upward motion.

3.5. Implementation

In order to conduct subjective quality assessments we have used, integrated, and implemented the following hardware/software components:

- Off-the-shelf *amBX system & SDK* [21] comprising left and right 2.1 sound speaker lights with a sub-woofer, a wall washer, a set of fans, and a wrist rumbler including an appropriate SDK in order to control these devices.
- *Sensory Effect Video Annotation (SEVino)* tool [2] allows for describing sensory effects for a video sequence. It is based on Java and provides means for simply entering and editing of sensory effects.
- *Sensory Effect Metadata Player (SEMP)* [3] is a DirectShow-based media player with support for sensory effects and the amBX system.
- *Sensory Effect Simulation (SESim)* tool [2] has been developed for the evaluation of SEM generated by SEVino.

Please note that ambient light devices are not controlled via SEM because within SEMP an automatic color calculation is deployed. The advantage of the automatic color calculation is that it reduces the description size because light effects do not have to be described explicitly which also speeds up the authoring process. However, different automatic color calculation methods may lead to different user experiences and therefore we have implemented four different algorithms that control the light devices: (1) *Average color in the RGB* color space: the average color is calculated based on the pixel value average. (2–4) *Dominant color in the RGB, HSV, and HMMD* [19] color space: these algorithms use the dominant color according to the RGB, HSV, and HMMD color spaces, respectively.

HSV and HMMD are used since these color spaces are closer to the human perception of color than RGB. However, the major problem with the color calculation is that it requires a lot of computational resources. In particular, the dominant color algorithm needs much more computational resources than the average color algorithm due to the management of color bins for determining the dominant color for a frame. Please note that the amBX system supports only RGB values which requires additional computational resources due to the back-transformation from HSV/HMMD to RGB. A detailed performance evaluation is given in [2] and only major findings are summarized here.

Using the average color for the automatic color calculation enables an immediate reaction to color changes in the content resulting in appealing effects with low computational requirements and, thus, suitable for real-time extraction.

The HSV and HMMD dominant color algorithms provide a smoother reaction to color changes in the content but have higher computational requirements. Therefore, real-time extraction is not achievable on low-end devices and, thus, additional metadata support would be required. That is, the color information is not extracted from the media resource but provided as metadata either within the Sensory Effect Metadata or as, e.g., MPEG-7 description.

4. Improving the Quality of Experience

In order to study the impact on the QoE when consuming multimedia assets annotated with sensory effects, we have conducted three subjective quality assessments so far with slightly different contexts and goals as well as partially utilizing different methods.

In all cases, we have adopted methods defined by ITU-T P.910 [22] and P.911 [23], respectively. The setup for our experiments, i.e., location, participants, apparatus, and procedure for evaluation, are described in detail in [3–5] and only briefly described here. For all subjective tests we invited around 20 participants, equally distributed among males and females, and not familiar with the subject which conforms to guidelines defined in [22,23]. The test sequences have been carefully selected in terms of content, genre, and qualities (when needed) and manually annotated with different sensory effects. For all tests, the same setup has been used which was inspired by and partially based on [24].

In the following subsections, we will provide a concise summary of the results obtained from these experiments including major findings.

4.1. Experiment I

In our first experiment [3], we demonstrated that sensory effects provide a vital tool for enhancing the user experience depending on the actual genre. Therefore, we gathered test sequences of different genres, i.e., action (Rambo 4, Babylon A.D.), news (ZIB Flash), documentary (Earth), commercials (Wo ist Klaus), and sports (Formula 1), and annotated them with various Sensory Effect Metadata, i.e., wind, vibration, and light effects. Note that light effects are not actually part of SEM but extracted automatically from the video content [2]. The sequences were chosen carefully to have all different types of effects within each sequence.

For the actual method, we adopted the Double Stimulus Impairment Scale (DSIS) also known as Degradation Category Rating (DCR) [23] and turned the five-level impairment scale into a new five-level enhancement scale. That is, the subjects rate on the enhancement of a stimulus annotated with sensory effects compared to a reference stimulus without sensory effects rather than on the impairment. The quality of the video content (independent of the sensory effects) was equal for both sequences of the same genre.

The detailed evaluation results are given in [3]. The mean opinion score (MOS) with a confidence interval of 95% is depicted in Fig. 2. The *x*-axis shows the name of the sequences. As one can see, two sequences were presented twice but not directly one after the other in order to test the reliability of the participants. Additionally, the order of the sequences was randomized for each participant.

The figure clearly shows the lower MOS for news compared to the higher MOS for action and documentary genres. In particular, the action, sports, and documentary genres benefit more from these additional effects. Interestingly, although Rambo 4 and Babylon A.D. are from the same genre, the results differ slightly. The commercial genre can also profit from the additional effects but not at the same level as documentary. Only the news genre will not profit from these effects. Furthermore, the figure also depicts that the two videos presented twice differ in the results which is also an indication that the test method may not have functioned properly for this kind of content and/or evaluation.

4.2. Experiment II

The aim of our second experiment [4] was to investigate the relationship of the QoE to various video bit-rates of multimedia contents annotated with sensory effects. In particular, we were interested in the subjective quality gap between video resources annotated with and without sensory effects at different bit-rates.

The overall setup of the second experiment was similar to the first one. The test stimuli comprise the two best performing video sequences from our first experiment. For each sequence, four versions with different bit-rates were prepared whereby only the video bit-rate was affected and the audio bit-rate remained constant for all versions of a given sequence. Additionally, each sequence has been annotated with sensory effects resulting in 16 different bit-streams to be evaluated. For the actual subjective assessment, we have adopted the Absolute Category Rating with Hidden Reference (ACR-HR) method using a five-point discrete scale from excellent to bad as defined in [22].

Like in the previous subsection, the detailed evaluation results are given in [4]. Thus, we will only concentrate on the MOS values depending on various bitrates as depicted in Fig. 3 (sequence Earth, i.e., the documentary, only due to space constraints; results for the other sequences are similar). Interestingly, the sequences with sensory effects have always a higher MOS than their counterparts without sensory effects and almost steadily increase for higher PSNR/bit-rates.

In general, the results confirm the observations from the previous experiment (cf. Section 4.1). Additionally, Fig. 3 also shows that the MOS of the lowest bit-rate version with sensory effects is always higher than the MOS of all higher bit-rate variants without sensory effects. Furthermore, we calculated the average difference between the two curves using the Bjontegaard Delta (BD) method [25] with the result that the sequence enriched

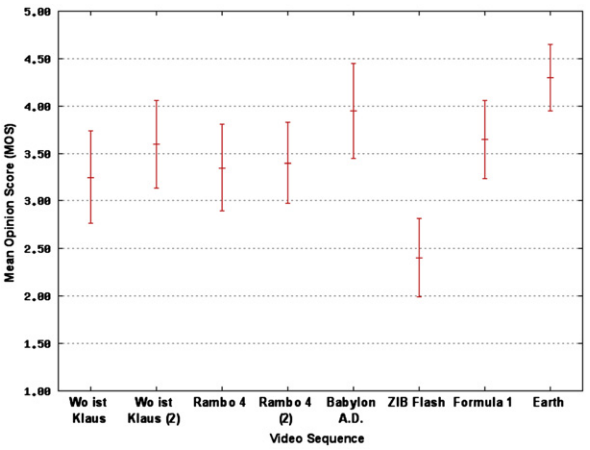

Fig. 2. MOS and confidence interval from Experiment I [3].

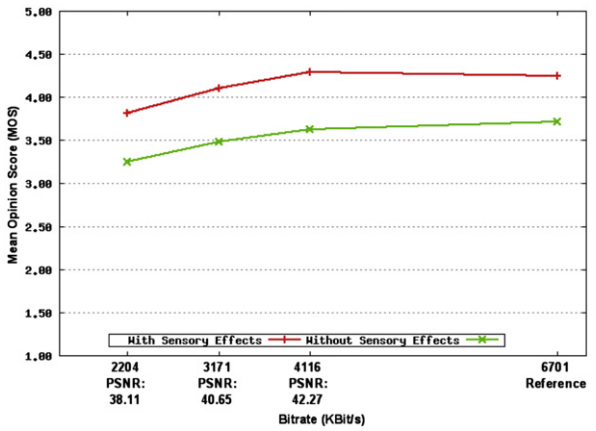

Fig. 3. MOS versus PSNR/bitrate for earth sequence from Experiment II [4].

with sensory effects is 0.6 MOS points higher than without sensory effects (0.5 MOS points on average for both sequences).

4.3. Experiment III

In the third experiment we modified the context in order to evaluate the sensory experience in the World Wide Web (WWW) [5]. Therefore, we have implemented a Web browser plug-in which is capable to render sensory effects. In its first version we have focused on light effects that can be automatically extracted from the video content without the need for additional metadata. Furthermore, we have conducted two formal subjective quality assessments: First, we investigated the benefit of Web videos (e.g., YouTube) annotated with sensory effects which is similar to Experiment I (cf. Section 4.1) but in the context of the Web. Second, as the color information is extracted directly from the video frames, we investigated the influence of the subjective quality when skipping pixels, entire rows, and frames in order to reduce the processing requirements at the browser.

For both tests a similar enhancement scale has been used as in one of our previous experiments [4]. The major difference to the previous experiment is that we use a continuous scale from 0 to 100 instead of a discrete five-level enhancement scale as defined in the Degradation Category Rating (DCR) method. The finer scale allowed us to receive more precise results of the user experience.

The first subjective test is based on the Degradation Category Rating (DCR) defined in the ITU-T Rec. P.911 [23] and more or less confirmed the results from Experiment I (cf. Section 4.1) but in the context of the Web. The aim of the second experiment was to test the influence when ignoring information (i.e., pixels, rows, and frames) for the automatic color calculation. The results of this experiment may be used to configure the plug-in based on the capabilities of the Web browser/client. For this subjective test we used the Absolute Category Rating with Hidden Reference (ACR-HR) [22] with the same modifications as for the first study (i.e., a voting scale from 0 to 100 instead of a discrete scale). For this user study we used only two videos from the action and the documentary genre but always with sensory effects enabled. Each video is shown multiple times and each time with different settings for the color calculation. The difference in the color calculation concerns the usage of frames and pixels. That is, we skipped up to two frames (FS=frame skip), ignored at most every second pixel within a row (PS=pixel skip), or ignored between zero and two rows entirely (RS=row skip). In total we had 18 video sequences that were randomly shown to the participants.

Fig. 4 presents the MOS and confidence interval (95%) for the action video. The number indicates the number of frames, rows, and pixels to be skipped, respectively. Interestingly, the results reveal that the ratings remain almost constant in case only entire frames are skipped (i.e., frame skip=0, 1, 2, row skip=0, frame skip=0). That is, users provide a lower rating when pixels and/or entire rows are skipped while voting almost constant when entire frames are skipped. That leads to the assumption that in case the client faces performance issues, the automatic color calculation should first skip frames before starting to skip pixels and/or entire rows within a frame. However, this behavior can be only partially confirmed when looking at the results for the other sequence not shown here and, thus, requires probably further tests.

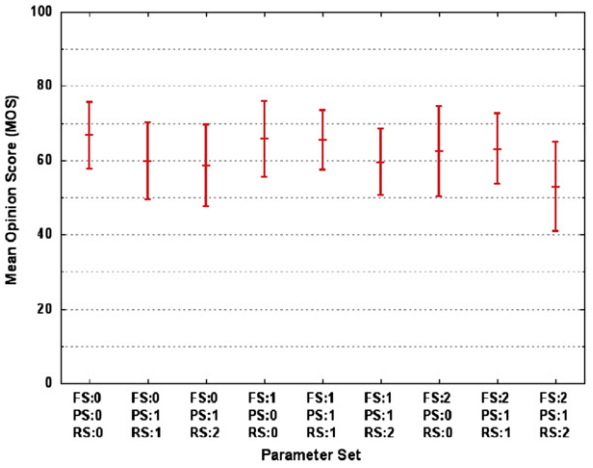

Fig. 4. MOS and confidence interval for each parameter set for Experiment III [5].

5. Open issues and research challenges

The main issue within this work item is definitely to establish a sound quality/utility model for sensory experiences, i.e., the main component in the call for a scientific framework as outlined in [1]. In [26] a theoretical framework is presented for the QoE in distributed interactive multimedia environments adopting an empirical mapping between QoE and QoS which may be also suitable for the mapping between sensory effects and human senses.

Yet another dimension is the investigation of the perception of sensory effects at the emotional level while consuming fictive (e.g., science fiction movies) versus non-fictive content (e.g., news, documentary). For example, we believe that there is a major difference when experiencing an earthquake (or the like) while watching the most recent fiction movie compared to a news report or documentary based on a true story.

From a technical point of view the (semi-)automatic generation of sensory effects seems to be a challenging task. Some authoring tools are presented in [2,7] though and Waltl et al. [2] describes also means for the automatic extraction of color information from the videos frame content in order to control the ambient light within a home environment. On the other hand, the extraction of, for example, wind and vibration effects (including intensity/strength and direction from where it shall be perceived) based on content analysis seems to be a huge challenge. However, everything that can be extracted automatically and does not have to be annotated manually which is cost-intensive would increase market adoption of such tools.

Finally, the (efficient) transport of SEM together with the multimedia content seems to be manageable and already partially addressed in the literature, e.g., [9]. However, the efficient rendering of sensory effects and

the synchronization with the actual multimedia content is not trivial, specifically for olfactory data [12]. Basically it introduces similar issues like lip synchronization for matching lip movements with voice. Finally, some devices (e.g., scent vaporizer, heater/cooler air condition) require start-/warm-up periods as well as a rundown time which needs to be considered as well and not fully studied so far.

6. Conclusions and future work

In this paper we have presented a comprehensive introduction into the concept of sensory experience, its assessment in terms of the QoE, and related standardization efforts. The core of the paper is the results from three subjective quality assessments with the following major findings: (1) sensory effects provide a vital tool for enhancing the user experience depending on the actual genre, specifically for action and also documentary. (2) Sensory effects may compensate for quality degradations within the actual audio/visual content and, thus, may help to reduce the overall bitrate of the multimedia transmission without affecting the QoE. (3) The automatic color extraction for controlling ambient light installations can be done in real-time. In case of low computational requirements (at the client) one may skip entire frames before skipping pixels or rows within a frame without affecting the QoE significantly.

However, the important step is now to establish a sound quality/utility model for the sensory experience which is a challenging task and cannot be done within a single step. Thus, we outline the individual steps of our future work in the following. First, we will build a database of test sequences with audio/visual contents annotated with sensory effects, both with single effect types and combinations thereof. Second, we will investigate the correlation mapping between sensory effect types and human senses by means of empirical studies that shall provide the basis for our quality/utility model. Third, we further improve the subjective quality assessment method based on previous results. Fourth, we will perform intensive tests with various setups in order to determine characteristics of the individual QoE dimensions, trying to be as generic as possible. This will hopefully lead to a generally accepted QoE model for sensory experience. A last and final step would be working towards a standardized quality assessment method by contributing our results to, for example, ITU-T SG12 and/or VQEG.

References

[1] E. Aarts, B. de Ruyter, New research perspectives on ambient intelligence, Journal of Ambient Intelligence and Smart Environments 1 (1) (2009) 5–14, doi:10.3233/AIS-2009-0001.

[2] M. Waltl, C. Timmerer, H. Hellwagner, A test-bed for quality of multimedia experience evaluation of sensory effects, in: T. Ebrahim, K. El-Maleh, G. Dane, L. Karam (Eds.), Proceedings of the First International Workshop on Quality of Multimedia Experience (QoMEX 2009), IEEE, Los Alamitos, CA, USA, 2009, pp. 145–150, doi: 10.1109/QOMEX.2009.5246962. URL 〈http://www.qomex2009.org〉.

[3] M. Waltl, C. Timmerer, H. Hellwagner, Increasing the user experience of multimedia presentations with sensory effects, in: R. Leonardi, P. Migliorati, A. Cavallaro (Eds.), Proceedings of the 11th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS'10), IEEE, Desenzano del Garda, Italy, 2010, pp. 1–4.

[4] M. Waltl, C. Timmerer, H. Hellwagner, Improving the quality of multimedia experience through sensory effects, in: A. Perkis, S. Möller, P. Svensson, A. Reibman (Eds.), Proceedings of the Second International Workshop on Quality of Multimedia Experience (QoMEX'10), IEEE, Trondheim, Norway, 2010, pp. 124–129 URL 〈http://www.qomex2010.org〉.

[5] M. Waltl, B. Rainer, C. Timmerer, H. Hellwagner, Sensory experience for videos on the web, in: L. Böszörmenyi, O. Marques, M. Lux, R. Klamma (Eds.), Proceedings of the Workshop on Multimedia on the Web (MMWeb) 2011, IEEE, Graz, Austria, 2011, pp. 1–3.

[6] B. de Ruyter, E. Aarts, Ambient intelligence: visualizing the future, in: AVI'04: Proceedings of the Working Conference on Advanced Visual Interfaces, ACM Press, New York, NY, USA, 2004, pp. 203–208, doi:10.1145/989863.989897.

[7] C.B. Suk, J.S. Hyun, L.H. Yong, Sensory effect metadata for SMMD media service, in: Proceedings of the 2009 Fourth International Conference on Internet and Web Applications and Services, IEEE Computer Society, Washington, DC, USA, 2009, pp. 649–654, doi: 10.1109/ICIW.2009.104. URL 〈http://dl.acm.org/citation.cfm?id=1585689.1586195〉.

[8] S. Pyo, S. Joo, B. Choi, M. Kim, J. Kim, A metadata schema design on representation of sensory effect information for sensible media and its service framework using UPnP, The 10th International Conference on Advanced Communication Technology, 2008, ICACT 2008, vol. 2, 2008, pp. 1129–1134, doi:10.1109/ICACT.2008.4493965.

[9] K. Yoon, B. Choi, E.-S. Lee, T.-B. Lim, 4-D broadcasting with MPEG-V, in: IEEE International Workshop on Multimedia Signal Processing (MMSP) 2010, 2010, pp. 257–262, doi:10.1109/MMSP.2010.5662029.

[10] A. Chang, C. O'Sullivan, Audio-haptic feedback in mobile phones, in: CHI'05 Extended Abstracts on Human Factors in Computing Systems, CHI EA'05, ACM, New York, NY, USA, 2005, pp. 1264–1267, doi: 10.1145/1056808.1056892.

[11] J.-H. Kim, H.-J. Kwon, K.-S. Hong, Location awareness-based intelligent multi-agent technology, Multimedia Systems 16 (4–5) (2010) 275–292.

[12] O. Ademoye, G. Ghinea, Synchronization of olfaction-enhanced multimedia, IEEE Transactions on Multimedia 11 (3) (2009) 561–565, doi:10.1109/TMM.2009.2012927.

[13] The International Multisensory Research Forum (IMRF). URL 〈http://www.imrf.info/〉 (accessed November 2011).

[14] M. Grega, L. Janowski, M. Leszczuk, P. Romaniak, Z. Papir, Quality of experience evaluation for multimedia services—Szacowanie postrzeganej jakości usług (QoE) komunikacji multimedialnej, Przegląd Telekomunikacyjny 81 (4) (2008) 142–153.

[15] D. Hands, A basic multimedia quality model, IEEE Transactions on Multimedia 6 (6) (2004) 806–816, doi:10.1109/TMM.2004.837233.

[16] T. Hoßfeld, D. Hock, P. Tran-Gia, K. Tutschku, M. Fiedler, Testing the IQX hypothesis for exponential interdependency between QoS and QoE of voice codecs iLBC and G.711, in: The 18th ITC Specialist Seminar on Quality of Experience, Karlskrona, Sweden, 2008.

[17] F. Pereira, A triple user characterization model for video adaptation and quality of experience evaluation, in: IEEE Seventh Workshop on Multimedia Signal Processing, 2005, 2005, pp. 1–4, doi:10.1109/MMSP.2005.248674.

[18] ISO/IEC 23005-3, Information Technology—Media Context and Control—Sensory Information, ISO/IEC JTC 1/SC 29/WG 11/N11425, Geneva, Switzerland.

[19] P. Salembier, T. Sikora, Introduction to MPEG-7: Multimedia Content Description Interface, John Wiley & Sons, Inc., New York, NY, USA, 2002.

[20] ISO/IEC 21000-7:2007, Information Technology–Multimedia Framework (MPEG-21)–Part 7: Digital Item Adaptation (accessed November 2007).

[21] amBX. URL 〈http://www.ambx.com〉 (accessed November 2011).

[22] ITU-T Rec. P.910, Subjective Video Quality Assessment Methods for Multimedia Applications, April 2008.

[23] ITU-T Rec. P.911, Subjective Audiovisual Quality Assessment Methods for Multimedia Applications, December 2008.

[24] R.L. Storms, M.J. Zyda, Interactions in perceived quality of auditory-visual displays, Presence: Teleoperators and Virtual Environments 9 (6) (2000) 557–580, doi:10.1162/105474600300040385.

[25] G. Bjontegaard, Calculation of average PSNR differences between RD curves, in: ITU-T VCEG Meeting VCEG-M33, Austin, USA, 2001.

[26] W. Wu, A. Arefin, R. Rivas, K. Nahrstedt, R. Sheppard, Z. Yang, Quality of experience in distributed interactive multimedia environments: toward a theoretical framework, in: Proceedings of the 17th ACM International Conference on Multimedia, MM'09, ACM, New York, NY, USA, 2009, pp. 481–490, doi:10.1145/1631272.1631338.