Sensory Effects for Ambient Experiences in the World Wide Web

Markus Waltl
markus.waltl@itec.uni-klu.ac.at

Christian Timmerer
christian.timmerer@itec.uni-klu.ac.at

Benjamin Rainer
benjamin.rainer@itec.uni-klu.ac.at

Hermann Hellwagner
hermann.hellwagner@itec.uni-klu.ac.at

Institute of Information Technology
Alpen-Adria-University Klagenfurt
Technical Report No TR/ITEC/11/1.13
July 2011

Sensory Effects for Ambient Experiences in the World Wide Web

Markus Waltl, Christian Timmerer, Benjamin Rainer, and Hermann Hellwagner
Klagenfurt University, Department of Information Technology (ITEC), Multimedia Communication (MMC)
Universitätsstraße 65-67
9020 Klagenfurt am Wörthersee, Austria
{*firstname.lastname*}@itec.uni-klu.ac.at

ABSTRACT
More and more content in various formats become available via the World Wide Web (WWW). Currently available Web browsers are able to access and interpret these contents (i.e., Web videos, text, image, and audio). These contents stimulate only senses like audition or vision. Recently, it has been proposed to stimulate also other senses while consuming multimedia content through so-called sensory effects. These sensory effects aim to enhance the ambient experience by providing effects, such as, light, wind, vibration, etc. The effects are represented as Sensory Effect Metadata (SEM) which is associated to multimedia content and is rendered on devices like fans, vibration chairs, or lamps. In this paper we present a plug-in for the Mozilla Firefox browser which is able to render such sensory effects that are provided via the WWW. Furthermore, the paper describes two user studies conducted with the plug-in and presents the results achieved.

Categories and Subject Descriptors
H.5.1 [**Multimedia Information Systems**]: Sensory Experience, World Wide Web.

General Terms
Measurement, Performance, Design, Experimentation, Human Factors, Standardization.

Keywords
World Wide Web, MPEG-V, Subjective Quality Assessment, Sensory Effects, Quality of Multimedia Experience

1. INTRODUCTION
Multimedia content (i.e., combinations of text, graphics, images, audio, and video) has become omnipresent in our life. Each day we consume dozens and more multimedia assets when reading electronic newspaper, listening to podcasts or Internet radio, and watching digital television (TV). Recently, W3C established a strategy that video should become a first class citizen on the Web [17]. Major contributions towards this goal are HTML5s' `video` tag and platforms like YouTube and the like. For example, YouTube started with High Definition (HD) some time ago and recently started to offer services beyond HD in the so-called Ultra-HD format (i.e., resolutions of 4K and beyond). In any case, this seems to be an important step towards increasing the user experience.

Another dimension that recently gained popularity is 3D thanks to the commercial success in cinemas and home entertainment. However, 3D graphics is already supported on the Web and the technical foundation for supporting 3D video on the Web is already in place [22].

In our work we target yet another dimension addressing human senses that go beyond audition and vision. The consumption of multimedia assets may stimulate other senses such as olfaction, mechanoreception, equilibrioception, or thermoreception, opening a number of new issues 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 on 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 rendered on appropriate devices (e.g., ambient lights, fans, vibration chairs, perfumer, water sprayer). The ultimate goal of this system 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, thus, resulting in a worthwhile, informative user experience.

In prior publications we have conducted formal subjective quality assessments in order to investigate the benefit of such sensory effects under different circumstances (e.g., genres, bitrates) with promising results [19][20]. In [15] we have investigated whether sensory effects are ready for the World Wide Web (and vice versa) and concluded that, yes, they are ready but it requires some implementation efforts and additional quality assessments to see how these sensory effects are perceived by the users of the Web content. Therefore, in this paper 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 annotated with sensory effects which is similar to one of our previous experiments but in the context of the Web. Second, as the color information is extracted directly from the video frame, we investigated the influence of the subjective quality when skipping pixels, entire rows, and frames.

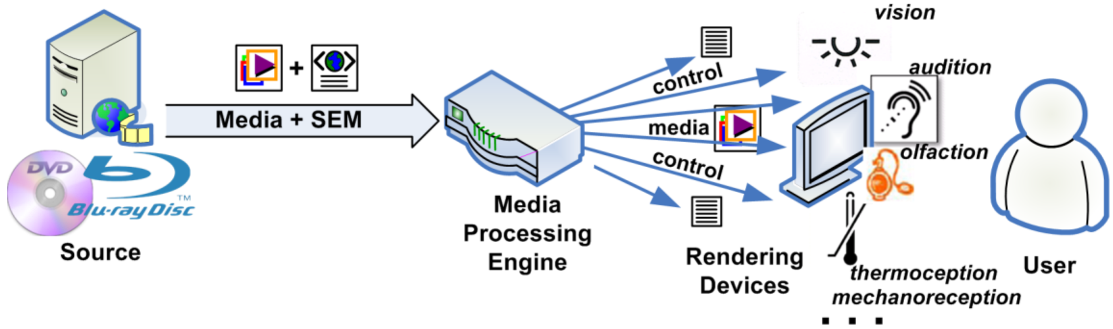


Figure 1. Concept of MPEG-V Sensory Effects [18].

The reminder of the paper is organized as follows. Section 2 gives an introduction to MPEG-V Sensory Information that provides an interoperable description format for describing sensory effects. The architecture of the Web browser plug-in we have used for the subjective tests is described in Section 3. Section 4 presents the test environment, its setup, and the conditions for both user assessments. The results for each user assessment are detailed in Section 5 and a concise discussion thereof is presented in Section 6. Section 7 describes related work and Section 8 concludes the paper including future work items.

2. SENSORY INFORMATION

This section comprises an updated version of [18] which reflects the changes adopted into the final version of the MPEG-V Part 3 (Sensory Information) standard [7] and provides the necessary background for the rest of the paper.

2.1 Sensory Effect Description Language

The *Sensory Effect Description Language (SEDL)* is an XML Schema-based language which enables one to describe so-called sensory effects such as light, wind, fog, vibration, etc. that trigger human senses. 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 (see Section 2.2). A description conforming to SEDL is referred to as *Sensory Effect Metadata (SEM)* and may be associated to 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 increase 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, giving her/him the sensation of being part of the particular media which shall result in a worthwhile, informative user experience.

The concept of receiving sensory effects in addition to audio/visual content is depicted in Figure 1. The *media* and the corresponding *SEM* may be obtained from a Digital Versatile Disc (DVD), Blu-ray Disc (BD), or any kind of online service (i.e., download/play or streaming). The *media processing engine* acts as the mediation device and is responsible for playing the actual media resource and accompanied sensory effects in a synchronized way based on the user's setup in terms of both media and sensory effect rendering. Therefore, the media processing engine may adapt both the media resource and the SEM according to the capabilities of the various *rendering devices*.

The current syntax and semantics of SEDL are specified in [7]. However, in this paper we provide an EBNF (Extended Backus–Naur Form)-like overview of SEDL due to the lack of space and 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 effect 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 [11] 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 [6]. 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 user's 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.

2.2 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 author's intention and be independent from the user's 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 automatically be 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.

2.3 Usage Example

In this section we provide a short example of a SEM description with an in-depth description how a *media processing engine* should handle this description to control sensory devices. Let us assume 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:mpeg:mpeg-v:01-SI-
LocationCS-NS: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:mpeg:mpeg-
v:01-SI-LocationCS-
NS: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 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. PLUG-IN ARCHITECTURE

The Mozilla Firefox *amBX Plug-in* provides a communication component between the Web browser (in our case Mozilla Firefox 3.6.10) and the *amBX System* [2] used in our experiments. The amBX System consists of two fans, a wrist rumbler, two sound speakers and a subwoofer, two front lights, and a wall washer.

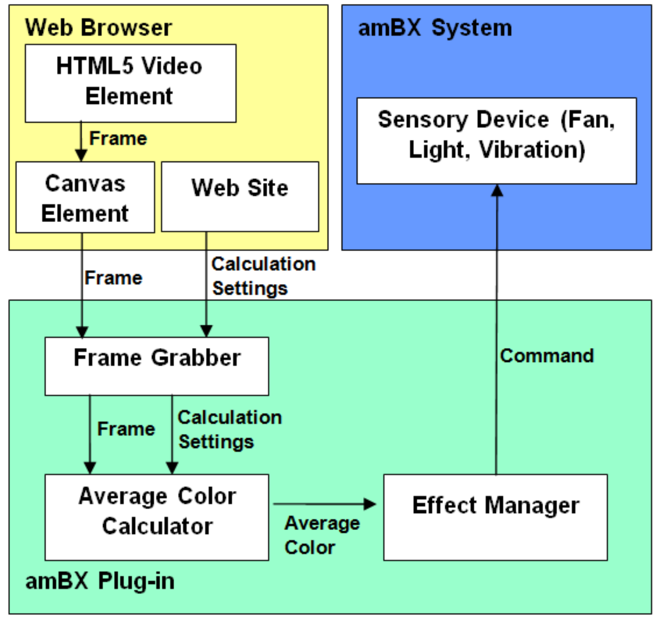


Figure 2. Architecture of the Mozilla Firefox amBX Plug-in.

The aim of this plug-in is to extract and interpret the SEM description from within Web site and synchronize the sensory effects with the contents of the Web site. The current version of the plug-in supports the automatic color extraction from a Web video only and hence does not yet require an actual SEM description which is planned for future work. It is written in C++ including Assembler code which is used for faster image processing via *Single instruction, multiple data (SIMD)* and *MultiMedia Extension (MMX)*. The plug-in utilizes the Gecko SDK 1.9.2 [4] and exploits HTML5s' `video` element [5]. The latter provides an easy way for playing videos within a Web site without the need for an additional player such as Flash or Quicktime. As the current version of the Gecko SDK does not support the direct extraction of the frame pixels from the video element, a `canvas` element is needed which is dynamically added to the Web site once the plug-in is enabled. That is, if the user does a right mouse click on the `video` element a new drop-down menu entry is added through the plug-in providing full control over the plug-in by the user. Finally, the `canvas` element provides the pixel information of the currently rendered frame which is accessible by the plug-in. The architecture of the plug-in is shown in Figure 2 comprising the Web browser which provides the input (i.e., *Frame*) to the plug-in which uses this information among configuration settings (i.e., *Calculation Settings*) in order to control the amBX system via *Commands* through the corresponding SDK. The calculation settings are embedded within the Web site and provided to the plug-in during its initialization. The calculation settings determine whether and how many frames, rows, or pixels within a row should be skipped in order to increase the performance of the color calculation. In this way, the settings for the color calculation can also be used to change the frequency of light changes. For example, if we take only every third frame for the average color calculation, light colors will change less frequently. Frames are retrieved every 30 milliseconds by the *Frame Grabber* which results in approximately 33 frames/s that are retrieved. Based on the frame skip value the *Frame Grabber* decides whether the current frame should be skipped or

Table 1. Web Video Sequences.

Sequence Name	Genre	Bit-rate (Kbit/s)	Length (sec)
Babylon A.D.	Action	2724	23
Big Buck Bunny	Cartoon	2109	25
Earth	Documentary	2321	21
BYU commercial	Sports	2474	23

not. If the current frame is not skipped it forwards the calculation settings and the frame to the *Average Color Calculator*. The *Average Color Calculator* splits the current frame into three parts which correspond to the setup of the lights of the *amBX System* (i.e., lights on the left, center, and right). For each of the three parts the average color will be calculated and the result is forwarded to the *Effect Manager*. The *Effect Manager* activates and sets the color for the corresponding lights. Note that the *amBX System* supports further sensory effects like wind and vibration effects through its corresponding devices. In this paper, however, we focus on the light effects because they can be automatically extracted from the video. Support for other effects (e.g., wind, vibration) could be added by a SEM description which can be embedded into the Web site as described in [15]. Finally, the plug-in is designed in a way so that it can be easily ported to any other Web browser by replacing the SDK, e.g., using WebKit [14] for supporting the Safari Web browser.

4. TEST ENVIRONMENT

This section describes the environment for the subjective tests which are based on ITU-T Rec. P.910 [9] and P.911 [10] respectively. In particular, we describe how we selected the number of participants and stimuli (cf. 4.1), the actual test setup (cf. 4.2), and the test method as well as experimental design for the two tests we have conducted (cf. 4.3). The results are defined in Section 5.

4.1 Participants and Stimuli

A subjective test should comprise around 10 to 20 participants in order to achieve relevant results. For our two subjective tests we invited the following number of participants: for the first study we invited 20 students (11 female and 9 male) between the age of 19 and 31 years. For the second study we invited 18 students (7 female and 11 male) between the age of 21 and 34 years. In the first study there were two and in the second study there were three participants that already took part in one of our previous experiments [19][20] and in the second study 6 participants were students of computer science. The rest of the participants were not familiar with the evaluation topic or were students from other areas (e.g., psychology).

In order to provide every participant the same test conditions we prepared the introduction for the assessment as a Web site which was shown before the subjective test. The introduction, as the whole assessment, was provided in English and German to avoid possible misunderstandings.

For both tests we used the same Web videos which are shown in Table 1 indicating the genre, bit-rate, and length. All Web videos had a resolution of 540p and as we only used light effects each Web video provides a variety of different colors. The color for the lights was calculated automatically every 0.03 seconds and mapped onto the lights as introduced in the previous section.

4.2 Test Setup

Both experiments were conducted in an isolated room under the same ambient conditions. We ensured that the following conditions prevailed before each session:

- 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.

The windows were covered for a better light contrast else the light effects could easily be overlooked.

The following hardware and software was used for conducting the subjective tests:

- Dell Optiplex 655: Pentium D 2.8 GHz with 2 GB RAM and NVidia Quadro NVS (64 MB)
- amBX Premium Kit (Fan, Vibrator, Light, Sound)
- 19’’ Monitor with a resolution of 1280x1024
- Windows XP SP3
- Mozilla Firefox 3.6.10
- Mozilla Firefox amBX Plug-in 1.5
- amBX Software (i.e., amBX System 1.1.3.2 and Philips amBX 1.04.0003)

The test computer is equipped with the amBX premium kit that comprises several sensory devices, such as, a wall washer light with controller unit, left & right 2.1 sound speaker lights and a sub woofer, a pair of fans, and a wrist rumbler. In the two subjective tests only the lights (wall washer, left & right lights) and sound devices were used.

The lights of the amBX System consist of high power RGB LEDs with over 16 million additive RGB colors. The colors respond instantly and they can vary continuously in intensity. The 2.1 sound system comes with 160 W music power, two 40 W satellite speakers (integrated within the light system) and a 80 W subwoofer operating in the frequency range of 35 Hz ~ 20 kHz.

The actual tests were divided into three parts with a total duration of around 20 to 25 minutes per subject. For the two tests the first and third part were identical only the second part differed between the two subjective tests. The *first part* comprised the introduction and general questions (i.e., gender, age, and field of study) about the subject. Each subject had to read the introduction which explained them the whole test procedure.

The actual subjective test was conducted in the *second part* of both studies. The participants sat (in a comfortable seat) in a distance of around three times the height of the monitor. A detailed description of the two different test methods and experimental designs can be found in Section 4.3.

Table 2. Five-level Continuous Enhancement Scale.

80 – 100	Big enhancement
60 – 80	Little enhancement
40 – 60	Imperceptible
20 – 40	Annoying
0 – 20	Very annoying

After the second part of the subjective test a post-experiment questionnaire was presented to the participants. In this part the participants had no time limit to answer these questions. Furthermore, they could ask questions about the questionnaire if something was not clear enough. The following questions were asked during the post-experiment part:

Q1. Where you mentally overloaded during any part of the experiment?

Q2. Have you ever participated in an experiment similar to this one?

Q3. Any other comments about what you liked or did not like, or things that should be changed during the course of this experiment?

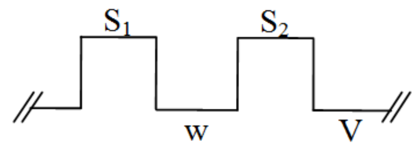
Finally, the overall test setup for both tests was inspired by and partially based on [13].

4.3 Test Methods and Experimental Designs

As stated before we conducted two different subjective tests in the area of sensory effects. For both tests a similar enhancement scale has been used as in one of our previous experiments [20]. 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. Nevertheless, we roughly divided the scale in the following levels as presented in Table 2.

4.3.1 Benefits of Sensory Effects in the WWW

The aim of the first experiment was to test if Web videos, from different genres, accompanied by sensory effects are enhancing the user experience in the WWW. The results of this experiment may indicate for which genre sensory effects are more useful than for other genres. The subjective test is based on the *Degradation Category Rating* (DCR) defined in the ITU-T Rec. P.911 [10]. The test method is shown in Figure 3 and the four video sequences from different genres are shown to the user in randomized order. Each video sequence is shown twice to the participants: once without sensory effects and once with sensory effects (i.e., light effects). The first video sequence (i.e., without sensory effects) acts as the reference for the second that is shown after a five seconds break. During the break a grey screen is shown to the user. After the video sequence with the sensory effects the participants vote the enhancement of the video with sensory effects compared to the video without sensory effects. The participants have 15 seconds for giving their vote on a slider with a range from 0 to 100 with the major steps as shown in Table 2.



S_1 … Test sequences without sensory effects
S_2 … Test sequences with sensory effects
V … Voting for corresponding sequence
w … Grey screen (5 seconds)

Figure 3. Testing Method for the First Study.

Table 3. Parameter Sets.

Parameter Set	Description
FS:0;PS:0;RS:0	Each frame, pixel of a row and row is used for the color calculation.
FS:0;PS:1;RS:1	Each frame is used but only every second pixel of a row and every second row are used for the color calculation.
FS:0;PS:1;RS:2	Each frame is used but only every second pixel of a row and every third row are used for color calculation.
FS:1;PS:0;RS:0	Only every second frame is used but from this frame each pixel of a row and each row are used for the color calculation.
FS:1;PS:1;RS:1	Every second frame is used. Furthermore, every second pixel of a row and every second row are used for the color calculation.
FS:1;PS:1;RS:2	Every second frame is used. From a row every second pixel is taken for the color calculation. Furthermore, only every third row is used.
FS:2;PS:0;RS:0	Only every third frame is used but from this frame each pixel of a row and each row are used for the color calculation.
FS:2;PS:1;RS:1	Every third frame is used. Furthermore, every second pixel of a row and every second row are used for the color calculation.
FS:2;PS:1;RS:2	Every third frame is used. From a row every second pixel is taken for the color calculation. Furthermore, only every third row is used.

4.3.2 Performance Measurements

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) [9] 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 movies from the action and the documentary genre but always with sensory effects enabled. Each video is shown 9 times and each time with different settings for the color calculation. The difference in the color calculation concerns the usage of frames and pixels. The parameter sets are described in Table 3. For example, we may skip entire frames (FS = frame skip), ignore pixels within a row (PS = pixel skip), or ignore entire rows (RS =

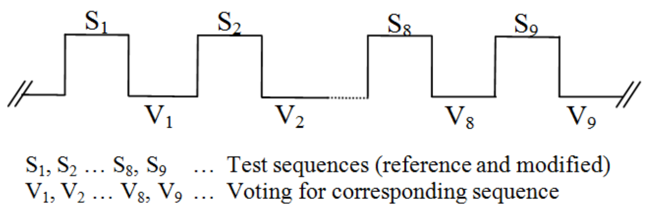


Figure 4. Test Method for the Second Study.

row skip). The number indicates the number of frames, rows, and pixels to be skipped respectively.

In total there are 18 video sequences that are randomly shown to the participants. Note that the two genres are mixed together which means that not all 9 video sequences of one genre are presented in a row. After each video sequence the participants have 15 seconds to give their vote about the overall quality of the experience. Figure 4 depicts the test method for the second user study.

5. RESULTS

This section comprises the test results for both subjective user studies. Please note that for both tests there were no outliers, according to [8], detected.

5.1 Benefits of Sensory Effects in the WWW

5.1.1 Evaluation Results

Figure 5 shows the mean opinion score (MOS) and confidence interval (95%) for all four genres. As someone can see the documentary was rated the lowest and sports the highest. But all four video sequences are situated in the overall area of *Imperceptible* (Score 40 – 60) or *Little enhancement* (Score 60 – 80). Interestingly, in our previous studies the documentary genre had always the highest ratings which are further discussed in the next section.

The detailed results of the first experiment is shown in Figure 6. We counted the votes for each rating category to provide a rough overview of the ratings per genre which clearly indicates that the majority of votes are in the upper region of the voting scale. Further, the figure shows that the sports sequence is rated much higher than, for example, the second best genre, i.e., action.

5.1.2 Post-Experiment Questionnaire Results

The results of the post-experiment questionnaire for the first user study are as follows. 15% of the participants stated that they were mentally overloaded by sensory effects (i.e., Q1). All of these participants commented on this question that too many effects were displayed in a short time period. These participants were not able to concentrate anymore on the video content because they were more focused on the lights.

Concerning the experience of the participants in user studies (i.e., Q2), 20% of the participants already took part in a similar study (e.g., study with sound and movement sensors or one of our earlier studies [19]). For Q3 it is only worth mentioning that 55% of the participants stated that the videos get more interesting and/or more intensive with sensory effects.

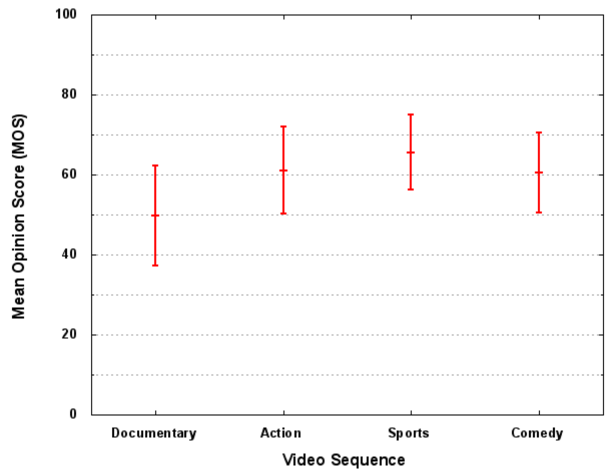


Figure 5. MOS and confidence interval for each genre.

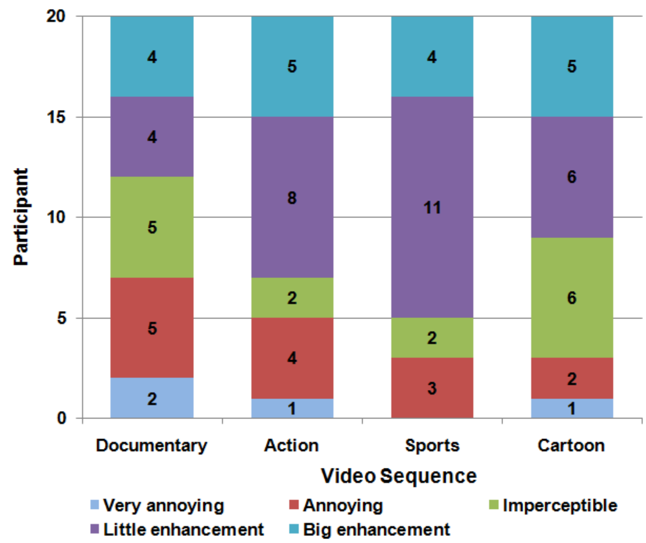


Figure 6. Evaluation Results for each Genre.

5.2 Performance Measurements

5.2.1 Evaluation Results

Figure 7 and Figure 8 depict the detailed voting results for each parameter settings as described in Table 3. Note that we also summed up the ratings for every major voting step.

One obvious observation is that the more information is skipped for calculating the color for the lights the lower the ratings with one exception which is the last bar of Figure 7. However, in general please note that there are some minor deviations but the majority of the votes show a clear tendency to lower ratings if less information is available. The "outlier" in Figure 7 as well as these minor deviations are subject to further investigations as part of our future work.

Figure 9 and Figure 10 present the MOS and confidence interval (95%) for both video sequences. Interestingly, the results as depicted in Figure 10 reveal that the ratings remain almost constant in case only entire frames are skipped (i.e., FS={0, 1, 2}, PS=0, RS=0). This aspect of the results is further discussed in the next section.

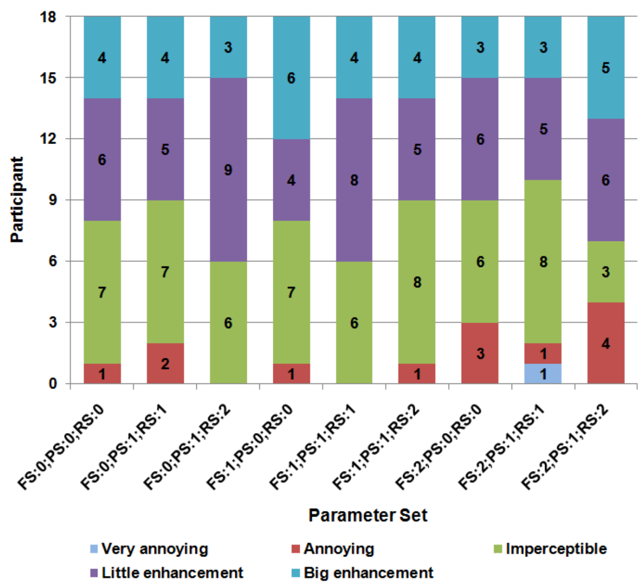


Figure 7. Evaluation Results for each Parameter Set for the Documentary Video.

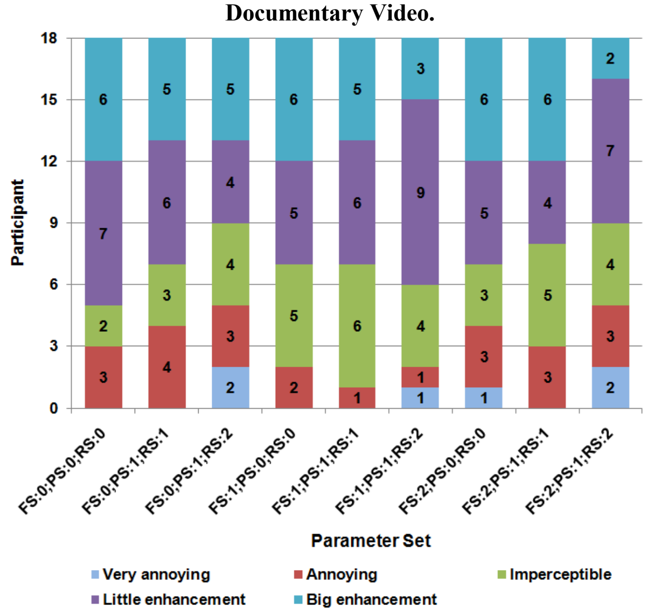


Figure 8. Evaluation Results for each Parameter Set for the Action Video.

Overall, the results are in the area of *Imperceptible* (Score 40 – 60) or *Little enhancement* (Score 60 – 80) similar to our first study and the majority of the results is in the area of little enhancement. However, it seems that in this test the documentary gets a higher rating compared to the previous test whereas the results for the action video more or less confirm the previous test.

5.2.2 Post-experiment Questionnaire Results

The results of the post-experiment questionnaire for the second user study are as follows. For Q1 two participants (11%) stated that they were mentally overloaded by the sensory effects. One of them stated that the field of view is distracted by the lights and they are focusing more on the effects than on the video. The other, mentioned that during the action sequence for him/her there was a sensory overload. 22% of the participants took already part in a similar user study (i.e., Q2). 75% of them already attended one of our previous user studies [19][20]. For Q3 no feedback worth mentioning here has been provided.

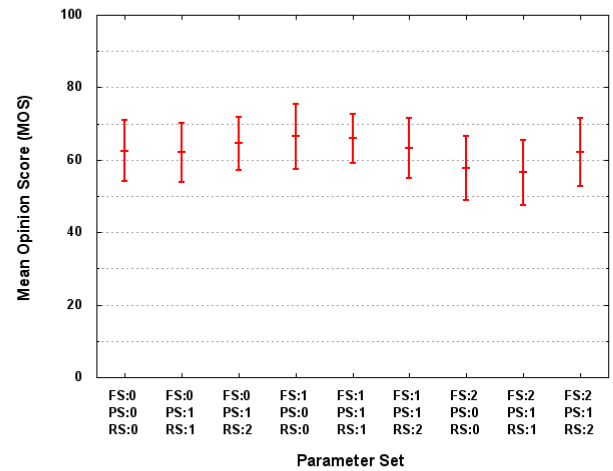


Figure 9. MOS and Confidence Interval for each Parameter Set for the Documentary Video.

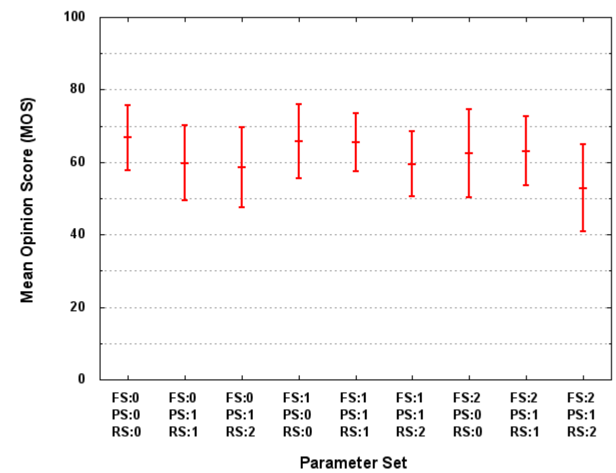


Figure 10. MOS and Confidence Interval for each Parameter Set for the Action Video.

6. DISCUSSION

In general, the results of both studies reveal that videos accompanied by sensory light effects enhance the user experience although not that much as we have shown in our previous study [19]. The major differences between this and the previous study was that the current study has been conducted in a Web context and with light effects only. Thus, one may conclude that the difference is partially due to the absence of additional sensory effects such as wind or vibration but also due to the different context. Adding additional sensory effects to the Web context is part of our future work.

One interesting observation is that the documentary video gets the lowest rating in our current study compared to the highest rating in one of our previous studies [20]. One explanation of this difference could be again the different context and the absence of additional effects. However, we also believe that this might be caused by a different editing sequence of the individual shots. In particular, we believe that sequences with short shots and a lot of shot transitions require a more sophisticated treatment of the automatic color calculation which is also an interesting topic for future work.

For the second experiment we find it rather surprising, specifically the results as shown in Figure 10, that 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 pixel and/or entire rows within a frame. However, this behavior can be only partially confirmed when looking at the results for the documentary video and, thus, requires probably some further test.

Finally, another reason for some of the issues identified here may be due to the lack of an appropriate test method and common data sets as pointed out already elsewhere [19].

7. RELATED WORK

Sensory effects and MPEG-V can be seen as part of Ambient Intelligence (AmI). With AmI researchers try to create intelligent environments that are sensitive and responsive to the presence of a person [1]. For example, there are so-called Ambient Displays (AD) or Ambient Media. Such ADs can be used to display information to the user through different types of displays (e.g., lamps, active wallpaper). MPEG-V does not only support information exchange from the virtual world to the real world (e.g., using sensory effects) via such ADs but also information exchange from the real world to the virtual world (e.g., sensors). There is already a numerousness research community for providing such displays to home or office environments. For example, the ambientROOM [21] provides a set of different types of ADs for giving the person inside the room information about activities of people outside the room (e.g., presence of people in a seminar room). As stated in [21] the room currently only supports awareness of people activities but can be extended to provide dynamic information, such as, weather, stock market or network traffic. There is also the so-called Infostudio [16] which is a course at the University of Sydney for encouraging students to build their own ADs for home environments. Infostudio uses TCP/IP connections and Telnet commands for controlling available devices. [12] introduces a CGI-based approach for controlling devices, such as, lamps or table fountains via a Web browser. Instead of using Telnet for controlling available devices, the paper introduces an architecture that uses a Web server for providing links to CGI-scripts. The Web browser sends an HTTP-GET-request to the Web server to activate the CGI-script. After activation the CGI-script manipulates associated ADs. For example, with this approach a user is able to get information through ADs about people visiting his/her personal Web site (e.g., which page was visited and in which part of the web site people were more interested).

Most of the work described above is based on user awareness and user-driven controlling of devices. MPEG-V and in particular sensory information provides additionally a content-driven approach. This means that the enriched content can enhance the user experience without user intervention but if desired user preferences or device capabilities can be taken into account.

8. CONCLUSIONS AND FUTURE WORK

In this paper we have presented our results of a formal subjective quality assessment in the area of sensory information and the World Wide Web. As there are currently no feasible assessment methods for sensory effects we slightly adopted existing methods (i.e., DCR and ACR-HR). The procedures used as well as the results are presented and discussed in detail in the previous sections. The results obtained indicate that sensory effects can improve the user experience in the World Wide Web. However, at this time a general conclusion about the improvement of the Quality of Experience cannot be stated due to the lack of appropriate subjective test methods and common test sequences.

Future work includes, in addition to what have been already identified when discussing the results, adding further sensory effects, in particular, vibration and wind effects that are provided via the Web site in form of an explicit SEM description. Furthermore, we are looking forward extending the support for other browsers than Mozilla Firefox such as Safari, Internet Explorer, or Chrome. Finally, we will monitor the development of the emerging HTML5 recommendation as well as future releases of the Gecko engine in order to optimize our implementation.

9. ACKNOWLEDGMENTS

This work was supported in part by the EC in the context of the NoE INTERMEDIA (NoE 038419) and the ALICANTE project (FP7-ICT-248652).

10. REFERENCES

[1] Aarts, E. H. L. 2004. Ambient intelligence: A multimedia perspective. *IEEE MultiMedia* (2004), pp. 12-19.

[2] amBX UK Ltd. http://www.ambx.com.

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

[4] Gecko SDK. https://developer.mozilla.org/en/Gecko_SDK

[5] Hickson, I. 2010. HTML5, Working Draft. W3C (June 24, 2010). http://www.w3.org/TR/2010/WD-html5-20100624/

[6] ISO/IEC 21000-7:2007. 2007. Information technology - Multimedia framework (MPEG-21) - Part 7: Digital Item Adaptation. (November, 2007).

[7] ISO/IEC 23005-3 FCD. 2009. Information technology – Media context and control – Sensory information. *ISO/IEC JTC 1/SC 29/WG 11/N10987*, (Xi'an, China, Oct., 2009).

[8] ITU-R Rec. BT.500-11. 2002. Methodology for the subjective assessment of the quality of television pictures (2002).

[9] ITU-T Rec. P.910. 2008. Subjective video quality assessment methods for multimedia applications (Apr. 2008).

[10] ITU-T Rec. P.911. 1998. Subjective audiovisual quality assessment methods for multimedia applications (Dec. 1998).

[11] Manjunath, B. S., et al. 2002. Introduction to MPEG-7: Multimedia Content Description Interface. *John Wiley and Sons Ltd.* (June, 2002).

[12] Schmidt, A., Gellersen, H., Beigl, M. 1999. Matching Information and Ambient Media. In *Proceedings of the Second international Workshop on Cooperative Buildings, integrating information, Organization, and Architecture* (October 01 - 02, 1999). N. A. Streitz, J. Siegel, V. Hartkopf, and S. Konomi, Eds. Lecture Notes In Computer Science, vol. 1670. Springer-Verlag, London, 140-149.

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

[14] The WebKit Open Source Project. http://webkit.org/

[15] Timmerer, C., Waltl, M., Hellwagner, H. 2010. Are Sensory Effects Ready for the World Wide Web?. In *Proceedings of the Workshop on Interoperable Social Multimedia Applications* (Barcelona, Spain, May 19-20, 2010). WISMA'10.

[16] Vande Moere, A. 2005. Infostudio: teaching ambient display design using home automation. In *Proceedings of the 19th Conference of the Computer-Human Interaction Special Interest Group (CHISIG) of Australia on Computer-Human Interaction* (Narrabundah, Australia, 2005). (OZCHI '05). pp. 1-4.

[17] W3C. 2008. Video in the Web Activity Statement. (2008). http://www.w3.org/2008/WebVideo/Activity.html.

[18] Waltl, M., Timmerer, C., Hellwagner, H. 2009. A Test-Bed for Quality of Multimedia Experience Evaluation of Sensory Effects. In *Proceedings of the 1st Int'l Workshop on Quality of Multimedia Experience* (San Diego, USA, July 29-31, 2009). QoMEX'09.

[19] Waltl, M., Timmerer, C., Hellwagner, H. 2010. Improving the Quality of Multimedia Experience through Sensory Effects. In *Proceedings of the 2nd Int'l. Workshop on Quality of Multimedia Experience* (Trondheim, Norway, June 21-23, 2010). QoMEX'10.

[20] Waltl, M., Timmerer, C., Hellwagner, H. 2010. Increasing the User Experience of Multimedia Presentations with Sensory Effects. In *Proceedings of the 11th Int'l. Workshop on Image Analysis for Multimedia Interactive Services* (Desenzano del Garda, Italy, April 12-14, 2010). WIAMIS'10.

[21] Wisneski, C., Ishii, H., Dahley, A., Gorbet, M., Brave, S., Ullmer, B., Yarin, P. 1998. Ambient Displays: Turning Architectural Space into an Interface between People and Digital Information. *Lecture Notes in Computer Science* (1998), vol. 1370, pp. 22-32.

[22] YouTube Blog. 2009. Now in 3D: Join the Experiment With Us!. (July 22, 2009). http://bit.ly/lZ6pd.

Institute of Information Technology
Alpen-Adria-University Klagenfurt
Universitaetsstr. 65-67
A-9020 Klagenfurt
Austria
http://www-itec.uni-klu.ac.at

ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ Alpen-Adria-University Klagenfurt