Interfacing with Virtual Worlds

Christian Timmerer[1], Jean Gelissen[2], Markus Waltl[1], and Hermann Hellwagner[1]

[1]Klagenfurt University, Klagenfurt, Austria; [2]Philips Research, Eindhoven, The Netherlands

E-mail: [1]*firstname.lastname*@itec.uni-klu.ac.at, [2]jean.gelissen@philips.com

Abstract: Virtual worlds (often referred to as 3D3C for 3D visualization & navigation and the 3C's of Community, Creation and Commerce) integrate existing and emerging (media) technologies (e.g. instant messaging, video, 3D, VR, AI, chat, voice, etc.) that allow for the support of existing and the development of new kinds of networked services. The emergence of virtual worlds as platforms for networked services is recognized by businesses as an important enabler as it offers the power to reshape the way companies interact with their environments (markets, customers, suppliers, creators, stakeholders, etc.) in a fashion comparable to the Internet and to allow for the development of new (breakthrough) business models, services, applications and devices. Each virtual world however has a different culture and audience making use of these specific worlds for a variety of reasons. These differences in existing Metaverses permit users to have unique experiences. In order to bridge these differences in existing and emerging Metaverses a standardized framework is required, i.e., MPEG-V Media Context and Control (ISO/IEC 23005), that will provide a lower entry level to (multiple) virtual worlds both for the provider of goods and services as well as the user. The aim of this paper is to provide an overview of MPEG-V and its intended standardization areas. Additionally, a review about MPEG-V's most advanced part – Sensory Information – is given.

Keywords: Virtual World, Interoperability, MPEG-V, Sensory Information

1 INTRODUCTION

Multi-user online virtual worlds, sometimes called Networked Virtual Environments (NVEs) or massively-multiplayer online games (MMOGs), have reached mainstream popularity. Although most publications (e.g., [1]) tend to focus on well-known virtual worlds like World of Warcraft, Second Life, and Lineage, there are hundreds of popular virtual worlds in active use worldwide, most of which are not known to the general public. These can be quite different from the above-mentioned titles. To understand current trends and developments, it is useful to keep in mind that there is a large variety in virtual worlds and that they are not all variations on Second Life.

The concept of online virtual worlds started in the late seventies with the creation of the text-based dungeons & dragons world Multi-User Dungeon (MUD). In the eighties, larger-scale graphical virtual worlds followed, and in the late nineties the first 3D virtual worlds appeared. Many virtual worlds are not considered games (MMOGs) since there is no clear objective and/or there are no points to score or levels to achieve. In this report we will use virtual worlds as an umbrella term that includes all possible varieties. See the literature for further discussion of the distinction between gaming/non-gaming worlds (e.g., [2]). Often, a virtual world which is not considered to be an MMOG does contain a wide selection of 'mini-games' or quests, in some way embedded into the world. In this manner a virtual world acts like a combined graphical portal offering games, commerce, social interactions and other forms of entertainment. Another way to see the difference: games contain mostly pre-authored stories; in virtual worlds the users more or less create the stories themselves. The current trend in virtual worlds is to provide a mix of pre-authored and user-generated stories and content, leading to user-modified content.

Current virtual worlds are graphical and rendered in 2D, 2.5D (isometric view) or 3D depending on the intended effect and technical capabilities of the platform, i.e., Web browser, gaming PC, average PC, game console, mobile phone, and so on.

Would it not be great if the real world economy could be boosted by the exponential growing economy of the virtual worlds by connecting the virtual - and real world? In 2007 the Virtual Economy in Second Life's alone was around 400 MEuro, a factor nine growth from 2006. The connected devices and services in the real world can represent an economy of a multiple of this virtual world economy.

In the future, virtual worlds will probably fully enter our lives, our communication patterns, our culture, and our entertainment never to leave again. It's not only the teenager active in Second life and World of Warcraft, the average age of a gamer is 35 years by now, and it increases every year. This does not even include role-play in the professional context, also known as serious gaming, inevitable when learning practical skills. Virtual worlds are in use for entertainment, education, training, getting information, social interaction, work, virtual tourism, reliving the past and forms of art. They augment and interact with our real world and form an important part of people's lives. Many virtual worlds already exist as games, training systems, social networks and virtual cities and world models. Virtual worlds will most likely change every aspect of our lives, e.g., the way we work, interact, play, travel and learn. Games will be everywhere and their societal need is very big, it will lead to many new products and it requires many companies.

Technology improvement, both in hardware and software, forms the basis. It is envisaged that the most important developments will occur in the areas of display technology, graphics, animation, (physical) simulation, behavior and artificial intelligence, loosely distributed systems and network technology. Furthermore, a strong connection between the virtual and the real world is needed to reach simultaneous

Corresponding author: Christian Timmerer, Klagenfurt University, Universitätsstrasse 65-67 9020 Klagenfurt, +43 463 2700-3621, christian.timmerer@itec.uni-klu.ac.at

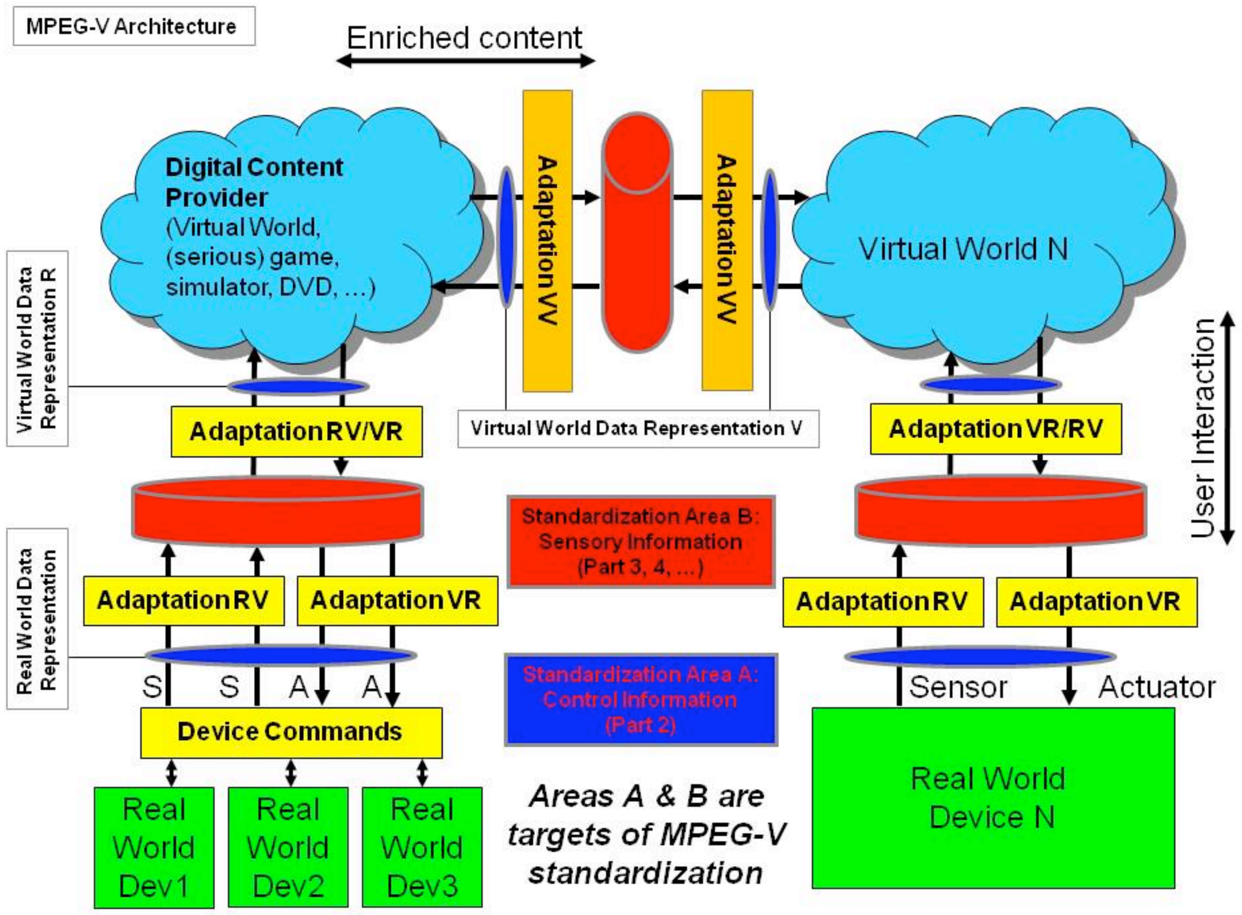
MPEG-V Architecture
Enriched content
Digital Content Provider
(Virtual World, (serious) game, simulator, DVD, …)
Adaptation VV
Adaptation VV
Virtual World N
Virtual World Data Representation R
Virtual World Data Representation V
User Interaction
Adaptation RV/VR
Adaptation VR/RV
Standardization Area B: Sensory Information (Part 3, 4, …)
Real World Data Representation
Adaptation RV
Adaptation VR
Adaptation RV
Adaptation VR
Standardization Area A: Control Information (Part 2)
S
S
A
A
Sensor
Actuator
Device Commands
Real World Dev1
Real World Dev2
Real World Dev3
Areas A & B are targets of MPEG-V standardization
Real World Device N

Figure 1. System Architecture of the MPEG-V Framework.

reactions in both worlds to changes in the environment and human behavior. Efficient, effective, intuitive and entertaining interfaces between users and virtual worlds are of crucial importance for their wide acceptance and use. To improve the process of creating virtual worlds a better design methodology and better tools are indispensible. For fast adoption of virtual worlds we need a better understanding of their internal economics, rules and regulations. And finally interoperability achieved trough standardization.

In particular, MPEG-V (ISO/IEC 23005) will provide an architecture and specifies associated information representations to enable the interoperability between virtual worlds, e.g., digital content provider of a virtual world, (serious) gaming, simulation, DVD, and with the real world, e.g., sensors, actuators, vision and rendering, robotics (e.g. for revalidation), (support for) independent living, social and welfare systems, banking, insurance, travel, real estate, rights management and many others. This bridging will provide a lower entry level to (multiple) virtual worlds both for the provider of goods and services as well as the user.

This paper is organized as follows. Section 2 describes the system architecture and gives an overview of MPEG-V. Section 3 reviews MPEG-V's most advanced part which is referred to as Sensory Information. In particular, the Sensory Effect Description Language (SEDL) is described as well as the Sensory Effect Vocabulary (SEV). Furthermore, a detailed usage example is given. Finally, the paper is concluded in Section 4.

2 SYSTEM ARCHITECTURE

The overall system architecture for the MPEG-V framework [3] is depicted in Figure 1 comprising two standardization areas. A control information and B sensory information. It is foreseen that standardization area B may be composed of multiple parts of the MPEG-V standard. The individual elements of the architecture have the following function:

Digital Content Provider. A provider of digital content, real time or non real time, of various nature ranging from an on-line virtual world, simulation environment, multi user game, a broadcasted multimedia production, a peer-to-peer multimedia production or 'packaged content like a DVD or game.

Virtual World Data Representation R and V. The native representation of virtual world related information that is intended to be exchanged with the real world and with another virtual world respectively (either exported or imported). On the other hand, the **Real World Data Representation** is referred to as the native representation of real world related

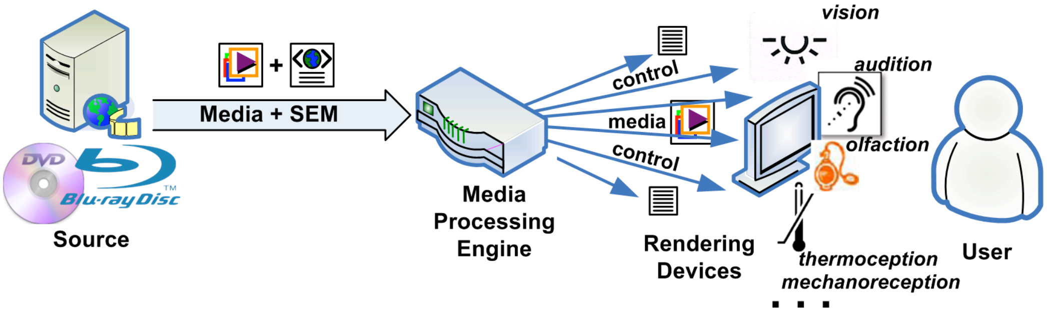

Figure 2. Concept of MPEG-V Sensory Effect Description Language [6].

information that is intended to be exchanged with the virtual world (either exported or imported).

The current architecture envisages the adaptation of the native representation of virtual world related information to the standardized representation format of MPEG-V in the standardization area B (cf. **Adaptation RV/VR** and **Adaptation VV**). This might be required for both the information that is intended to be exchanged with the real world and with another virtual world. Furthermore, this adaptation might be required bidirectional, i.e., from the standardized representation into the native representation and vice versa. Examples of these representation formats include effect information, haptic/tactile information, emotion information, etc. and are collectively referred to as **Sensory Information**.

In addition to the above-mentioned adaptations, the MPEG-V standard foresees further adaptations between the standardization areas A and B which are defined as **Adaptation RV** and **Adaptation VR** respectively. This kind of adaptation becomes necessary due to the possible mismatch of data representations in the virtual worlds and the real world. In particular, standardization area A – **Control Information** – is concerned about the description of the capabilities of real world devices including the user's preferences and device commands how to control these devices. The control information is bi-directional as it conveys information from the real world towards the virtual world (i.e., capabilities and preferences) and vice versa (i.e., the actual control). On the other hand, standardization area B is related to the virtual world data representation. However, a one-to-one mapping between the data representation of virtual worlds and real world devices is impractical and, thus, adaptation becomes necessary which also needs to be provided in both directions.

Finally, **Real World Device S** is referred to as a sensor (e.g., a temperature, light intensity, blood pressure, heartbeat) and **Real World Device A** is defined as an actuator (e.g., a display, speaker, light speaker, fan, robot, implant). Note that real world devices can contain any combination of sensors and actuators in one device.

Currently, the MPEG-V standard is at working draft level but is expected to become an international standard in late 2010. It currently comprises the following parts:

- Part 1: *Architecture* [3] as described in this section.
- Part 2: *Control information* covering standardization area A.
- Part 3: *Sensory Information* [4] which is part of standardization area B provides means for describing sensory effects as described in the next section. This is also the most advanced part of MPEG-V. Furthermore, haptic, tactile, and emotion information also falls in this standardization area but lacks of details at the time of writing this paper.
- Part 4: Avatar characteristics are also related to standardization area B and provides data representation formats to describe avatars that are intended to be exchanged with another virtual worlds.

In the following we will provide details about the sensory information and, in particular, how to describe sensory effects and how they shall be rendered within the end users premises. For further information concerning MPEG-V the interested reader is referred to the MPEG Web site [5].

3 SENSORY INFORMATION

3.1 Sensory Effect Description Language

Note that this section represents an updated version of what can be found in [6].

The Sensory Effect Description Language (SEDL) [4] 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 3.2). 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 2. 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* – sometimes also referred to as RoSE 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 [4]. 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 ::= [autoExtraction]
    [DescriptionMetadata](Declarations|
    GroupOfEffects|Effect|ReferenceEffect)+
```

SEM is the root element which may contain an optional *autoExtraction* attribute and *DescriptionMetadata* 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 [7] 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 EffectDefinition
  EffectDefinition (EffectDefinition)*
```

A *GroupOfEffects* starts with a *timestamp* which provides information about the point in time when this group of effects should become available for the application. This information can be used for rendering purposes and synchronization with the associated media resource. Therefore, the so-called XML Streaming Instructions as defined in MPEG-21 Digital Item Adaptation [8] have been adopted which offer this functionality. Furthermore, a *GroupOfEffects* shall contain at least two *EffectDefinition* for which no timestamps are required as they are provided within the enclosing element. The actual *EffectDefinition* comprises all information pertaining to a single sensory effect.

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

An *Effect* is used to describe a single effect with an associated *timestamp*.

```
EffectDefinition::=[SupplementalInformation]
  [activate][duration][fade-in][fade-out]
  [alt][priority][intensity][position]
  [adaptability][autoExtraction]
```

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 defined as follows: *activate* describes whether the effect shall be activated; *duration* describes how long the effect shall be activated; *fade-in* and *fade-out* provide means for fading in/out effects respectively; *alt* describes an alternative effect identified by a URI (e.g., in case the original effect cannot be processed); *priority* describes the priority of effects with respect to other effects in the same group of effects; *intensity* indicates the strength of the effect in percentage according to a predefined scale/unit (e.g., for wind the Beaufort scale is used); *position* describes the position from where the effect is expected to be received from the user's perspective (i.e., a three-dimensional space is defined in the standard); *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.

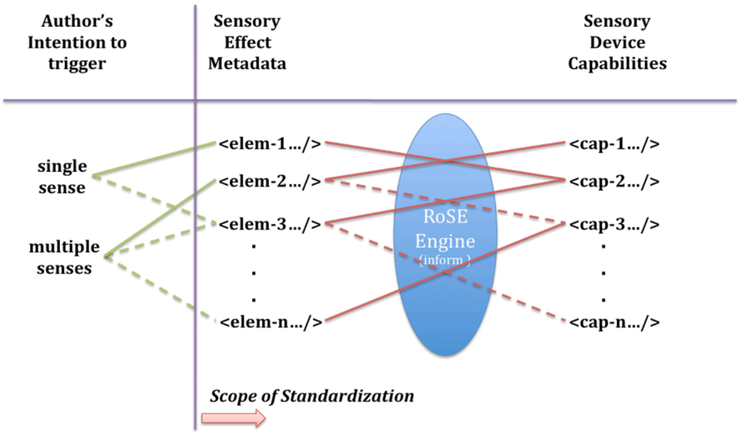

Figure 3. Mapping of Author's Intentions to Sensory Effect Metadata and Sensory Device Capabilities [4].

3.2 Sensory Effect Vocabulary

The Sensory Effect Vocabulary (SEV) defines a clear set of actual sensory effects to be used with the Sensory Effect Description Language (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. Furthermore, the effects are defined in a way to abstract from the authors intention and be independent from the end user's device setting as depicted in Figure 3. The sensory effect metadata elements or data types are mapped to commands that control sensory devices based on their capabilities. This mapping is usually provided by the RoSE 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 effects.

Light, colored light, flash light for describing light effects with the intensity in terms of illumination expressed in [lux]. For the color information, a classification scheme (CS) is defined by the standard comprising a comprehensive list of common colors. Furthermore, it is possible to specify the color as RGB. The flash light effect extends the basic light effect by 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.

Finally, the **color correction** provides means to define parameters that may be used to adjust the color information in a media resource to the capabilities of end user devices. Furthermore, it is also possible to define a region of interest where the color correction shall be applied in case this desirable (e.g., black/white movies with one additional color such as red).

3.3 Usage Example

In this section we will provide an example of Sensory Effect Metadata with an in-depth description how it shall be used by a media processing engine to control the available sensory devices. Lets assume we have a movie with windy scenes, possibly at different temperatures and also in combination with different vibrations (e.g., earthquake, turbulences in an airplane, etc.). Additionally, we may observe different illumination conditions with different colors. In previous work [6] we have shown that it is feasible to extract the color information directly from the media resource for steering additional light sources which might be deployed around the display from which the movie is expected to be received. Interestingly, the color information can be extracted also from certain regions of the movie which can be associated to certain light sources (e.g., left part of the movie is associated to the light sources left to the display from the users perspective). Please note that using the colored light effect and exploiting the position attribute as indicated in the following excerpt (Listing 1) can also describe this kind of effect.

Listing 1. Example for a Colored Light Effect.

```
<sedl:Effect xsi:type="sev:LightType"
  color="urn:mpeg:mpeg-v:01-SI-ColorCS-
NS:alice_blue"
  position="urn:mpeg:mpeg-v:01-SI-
PositionCS-NS:left:front:*"
  duration="..." si:pts="..." .../>
```

The color attribute refers to a CS term describing the color Alice blue (i.e., #F0F8FF) and the position attribute defines that the effect shall be perceived from the front/left from the user's perspective. That is, the light source left to the display should render this effect. The other attributes like duration and presentation time stamp (pts) will be described in the following excerpts.

A light breeze during a warm summer full moon night could be defined by combining the wind (i.e., light breeze), temperature (i.e., warm summer), and light effects (i.e., full moon night) as shown in Listing 2.

Listing 2. Example for Group of Effects.

```
<sedl:GroupOfEffects si:pts="3240000"
  duration="100" fade-in="15" fade-out="15"
  position="urn:mpeg:mpeg-v:01-SI-
PositionCS-NS:center:*:front">
 <sedl:Effect xsi:type="sev:WindType"
   intensity="0.0769"/>
 <sedl:Effect
   xsi:type="sev:TemperatureType"
   intensity="0.777"/>
 <sedl:Effect xsi:type="sev:LightType"
   intensity="0.0000077"/>
</sedl:GroupOfEffects>
```

The si:pts attribute indicates the start of the effect according to a predefined time scheme and the duration attribute defines how long it shall last. Furthermore, the effect's intensity should be reached within the time period as defined by the fade-in attribute. The same approach is used when the effect is about to finish (cf. fade-out attribute).

The group of effects comprises three single effect elements.

The first element, i.e., sev:WindType, is responsible to render a light breeze which is about Beaufort one (out of 13 possible values on this scale) that results in an intensity value of 0.0769 (approx. 7.69%). The rendering of such an effect could be achieved by fans (or ventilators) which are deployed around the user. A simple deployment would have two fans, one right and the other one left of the display. The media processing engine will map the intensity from the effect description to the capabilities of the fans which are ideally described using the same scale as for the effect description. On the other hand, if the fans can be controlled only at fixed intervals, the intensity value could be directly mapped to these intervals.

A warm summer could be characterized by 25°C and is signalled by means of the second element with sev:TemperatureType. In this case the domain has been chosen from the min/max measured temperatures on earth which are in the range of about [-90, +58]. Thus, the intensity of 25°C is represented as 0.777 (approx. 77.7%). An air condition could be used to render this type of effect but appropriate handling time needs to be taken into account.

The last effect, i.e., sev:LightEffect, shall render a full moon night which can be commonly described as one lux in terms of illumination. The domain defined in the standard has a range of [10^{-5}lux, 130klux] which corresponds to the light from Sirius, the brightest star in the night sky and direct sunlight respectively. Consequently, the intensity of this effect will be represented as 0.0000077 (approx 0.00077%). There are multiple devices that could render this effect such as various lamps, window shades, or a combination thereof. The standard deliberately does not define which effects shall be rendered on which devices which is left open for industry competition and, in particular, for media processing engine manufacturers.

Finally, the movie might include scenes like an earthquake or turbulences in an airplane which calls for a vibration effect as shown in Listing 3.

Listing 3. Example for a Vibration Effect.

```
<sedl:Effect xsi:type="sev:VibrationType"
  intensity="0.56" duration="..."
  si:pts="..." .../>
```

Assuming we would like to generate a vibration that is comparable to 5.6 on the Richter magnitude scale, this would result in an intensity of 0.56 (i.e., 56%) if we consider 10 as the maximum although no upper limit is defined. However, an earthquake with this intensity has never been recorded and it is unlikely that such a similar effect shall be created for this kind of application. A device that could render such an effect could be a TV/armchair equipped with additional vibration engines that may be configured at different strengths. The mapping from the effect's intensity value to the device's capabilities is similar to that from the previous effects.

4 CONCLUSION

In this paper we have presented an overview of MPEG-V which is an emerging standard for interfacing with virtual worlds. In particular, we have motivated the need for standardized interfaces that allow for inter-virtual world communication and also virtual-real world communication. Furthermore, we provided a detailed overview of Part 3 of MPEG-V, entitled Sensory Information, which is the most advanced part so far. Currently, it comprises means for describing sensory effects that may be perceived in conjunction with the traditional audio-visual media resources in order to increase the Quality of Experience.

The development aspects of the MPEG-V standard are discussed within a so-called Ad-hoc Group (AhG) that is open to the public and interested parties are invited to join this exciting activity. Details about the AhG on MPEG-V can be found at the MPEG Web site [9].

References

[1] W. Roush, "SecondEarth", *TechnologyReview*, July/August 2007.

[2] M. Papastergiou, "Digital Game-Based Learning in high school Computer Science education: Impact on educational effectiveness and student motivation", *Computers & Education*, vol. 52, no. 1, January 2009, pp. 1–12.

[3] Jean. H. A Gelissen (ed.), "Working Draft of ISO/IEC 23005 Architecture," *ISO/IEC JTC 1/SC 29/WG 11/N10616*, Maui, USA, April 2009.

[4] C. Timmerer, S. Hasegawa, S.-K. Kim (eds.) "Working Draft of ISO/IEC 23005 Sensory Information," *ISO/IEC JTC 1/SC 29/WG 11/N10618*, Maui, USA, April 2009.

[5] MPEG Web site, *MPEG-V*, http://www.chiariglione.org/mpeg/working_documents.htm#MPEG-V (last accessed: May 2009).

[6] M. Waltl, C. Timmerer, H. Hellwagner, "A Test-Bed for Quality of Multimedia Experience Evaluation of Sensory Effects", *Proceedings of the First International Workshop on Quality of Multimedia Experience (QoMEX 2009)*, San Diego, USA, July, 2009.

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

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

[9] ISO/MPEG, "Ad-hoc Group on MPEG-V", ISO/IEC MPEG/N10681, Maui, USA, April 2009. http://www.chiariglione.org/mpeg/meetings.htm (last accessed: May 2009).