INTERNATIONAL STANDARD

ISO/IEC 23005-1

Second edition
2014-01-15

Information technology — Media context and control —

Part 1: Architecture

Technologies de l'information — Contrôle et contexte de supports —

Partie 1: Architecture

Reference number
ISO/IEC 23005-1:2014(E)

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

ISO/IEC 23005-1 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

This second edition cancels and replaces the first edition (ISO/IEC 23005-1:2011), which has been technically revised.

ISO/IEC 23005 consists of the following parts, under the general title *Information technology — Media context and control*:

— *Part 1: Architecture*

— *Part 2: Control information*

— *Part 3: Sensory information*

— *Part 4: Virtual world object characteristics*

— *Part 5: Data formats for interaction devices*

— *Part 6: Common types and tools*

— *Part 7: Conformance and reference software*

Introduction

The usage of multimedia content is becoming omnipresent in everyday life, in terms of both consumption and production. On the one hand, professional content is provided to the end user in high-definition quality, streamed over heterogeneous networks, and consumed on a variety of different devices. On the other hand, user-generated content overwhelms the Internet with multimedia assets being uploaded to a wide range of available Websites. That is, the transparent access to multimedia content, also referred to as Universal Multimedia Access (UMA), seems to be technically feasible. However, UMA mainly focuses on the end-user devices and network connectivity issues, but it is the user who ultimately consumes the content. Hence, the concept of UMA has been extended to take the user into account, which is generally referred to as Universal Multimedia Experience (UME).

However, the consumption of multimedia assets can also stimulate senses other than vision or audition, e.g. olfaction, mechanoreception, equilibrioception, or thermoception. That is, in addition to the audio-visual content of, for example, a movie, other senses shall also be stimulated giving the user the sensation of being part of the particular media which shall result in a worthwhile, informative user experience.

This motivates the annotation of the media resources with metadata as defined in this part of ISO/IEC 23005 that steers appropriate devices capable of stimulating these other senses.

ISO/IEC 23005 (MPEG-V) provides an architecture and specifies associated information representations to enable the interoperability between virtual worlds, for example, digital content provider of a virtual world, (serious) gaming, simulation, DVD, and with the real world, for example, 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.

Virtual worlds[1)] (often referred to as 3D3C for 3D visualization and 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 social networks. The emergence of virtual worlds as platforms for social networking is recognized by businesses as an important issue for at least two reasons:

a) 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;

b) it allows 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. Resistance to real-world commercial encroachment still exists in many virtual worlds where users primarily seek an escape from real life. Hence, marketers should get to know a virtual world beforehand and the rules that govern each individual universe.

Although realistic experiences have been achieved via devices such as 3-D audio/visual devices, it is hard to realize sensory effects only with presentation of audiovisual contents. The addition of sensory effects leads to even more realistic experiences in the consumption of audiovisual contents. This will lead to the application of new media for enhanced experiences of users in a more realistic sense.

Such new media will benefit from the standardization of a control and sensory information which can include sensory effect metadata, sensory device (actuator) capabilities/commands, user's sensory preferences, and various delivery formats. The MPEG-V architecture can be applicable for various business models for which audiovisual contents can be associated with sensory effects that need to be rendered on appropriate sensory devices (actuators).

Multi-user online virtual worlds, sometimes called Networked Virtual Environments (NVEs) or massively-multiplayer online games (MMOGs), have reached mainstream popularity. Although most

1) Some examples of virtual worlds are: *Second Life* (http://secondlife.com/), IMVU (http://www.imvu.com/) and Entropia Universe (http://www.entropiauniverse.com/).

publications 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 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 70s with the creation of the text-based Dungeons and Dragons world 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. 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, 2.5 D (isometric view) or 3D, depending on the intended effect and technical capabilities of the platform: 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* 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.

Virtual worlds have entered 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, obtaining 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 change every aspect of our lives: the way we work, interact, play, travel and learn. Games will be everywhere and their societal need is very big and will lead to many new products and require many companies.

Technology improvement, both in hardware and software, forms the basis of this. It is envisaged that the most important developments will occur in the areas of display technology, graphics, animation, (physical) simulation, behaviour and artificial intelligence, loosely distributed systems and network technology.

The figures in this part of ISO/IEC 23005 have been reproduced here with the permission of Samsung, Sharp Electronics, ETRI, University of Klagenfurt, Institute of Science and Technology, Myongji University, Institut national des télécommunications and the partners of the ITEA2 project Metaverse1: Philips, Forthnet S.A., Alcatel-Lucent Bell N.V., Innovalia, Alcatel-Lucent France, Technicolor, Orange Labs, DevLab, CBT, Nextel, Carsa, Avantalia, Ceesa, Virtualware, I&IMS, VicomTECH, E-PYME, CIC Tour Gune, Artefacto, Metaverse Laboratories, Technical University Eindhoven, Utrecht University, University of Twente, Stg. EPN, VU Economics & BA, VU CAMeRA, Ellinogermaniki Agogi, IBBT-SMIT, UPF-MTG, CEA List and Loria/Inria Lorraine.

Information technology — Media context and control —

Part 1:
Architecture

1 Scope

This part of ISO/IEC 23005 specifies the architecture of MPEG-V (media context and control), and its three associated use cases of information adaptation from virtual world to real world, information adaptation from real world to virtual world, and information exchange between virtual worlds.

2 Terms and definitions

2.1
Device Command
description of controlling actuators used to generate **Sensory Effects.**

2.2
R→V Adaptation

procedure that processes the Sensed Information from the real world in order to be consumed within the virtual world's context; takes the Sensed Information with/without the Sensor Capabilities from Sensors, the Sensor Adaptation Preferences from Users, and/or the Virtual World Object Characteristics from a Virtual world; controls the Virtual World Object Characteristics or adapts the Sensed Information by adapting the Sensed Information based on the Sensor Capabilities and/or the Sensor Adaptation Preferences

2.3
Sensed Information
information acquired by sensors

2.4
Sensor
device by which user input or environmental information can be gathered

EXAMPLE Temperature sensor, distance sensor, motion sensor, etc.

2.5
Sensor Adaptation Preferences
description schemes and descriptors to represent (user's) preferences with respect to adapting sensed information

2.6
Sensor Capability
description of representing the characteristics of sensors in terms of the capability of the given sensor such as accuracy, or sensing range

2.7
Sensory Device
consumer device by which the corresponding sensory effect can be made

Note 1 to entry: Real world devices can contain any combination of sensors and actuators in one device.

2.8
Sensory Device Capability
description of representing the characteristics of actuators used to generate sensory effects in terms of the capability of the given device

2.9
Sensory Effects
effects to augment perception by stimulating human senses in a particular scene

EXAMPLE Scent, wind, light, haptic [kinesthetic-force, stiffness, weight, friction, texture, widget (button, slider, joystick, etc.), tactile: air-jet, suction pressure, thermal, current, vibration, etc. Note that combinations of tactile display can also provide directional, shape information].

2.10
Sensory Effect Metadata
metadata that defines the description schemes and descriptors to represent sensory effects

2.11
User's Sensory Preferences
description schemes and descriptors to represent (user's) preferences with respect to rendering of sensory effect

2.12
User
the end user of the system

2.13
Virtual World
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

2.14
V→R Adaptation
procedure that processes the Sensory Effects from the Virtual World in order to be consumed within the real world's context; takes Sensory Effect Metadata from a Virtual World, Sensory Device (Actuator) Capabilities from the Sensory Devices (Actuators), the User's Sensory Preferences from users, and/or the Sensed Information as well as the Sensor Capabilities from Sensors as inputs; generates the Device Commands by adapting the Sensory Effects based on the Sensed Information, the Capabilities and/or the Preferences

2.15
VW Object Characteristics
description schemes and descriptors to represent and describe virtual world objects (from the real world into the virtual world and vice versa)

3 MPEG-V System Architecture

A strong connection (defined by an architecture that provides interoperability trough standardization) between the virtual and the real world is needed to reach simultaneous reactions in both worlds to changes in the environment and human behaviour. 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 world technologies we need a better understanding of their internal economics, rules and regulations. The overall system architecture for the MPEG-V framework is depicted in Figure 1.

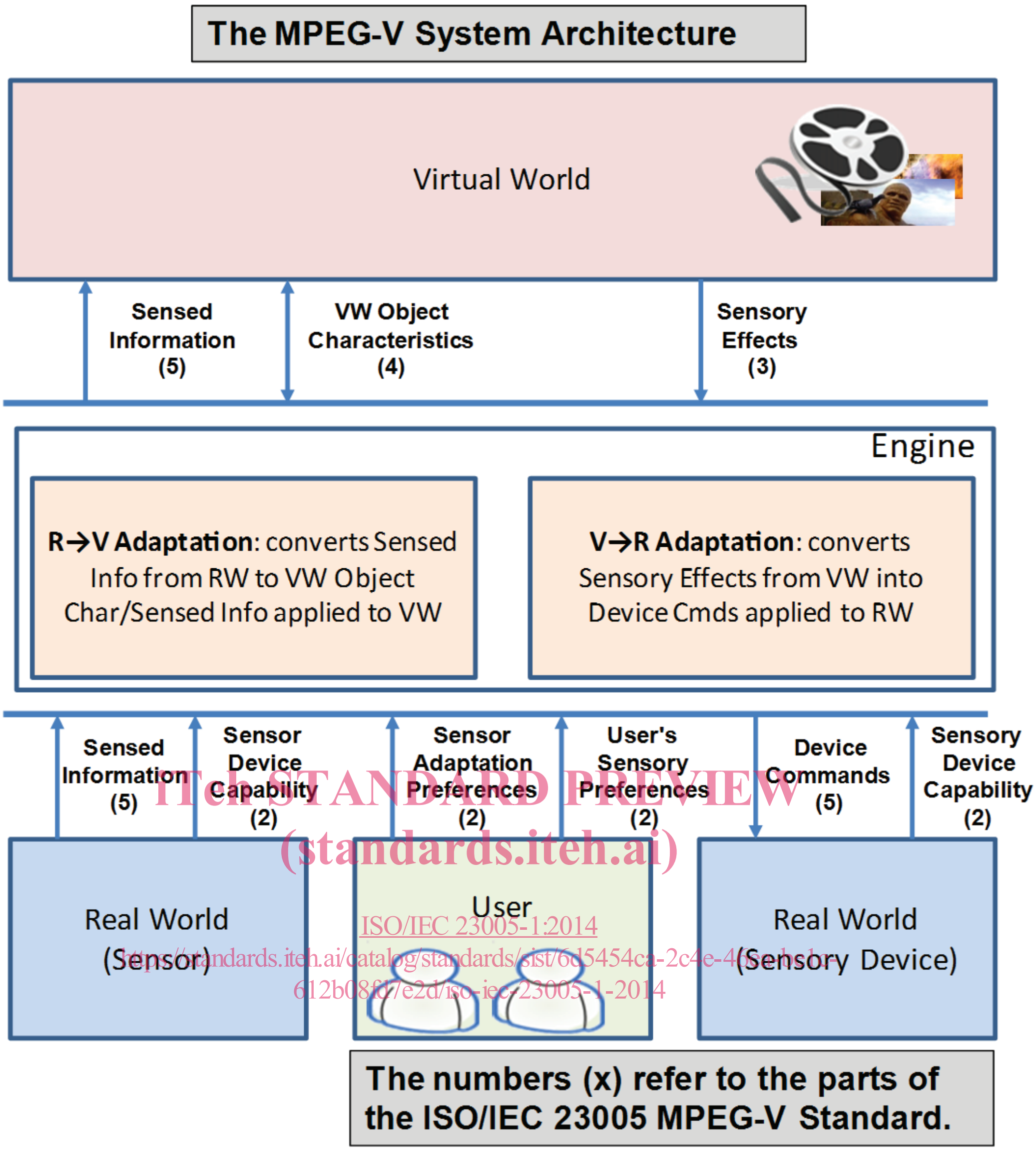

Figure 1 — System Architecture of the MPEG-V Framework

The MPEG-V System Architecture can be used to serve three different media exchanges. There are two types of media exchanges occurring between real world and virtual world, i.e. the information exchange from real world to virtual world and the information exchange from virtual world to real world. An additional type of media exchanges is the information exchange between virtual worlds. The three media exchanges are defined as use cases in Clause 4.

It is important to note that *Sensory Effect Metadata*, *Sensory Device Capability*, User's *Sensory Preferences*, *Device Commands*, *Sensed Information*, *Sensor Device Capability*, *Sensor Adaptation Preferences*, *and Virtual World Object Characteristics* are within the scope of standardization and, thus shall be normatively specified. On the other side, the *V→R Adaptation Engine*, *V→R Adaptation Engine*, *Virtual World*, as well as *Devices (Sensors and Sensory devices)* are informative and are left open for industry competition.

Metadata within the scope is formed other parts of the ISO/IEC 23005. *Sensor Device Capability*, *Sensory Device Capability*, *Sensor Adaptation Preferences*, and User's *Sensory Preferences* are specified in Part 2: Control information. *Sensory Effect Metadata* is specified in Part 3: Sensor information. *Virtual World Object Characteristics* is specified in Part 4: Virtual world object characteristics. *Device Commands* and *Sensed Information* are specified in Part 5: Formats for interaction devices.

4 Use cases

The three media exchanges require information adaptations in order for a targeting world to adapt information based on capabilities and preferences: information adaptation from virtual world to real world, information adaptation from real world to virtual world, and information adaptation between virtual worlds.

4.1 Information adaptation from virtual world to real world

4.1.1 System Architecture for information adaptation from virtual world to real world

The system architecture for the information adaptation from virtual world to real world is depicted in Figure 2. It represents V→R adaptation comprising *Sensory Effect Metadata*, *VW Object Characteristics*, *Sensory Device Capability (actuator capability)*, *Device Commands*, User's *Sensory Preferences*, and a *V→R Adaptation* Engine which generates output data based on its input data.

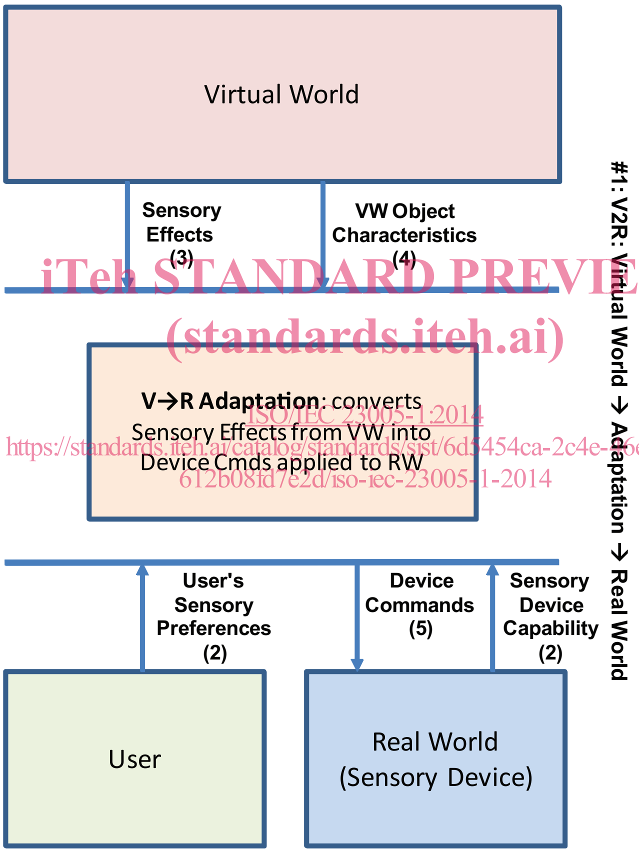

Figure 2 — (Possible) System Architecture for information adaptation from virtual world to real world

A *virtual world* within the framework is referred to as an entity that acts as the source of the *sensory effect metadata* and *VW Object Characteristics* such as a broadcaster, content creator/distributor, or even a service provider. The *V→R Adaptation Engine* is an entity that takes the *sensory effect metadata*, the *sensory device (actuator) capability* and the user's *sensory preferences* as inputs and generates the *device commands* based those in order to control the consumer devices enabling a worthwhile, informative experience to the user.

Real world devices (sensory devices) are entities that act as the sink of the *device commands* and act as the source of *sensory device (actuator) capability*. Additionally, entities that provide user's *sensory preferences* towards the *RoSE* engine are also collectively referred to as *real world devices*. Note that *sensory devices* (actuators) are sub-set of *real world devices* including fans, lights, scent devices, human input devices such as a TV set with a remote control (e.g. for preferences).

The actual *sensory effect metadata* provides means for representing so-called *sensory effects*, i.e. effects to augment feeling by stimulating human sensory organs in a particular scene of a multimedia application. Examples of *sensory effects* are scent, wind, light, etc. The means for transporting this kind of metadata is referred to as sensory effect delivery format which, of course, could be combined with an audio/visual delivery format, e.g. MPEG-2 transport stream, a file format, or Real-time Transport Protocol (RTP) payload format, etc.

The *sensory device capability* defines description formats to represent the characteristics of sensory devices (actuators) in terms of which sensory effects they are capable to perform and how. A *sensory device* (actuator) is a consumer device by which the corresponding *sensory effect* can be made (e.g. lights, fans, heater, fan, etc.). *Device commands* are used to control the *sensory devices (actuators)*. As for *sensory effect metadata*, also for *sensory device (actuator) capability* and *device commands* corresponding means for transporting this assets are referred to as sensory device capability/commands delivery format respectively.

Finally, the user's *sensory preferences* allow for describing preferences of the actual (end) users with respect to rendering of *sensory effects* for also a delivery format is provided.

4.2 Information adaptation from real world to virtual world

4.2.1 System Architecture for information adaptation from real world to virtual world

The system architecture for information adaptation from real world to virtual world is depicted in Figure 3. It represents R2V adaptation comprising *VW Object Characteristics*, *Sensed Information*, *Sensor Capability*, *Sensor Adaptation Preferences*, and an *R→V Adaptation Engine* which generates output data based on its input data.

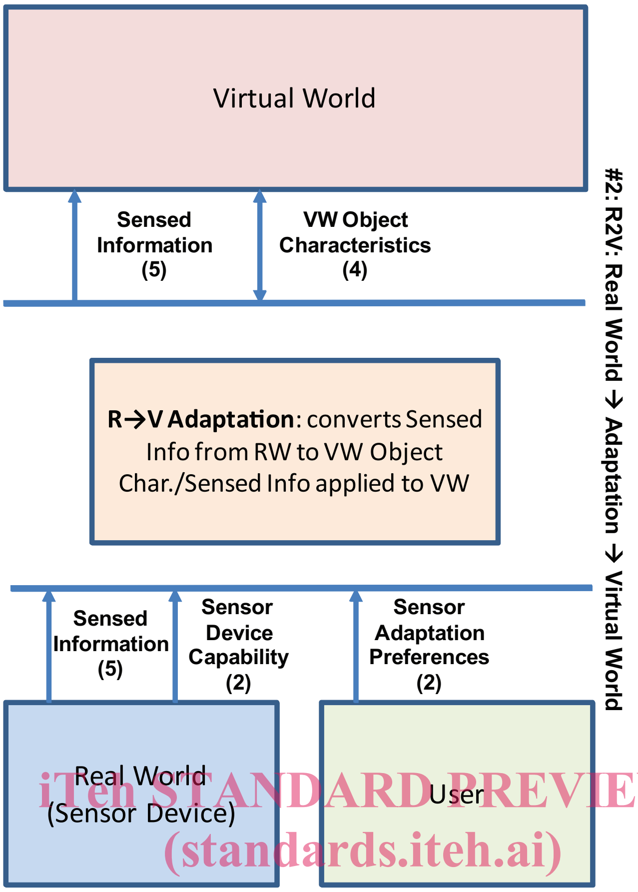

Figure 3 — (Possible) System Architecture for information adaptation from real world to virtual world

Entity that processes the Sensed Information from the real world in order to be consumed within the virtual world's context; takes the Sensed Information with/without the Sensor Capabilities from Sensors, the Sensor Adaptation Preferences from Users, and/or the Virtual World Object Characteristics from a Virtual world; controls the Virtual World Object Characteristics or adapts the Sensed Information by adapting the Sensed Information based on the Sensor Capabilities and/or the Sensor Adaptation Preferences.

There are two possible implementations to adapt information from real world to virtual world. In the first system implementation, *R→V adaptation* takes the *Sensor Capabilities* as inputs, the *Sensed Information* from *Sensors*, and *Sensor Adaptation Preferences* from *Users*; adapts the *Sensed Information* based on the *Sensor Capabilities* and/or *Sensor Adaptation Preferences*.

In the second system implementation, R→V adaptation takes the *Sensed Information* with/without the *Sensor Capabilities* from *Sensors*, the *Sensor Adaptation Preferences* from *Users*, and/or the *Virtual World Object Characteristics* from a *Virtual world*; controls the *Virtual World Object Characteristics* adapting the *Sensed Information* based on the *Sensor Capabilities* and/or the *Sensor Adaptation Preferences*.

4.3 Information exchange between virtual worlds

4.3.1 System Architecture for exchanges between virtual worlds

The system architecture for information exchange between virtual worlds is depicted in Figure 4. It represents information exchange comprising *VW Object Characteristics* which generates exchangeable information within virtual worlds.

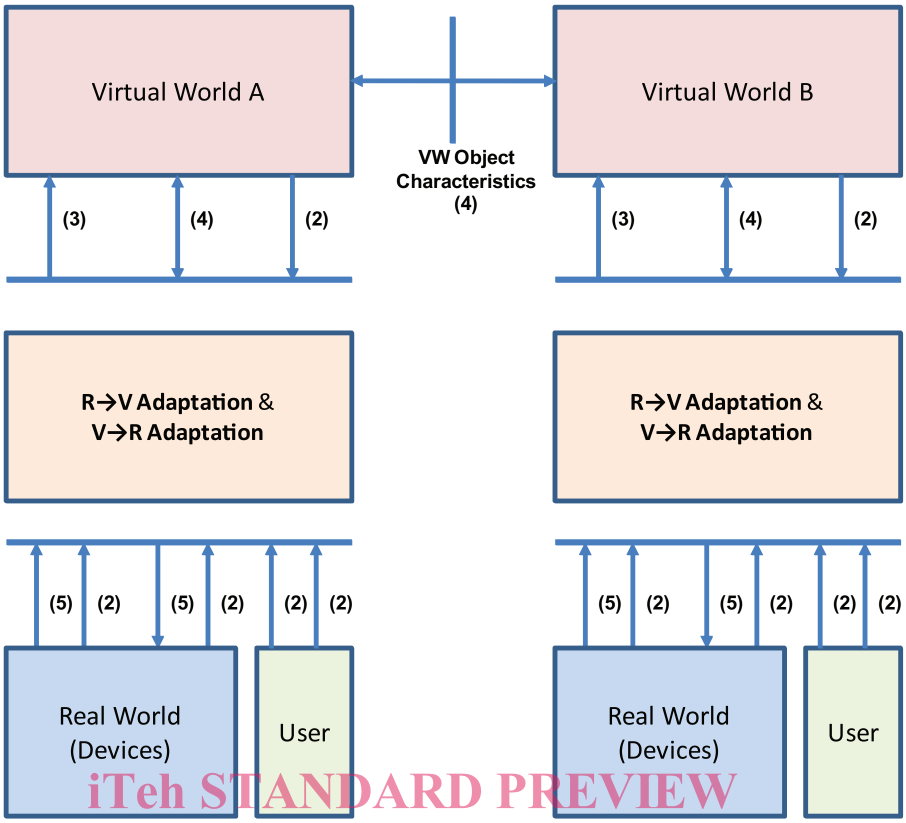

Figure 4 — (Possible) System Architecture for (bidirectional) exchange of information between virtual worlds

V→V adaptation adapts proprietary virtual world object characteristics from a Virtual World to *VW Object Characteristics* and sends the *VW Object Characteristics* from the *Virtual World* to another *Virtual World* to support interoperability. Based on the data provided in *Virtual World Object Characteristics*, the *Virtual World* will internally adapt its own representation for virtual object/avatar.

5 Instantiations

5.1 Instantiation A: Representation of Sensory Effects (RoSE)

5.1.1 System Architecture for Representation of Sensory Effects

The system for representation of sensory effects is partly instantiated from the system architecture of information adaption from virtual world to real world. The overall system architecture for Representation of Sensory Effects (RoSE) is depicted in Figure 5 comprising Sensory Effect Metadata, Sensory Device (actuator) Capability, Device Commands, User's Sensory Preferences, and a so-called RoSE Engine which generates output data based on its input data.

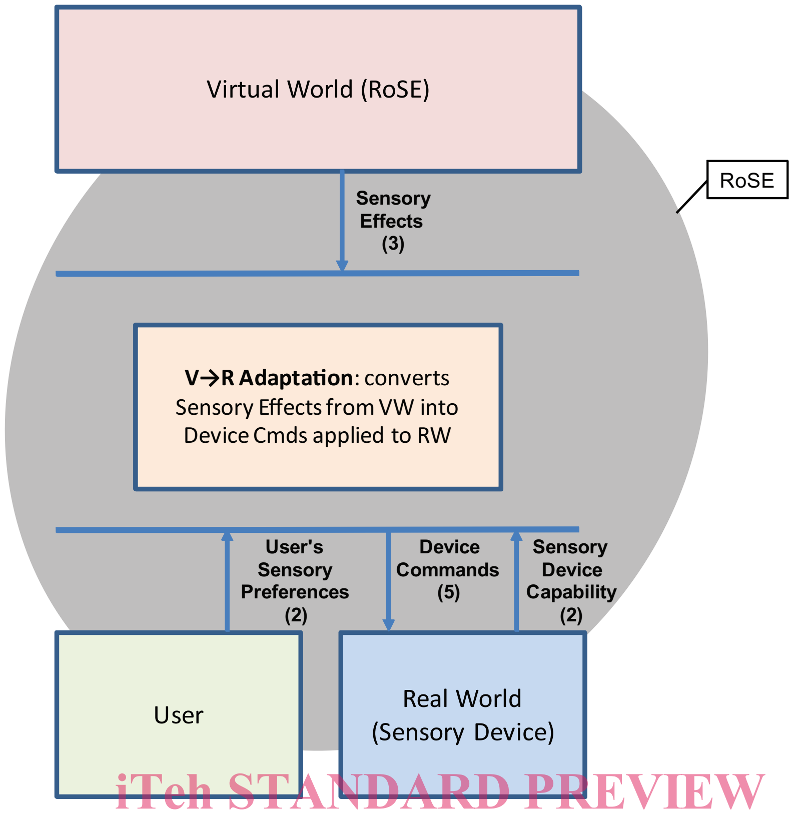

Figure 5 — RoSE System Architecture

A *provider* within the *RoSE* framework is referred to as an entity that acts as the source of the *sensory effect metadata* such as a broadcaster, content creator/distributor, or even a service provider. The *RoSE Engine* is an entity that takes the *sensory effect metadata*, the *sensory device (actuator) capability* and the user's *effect preferences* as inputs and generates the *device commands* based those in order to control the consumer devices enabling a worthwhile, informative experience to the user.

Consumer devices are entities that act as the sink of the *device commands* and act as the source of *sensory device (actuator) capability*. Additionally, entities that provide user's *sensory preferences* towards the *RoSE* engine are also collectively referred to as *consumer devices*. Note that *sensory devices (actuators)* are sub-set of *consumer devices* including fans, lights, scent devices, human input devices such as a TV set with a remote control (e.g. for preferences).

The actual *sensory effect metadata* provides means for representing so-called *sensory effects*, i.e. effects to augment feeling by stimulating human sensory organs in a particular scene of a multimedia application. Examples of *sensory effects* are scent, wind, light, etc. The means for transporting this kind of metadata is referred to as sensory effect delivery format which, of course, could be combined with an audio/visual delivery format, e.g. MPEG-2 transport stream, a file format, or Real-time Transport Protocol (RTP) payload format, etc.

The *sensory device (actuator) capability* defines description formats to represent the characteristics of sensory devices (actuators) in terms of which sensory effects they are capable to perform and how. A *sensory device (actuator)* is a consumer device by which the corresponding *sensory effect* can be made (e.g. lights, fans, heater, fan, etc.). *Device commands* are used to control the *sensory devices (actuators)*. As for *sensory effect metadata*, also for *sensory device (actuator) capability* and *device commands* corresponding means for transporting this assets are referred to as sensory device (actuator) capability/commands delivery format respectively.

Finally, the user's *sensory preferences* allow for describing preference of the actual (end) users with respect to rendering of *sensory effects* for also a delivery format is provided.

5.1.2 Instantiation A.1: Multi-sensorial Effects

Traditional multimedia with audio/visual contents have been presented to users via display devices and audio speakers as depicted in Figure 6. In practice, however, users are becoming excited about more advanced experiences of consuming multimedia contents with high fidelity. For example, stereoscopic video, virtual reality, 3-dimensional television, multi-channel audio, etc. are typical types of media increasing the user experience but are still limited to audio/visual contents.

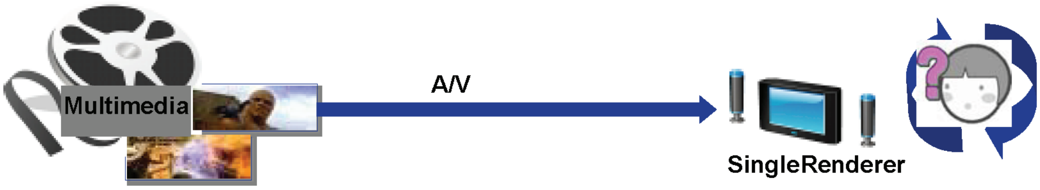

Figure 6 — Traditional Multimedia Consumption

From a rich multimedia perspective, an advanced user experience would also include special effects such as opening/closing window curtains for a sensation of fear effect, turning on a flashbulb for lightning flash effects as well as fragrance, flame, fog, and scare effects can be made by scent devices, flame-throwers, fog generators, and shaking chairs respectively. Such scenarios would require enriching multimedia contents with information enabling consumer devices to render them appropriately in order to create the advanced user experience such as described above. Figure 7 shows an example configuration adopting a multimedia multiple device (MMMD) approach for an advance user experience compared to the multimedia single device (MMMD) approach as illustrated in Figure 7. In this configuration, the multimedia contents are not rendered by a single device but with multiple devices in a synchronized manner.

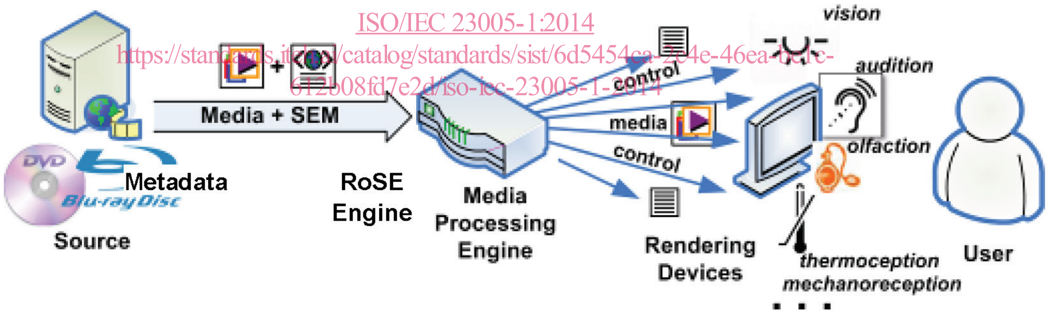

Figure 7 — RoSE-enabled Multimedia Consumption for Advanced User Experience

From a technical perspective, this requires a framework for the *Representation of Sensory Effects (RoSE)* information which may define metadata about special or sensory effects, characteristics of target devices, synchronizations, etc. The actual presentation of the RoSE information and associated audio/visual contents allows for an advanced, worthwhile user experience.

5.1.3 Instantiation A.2: Motion effects

One of the important sensory effects that we should not ignore is the effect related to the motion. The motion effect gives a user a similar feeling on the movement like the actor/actress feels in the movie. The motion effect is popular sensory effects commonly used in such a places like theme park, game room, and the movie theatre now a day. The motion effect is usually provided by the motion chair. The motion chair usually has motor(s) and axis underneath or above the chair. The number of motor and the length of axes determine the range and depth of the movement of the chair. There are a lot of manufacturers of motion chair in the world and each of them has its own mechanical characteristics as shown in Figure 8 below.