Università degli Studi di Cagliari

DOTTORATO DI RICERCA

INGEGNERIA ELETTRONICA E INFORMATICA

Ciclo XXXI

QUALITY OF EXPERIENCE METHODS AND MODELS FOR MULTI-SENSORIAL MEDIA

Settore scientifico disciplinare di afferenza

ING-INF/03 TELECOMUNICAZIONI

Presentata da Lana Jalal

Coordinatore Dottorato Prof. Fabio Roli

Tutor Prof. Maurizio Murroni

Esame finale anno accademico 2017 – 2018

Tesi discussa nella sessione d’esame Gennaio-Febbraio 2019

PH.D. IN ELECTRONIC AND COMPUTER ENGINEERING
Dept. of Electrical and Electronic Engineering
University of Cagliari

Quality of Experience Methods and Models for Multi-Sensorial Media

LANA JALAL

Advisor: Prof. Maurizio Murroni

Curriculum: ING-INF/03 Telecomunicazioni

XXXI Cycle

2018

Acknowledgements

First of all, I would like to express my warm thanks to my supervisor, Prof. Maurizio Murroni. His precious advice and support always encouraged me in the right direction of my research goal. It would not be possible for me to finish my work without his kind helps and wise suggestions.

Special thanks go to the research programme Erasmus Mundus Marhaba project, for giving me the opportunity to make important experiences and enrich my cultural and human awareness.

Last, but not least, I would love to thank my family members: my father, mother, sister, brothers and my beloved husband for their support and care.

To the soul of my brother, despite the loss of your physical presence, you are still alive in our hearts and minds.

List of Publications

1. L. Jalal, M. Anedda, V. Popescu, M. Murroni, “ Internet of Things for Enabling Multi Sensorial TV in Smart Home” 2018 IEEE Broadcast Symposium, Virginia, USA, October 2018, in press.
2. L. Jalal, V. Popescu, M. Murroni, “Experimental Setup for IoT - Based Multi Sensorial Media” Baltic URSI Symposium (URSI), Poznan, Poland, pp. 80–84, 2018.
3. L. Jalal, M. Anedda, V. Popescu, M. Murroni, “QoE Assessment for Broadcasting Multi Sensorial Media in Smart Home Scenario” IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Valencia, Spain, pp.1–5, 2018.
4. L. Jalal, M. Anedda, V. Popescu, M. Murroni, “QoE Assessment for IoT-Based Multi Sensorial Media Broadcasting.” IEEE Transactions on Broadcasting, vol. 64, no. 2, pp. 552–560. June 2018
5. L. Jalal, V. Popescu, M. Murroni, “IoT architecture for multisensorial media.” IEEE URUCON, Montevideo, Uruguay, pp. 1–4, 2017.
6. L. Jalal, M. Murroni, “Enhancing TV broadcasting services: A survey on mulsemedia quality of experience.” IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Cagliari, Italy, pp. 1–7, 2017.
7. L. Jalal, M. Murroni, “A nonlinear quality of experience model for high dynamic spatio-temporal mulsemedia.” 9th International Conference on Quality of Multimedia Experience (QoMEX) Erfurt, Germany, pp. 224–229, 2017.
8. L. Jalal, V. Popescu, M. Murroni, “Quality-of-experience parameter estimation for multisensorial media using Particle Swarm Optimization.” 2017 International Conference on Optimization of Electrical and Electronic Equipment (OPTIM) & 2017 Intl Aegean Conference on Electrical Machines and Power Electronics (ACEMP), Brasov, Romania, pp. 965–970, 2017.

Table of Contents

Acknowledgements .... **ii**

List of Publications .... **iii**

Table of Contents .... **iv**

List of Figures .... **viii**

List of Tables .... **xi**

List of Abbreviations .... **xii**

Chapter 1 General Introduction **1**

1.1 Introduction 1

1.2 Aims of the Thesis 4

1.3 Thesis Outline 6

Chapter 2 Background and Related Work **8**

2.1 Multi-Sensorial Media 8

2.2 Multi-Sensorial Media Implementation 9

2.2.1 Authoring Tools for Multi-Sensorial Media 9

2.2.2 Sensory Effect Devices 12

2.2.2.1 Light Devices 12

2.2.2.2 Scent Devices 14

2.2.2.3 Vibration Devices 14

2.2.3 Multi-Sensorial Media Synchronization 15

2.2.4 Multi-Sensorial Media Dataset 16

2.3 Multi-Sensorial Media Subjective Evaluation 18

2.3.1 Assessment Parameters 18

2.3.2 Subjective Assessment Methods 21

2.3.2.1 Absolute Category Rating 22

2.3.2.2 Absolute Category Rating with Hidden Reference 22

2.3.2.3 Degradation Category Rating 23

2.3.2.4 Double Stimulus Continuous Quality Scale 24

2.3.2.5 Single Stimulus Continuous Quality Scale 25

2.4 Influence of Sensory Effects 25

2.4.1 Viewing Experience and Sense of Reality 25

2.4.2 Video Quality 26

2.4.3 Sensory Effects and Emotions 27

2.5 Multi-Sensorial Media Broadcasting 27

Chapter 3 QoE Assessment for IoT Based Multi-Sensorial Media 30

3.1 Multi-Sensorial Media TV in Smart Home 30

3.2 IoT Architecture and Implementation 33

3.2.1 Architecture 33

3.2.2 Implementation 35

3.3 Experimental Setup.. 36

3.3.1 Test Environment.. 36

3.3.2 Participants 39

3.3.3 Assessment Procedure.. 40

3.3.4 Multi-Sensorial Video Sequences....................................... 42

3.4 Experimental Results and MOS Ratings 45

Chapter 4 Impact of Sensory Effects on User Quality of Experience........... 49

4.1 Sensory Effects Impact on User Experience Study................................... 49

4.2 User Perception and Preference of Multi-Sensorial Media 50

4.2.1 Sensory Effects Enhance the Sense of Reality................................. 52

4.2.2 Sensory Effects are Distracting 53

4.2.3 The Intensity of Sensory Effect... 54

4.2.4 Impact of Sensory Effects on User Enjoyment 55

4.2.5 The Sensory Effects Enhance the Audiovisual Content 56

4.2.6 Sensory Effects and Emotions.. 57

4.2.7 Impact of Multi-Sensory Sequence Duration................................... 57

4.2.8 Experience Additional Sensory Effects 58

4.2.9 User Preferable Sensory Effect... 59

4.2.10 The Impact of Light Effect on User Experience............................... 60

4.2.11 The Impact of Airflow on User Experience 61

4.2.12 The Impact of Vibration on User Experience 63

4.3 Summary 64

Chapter 5 QoE Model for Multi-Sensorial Media TV in Smart Home 66

5.1 Quality of Experience and Sensory Effects Models 66

5.2 Quality of Experience Model for Multi- Sensorial Media TV 68

5.3 Multi-Sensorial QoE Model Validation 69

5.4 Parameter Estimation 73

5.4.1 Multi-Sensorial QoE Model Parameter Estimation Using PSO 75

5.5 Multi-Sensorial QoE Model Estimated Response 77

5.5.1 Model Estimated Response for the First Empirical Dataset 79

5.5.2 Model Estimated Response for the Second Empirical Dataset 82

Chapter 6 Conclusions and Suggestions for Future Works 86

6.1 Conclusions 86

6.2 Suggestions for Future Works 89

References 91

List of Figures

Figure 2-1: Multi-sensorial media concept 9

Figure 2-2: amBX System[31] 13

Figure 2-3: Cyborg Gaming Lights [31] 13

Figure 2-4: Vortex Active [31] 14

Figure 2-5: Multi-sensorial media broadcasting architecture [60] 29

Figure 3-1: IoT architecture for multi-sensorial media 35

Figure 3-2: Setup of the proposed geometry 37

Figure 3-3: Equipment used in the assessment: (a) Arduino microcontroller; (b) RGB smart LED lights; (c) Smartphone camera; (d) Smartphones 38

Figure 3-4: Panoramic of the test environment 38

Figure 3-5: ACR assessment method 41

Figure 3-6: The test environment during the assessment 42

Figure 3-7: Snapshot of the multi-sensorial video sequences: (a) 2012; (b) Berrecloth; (c) Bridgestone; (d) Pastranas; (e) Ice ski 2; (f) Skyfall; (g) Fireworks; (h) Earth; (i) Elysium; (j) Ice ski 1 44

Figure 3-8: Mean opinion score and confidence intervals (95%) 47

Figure 3-9: Ratings 47

Figure 3-10: %GOB and %POW 48

Figure 4-1: Sensory effects enhance the sense of reality 53

Figure 4-2: Sensory effects are distracting 53

Figure 4-3: Intensity of light effect 54

Figure 4-4: Intensity of airflow effect 55

Figure 4-5: Intensity of vibration effect 55

Figure 4-6: Sensory experience is enjoyable 56

Figure 4-7: Sensory effects enhance the audiovisual content 56

Figure 4-8: Emotions increased with sensory effects 57

Figure 4-9: Do you prefer a longer sequence? 58

Figure 4-10: User interest in experience thermal effect 59

Figure 4-11: User interest in experience scent effect 59

Figure 4-12: Which sensorial effect do you prefer? 60

Figure 4-13: Light effect is annoying 61

Figure 4-14: Light effect enhances sense of reality 61

Figure 4-15: Airflow effect enhances sense of reality 62

Figure 4-16: Airflow effect is annoying 62

Figure 4-17: Vibration effect enhances sense of reality 63

Figure 4-18: Vibration effect is annoying 63

Figure 5-1: Snapshot of the action sequence 2012 71

Figure 5-2: Snapshot of the documentary sequence Earth 72

Figure 5-3: Snapshot of the sport sequence Berrecloth 72

Figure 5-4: Snapshot of the commercial sequence Bridgestone 73

Figure 5-5: Snapshot of the sport sequence Pastranas 73

Figure 5-6: The convergence of MSE for the PSO algorithm 78

Figure 5-7: The estimated response by the proposed model compared to the MOS for the action sequence 2012 79

Figure 5-8: The estimated response by the proposed model compared to the MOS for the sport sequence Pastranas 80

Figure 5-9: The estimated response by the proposed model compared to the MOS for the sport sequence Berrecloth... 80

Figure 5-10: Comparison between the linear and the proposed model in term of MSE 81

Figure 5-11: MOS and confidence intervals (95%) for the action sequence 2012......... 82

Figure 5-12: MOS and confidence intervals (95%) for the documentary sequence Earth 83

Figure 5-13: MOS and confidence intervals (95%) for the sport sequence Berrecloth.. 83

Figure 5-14: MOS and confidence intervals (95%) for the commercial sequence Bridgestone... 84

Figure 5-15: MOS and confidence intervals (95%) for the sport sequence Pastranas ... 84

Figure 5-16: The overall estimated response by the proposed model for the second empirical dataset 85

List of Tables

Table 1: Five-level impairment scale 23

Table 2: Five-level continuous quality scale 24

Table 3: List of hardware/software deployed 39

Table 4: Five-level rating scale 41

Table 5: Video sequences details 44

Table 6: Video sequences scenario 45

Table 7: Hardware and software components used in the subjective test [77] 70

Table 8: PSO parameters setting for the experiment 78

Table 9: Comparison between the linear and the proposed model in term of R^2 81

List of Abbreviations

ACR	Absolute Category Rating
ACR-HR	Absolute Category Rating with Hidden Reference
AD	Ambient Displays
AmI	Ambient Intelligence
CI	Confidence Intervals
DCR	Degradation Category Rating
DSCQS	Double Stimulus Continuous Quality Scale
GA	Genetic Algorithm
HS	Home Server
ICT	Information and Communication Technology
IoT	Internet of Things
ITU	International Telecommunication Union
LS	Least Square
MEs	Micro Engines
MEW	Multiplicative Exponential Weighting
MLR	Multiple Linear Regression
MOS	Mean Opinion Score
MPEG	Moving Picture Experts Group
MR	Multiple Regression
MSE	Mean Squared Error
mulsemedia	Multiple Sensorial Media

NLR	Nonlinear Regression
OTT	Over The Top
PCC	Pearson Correlation Coefficient
PSO	Particle Swarm Optimization
PSQA	Pseudo Subjective Quality Assessment
QoE	Quality of Experience
QoS	Quality of Service
RoSE	Representation of Sensory Effects
RWOs	Real World Objects
SA	Simulated Annealing
SDSCE	Simultaneous Double Stimulus for Continuous Evaluation
SEDL	Sensory Effect Description Language
SEM	Sensory Effect Metadata
SER	Sensory Effect Renderer
SESim	Sensory Effect Simulator
SEVino	Sensory Effect Video Annotation
SFN	Single Frequency Network
SS	Single Stimulus
SSCQE	Single Stimulus Continuous Quality Evaluation
SSDP	Simple Service Discovery Protocol
VOs	Virtual Objects

Chapter 1

General Introduction

1.1 Introduction

Multimedia content is increasingly used in every area of our life [1]. Usually, multimedia contents only stimulate the visual and/or the hearing system of the end user. For instance, in [2] slow motion effect is given to video sequences to impact on the viewing perception of the users. Researchers try to simulate the other human senses by enriching multimedia with additional effects such as light, airflow, vibration, scent, temperature, etc. The main point of adding effects is to give the user the sensation of being part of the multimedia content, so as to increase the user's Quality of Experience (QoE). QoE can be defined as the degree of delight or annoyance of the user of an application or service. It results from the fulfilment of his/her expectations with respect to the utility and/or enjoyment of the application or service in the light of the user's personality and current state [3]. The parameters that affect the QoE can be classified into three groups:

1. The quality of the (video/audio) content at the source, which relates to the kind of codec used.
2. Quality of Service (QoS): "[The] Totality of characteristics of a telecommunications service that bear on its ability to satisfy stated and implied needs of the user of the service."[4]. The QoS parameters that affect the performance of streaming services most are bandwidth, delay, jitter, and packet loss.

3. Human influence factor which present any variant or invariant property or characteristic of a human user. The characteristic can describe the demographic and socioeconomic background. Human perception is usually captured by Mean Opinion Score (MOS), which reflects the evaluation of some test panel [5].

QoE metrics can be classified in to subjective and objective methods:

- Subjective methods are conducted to obtain information on the quality of multimedia services using opinion score, a typical example of qualitative metrics (subjective) is the MOS. Another example of a qualitative metric is the R-factor, which can be used in a manner similar to the MOS. Subjective evaluation of speech quality uses the R-factor in voice transmission models [6]. Subjective QoE measurement is time consuming, tedious, expensive and is not particularly applicable in a production environment.
- Objective methods are used to estimate the network performance using models that approximate the results of subjective quality evaluation, according to the level at which the input information is extracted [7].

Adding sensory effects to traditional multimedia present new topics, such as the evaluation of the QoE for video sequences with sensory effects, also called Multiple Sensorial Media (mulsemedia). Mulsemedia is a combination of traditional media with multiple sensory effects that aim to stimulate other human senses [8]. The resulting QoE is referred to as mulsemedia QoE.

Many questions rise in this emerging domain of study:

- Do sensory effects improve the viewing experience?
- What is the impact of sensory effects on the quality of audiovisual sequences?
- What is the impact of the sensory effects on perception and understanding of the users of multimedia content?
- How to bring sensory effects to the multimedia field?
- How to evaluate the user QoE with effects?
- How to render the multi-sensory media content on the consumer devices?

One of the goals of next generation TV broadcast services is to provide realistic media contents to the users. The user's sense of reality can be reinforced by adding to conventional media multiple sensorial effects, through five-sense stimulus (i.e., taste, sight, touch, smell, and hearing). In a smart TV broadcasting context, especially in a home environment, to deliver the additional effects, customary devices (e.g., air conditioning, lights, etc.), provided of opportune smart features, have to be preferred to ad-hoc devices, often deployed in other applications as for example in gaming systems. In this context, a key issue is the interconnection among the smart TV and the customary devices that deliver the additional sensorial effects to the user.

Furthermore, in the last years the concept of smart home has gained attention from the Information and Communication Technology (ICT) community. There has been a massive interest in the ability of embedded devices, sensors and actuators to communicate and create a ubiquitous cyber-physical world. Smartness has been extended to customary devices traditionally populating the users' houses, such as domestic appliance, for instance. Finally, today's new technologies enable also the interaction with the residential environment: control of utilities (e.g., lighting, heating, ventilation, air conditioning, automated window treatments, pool and spa controls), control of security (e.g., garage and access controls), control of home appliances locally or remotely from a smartphone [9].

A crucial role on the rapid evolution of this scenario has been played by the Internet of Things (IoT) paradigm [10] and by the recent development of short-range mobile communication technologies that together with an improved energy-efficiency are expected to create a pervasive connection of "things" [10]. The drastic increase in the number of smart devices and sensors connected to the IoT has the potential to change how consumers interact with networked technology, including media and entertainment platforms [11]. This represents an interesting opportunity for the entertainment industry to include the growing volume of customer interaction that comes with IoT in order to create more responsive and interactive applications, redefining the level of interaction between entertainment providers and their customers [12], [13]. There is now the condition for the TV service broadcasters to redefine their content and products, and have chances to reach also the traditional user who will be more and more immersed in a smart home scenario, surrounded by customary devices able to cooperate and provide enhanced TV experience.

1.2 Aims of the Thesis

To answer the questions rise in this emerging domain of study the primary objectives of the thesis can be summarized as follows to:

- Propose an IoT based architecture for multi-sensorial media delivery to TV users in a home entertainment scenario, the synchronization requirement between media and devices is analyzed and the architecture of the system is defined accordingly. Furthermore, a prototype is implemented in a real smart home scenario with real customary devices, which allowed a subjective test measurement campaign to assess the QoE of the users and the feasibility of the proposed multi-sensorial media TV service.
- Analyze users' perception of multi-sensorial media, in particular the influence of three sensory effects (i.e., airflow, vibration, and light) on user enjoyment, annoyance, emotions, and if the user like to experience additional sensory effects when viewing multi-sensory content. Furthermore investigate if multi-sensorial media can enhance the user sense of reality, as well the audiovisual content, and which effect has the highest impact on user experience for an IoT based multi-sensorial media in smart home.
- Propose a novel parametric model suitable for the estimation of the QoE for multi-sensorial media TV in smart home. The proposed model can be used by TV service providers to predict the enhancement produced by adding the effects to conventional TV services and plan for delivery multi-sensorial media as new advanced TV service. The parameter estimation of the model relies on Particle Swarm Optimization (PSO) which has been successfully applied to the optimization of nonlinear problems. A comparative analysis of the performance and the models' prediction accuracies of the proposed model with the state of the art model for the QoE have been carried out based on the same MOS dataset to assess the effectiveness of the former with respect to the latter.

1.3 Thesis Outline

- Chapter 2 provides an updated overview of the researches achieved in the domain of QoE for multi-sensorial media services. Results are summarized based on subjective quality assessments for audiovisual sequences enriched with effects, such as ambient light, airflow and vibration. The aim of providing this survey is to understand the role that multi-sensorial media could play in QoE enhancement in future services.
- Chapter 3 introduces IoT architecture for enabling the multi-sensorial media home TV services. For this purpose, to deliver effects, customary home devices jointly participate for the creation of the extended media experiences. A set-up based on Arduino device enhanced with switching capabilities has been implemented to test the proposed architecture. A measurement campaign has been then executed on a population of 40 users of various gender, age and instruction, to assess the QoE based on MOS. The goal of the measurement was to evaluate if the multi-sensorial media implemented on home customary devices can positively impress the common TV users.
- Chapter 4 presents a study of sensory effects (i.e., airflow, lights and vibration) impact on user experience in term of user enjoyment, annoyance, emotions, and if the user like to experience additional sensory effects (i.e., hot, cold, dry, and wet) or olfactory sensations. User perception and preference, analyzed based on the results obtained from the subjective assessment for the IoT based multi-sensorial media in smart home. Moreover, the results obtained reveal if multi-sensorial media can enhance the user sense of reality, as well the audiovisual content, and which sensory effect is user preferable effect.

- Chapter 5 presents a proposed parametric model for predicting the QoE for multi-sensorial media TV in smart home scenario. The model has been validated according to different MOS datasets from two different subjective assessments. The first dataset used to validate the state of the art model for the QoE of multi-sensorial media, the performance of the proposed model has been compared to the performance of the state of the art model for the first dataset. The second dataset used to validate the performance of the proposed model obtained from our subjective assessment for the multi-sensorial media TV. Furthermore this chapter illustrates the model estimated response and the evaluated response from the subjective assessment for the IoT based multi-sensorial media in smart home.
- Chapter 6 highlights the conclusions of the thesis achieved with respect to the specified research objectives, and directions for future researches are given.

Chapter 2
Background and Related Work

This chapter highlights the role of multi-sensorial media in Quality of Experience (QoE) enhancement by gathering and classifying the results obtained in this emerging field of research, and identifying the key challenges in this domain for future research. This chapter structured as follows: in section 2.1 the concept of multi-sensorial media is explained. Section 2.2 presents the implementation of multi-sensorial media. In section 2.3 the subjective assessment parameters and assessment methods are presented. Section 2.4 describes the impact of sensory effects on the QoE, through subjective quality assessments of multi-sensorial media content. Section 2.5 illustrates the framework to enable broadcasting with multi-sensorial media.

2.1 Multi-Sensorial Media

The concept of receiving sensory effects with audiovisual content is shown in Figure 2-1. The processing terminal is responsible for managing the actual media audiovisual resource associated with Sensory Effect Metadata (SEM) in a synchronized way based on user's setup in terms of both media and sensory effect rendering [5], [14], [15]. SEM is a description of supplementary effects based on Sensory Effect Description Language (SEDL), which is an XML-based language used to describe sensory effects. Media and effect renders are used to reproduce audiovisual media and supplementary effects that enable the stimulation of senses other than audition and vision [14]-[18]. For example, a mobile phone vibration, fan/ventilator, heater/cooler, can be used to address haptic sensations, whereas vaporizer devices can stimulate the olfactory system [16]. The stimulation of the visual system can be further enhanced using ambient lighting devices. The main point of adding effects is to give the user the sensation of being part of the multimedia content, so as to enhance user's viewing experience by increasing the sense of reality. The sensory effects role on viewing experience and user enjoyment was demonstrated in [18]-[20], where authors show that the viewing experience can be improved by adding effects to the multimedia content.

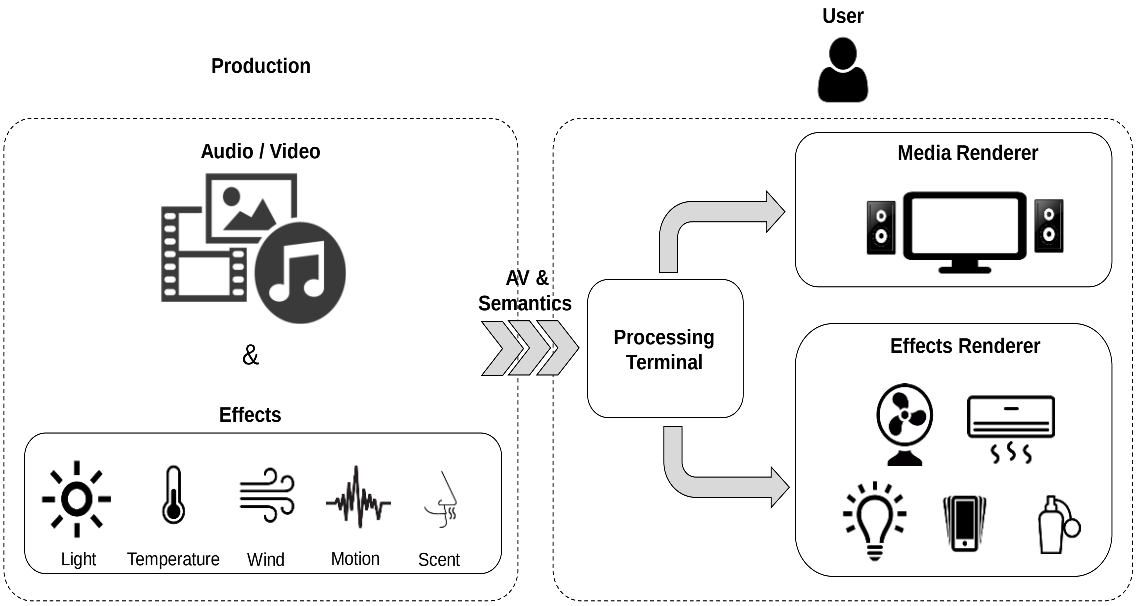


Figure 2-1: Multi-sensorial media concept

2.2 Multi-Sensorial Media Implementation

2.2.1 Authoring Tools for Multi-Sensorial Media

In Moving Picture Experts Group (MPEG), an exploratory activity of creating a standard enabling integrated user experiences of complex sensory effects, including more than video and audio like wind, vibration, light, and temperature effects was initiated in 2007 and called RoSE (Representation of Sensory Effects). Activities on RoSE were merged in 2008 with MPEG-V standard referred to as media context and control, which allows the annotation of multimedia content with additional effects and became a core application domain of MPEG-V, which is divided into seven parts, as follows [14]:

- 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: Reference Software and Conformance;

Sensory effects and MPEG-V can be seen as part of Ambient Intelligence (AmI) which aims at creating intelligent environments that are sensitive and responsive to the presence of a person. For example, is the so-called Ambient Displays (AD) that can be used to display information to the user through different devices like lamps or active wallpapers. In the context of MPEG-V, this kind of interaction is classified as information exchange between the real and the virtual world, which includes sensory effects [15]. A video content is authorized with a sensory effect description as defined in part 3 of MPEG-V standard.

Ordinary users may have a difficulty in making sensory effect description because of the lack of XML knowledge. For this purpose, there are several authoring, simulation, and rendering tools developed for sensory experiences as described below, one important thing in creating SEM is that, they should be easy to use and simple.

- A Sensory Effect Video Annotation (SEVino) tool developed in [21], which is written in Java, and provides means for adding sensory effects to multimedia content.
- A GUI based authoring tool called “RoSE Studio” was developed by [14], which provides convenience to users in generating sensory effect description.
- A realistic media authoring tool proposed in [22] based on MPEG-V standard. The proposed authoring tool consists of two types: the SEM generation tool and the SEM authoring tool.
- A Sensory Effect Simulator (SESim) tool presented in [21] for simulating SEM descriptions including various sensory effects, i.e., light, wind, vibration, fog, scent, temperature, and water-sprayer. The simulator gives the ability for testing the SEM descriptions without the need of having the rendering devices.
- A platform to play multi-sensorial media presented in [23], which comprised of: (a) PlaySEM SE Video Player, responsible for video playback and SEM input; (b) PlaySEM SER (Sensory Effect Renderer), responsible for converting SEM data into commands and controlling the devices that perform sensory effects rendering.
- A Web browser plug-in tool called AmbientLib presented in [24], which offers the possibility to enhance the user experience by sensory effects in the context of the World Wide Web.
- As there is a lack of suitable software for testing sensory effects (i.e., multimedia player), [25] present a multimedia player SEMP which supports sensory effects. The tools, presented in [21], [25], are open source and can be downloaded from the sensory experience lab website [26].

2.2.2 Sensory Effect Devices

The following subsections introduce some devices already available and researches have been performed with these devices. One of the challenges for future research when implementing multi-sensorial media applications, is that adopt further multi-sensorial media rendering devices that can enhance sense of reality.

2.2.2.1 Light Devices

The Ambilight Television Set [27] is one of the light-based devices, which consists of lights integrated in the frame of the television set. Experiments using the Ambilight TV had shown that, additional light effects reduce eye strain due to a smoother lighting difference between display and background. Moreover, the experiments indicate that additional light effects are more pleasant for the eyes with nature and sports films than with action films.

For computer games, there are two devices available. The first device is the amBX System [28] which is a follow-up of the Ambilight Television Set and adds to the light effects additional wind and vibration effects by means of respective instruments (i.e., fans and wrist rumbler) as shown in Figure 2-2 . It comprises a wall washer light with controller unit, left and right 2.1 sound speaker lights and a subwoofer, a set of fans, and a wrist rumbler. The wall washer and the lights of the speakers contain high power RGB LEDs with over 16 million additive RGB colours. The LEDs provide instant response and one can vary the intensity continuously. The integrated 2.1 sound system provides 160 W music power through the two speakers (2x 40 W) and the subwoofer (80 W). All devices operate in the frequency range of 35 Hz - 20 kHz. The two fans have a variable speed control with up to 5,000 rotations per minute (rpm). Finally, the wrist rumbler consists of two integrated motors that allow variable rotation speed with deferent patterns [29].

Figure 2-2: amBX System[31]

The second device is the Cyborg Gaming Lights [30], which provide light effects for games as shown in Figure 2-3. These lights are a follow-up of the amBX System and provide a color spectrum of 16 million colors. They have more intense and colorfast lights than the amBX System. These lights can be combined with the amBX System to provide a more immersive gaming experience [31].

Figure 2-3: Cyborg Gaming Lights [31]

2.2.2.2 Scent Devices

Scent-based devices are devices that emit scent during the consumption of the content (e.g., a movie or a game). Research on the use of such devices started in cinemas in the 1950s. The article *The Lingering Reek of ’Smell-O-Visio*n’ [32] provides an insight why such innovations as scent did not get successful in early years. In recent years, the topic of scent-supported movies or devices was taken up again. In [16] the author described why smell is difficult and in which area with which devices smell can be used. In addition, the effect of olfactory data on multimedia evaluated in [33]-[35]. Furthermore, many companies are providing scent-based devices such as the vortex active from Dale Air [36] as shown in Figure 2-4, for enriching movies, websites, etc., the Game Skunk from Sensory Acumen Inc. for games, or ScentScape from Scent Sciences for gaming, movies, etc., details about these prototypes and further devices and companies are showed in [37].

Figure 2-4: Vortex Active [31]

2.2.2.3 Vibration Devices

Vibration devices can be wearable, handheld, desktop devices or haptic seats. Wearable devices are designed to be worn by the user while he experiences the audiovisual content. Typically they are composed of several vibrotactile actuators embedded into clothe like vibrotactile glove [38], vibrotactile jacket or vest [39], [40].

The second type of device corresponds to the handheld devices. In this case the user experiences haptic feedback through a portable device held in his hand like a mobile phone [41]. The third type is the desktop device, like the Philips amBX system, also it provides wind, light effects, and add vibrations to a keyboard [29]. The fourth type of device relates to haptic seats like vibrotactile blanket, chair, couch and moving chair [42], [43], while seated on a modified chair, the user passively feels the haptic effects,

2.2.3 Multi-Sensorial Media Synchronization

Synchronization between sensorial effects and video content, when implementing multi-sensorial media applications is one of the challenges in this emerging domain and should be carefully designed same as audio/video synchronization.

Effect render devices must be activated at the same time of the scene to be played to give the users the sense of reality. A synchronization algorithm that calculates the devices activation time proposed in [44] and had been tested in an experiment room, in the home server event occurs according to the time line of the main track media. According to the this algorithm , when the synchronization manager sends control command to the device controller at device activation time, the command is analyzed and decomposed into control command type, interface, control value, and start time.

In the test room multi-track media used to display three-dimensional media, and there are various kind of effect devices that are capable of making real-sense representation. The test methods for the multi-track media are as follows. First each track displayed without carrying out the synchronization algorithm, people feel awkward when synchronization algorithms did not work. Then time gap between home server and each client compared whenever home server broadcasts synchronization time in every 10 seconds, the error can be reduced to 10% compared to not using the synchronization algorithm.

There are various types of delays should be considered to assess how much a synchronism can be accepted when enriching multimedia with sensory effects, delays involved with the decoding and rendering of the sensory effects. When these types of delays are not properly considered, there can be unexpected delays disabling the synchronous play of audiovisual content and sensory effect, the impact of synchronism between the sensory effects and multimedia content had been investigated in [45]-[48]. According to [49] haptic effect could be presented with a delay by up to 1 second behind the video content in order to be acceptable by most of the users; in contrast airflow effect could be released either 5 seconds ahead of or 3 seconds behind the video content to achieve the acceptable level. The results in [34] indicate that the time window for releasing a certain scent ranges from about 30 seconds before to up to 20 seconds after the content is displayed to which the scent would match.

2.2.4 Multi-Sensorial Media Dataset

The most existing datasets do not provide suitable video content for performing evaluations of multi-sensorial media. For example, the default test sequences used for video quality evaluation cannot be enriched with additional effects, as they are too short and/or difficult to annotate due to the lack of appropriate effects such as vibration or airflow or are not suitable for sensory effects (e.g., in-door scenes).

A dataset had been provided by [50] that have the advantage of providing the SEM descriptions for a number of sequences from different category. Thus, the time consuming procedure of generating the SEM descriptions can be omitted. The sensory effect dataset can be used both for conducting traditional video quality assessments based on different qualities and for evaluating the impact of sensory effects on different conditions.

The dataset collected in total 76 video sequences, i.e., 38 action, 12 documentary, 8 sports, 5 news, and 13 commercial sequences, and enriched with wind and vibration effects. The light effects are calculated automatically from the video content which is always performed by the processing engine (e.g., SEMP). The reason for this decision was to keep the SEM descriptions small. Additionally, the manual description of light effects would be too complex. Moreover, some sequences were selected with respect to the impact on the viewer's emotional states (i.e., fiction vs. reality). Some sequences are replicated in the dataset because some sequences were used in various assessments within different contexts (e.g., Web-based assessments) and different evaluation scenarios (e.g., influence of sensory effects on the perceived video quality). Hence, diverse representations had to be generated. For example, in scenarios that comprise playback on TV sets or local playback on a computer, a subjective quality assessment needs to provide high resolutions (i.e., 720p spatial resolution upwards) and high bit-rates (i.e., 4 Mbit/s upwards). On the other hand, in Web-based assessments, lower resolutions (i.e., 720p spatial resolution downwards) and bit-rates (i.e., 2 Mbit/s downwards) are enough. Furthermore, the dataset can be used for evaluating the impact of sensory effects on different conditions (e.g., emotions, perceived video quality, QoE). For performing such evaluations, additional sensory devices (e.g., fans, lamps, or vibration chairs) are mandatory for rendering sensory effects (i.e., light, wind, and vibration). One example for a collection of such devices is the amBX System which was used throughout collecting this dataset. As the dataset is based on the MPEG-V standard, it can be used with any device or software supporting MPEG-V for enhancing the viewing experience.

2.3 Multi-Sensorial Media Subjective Evaluation

Subjective assessment is time and cost consuming. However, a higher number of experimental subjects would result in more accurate results. The trade-off between the performance and experimental expense should be considered.

The subjective evaluation procedure must be based on the International Telecommunication Union (ITU) -T recommendation BT.500-13 [51], all the sessions of the experiments must be conducted in an isolated room under the same ambient conditions. Before the session, the following conditions should prevail [52]:

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

2.3.1 Assessment Parameters

To conduct a subjective quality assessment, a variety of different parameters have to be taken into account. For example, the number of participants and what level of knowledge (i.e., expert or non-expert) they have are two important parameters [31]. The most important parameters for a subjective quality assessment are described below:

- The most important parameter is the number of participants. In [53], the number of participants is mentioned as 4 to 40 participants. [51] States other numbers of participants but should not be fewer than 4 participants because in that case the results are not statistically relevant [53]. Furthermore, going beyond the number of 40 participants does not necessarily lead to better results because the variation in the results is minimal. A good number for an assessment is 16 to 24. The higher the number of participants, the more significant the results are [54].
- The second parameter is the type of participants. The participants can either be experts or non-experts. Which type of participants to select depends on what one wants to evaluate. Experts are a useful resource for algorithm development (e.g., new video codec). They know where to look at and how to evaluate the technical aspects of the algorithms [54]. However, experts are poor in evaluating the system from a general perspective. To evaluate a system for the market, non-experts should be used. Non-experts represent the general public in a subjective quality assessment. They are able to recognize artifacts or problems that an expert might not be able to detect because non-experts have no pre-determined way of looking at the content. It is important to mention that each participant, both expert and non-expert, has to be screened before the subjective quality assessment. According to [54], the two most important factors to screen for are: color blindness and visual acuity. For both factors, standardized methods (i.e., for detecting color blindness the

Ishihara test [54], [55] and for detecting visual acuity the Snellen Eye Chart [54], [56]) should be used.

- The third parameter is the viewing conditions under which the assessment should be conducted. In [53], a number of different viewing condition parameters are defined; for example, viewing distance, peak luminance of the screen, ratio of luminance of inactive screen to peak luminance, background room illumination. Most of these parameters cannot be easily evaluated, e.g., the evaluation of the luminance needs special equipment. Thus, at least the following two conditions, mentioned in [53], have to hold. First, the participants should sit at a distance of 1 to 8 times the height of the screen (i.e., normally around 90 cm). Second, if the test material (i.e., video, image) is displayed in a window on the screen, the visible background should be 50% grey.
- The fourth parameter for an assessment is the test material used. The test material can either be videos, images, audio, or a combination of them. The test material should have a length of 10 seconds to 30 minutes but not more [51], [53]. The length of the test material depends on the selected assessment method. The person conducting the assessment is advised to use test material from standardized sources. The standardized test material allows better comparison between different algorithms. Regarding the number of used test contents, there is no default number of test stimuli. [54] Suggests a number of 8 to 16 different stimuli to achieve good results.
- The last parameter for an assessment is the length of the assessment itself. The length depends on the selected assessment method and can range from a maximum duration of 30 minutes to 90 minutes [51], [53]. The 90 minute

sequences are only used for the Single Stimulus Continuous Quality Evaluation (SSCQE) and for the Simultaneous Double Stimulus for Continuous Evaluation (SDSCE) assessment methods [51].

Besides the different parameters for an assessment, the reporting of the results is important. The ITU specifies in [51] and [53] the following items that should be reported:

- Details of the test configuration
- Details of the test material
- Type of picture source (e.g., the camera used) and display monitors
- Number and type of assessors
- Reference system used (if any)
- The grand mean score for the experiment
- Original and adjusted mean scores (if one or more outliers were detected according to, e.g., [51])
- 95% Confidence Intervals (CI)

2.3.2 Subjective Assessment Methods

This section briefly describes a number of different approved subjective quality assessment methods. For more detailed description of each assessment method see [53], [54].

2.3.2.1 Absolute Category Rating

The Absolute Category Rating (ACR) [53] or also called Single Stimulus (SS) method presents to the participant one test sequence at a time. After each sequence, the participant has to rate the overall quality of this sequence using a discrete rating scale, with a range from 1 to 5 or 1 to 9 for bad, poor, fair, good, and excellent. In the ACR method, the length of the test sequences should be around 10 seconds. Depending on the selected test material, the length of the test sequences can be increased or decreased. Furthermore, the time for rating a test sequence is specified as less than or equal to 10 seconds. The maximum duration of this assessment method should be around 30 minutes.

2.3.2.2 Absolute Category Rating with Hidden Reference

The Absolute Category Rating with Hidden Reference (ACR-HR) [53] is similar to the ACR method. The major difference between this method and the ACR method is that in the ACR-HR method a reference version of each test sequence has to be included. As the name of the method indicates, the participants do not know which test sequence is the reference. Thus, the participants rate reference and processed test sequence independently. The assessment procedure and the voting scale are the same as in the ACR method. The ACR-HR has all advantages of the ACR method (e.g., speed). Furthermore, ACR-HR has the advantage, due to the hidden reference, that the reference video quality does not influence the final score. This method should be used in large experiments but only if the reference videos are at least of good quality. Furthermore, the ACR-HR method is not suitable for some types of impairments (e.g., dulled colors) [53].

2.3.2.3 Degradation Category Rating

The Degradation Category Rating (DCR) [53] also called Double Stimulus Impairment Scale (DSIS); this method presents a pair of test sequences to the participant. The first sequence is always the reference sequence and the second is the processed test sequence. Table 1 shows the five-level impairment scale used for rating in the DCR method. The length of the test sequences should be around 10 seconds but, if necessary, the length can be increased or decreased. A short intermission between the reference sequence and the processed sequence should be made. This intermission should last around 2 seconds. After the second sequence of a pair, the participant should rate the impairment of the second sequence with respect to the first (reference) sequence. The DCR should be used for testing the fidelity of transmitted content over a distribution channel (e.g., broadcasting, Internet) with respect to a reference. Furthermore, it should be used for evaluating systems that should provide high quality output (e.g., video conference and telephony) in the context of multimedia communication.

Table 1: Five-level impairment scale

Value	Description
5	Imperceptible
4	Perceptible but not annoying
3	Slightly annoying
2	Annoying
1	Very annoying

2.3.2.4 Double Stimulus Continuous Quality Scale

The continuous subjective quality assessment methods use a continuous rating scale from 0 to 100 or similar ranges. The usage of such a continuous quality scale allows a more fine granular rating of the sequences. Usually, the value range is divided into five equal intervals. The Double-Stimulus Continuous Quality-Scale (DSCQS) [51] provides the participant with two test sequences. One of the sequences is the original sequence directly retrieved from the source (e.g., camera, live-stream). The second sequence is an impaired version of the original sequence. The participants may not know which version is the reference. Participants are asked to rate the quality of each sequence using the scale displayed in Table 2. In [51], two different variants for the rating procedure are defined. Variant 1 allows the participants to switch between the two sequences as desired. A participant can switch between the two sequences until he/she has determined the quality of each sequence. Variant 2 displays both sequences separately to the participants. After the participants have seen the sequences, they are presented again. During the second presentation they provide their voting. The DSCQS is useful for evaluating the quality of stereoscopic image coding or for evaluating the quality of systems relative to a reference [51].

Table 2: Five-level continuous quality scale

Value	Description
80-100	Excellent
60-80	Good
40-60	Fair
20-40	Poor
0-20	Bad

2.3.2.5 Single Stimulus Continuous Quality Scale

In the Single Stimulus Continuous Quality Scale (SSCQS) method each sequence rated separately using the continuous ratings (i.e., scales ranging from 0 to 100), this method can be selected if none of the above methods are suitable for the planned evaluation [53].

2.4 Influence of Sensory Effects

In order to highlight the role of sensory effects on the QoE when consuming multi-sensorial media, the following subsections present the results of some of the literature subjective quality assessments for multi-sensorial media.

2.4.1 Enhancing the Viewing Experience and Sense of Reality

The sensory effect role on viewing experience demonstrated in [18], evaluation results show that the viewing experience can be improved by adding effects to the multimedia content. For the subjective assessment, 25 participants (12 male, 13 female) between 20 and 31 years old were invited. Six sequences from different category were showed to the participants, with three effects wind, vibration, and light. Each sequence presented twice, without effects and then with effects. The experimental result of this study shows that multi-sensorial media increase the QoE for action and sport category, but for other category can be bothering.

Another subjective test on the effect of multi-sensorial media on the sense of reality and user enjoyment was performed in [19]. Eighteen participants (7 female, 11 male) from various backgrounds and various areas of major interest between 20 and 36 years old took part in this test. The participants were invited to watch 12 video clips from two different movies with haptic and airflow effects. The majority of participants polled (almost 70%) considered both haptic and airflow effects in multi-sensorial media enhance the sense of reality and enjoyment levels.

2.4.2 Video Quality

The impact of sensory effects on the quality of the perceived video was investigated in [57] through subjective assessment, in the experiment, 24 participants (13 male, 11 female) between 18 and 37 years old were invited. Two video sequences with different bit-rates (i.e., 2154 Kbit/s, 3112Kbit/s, 4044 Kbit/s, and 6315Kbit/s) were presented to the participants, with light, wind, and vibration effects, and then without effects. The assessment results show that sensory effects have an important role on perceived video quality. On the other hand, the rate of multi-sensorial media sequences with low bit-rate is higher than the rate of the same sequences with higher bit-rate but without sensory effects.

Another subjective test was carried out in [58]; in this assessment 128 participants with an average age of 26 years old were invited to the test, each participant watched 16 unique multimedia sequences taken from two different video quality levels with a resolution of 720p and 480p, annotated with haptic, olfaction and airflow effects. One of the assessment conclusions is that without effects, the large majority of participants have noticed the differences in multimedia quality. However, when delivering the media content with the effects, there is no statistical difference between user enjoyment levels when exposed to average and high quality sequences. In addition, the participant’s enjoyment levels were maintained high when lower multimedia quality sequences have been used in conjunction with multiple sensorial effects. This is as the sensory effects have partly masked the video quality decrease.

2.4.3 Sensory Effects and Emotions

The impact of sensory effects on the emotions was investigated in [20]. The assessment was conducted in three different geographic locations, to determine the differences in participant's perception depending on their location. In this study 26 students (18 female and 8 male) invited, aged between 20 and 57 years old who participated at AAU Klagenfurt, Austria. For the user study at RMIT University, Australia, 21 students and staff were invited (12 female and 9 male) aged between 22 and 58 years old. For the user study at UoW University, Australia, 21 students and staff were invited (6 female and 15 male) aged between 22 and 63 years old. Fifteen video sequences from different category presented, with a resolution of 720p and annotated with light, wind, and vibration effects.

The assessment results show that, when multi-sensorial media content used, active emotions (e.g., worry, fun) increased in their intensity with sensory effects. Concerning the QoE, the results show that the difference in the multi-sensorial media QoE in different locations is minor.

2.5 Multi-Sensorial Media Broadcasting

The framework to enable broadcast providing sensory experience in addition to the audio and visual experience has been presented as 4-D broadcasting based on MPEG-V standard. In [59] a 4-D broadcasting framework proposed, the authors of the 4-D content construct the sensory effect metadata using MPEG-V Part 3, and synchronized it with the content. The authored SEM is multiplexed with the content and broadcasted on-air using MPEG-2 transport stream. The receiver processes the signal and extracts the sensory effect metadata through Part 3. The devices used to render the effects around the user environment are detected using Simple Service Discovery Protocol

(SSDP) and the device capabilities are written using MPEG-V Part 2. The adaptation engine, then, generates a sequence of adequate device commands using MPEG-V Part 5, so that the devices can generate the sensory effects properly, creating 4-D effects.

The architecture for the multi-sensorial media broadcasting system, and how to represent the real-sense media in multiple devices was also suggested in [60]. The system divided into two main parts as shown in Figure 2-5. First part is production and delivery network part that packetizes multiple audiovisual contents into MPEG-2 transport stream, and combines SEM together when user wants to input sensory effects in real time. Second part is home network part that receives real-sense media. The Home Server (HS) plays the audiovisual content related with the topic in digital TV by using H/W decoder located in the HS; the rest tracks are relayed to the laptop and the smartphone connected with the HS; SEM is decoded by the SEM decoder, translated into device control variables after analysis, and synchronized with audiovisual content.

A real sensory aggregator able to collect sensory effects and send it formatted as MPEG-V standard to the processing terminal via the 4-D broadcasting system was proposed in [61]. To provide the 4-D broadcasting service, there are several function modules to control collecting the sensory effects information. The aggregator has several sensors and communication interfaces for the 4-D broadcasting service. Although the proposed aggregator can collect real-sense effects, there are further requirements to minimize the aggregator and to support more sensors for gathering further real-sense effects.

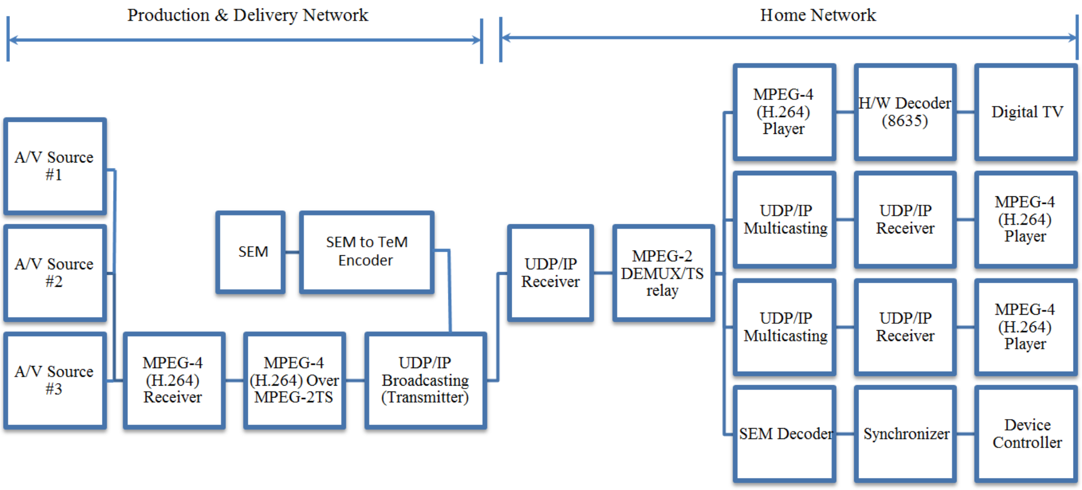


Figure 2-5: Multi-sensorial media broadcasting architecture [60]

Chapter 3

QoE Assessment for IoT Based Multi-Sensorial Media

This chapter presents an Internet of Things (IoT) architecture to enable multi-sensorial media home TV services based on cloud IoT platform. A Quality of Experience (QoE) assessment campaign has been performed based on subjective tests. Section 3.1 discusses the provision of multi-sensorial media TV, section 3.2 introduces the proposed IoT architecture and gives the details of the implementation. Section 3.3 describes the experimental setup to assess the quality of experience. Results are discussed in section 3.4.

3.1 Multi-Sensorial Media TV in Smart Home

During the last decade, the evolution of TV market has been terrific. Broadcasters have been facing new challenges to cope with an increasing demand of new services from user's side. With the convergence of second-screen adoption and the abundance of real-time news consumption via social channels, the broadcast landscape underwent a major transformation. Viewers have begun to demand highly customized experiences that meet their individual needs.

In short, the evolving needs of the viewer seem to be in the future of broadcast television. In the next years, it is likely that this will become even more evident, with more people demanding customized television experiences through user-generated content and the option of micro bundled packages. To keep up, broadcasters must stay current with the latest innovations to engage with their customers.

Despite the increasing market of handled devices such as smartphones and tablets, and consequent demand of spontaneous access to video content form mobile broadband users, the total minutes watching video per week of traditional home TV is still predominant [62]. The global service providers' offer of advanced whole-home video delivery enables consumers to new services. Over The Top (OTT) content providers are offering movies and TV shows for either download or direct streaming over the internet, the type of shows that consumers prefer to watch on a big-screen high definition TV. Within this framework, home entertainment systems have known for the past few years a constant evolution in size and complexity, delivering new levels of experience and adventure to consumers. To adapt to users' need the home entertainment sector developed a true dedicated electronic playground, with large-screen displays, consoles for gaming, audio gear, and docking stations, generally managed through a single remote control giving the complete command to the user. Technology companies have been announcing linkage of TV screens, PCs, video recorders, game consoles, and other electronic devices together in the same home network, allowing the user to share content among these devices. On the other hand, due to the complexity and cost, the home entertainment products reached only a niche of the population. This slow down the evolution of the TV broadcasting services, which are still based on content media not able to exploit all the features that home entertainment systems could provide, being intended for traditional TV services.

In parallel to the development of the home entertainment environment, there has been a growing interest in the ability of embedded devices, sensors and actuators to communicate and create a ubiquitous cyber-physical world. The growth of the IoT paradigm which is a world-wide network of interconnected objects, uniquely addressable, based on standard communication protocols [10], and the rapid development of short range mobile communication technologies together with an improved energy-efficiency is expected to create a pervasive connection of “things” [10]. The drastic increase in the number of smart devices and sensors connected to the IoT has the potential to change how consumers interact with networked technology, including media and entertainment platforms. This represents an interesting opportunity for the entertainment industry to include the growing volume of customer interaction that comes with IoT in order to create more responsive and interactive applications, redefining the level of interaction between entertainment providers and their customers. In short, IoT and the resources in the cloud that can connect devices to real-time computational engines are rapidly evolving and will create an immersive environment that can augment our array of experiences [10].

This study proposes an IoT based architecture to enable multi-sensorial media home TV services. For this purpose, to deliver effects, customary home devices jointly participate to the creation of the extended media experiences. The proposed architecture relies on the IoT platform Lysis [63], which is a cloud-based platform for the deployment of IoT applications. The major features that have been followed in its design are the following: each object is an autonomous social agent; the Platform as a Service (PaaS) model is fully exploited; re-usability at different layers is considered; the data is under control of the users. The first feature has been introduced by adopting the social IoT concept, according to which objects are capable of establishing social relationships in an autonomous way with respect to their owners with the benefits of improving the network scalability and information discovery efficiency. The major components of PaaS services are used for an easy management and development of applications by both users and programmers. The re-usability allows the programmers to generate templates of objects and services available to the whole Lysis community. The data generated by the devices is stored at the objects owners cloud spaces.

In Lysis the smart TV and the rendering devices, such as remote switches for air conditioning, lighting and vibration are represented as Virtual Objects (VOs) [64]. Micro Engines (MEs) combine and control VOs so that the requirements in terms of synchronization between media and devices are fulfilled.

Synchronization between sensory effects and video content, when implementing multi sensorial media applications is a challenging task, the impact of synchronism between the sensory effects and multimedia content had been investigated by [45]-[48]. According to [49] haptic effect could be presented with a delay up to 1second behind the video content in order to be acceptable by most of the users; in contrast airflow effect could be released either 5 seconds ahead of or 3 seconds behind the video content to achieve the acceptable level. The results in [34] indicate that the time window for releasing a certain scent ranges from about 30 seconds before to up to 20 seconds after the content is displayed.

3.2 IoT Architecture and Implementation

3.2.1 Architecture

The proposed architecture for multi-sensorial media relies on a cloud IoT platform named Lysis [63], which foresees four layers, as depicted in Figure 3-1.

- Physical layer: cloud implemented, this layer includes objects capable of accessing the internet, called Real World Objects (RWOs) due to their direct connection with the physical environment where they sense and act. For this particular scenario, the RWOs are either electronic devices with processing capabilities and integrated peripherals, such as smartphones, or computational platforms equipped with switching capabilities, for example Raspberry Pi or Arduino platforms, able to switch on and off rendering devices. The physical layer communicates with the upper layers using standard wired or wireless communication (e.g., Wi-Fi, Bluetooth, LTE, USB, Gigabit Ethernet, etc.) methods and data protocols (i.e., HTTP and MQTT).

- Virtualization layer: for de-coupling the hardware part from the cloud-based software representation, most IoT solutions introduce the VO concept as a digital counterpart of any entity in the real world [47], [48], so that each object in the physical layer is represented by a virtualization. The VO is a key part of the overall solution and depicts the RWO in terms of semantic description and functionalities. It is equipped with two interfaces, which allow for a standardized communication procedure: on one side, it enables the VO to communicate with the aggregation layer, while on the other side it represents the access point to the real world, providing the connection with the RWO. For our specific purpose, the virtualization layer is implemented by means of a software driver installed on the RWO, as detailed in the next section.
- Aggregation layer: this layer is responsible for the aggregation of data coming from one or more VOs in order to ensure a high re-usability level. The ME is a mash-up of one or more VOs and even other MEs, in charge for getting and processing data from VOs into high-level services requested by the higher layers (application layer).
- Application layer: at this level, user applications are responsible for the final processing and presentation. The deployment and execution of applications is based on the use of one or more MEs.

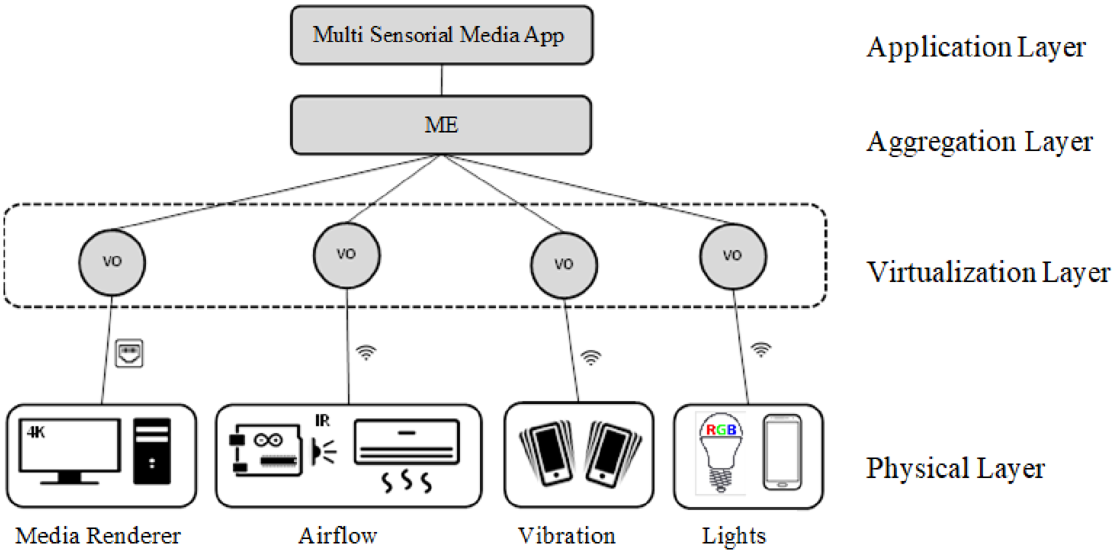


Figure 3-1: IoT architecture for multi-sensorial media

3.2.2 Implementation

For the media renderer the RWO is a desktop PC connected to the TV via a 4K HDMI cable and connected through Gigabit Ethernet. For the implementation of the multi-sensory effects on the proposed architecture, we rely on home customary devices. Specifically, the devices involved in this architectural implementation are: the fan of an air conditioning wall-mounted split to reproduce airflow effect, an RGB smart LED light system with integrated Wi-Fi connection as to the light enhancement effect and the integrated call vibration feature of a set of smartphones to provide haptic effect. The virtualization of the smartphones is done through application software opportunely developed and running on Android operative system. The air conditioning fan is controlled via an Infrared (IR) remote using an Arduino board with an IR sensors which represents the system's RWO. The RGB Smart LED lights are connected via Wi-Fi to a smartphone running an application able to automatically extract light effects form the phone camera while placed in front of the TV. In this case the RWO is the smartphone controlling the RGB Smart LED lights. Using a software driver for the Arduino board and an iOS app for the smartphone, we were able to virtualize the RWOs and make them accessible through the virtualization layer (i.e., VOs) to the upper layers of the proposed architecture. This allowed us to have full control of the system and respect the synchronization constraints specified in section 3.1. The communication with the RWO was implemented using the MQTT data protocol over the various communication standards which assured a low latency, with values lower than 1second for this specific implementation.

3.3 Experimental Setup

3.3.1 Test Environment

The measurement tests have been performed at the QoE Lab of the Department of Electrical and Electronic engineering of the University of Cagliari, Italy. The QoE lab is a 4×4×2.70 m (l×w×h) separate room furnished with a three seat sofa, and parquet floor, and equipped with an Haier inverter technology air conditioner wall split [65], a three RGB smart Philips LED lights system [66], a SAMSUNG TV UHD 4K Flat Smart JU6800 Series 6 with a 60-inch diagonal [67] and Wi-Fi internet connection, with the purpose of replicating the living room environment in a smart home scenario.

The setup of the test environment was performed according to ITU-T Recommendation P.911 [68]. Our tests involved two assessors per session, simultaneously rating the test video sequences with multi-sensory effects. The participants sat down in the sofa in front of the air conditioning wall split fan, which placed above the smart TV at a height of 2.5 m. The monitor has been calibrated before the startin of the test.

A sketch of the setup geometry is shown in Figure 3-2. The distance d from the monitor is 2.5 times the height H of the video monitor (2.5 H), 186 cm for position 1 and 2 (i.e., angle α ±24°). Vision angle is referred to the surface normal to center screen angle. An Arduino microcontroller board control the air conditioner fan via Infrared (IR) sensors as illustrated in Figure 3-3(a).Three RGB smart Philips LED lights are placed behind the monitor, to give the feeling that the lights integrated in the frame of the monitor, as shown in Figure 3-3(b). The RGB smart Philips LED lights are piloted by a smartphone placed behind the sofa with the camera faced in front of the TV screen as shown in Figure 3-3(c). Each assessor is provided with a smartphone with vibration call feature as shown in Figure 3-3(d). The participants can hold the smartphone in their hand, or placed it in their pockets, or leaning it beside them in the sofa. Figure 3-4 shows a real panoramic of the QoE lab. The hardware and software deployed to implement the system are presented in Table 3.

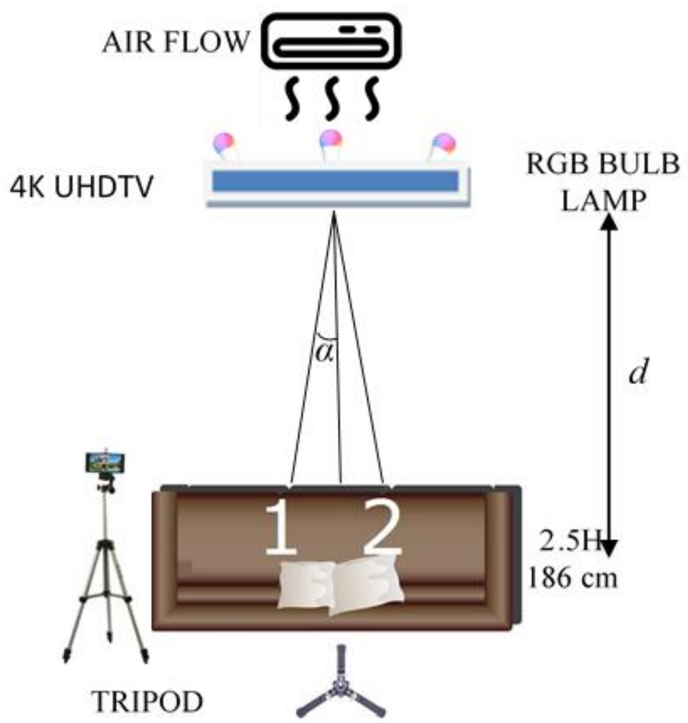


Figure 3-2: Setup of the proposed geometry

(a) (b) (c) (d)

Figure 3-3: Equipment used in the assessment: (a) Arduino microcontroller; (b) RGB smart LED lights; (c) Smartphone camera; (d) Smartphones

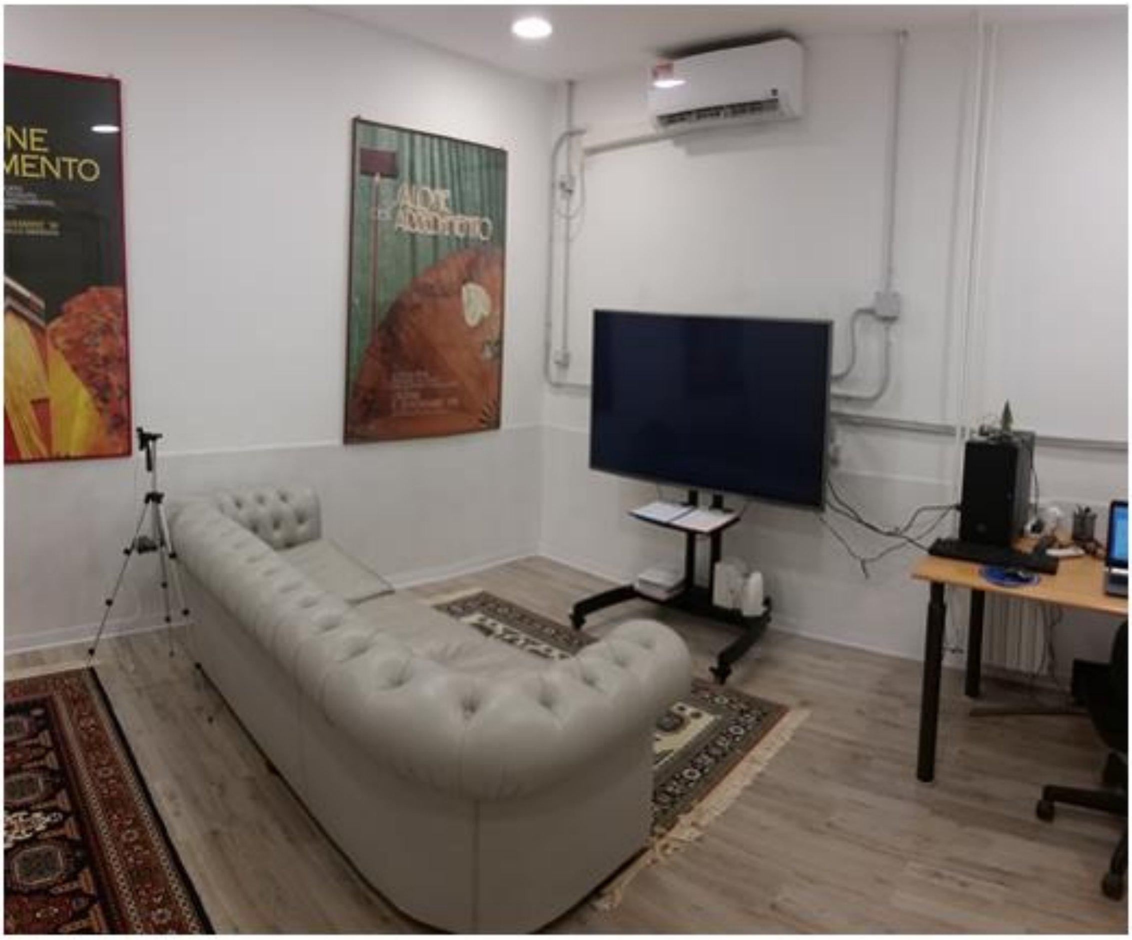


Figure 3-4: Panoramic of the test environment

Table 3: List of hardware/software deployed

Item	Description
Samsung Serie 6 JU6800	Professional high-performance 4K monitor with 60-inch diagonal
Haier Model AS09BS4HRA	air conditioner wall split INVERTER technology 9000 btu
Philips hue personal wireless lighting RGB Smart LED	Three RGB smart LED lights
Android/iOS Smartphones	1 iOS Smartphone to synchronize the RGB smart LED lights with the video sequences by adding related colours. 2 AndroidSmartphones to generate the haptic effect
Arduino MEGA 2560	Arduino microcontroller equipped with IR sensors and Wi-Fi shield
PlaySEM/SER [23]	PlaySEM SE Video Player PlaySEM SER
Desktop PC	CPU: Intel Core I7-7700k RAM: 2x16GB 3000MHz Video Board: GTX 1080 Ti 11GB GDDRSX Mother Board Asus ROG Strix 270I Gaming Antec H1200-Pro cooling Corsair CX850M 850 Watt supplier Hard drive: SSD 480 GB

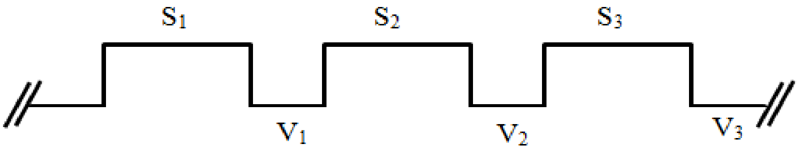
3.3.2 Participants

40 participants (31 males and 9 females) from various backgrounds, between 22-50 years old, with the average age 32 years had been invited to this assessment; only one participant took part in a similar assessment. For each participant, the following information was asked: age, gender, education, and occupation. Prior to a session, the

participants were screened for normal visual acuity on the Snellen chart [69]. A person taking the test covers one eye from 3 meters away, and reads aloud the letters of each row, beginning at the top. The smallest row that can be read accurately indicates the visual acuity in that specific eye. Moreover, the participants were tested through the Ishihara color test [55] to detect color blindness. The Ishihara color test consists of 38 so called pseudo isochromatic plates, each of them showing either a number or some lines. According to what you can see and what not, the test gives feedback of the degree of your red-green color vision deficiency. According to this test, the observer could be "none", "weak", "moderate" or "strong" red-green colorblind. All observers reported normal or corrected-to-normal vision, and had no color vision deficiency.

3.3.3 Assessment Procedure

The Absolute Category Rating (ACR) method or also called the Single Stimulus (SS) method as defined by ITU-R Rec. BT.500-13 [51] was used in this assessment, as shown in Figure 3-5. In the assessment 22-40 seconds video sequences are shown randomly interleaved with 5 seconds of grey screen used by assessors to rate the video sequences. In our assessment we did not imply the use of reference sequences to be shown to the observers. This is not a limitation in this particular assessment scenario where the aim is to evaluate the delight/annoyance caused by adding multi-sensory effects to conventional TV services. Observers are supposed to be familiar with conventional TV since they experience it daily. The use of ACR allowed reducing the overall time of the assessment for each participant pair to less than 30 minutes, thus avoiding lack of concentration due to user tiredness.



S_1, S_2, S_3,… Video Sequences

V_1, V_2, V_3,…Gray screen 5 seconds vote for video sequence

Figure 3-5: ACR assessment method

Furthermore, the ACR defines sequences of around 10 seconds but in our assessment the length was increased in order to allow for more sensory effects within one sequence. The rating scale used in the assessment is based on Mean Opinion Score (MOS) as defined in the ITU-T Rec. P.911. The ITU-T Rec. P.911 defines five-level rating scale as reported in Table 4. Each participant in the assessment asked to give his/her rating for each video sequences.

Table 4: Five-level rating scale

5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

Before the start of the assessment session there was an oral presentation, prepared to make the participants familiar with this type of assessment, and explaining the rating scale. Figure 3-6 shows the test environment during the assessment.

Figure 3-6: The test environment during the assessment

3.3.4 Multi-Sensorial Video Sequences

The participants watched 10 video sequences with different resolutions and bit-rates enriched with light (L), vibration (V) and airflow (A) sensory effects, hereinafter referred to as multi-sensorial sequences. L effects is extracted automatically from the video content as described in section 3.2.2, whereas V and A effects are manually annotated to the videos using the sensory effect video annotation tool [21]. Figure 3-7 shows a snapshot from each video sequence. The sequences have been selected from the category action, sport, documentary, and commercial. The news category was not included in this assessment, since according to [18] sensory effects have low influence on news category.

Table 5 describes the details of each multi-sensorial sequence with the resolution, bit-rates, category, and duration. The effects added to the sequence, and the video scenario description showed in Table 6.

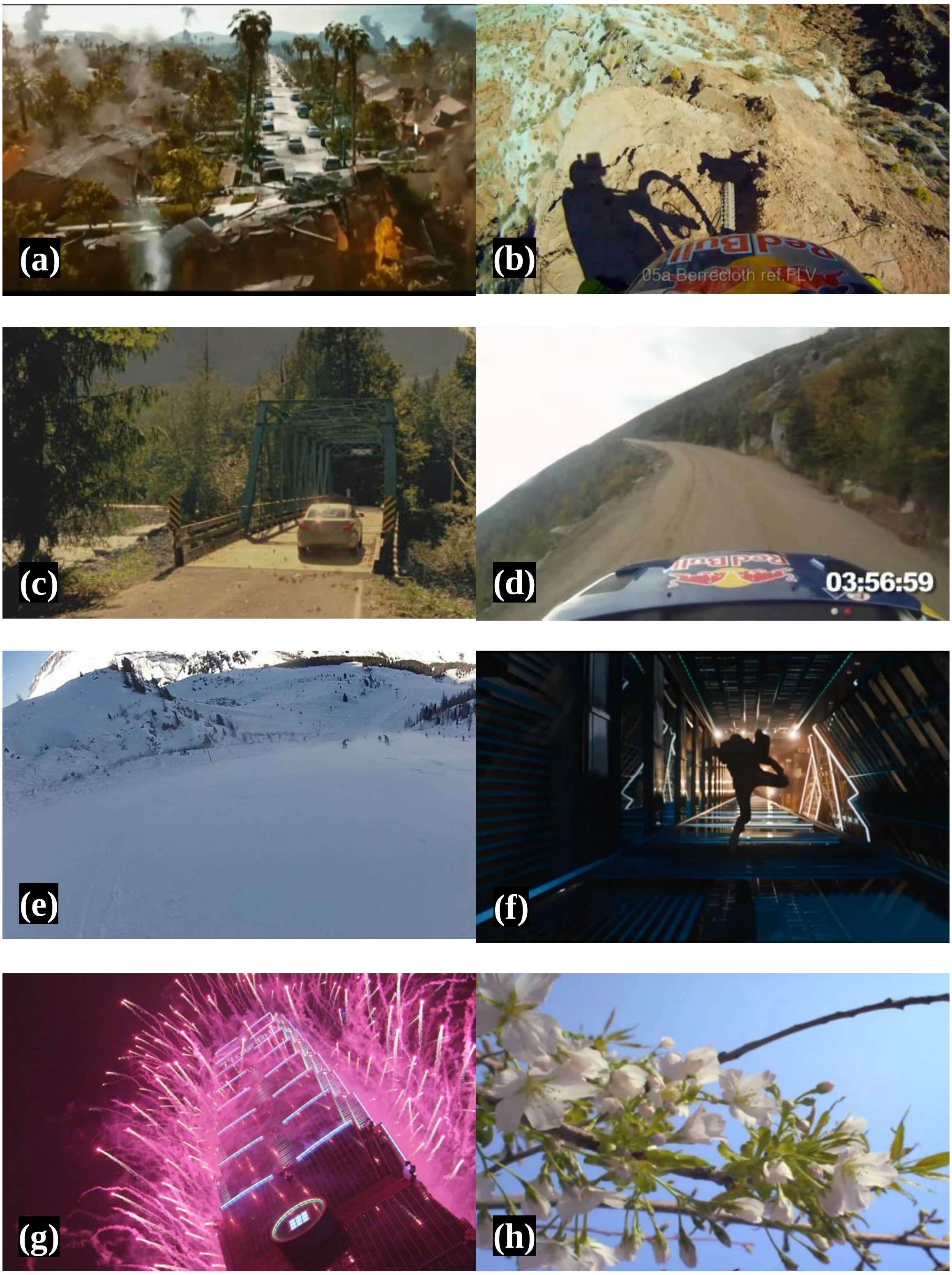

(a)
(b)
05a Berrecloth ref.FLV
(c)
(d)
03:56:59
(e)
(f)
(g)
(h)

Figure 3-7: Snapshot of the multi-sensorial video sequences: (a) 2012; (b) Berrecloth; (c) Bridgestone; (d) Pastranas; (e) Ice ski 2; (f) Skyfall; (g) Fireworks; (h) Earth; (i) Elysium; (j) Ice ski 1

Table 5: Video sequences details

Video Number	Video Sequence	Resolution	Bit-rate (Kbit/sec)	Duration (sec)
S1	Ice ski 1	1920x1080	15411	22
S2	Ice ski 2	1920x1080	14979	23
S3	Skyfall	3840x2160	12703	33
S4	Fireworks	3840x2160	10207	35
S5	Elysium	3840x1610	11364	40
S6	2012	1280x720	2186	30
S7	Pastranas	1280x720	2619	32
S8	Berrecloth	1280x720	3552	32
S9	Earth	1280x720	4116	21
S10	Bridgestone	1280x720	2421	30

Table 6: Video sequences scenario

Video Number	Video Sequence	Category	Effects	Scenario
S1	Ice ski 1	Sport	L, A	ice skiing
S2	Ice ski 2	Sport	L, A, V	subjective view, ice skiing, falling down
S3	Skyfall	Action	L, A, V	subway crash, car crash, falling down, wind, gun shots, explosion
S4	Fireworks	Documentary	L	different color fireworks
S5	Elysium	Action	L, A, V	car crash, shots, wind, explosion
S6	2012	Action	L, A, V	earth quick, tornado
S7	Pastranas	Sport	L, A, V	Rally
S8	Berrecloth	Sport	L, A, V	subjective view, bicycling down the rock cliffs
S9	Earth	Documentary	L, A, V	wind, animal jump
S10	Bridgestone	Commercial	L, A, V	windy weather, car moving

3.4 Experimental Results and MOS Ratings

From the 40 participants took part in this assessment, 8 outliers had been eliminated according to the procedure described in [51]. The reliability of the ratings given by a subject has been detected by checking the correlation between the average ratings and the i^{th} rating. An outlier is an observation that appears to deviate markedly from other observation in the sample and may indicate bad data. Evaluators were managed accordingly to the Pearson Correlation Coefficient (PCC). PCC is an index of the strength and direction of a linear relationship between two interval level variables. For PCC<0.75 the evaluator is considered as outlier.

The MOS and confidence intervals (CI) 95% for each multi-sensorial sequence are shown in Figure 3-8. The participant's ratings for each video are shown in Figure 3-9.

The ratings in percent of Good or Better (%GOB) using the values from Good and Excellent, the percent of Poor or Worse (%POW) using the values from Poor and Bad, and the rest percent (%Rest) for fair are described in Figure 3-10. It can be notes as for high resolution video sequences the perceived quality is higher. This was expected since the sight sense in human being is well known to be predominant. Nonetheless, greater results are achieved with sequences showing high dynamic motion in nature environments (e.g., Ice Ski 1, Bridgestone), but degrades when in similar scenario the subjective view is included (e.g., Ice Ski 2, Berrecloth). This can be justified by the fact that in subjective view sequence the expectation of the observer is higher, and the impact of the effects delivered by the customary devices deployed is perceived as week.

Concerning the sensory effects impact, the results show that it’s different from a video category to another category. The experiment results indicate that additional light effects are more pleasant for the eyes with nature and sports videos than with action videos. This is due to the fact that additional light effects reduce eye strain due to a smoother lighting difference between display and background which is more accentuated on high dynamic video sequences.

The majority of participants consider that the airflow and vibration effects in the multi sensorial media, improve the sense of reality. Also, most participants agree that both effects result in an enjoyable experience.

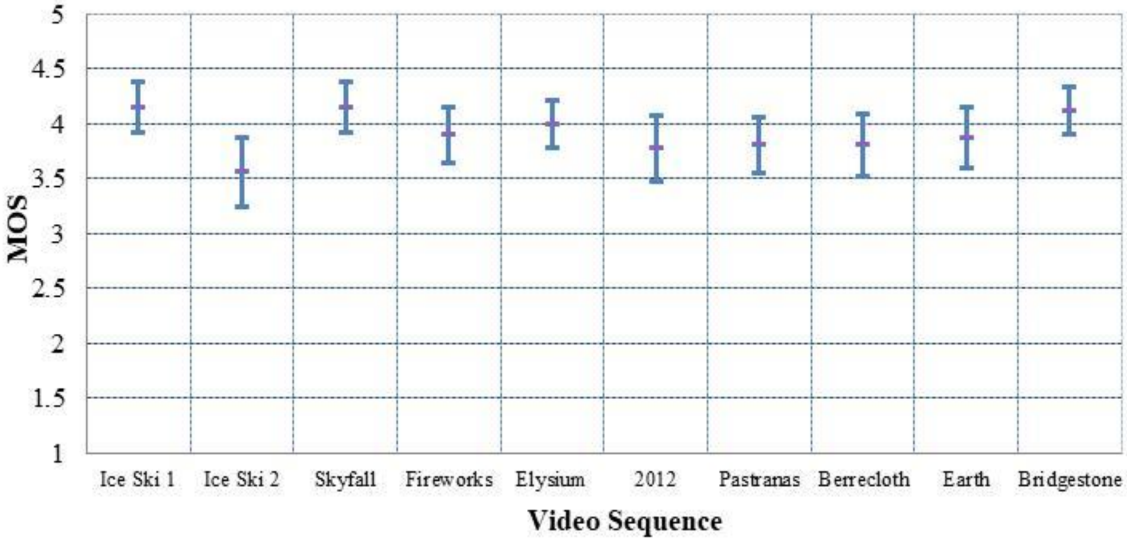


Figure 3-8: Mean opinion score and confidence intervals (95%)

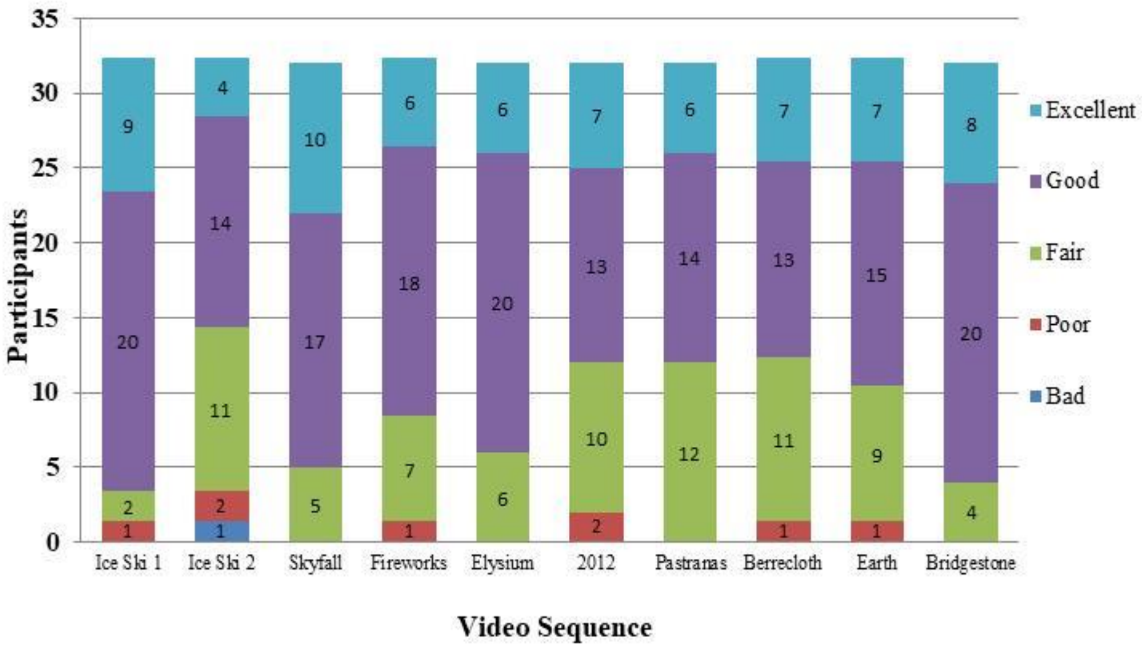


Figure 3-9: Ratings

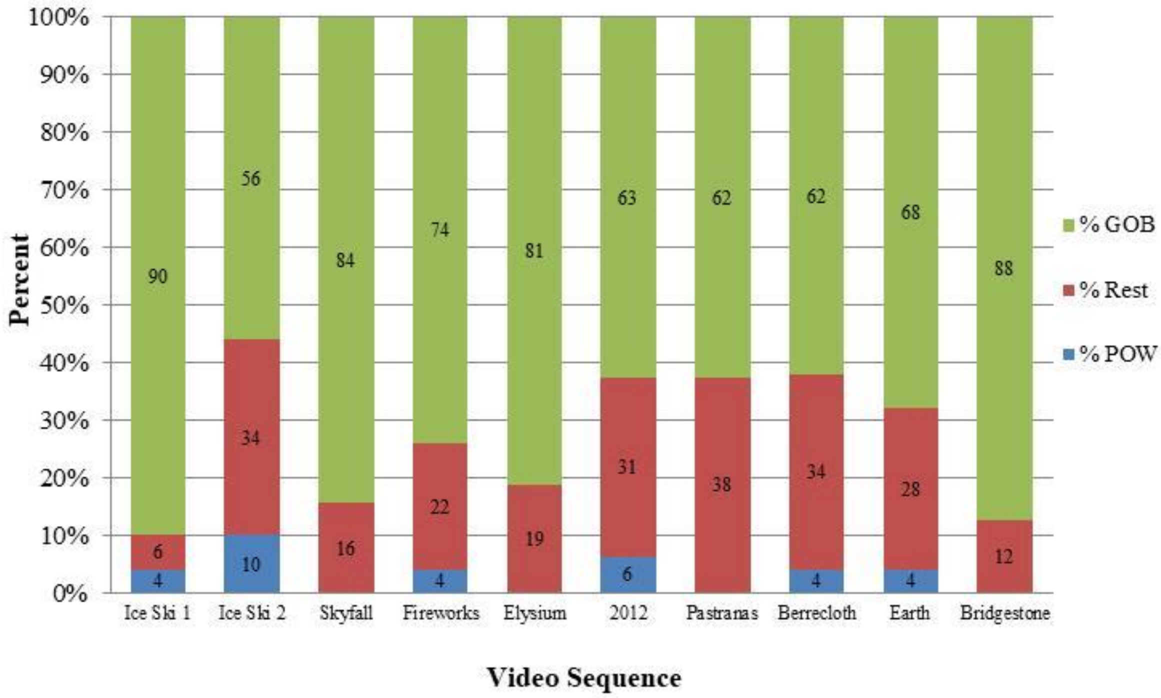


Figure 3-10: %GOB and %POW

Annotating video sequences with sensory effect depends on the presented multimedia content. For example, the video sequences Skyfall and 2012 belong to the same category but result in different MOS ratings. As a result, the usage of sensory effects for the selected video content is crucial as it influences the rating of the participants.

Finally in this chapter, the proposed IoT architecture and its implementation are a first step towards a common test environment which can be used for achieving comparable results throughout different subjective quality assessments; we provided the experimental setups for conducting subjective quality assessment. That is, we illustrated the test environment, suggested the type of the test sequences, the number of participants, and test method. Furthermore, we explained the experimental results and how outliers can be detected and eliminated.

Chapter 4
Impact of Sensory Effects on User Quality of Experience

. This chapter provides the overall evaluation results of an Internet of Things (IoT) based multi-sensorial media delivery to TV users in home entertainment scenario, through the subjective test measurement campaign described in chapter 3. The impact of sensory effects on user experience is investigated in section 4.1. User perception of multi-sensorial media is analyzed in section 4.2.

4.1 Sensory Effects Impact on User Experience Study

After displaying all the multi-sensorial sequences the participants had to answer the post-experiment questions, in which they were asked to comment on their experience regarding to the multi-sensorial media.

40 participants (31 males and 9 females) from various backgrounds, between 22-50 years old, (45% between 20-30 years, 35% between 30-40 years, and 20% between 40-50 years), with the average age 32 years took part in this study. Each session involved two participants that gave his/her rating for each sequence, which allowed reducing the number of sessions required to obtain reliable statistics.

The questionnaire included questions regarding to participants' preference levels of multi-sensorial media, including haptic, lights, olfaction, airflow, temperature, and humidity effects. Participants were asked to rate their response to investigate the impact of multi-sensorial media on user experience in terms of:

- Improvement in sense of reality.
- Sensory effects are distracting.
- The impact of sensory effect intensity on user experience.
- The impact of sensory effects on user enjoyment.
- The sensorial effects enhance the audiovisual content.
- If users emotions (i.e., surprise, fun and worry) increase in intensity by adding sensory effects rather than the absence of sensory effects.
- The sequence duration is enough to appreciate audiovisual content with sensory effects.
- If users like to experience thermal effects (i.e., hot, cold, dry, and wet) or olfactory sensations when viewing multi-sensory content.
- Which sensorial effect is user preferable effect?
- The impact of each single effect (i.e., airflow, vibration, and light) on user experience.

The participants also asked to give their comments and feedback about what they liked or did not like, or other things that could be changed to enhance the assessment.

4.2 User Perception and Preference of Multi-Sensorial Media

This section presents the overall evaluation results of the impact of multi-sensorial media on user experience, through the impact of each sensory effect and user opinions with respect to the post-experiment questions. From the 40 participants took part in this assessment, 8 outliers had been eliminated, data screening and analysis technique, are according to the procedure described in [51]. The following questions were asked in the post-experiment questionnaire:

Q1: Have you ever participated in an experiment similar to this one?

Q2: The sensory effects improve the sense of reality when watching the audiovisual content?

Q3: The sensory effects are distracting?

Q4: Did you direct your attention to any specific sensory effect when determining the quality of experience?

Q5: The intensity of the airflow effect is?

Q6: The intensity of the vibration effect is?

Q7: The intensity of the light effect is?

Q8: Did you enjoy the multi-sensory experience?

Q9: The sensorial effects enhance the audiovisual content?

Q10: Did your emotions (i.e., surprise, fun and worry) increase in intensity by adding sensory effects rather than the absence of sensory effects?

Q11: Would you prefer a longer sequence to better appreciate audiovisual content with sensory effects?

Q12: Would you like to experience olfactory sensations when viewing multi-sensory contents?

Q13: Would you like to experience thermal effects (i.e., hot, cold, dry, and wet) when viewing multi-sensory content?

Q14: Which sensorial effect do you prefer (or you like the best)?

Q15: The light effect is annoying?

Q16: The light effect increases the sense of reality?

Q17: The airflow effect is annoying?

Q18: The airflow effect increases the sense of reality?

Q19: The vibration effect is annoying?

Q20: The vibration effect increases the sense of reality?

The answers for each question include five options: Strongly Disagree, Disagree, Neutral, Agree, and Strongly Agree. The assessment overall time was about 30 minutes for each participant.

4.2.1 Sensory Effects Enhance the Sense of Reality

Users' opinions if sensory effects (i.e., airflow, vibration, and light) enhance the sense of reality are shown in Figure 4-1, most of the participants 85% felt that sensory effects enhance the sense of reality, when watching multi-sensorial sequence, 10% had neutral opinion, and 5% did not fell that sensory effects enhance the sense of reality.

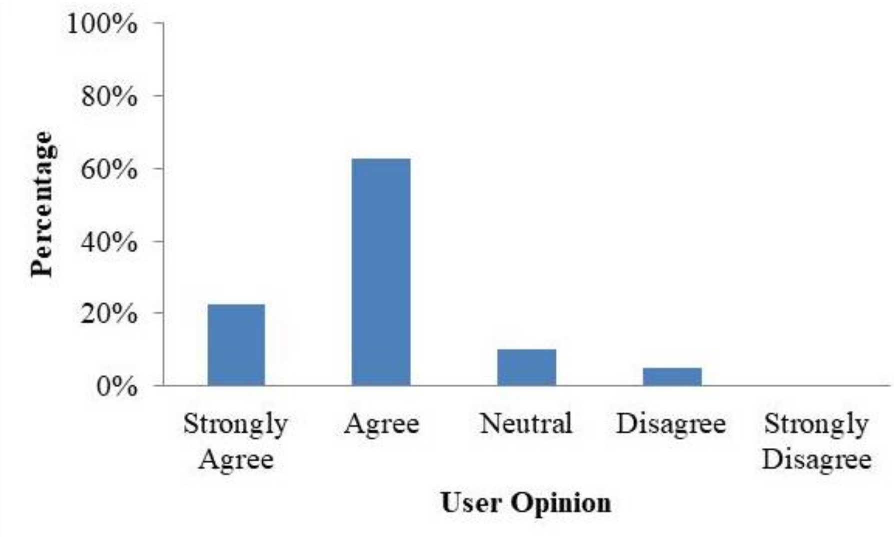


Figure 4-1: Sensory effects enhance the sense of reality

4.2.2 Sensory Effects are Distracting

The subjective assessment results indicate that 67.5% of the participants did not consider the effects distracting during the watching time, 17.5% had neutral opinion, and 15% distracted by the effects as shown in Figure 4-2. We asked the participants if they direct their attention to any specific sensory effect when evaluating their experience, 82.5% answered by Yes, and 17.5% answered by NO.

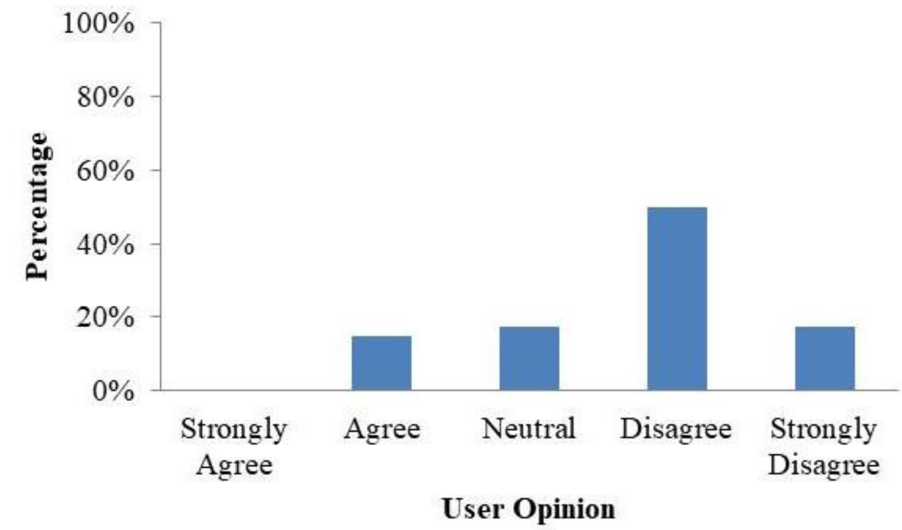


Figure 4-2: Sensory effects are distracting

4.2.3 The Intensity of Sensory Effect

Users' opinions regarding to the intensity of each sensory effect are shown in Figure 4-3, Figure 4-4 and Figure 4-5. User answer regarding to the intensity of the sensory effects is: (Too Weak, Weak, Fine, Strong, Too Strong).

The participants' opinion about the light intensity during the assessment was 72.5% of the participants considered the intensity as fine, 22.5% felt it was strong, and 5% found it weak-too weak intensity. The impact of airflow intensity on user experience was 45% of the participants considered the intensity as fine, 17.5% found it strong-too strong, and 37.5% found the intensity weak-too weak. Results obtained regarding to vibration intensity indicate that 40% of the participants found it strong-too strong, 32.5% considered it as fine, and 27.5% felt it as weak-too weak intensity.

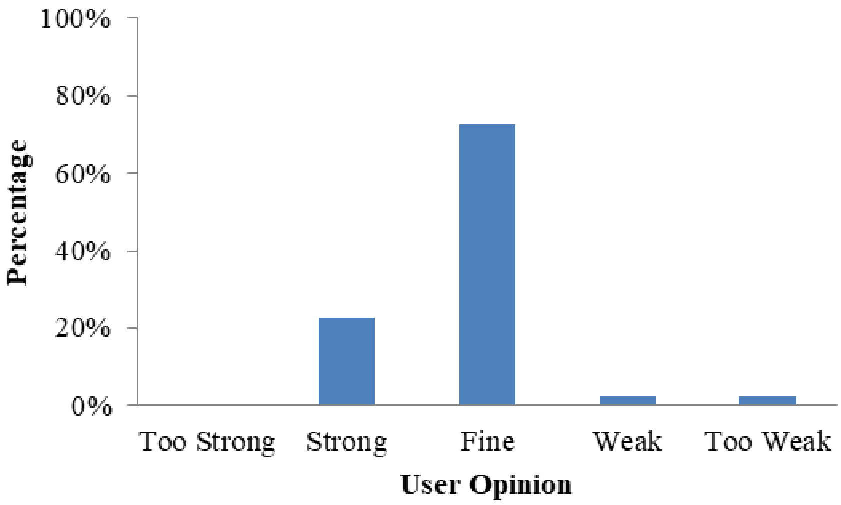


Figure 4-3: Intensity of light effect

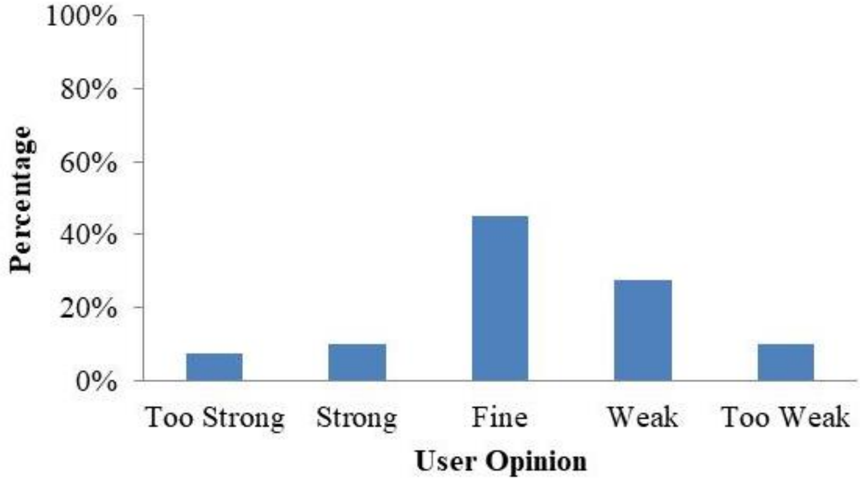


Figure 4-4: Intensity of airflow effect

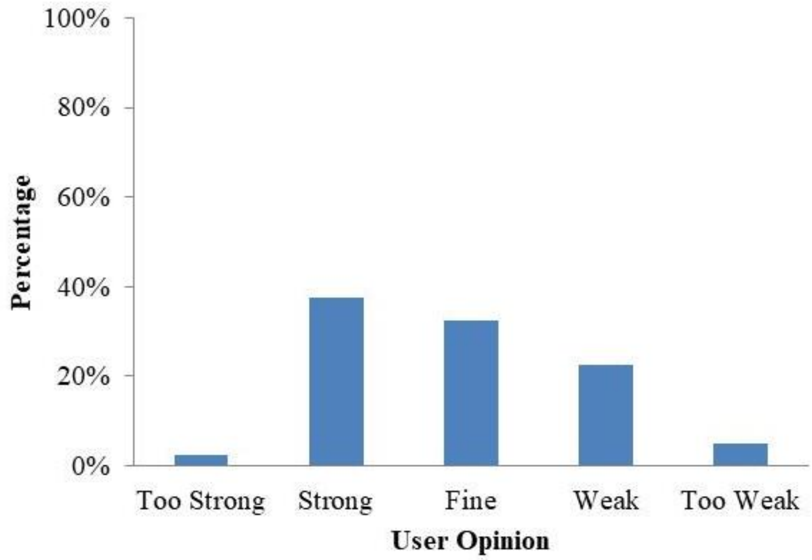


Figure 4-5: Intensity of vibration effect

4.2.4 Impact of Sensory Effects on User Enjoyment

User enjoyment is important to indicate the Quality of experience (QoE). Results obtained from the question "Did you enjoy the multi-sensory experience" analyzed and indicate that 80% of the participant enjoyed the multi-sensorial media, 12.5% had neutral opinion, and 7.5% did not enjoy the multi-sensorial media experience as shown in Figure 4-6.

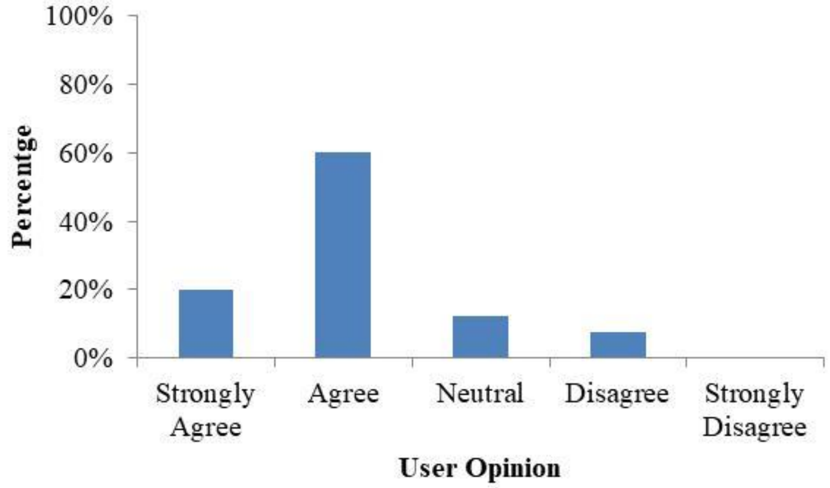


Figure 4-6: Sensory experience is enjoyable

4.2.5 The Sensory Effects Enhance the Audiovisual Content

Videos with different resolution and bit-rate showed to the participants during the subjective assessment. We asked the participants if the sensory effects can enhance the audiovisual content, the majority of the participants 85% agreed, 10% had neutral opinion, and 5% did not experience any enhancement as shown in Figure 4-7.

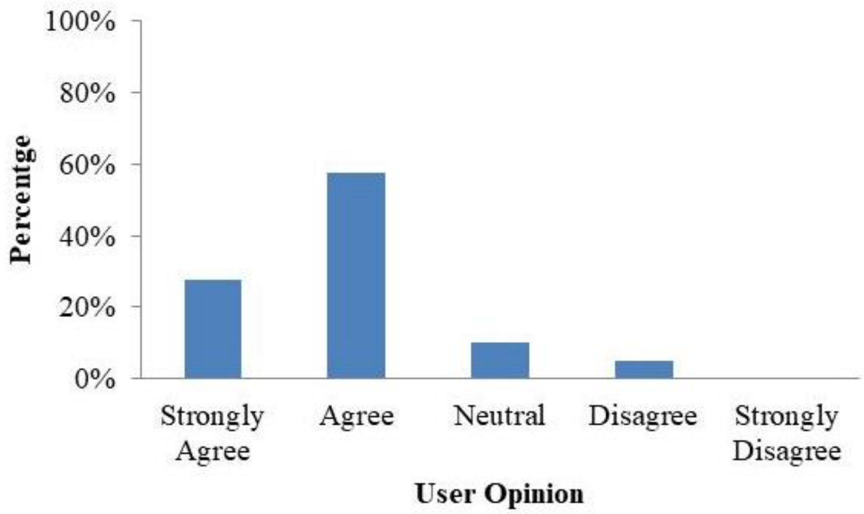


Figure 4-7: Sensory effects enhance the audiovisual content

4.2.6 Sensory Effects and Emotions

In our assessment, the participants watched 10 multi-sensorial video sequences with different scenario and from different category, so they can experience different emotions (i.e., surprise, anger, fear, fun, and worry). The results obtained revealed that 70% of the participants felt their emotions increased with the effects during the watching time, while 12.5% had neutral opinion, and 17.5% did not feel any change in their emotions as illustrated in Figure 4-8.

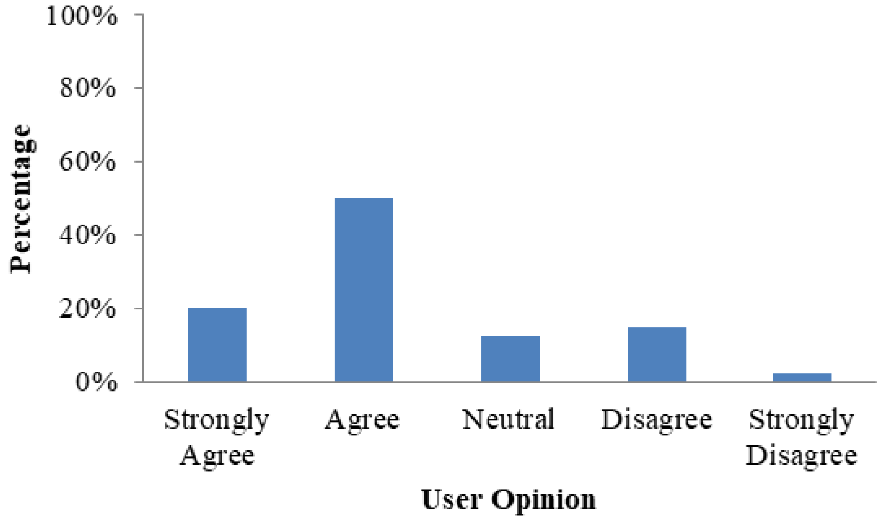


Figure 4-8: Emotions increased with sensory effects

4.2.7 Impact of Multi-Sensory Sequence Duration

The Absolute Category Rating (ACR) method defines sequence duration around 10 seconds, but in our assessment the length increased in order to allow for more sensory effects within one sequence, we asked the participants if they prefer a longer sequence to better appreciate the multi-sensorial sequence, most of the participant 77.5% tend to prefer longer sequence as illustrated in Figure 4-9, and for 22.5% of the participants the sequence duration was enough.

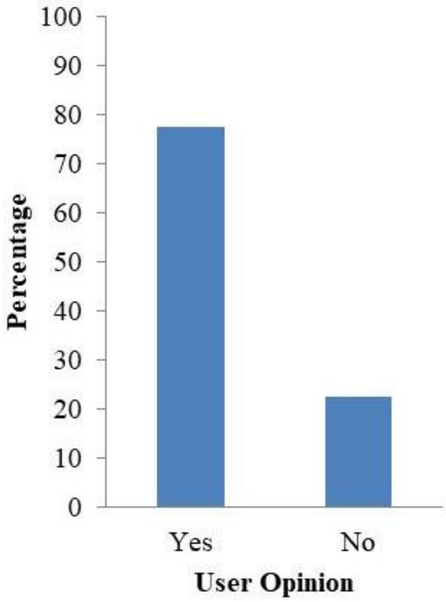


Figure 4-9: Do you prefer a longer sequence?

4.2.8 Experience Additional Sensory Effects

Result obtained regarding to user opinion to experience thermal effects (i.e., hot, cold, dry, and wet) or olfactory sensations analyzed and illustrated in Figure 4-10 and Figure 4-11. Regarding to thermal effects, result show that 77.5% of the participants interested in experience thermal effect, 10% had neutral opinion, and 12.5% show no interest in experience thermal effect. User opinions about scent effect showed that 75% of the participants interested in experience scent effect, 17.5% had neutral opinion, and 7.5% did not like to experience the scent effect as additional sensory effect.

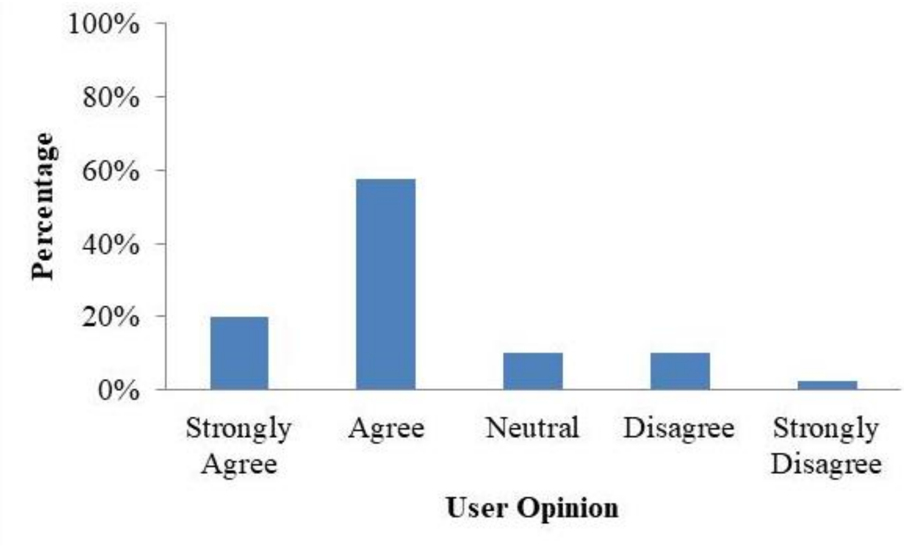


Figure 4-10: User interest in experience thermal effect

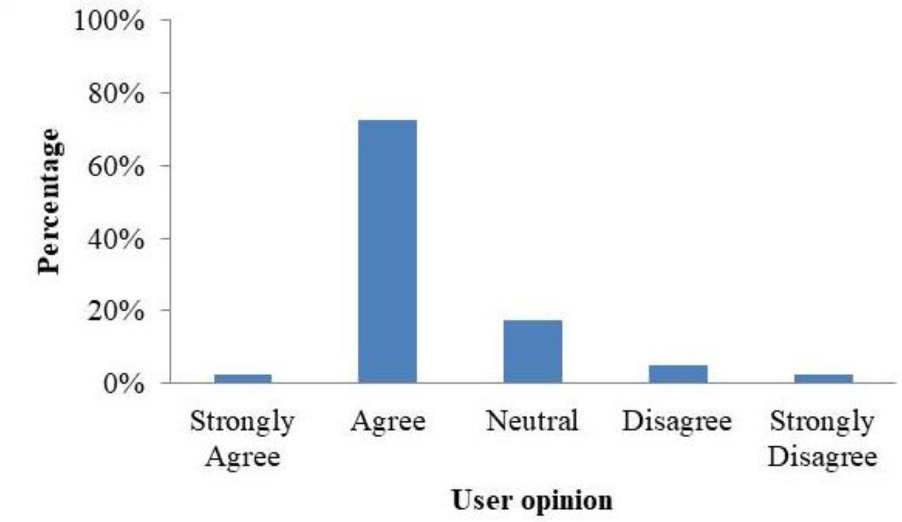


Figure 4-11: User interest in experience scent effect

4.2.9 User Preferable Sensory Effect

To identify which effect has the highest impact on user experience, we asked the participants to select their preferable effect. The result revealed that 67.5% of the participants considered the light effect as their preferable effect, 22.5% select vibration effect, and finally 10% select airflow as the effect with highest impact on their experience as illustrated in Figure 4-12.

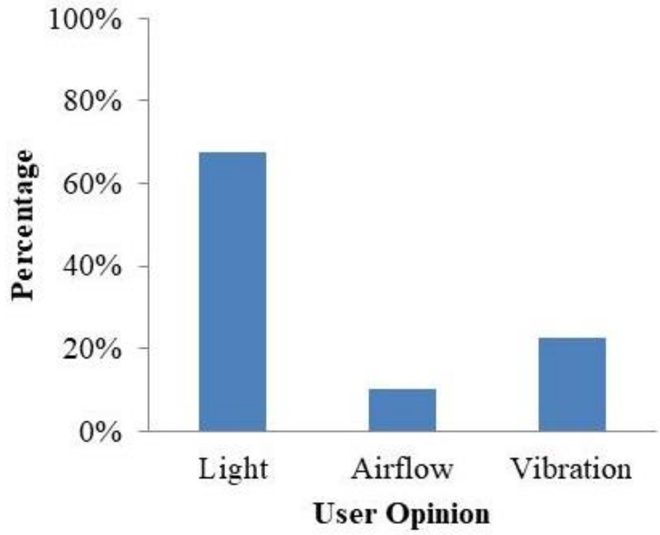


Figure 4-12: Which sensorial effect do you prefer?

4.2.10 The Impact of Light Effect on User Experience

In order to reveal the impact of light effect on user experience the participants asked to give their opinion regarding to the annoyance of light effect, and if the light effect enhance the sense of reality as shown in Figure 4-13 and Figure 4-14. Users' opinions to the questionnaire regarding to the impact of light effect on the experience analyzed and gave the following results: 95% of the participants did not consider light effect annoying, 2.5% had neutral opinion, and 2.5% annoyed by the light effect. Regarding to the sense of reality, results obtained revealed that 85% of the participants agree - strongly agree that light effect enhance the sense of reality, 10% had neutral opinion, and 5% did not consider light effect enhance the sense of reality.

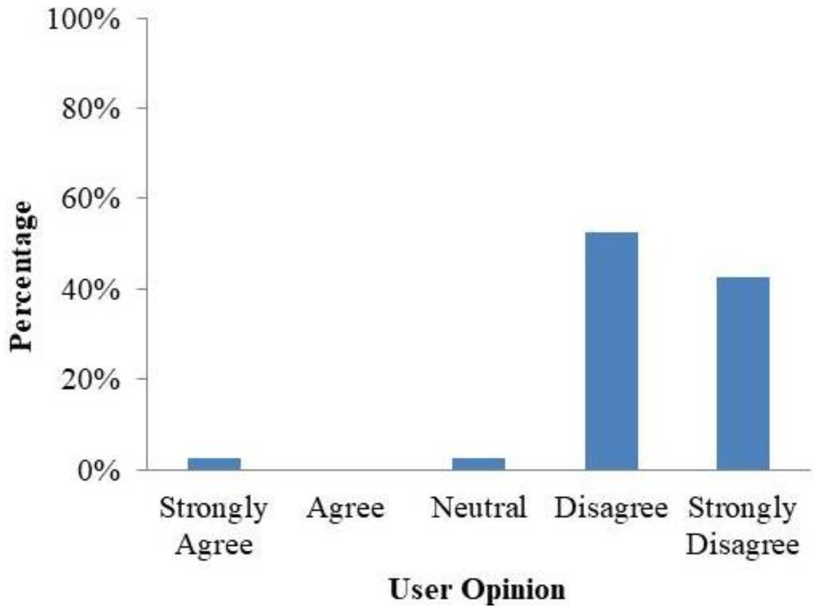


Figure 4-13: Light effect is annoying

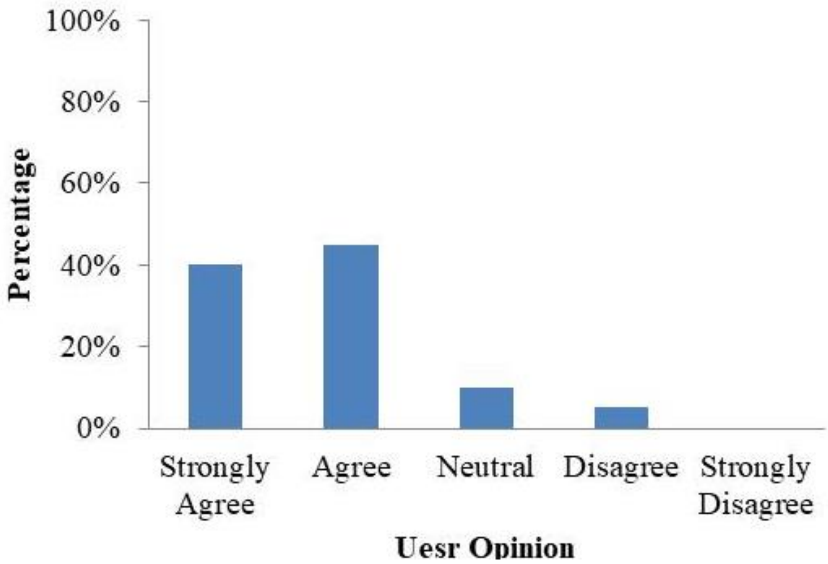


Figure 4-14: Light effect enhances sense of reality

4.2.11 The Impact of Airflow on User Experience

The impact of airflow effect on the perceived sense of reality, and annoyance collected and analyzed. Most of the participant 60% considered that the airflow enhances the sense of reality, 30% had neutral opinion, and 10% did not feel the enhancement in the sense of reality as illustrated in Figure 4-15. Participants' response to the annoyance of the airflow effect indicates that 65% of the participant did not experience any annoyance by the airflow, 20% had neutral opinion, and 15% annoyed by the airflow effect as shown in Figure 4-16.

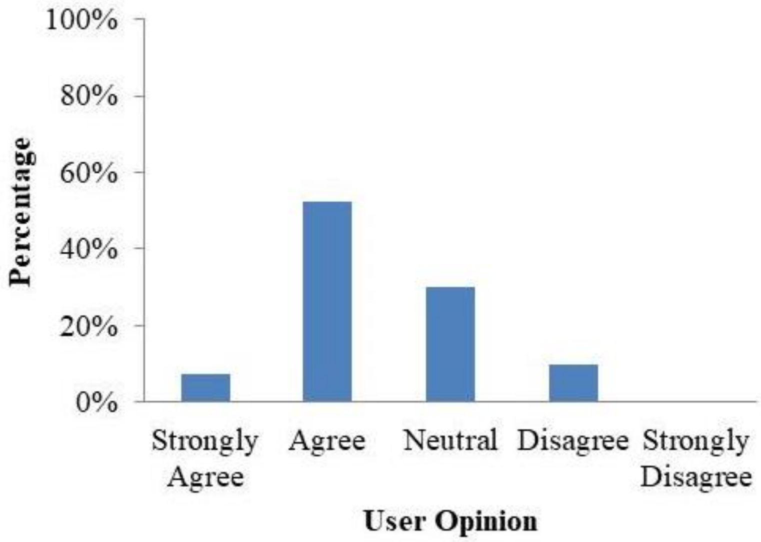


Figure 4-15: Airflow effect enhances sense of reality

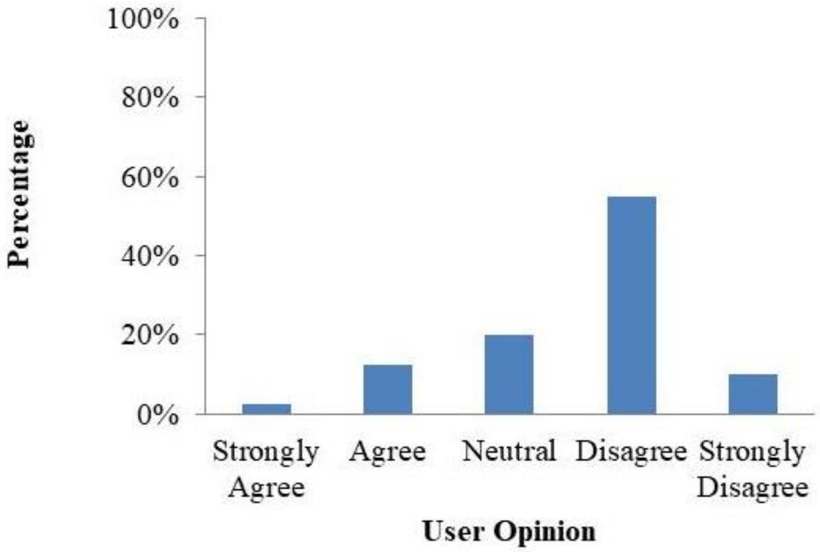


Figure 4-16: Airflow effect is annoying

4.2.12 The Impact of Vibration on User Experience

Users' opinions regarding the impact of vibration effect analyzed and illustrated in Figure 4-17 and Figure 4-18. The majority of the participants 40% agreed - strongly agreed that vibration effect enhance the sense of reality, 22.5% had neutral opinion, and 37.5% disagreed. Regarding to the annoyance caused by the vibration effect, results revealed that 45% of the participants did not consider vibration effect annoying, 22.5% had neutral opinion, and 32.5% found it annoying.

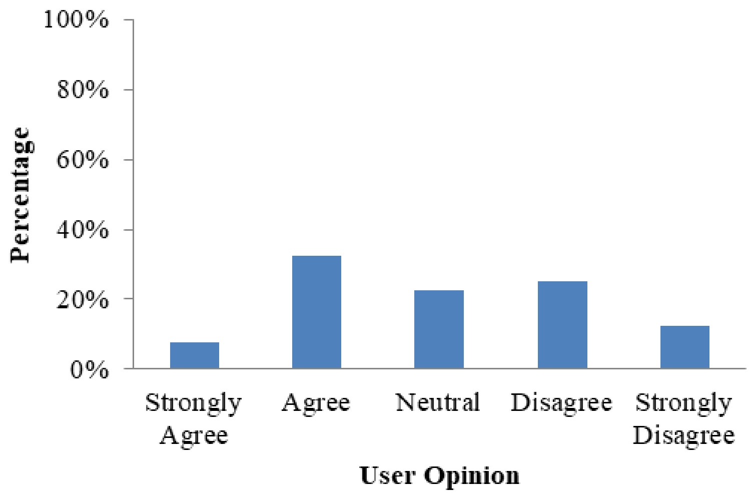


Figure 4-17: Vibration effect enhances sense of reality

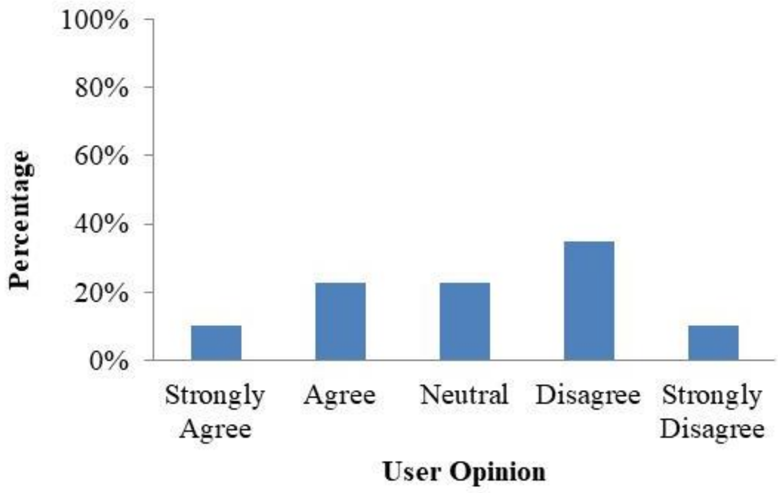


Figure 4-18: Vibration effect is annoying

4.3 Summary

In this chapter, we introduced the subjective assessment evaluation result that comprises the impact of sensory effects on user experience through a number of different video sequences from different category. The video sequences were annotated with light, airflow and vibration effects. Furthermore, the presented sequences in the subjective quality assessment provided with SEM descriptions are based on the most current version of the MPEG-V standard and, thus, can be used with any MPEG-V compliant software or device.

Additionally, we presented in this chapter the various post experiment questions results for the subjective quality assessment for video sequences enriched with sensory effects to investigate user perception of multi-sensorial media content. In particular, the impact of the intensity of light, vibration and airflow effects on user perceived experience, the impact of these effects on user enjoyment, annoyance, emotions, and if the user like to experience additional sensory effects when viewing multi-sensory content are studied. Furthermore investigate if multi-sensorial media can enhance the user sense of reality, as well the audiovisual content, and which effect has the highest impact on user experience for IoT based multi-sensorial media in smart home.

Majority of the participants 80% enjoyed the multi-sensorial experience with the three effects. Furthermore most of the assessors 85% consider that sensory effects enhance the sense of reality and 85% agreed that sensory effects enhance the audiovisual content. In contrast, the majority of the participants (i.e., 95% for light effect, 65% for airflow, and 45% for vibration) did not experience any annoyance as a result of the sensory effects, and 67.5% did not distract by the sensory effects. In addition, results reveal that users assess the intensity of the sensory effect during the assessment as fine (i.e., 72.5% for light effect, 45% for airflow, and 32.5% for vibration). Also 70% of the participants felt their emotions increased with the multi-sensorial media. Most of the participants 77.5% prefer a longer multi-sensorial sequence to better appreciate the impact of sensory effects. In our assessment 67.5% of the participants select the light effect as their preferred sensory effect. The assessment results show that most of the participants were interested in experience additional sensory effects (i.e., 77.5% for thermal effect and 75% for scent effect).

Furthermore, the participants represent a large age group and, thus, there is no need to perform the assessment with other age group to provide a broader range for the results. Additionally, the used devices are not very specific, but other systems rendering sensory effects are able to provide different results.

Chapter 5

QoE Model for Multi-Sensorial Media TV in Smart Home

Based on the results obtained from the subjective assessment presented in chapter 3, this chapter introduces a parametric Quality of Experience (QoE) model for multi-sensorial media TV in smart home scenario. The model instantiated and validated according to two different MOS datasets from different assessments. Section 5.1 gives an overview of some multimedia QoE models. Section 5.2 introduces the multi-sensory model. The validation of the proposed model illustrated in section 5.3. Model parameter estimation explained in section 5.4. Finally section 5.5 presents the model estimated responses.

5.1 Quality of Experience and Sensory Effects Models

In general there are three objective possible methodologies for measuring QoE:

1. The no-reference model has no knowledge of the original stream or source file and tries to predict QoE by monitoring several Quality of Service (QoS) parameters in real-time.
2. The reduced-reference model has some limited knowledge of the original stream and tries to combine this with real-time measurements to reach a prediction on the QoE.
3. The full-reference model assumes full access to the reference video, possibly combined with the measurements conducted in a real-time environment.

The first model fits under the umbrella of the second, which on its turn can be brought under the third. The full-reference model therefore should be able to give the best accuracy, but it is a method that can only be applied if one has control over both end systems. A no-reference model can be more easily adopted, but might not always give accurate results [5].

The current approaches for evaluating the QoE aim to map QoS to QoE [70], [71] or calculate the QoE from the audiovisual services [72], [73] and do not consider any additional effects like sensory effects. Exponential interdependency of QoE and QoS (IQX hypothesis) was introduced in [74]. The IQX hypothesis is formulated with QoE and QoS parameters, thus, providing an exponential function. That is, if the level of satisfaction decreases, the level of disturbance increases. The authors defined this function as an exponential because a small disturbance drastically decreases the satisfaction.

Additional QoE model presented in [75], which is triple user characterization model and consider three dimension into account, first is the sensorial quality which represents the quality of content sharpness, brightness, number of artifacts, blurriness, etc., second is the perceptual quality that depicts the amount of knowledge a user may acquire, and finally the emotional quality which depicts the satisfaction in terms of emotional experience. The model mainly addresses adaptation and presentation issues without addressing sensory effects. A Pseudo Subjective Quality Assessment (PSQA) presented in [76], which is a hybrid approach between objective and subjective evaluations. The results of the subjective assessment are used to train a learning tool that provides the relation between the parameters causing the distortion of the video sequences and the perceived quality.

The need for multi-sensorial media QoE models is a challenge for future research, since the evaluating of the multi-sensorial media QoE using subjective quality assessment is a time and cost consuming task although allowing for the definition of statistical prediction models.

A linear model for multi-sensory media introduced in [77] based on Mean Opinion Score (MOS) quality assessment [51]. This model evaluates the QoE of multi-sensory media from the quality of audio and video contents. The model proposed in [77] has been validated on three highly dynamic spatio-temporal multimedia sequences enriched with three sensory effects, namely wind, vibration and lights. Multiple Linear Regression (MLR) and the Least Square (LS) estimator method were employed to validate the model and to estimate the weights. From the result of the study, the authors conclude that the relationship between the QoE with effects and the number of effects is a linear relationship.

Another model proposed in [78] to estimate user QoE of olfaction-enhanced multimedia. The model instantiated a Multiplicative Exponential Weighting method (MEW), considering three factors, the system factors: inter-media skew for olfaction enhanced multimedia, content factors: impact of scent type (pleasant vs. unpleasant), and human factors: the influence of age and gender (human factors) on the user ability to detect skew.

5.2 Quality of Experience Model for Multi- Sensorial Media TV

This study proposes the following model for QoE for multi-sensorial media TV in smart home:

$$\mathrm{QoE_{eff}} = \mathrm{QoE_{av}} * \delta + \left(\sum \mathrm{b_i} * \mathrm{QoE_{av}^{w_i}}\right) \tag{1}$$

where QoE_{eff} is the quality with sensory effects, QoE_{av} is the multimedia QoE, (i.e., the quality of audio and video contents), w_i represents the weighting factor for each sensory effect, which can be airflow, light, or vibration, b_i is a variable with the value of 0 or 1 used to indicate the presence or not of the sensory effect, and δ is for tuning. The objective of choosing this model is the best and simplest model that adequately fits to our dataset.

Nonlinear Regression (NLR) [79] was employed to validate the model. Usually, NLR model arise when the relationship between the predictors and the response follows a particular functional form. In NLR, the equation output nonlinearly depends on one or more unknown model parameters.

The procedures to estimate the parameters are very important in many scientific fields for the development of mathematical models. Meanwhile all the process depends on parameters values obtained from experimental observation. The situation is made much more difficult and more challenging when the output of the model is depend nonlinearly on the model parameters. In order to overcome these difficulties, the Particle Swarm Optimization (PSO) algorithm is considered [80]. According to that, the parameter estimation of the proposed model has been based on PSO algorithm.

5.3 Multi-Sensorial QoE Model Validation

The model has been validated according to different empirical dataset MOS from two different subjective assessments:

- The first MOS dataset from the subjective assessments used to validate the linear model [77]. This allowed performing a fair comparison between the performance of the linear and our proposed model. In the assessment 32 students (6 female and 26 male) were invited to participate in the subjective test. Three highly dynamic spatio-temporal audiovisual sequences, one from the action category (2012), and two from the sports category (Pastranas, and Berrecloth), are showed to the participants, enriched with three sensory effects, light, airflow and vibration. Effects were also combined, thus creating seven different configurations. The resolution of each sequence is 720p. The assessment was performed in an isolated room, the details of the hardware and software components that used to perform the assessment are presented in Table 7. The assessment duration was around 15 minutes and was organized in four stages [77]. The first stage explains the test procedure and how the QoE will be assessed. The second stage collects information about the participants. The third stage is the subjective assessment, which is the main evaluation stage. In the last stage the participants were asked if they had took part in similar test and other questions related to the assessment, to provide a feedback.

Table 7: Hardware and software components used in the subjective test [77]

amBX Premium Kit (Fan, Vibration Panel, Light, Sound)
24” Monitor with a resolution of 1400 × 1050
Mozilla Firefox 6 & 8 in full-screen mode
Ambient Library 1.5 & Web browser plug-in 1.5
amBX Software (amBX System 1.1.3.2 and Philips amBX 1.04)
Dell Optiplex 655: Pentium D 2.8 GHz w/1 GB RAM & ATI Radeon HD 5450
Windows XP SP3

- The second dataset used to validate the model is the subjective assessment data MOS from our assessment as described in chapter 3, when 40 participants watched multi-sensorial video sequences taken from [50], enriched with three sensory effects, light (L), airflow (A) and vibration (V). Effects were also combined, creating different test cases. Figure 5-1, Figure 5-2, Figure 5-3, Figure 5-4, and Figure 5-5 show a snapshot from each video sequence (i.e., 2012 from action category, Berrecloth and Pastranas from sport category, Bridgestone from commercial category, and Earth from documentary category). In the assessment 5 video sequences presented to the participants in seven different test cases: L, A, V, L+A, L+V, A+V, L+A+V.

Figure 5-1: Snapshot of the action sequence 2012

Figure 5-2: Snapshot of the documentary sequence Earth

Figure 5-3: Snapshot of the sport sequence Berrecloth

Figure 5-4: Snapshot of the commercial sequence Bridgestone

Figure 5-5: Snapshot of the sport sequence Pastranas

5.4 Parameter Estimation

Parameter estimation procedures are very important in many scientific fields for the development of mathematical models, since all of the process depend on model parameters values obtained from experimental data. Difficulties in parameter estimation and statistical analysis of parameters are due to the large number of parameters and multi modal nature. In order to overcome these difficulties, the use of a powerful meta-heuristic method such as PSO algorithm may be considered [81]. PSO was introduced in 1995 by Kennedy and Eberhart [82] and has been successfully applied in engineering design and optimization, electronic system, robot control and navigation, and in industrial production optimization [83]. In [84]-[86], authors proposed a modified PSO for intelligent mobile robot navigation. In the literature, there are numerous articles about using PSO for parameter estimation. Reliable parameter estimation approach based on PSO algorithm for nonlinear regression model was developed in [80] and tested on well-known 28 nonlinear regression models. The results show that PSO is an efficient method for handling the problems of parameter estimation of the nonlinear regression models. PSO was successfully applied in [87] to obtain the parameters of linear regression models for the symbolic interval-values data. In the results, the proposed method presents a satisfactory performance. The capability of PSO, Genetic Algorithm (GA) [88]-[90], and Multiple Regression (MR) in estimation of soil mechanical resistance explored in [91]. The comparisons between the three models mentioned above indicate that PSO clearly outperforms GA and MR model, and the results obtained from PSO are in agreement with the experimental results. PSO algorithm and Simulated Annealing (SA) algorithm are used in [92] to optimize the coverage of television broadcasting Single Frequency Network (SFN) while minimizing the interference degree, using subjective evaluation criteria for quantifying the reception quality [93]. The results show that PSO algorithm, increase overall coverage and reduce interference in critical directions.

5.4.1 Multi-Sensorial QoE Model Parameter Estimation Using PSO

This section describes the implementation of PSO to estimate the optimal values of the parameters in the proposed model. The PSO known as an optimizer is a population-based, self-adaptive search optimization technique. The PSO consists of a set of solutions (particles) called population. Each solution consists of a set of parameters and represents a point in multidimensional search space [80].

Initially, we define the swarm size, p (i.e., number of particles), and the maximum number of iterations, t_max, The independent variables V, L, and A represent vibration, light, and airflow effects, and the dependent variable y represents the QoE_{eff}. The outcome of the PSO algorithm will be the MOS. In the proposed model the influence of independent variables (i.e., effects) on the dependent variable (QoE_{eff}), is represented by the coefficients w_L, w_A, w_V. The position of each particle and its velocity within the swarm are denoted by y_j, and v_j, respectively, where the index j is the particle number. For each particle at the 1st iteration (k = 1), the initial values of velocity v_j^1, weights w_{Lj}^1, w_{Aj}^1, w_{Vj}^1 and position y_j^1 are randomly selected. Particles are moved iteratively to find the new position in the one dimension search space.

At each iteration k the fitness value f_j evaluated, which is the difference value between the previous position and the current position. Then, the best location visited by each particle ($pbest_j^k$), and the best position in the whole swarm ($gbest^k$) is determined. Therefore, if the value of f_j^k is better than the best f_j ($pbest_j^k$) in the history then, set the current value as the new $pbest_j^k$, the particle with the best fitness value achieved among all particles in the swarm is selected as the $gbest^k$. Particles update their velocity according the following equation [94]:

$$v_j^{k+1}=I*v_j^k+c_1*r_1*(pbest_j^k-y_j^k)+c_2*r_2*(gbest^k-y_j^k) \quad \textbf{(2)}$$

where v_j^k and v_j^{k+1} are the current and updated particle's velocity, I is the inertia weight which is a constant, r_1 and r_2 are random variables [0,1], c_1 (self-confidence factor) and c_2 (swarm-confidence factor) are constants. The cognitive component term $pbest_j^k-y_j^k$, represents the best solution found by each particle. The social component term $gbest^k-y_j^k$, referred to the best solution in the whole swarm.

At iteration k+1, for each particle the weights $w_{i,j}^k$ are updated, where i ϵ{vibration (V), light (L), and airflow (A)}. The new updated weights values $w_{i,j}^{k+1}$ are calculated according to the following equation:

$$w_{i,j}^{k+1}=w_{i,j}^k+v_j^{k+1} \quad ,j=1,\ldots,p \quad \textbf{(3)}$$

Then, for each particle the new updated weights $w_{i,j}^{k+1}$ from Equation (3) are substituted in Equation (1) to compute the new position y_j^{k+1}. Each individual particle keeps searching for the individual and global best position based on updating the velocities. This process is continues until the optimal parameter values of the proposed model are achieved or maximum iteration number t_max is reached. The process of PSO algorithm is summarized in Algorithm1.

Algorithm 1 : PSO Algorithm

Inputs:
  $Dataset((Lm, Am, Vm), y)$, $_{test\,cases\,(m:1\to 7)}$
Initialize:
  p, t_max, c_1, c_2, $and\,I$
for $j : 1 \to N$ **do**
  Initialize:
  $y_j^1, v_j^1, w_{Lj}^1, w_{Aj}^1, w_{Vj}^1$
end for
while optimum parameter values or t_max is not attained **do**
  for $j : 1 \to N$ **do**
    $k : 1 \to t_{max}$
    compute y_j^k
    compute f_j^k
    if $f_j^k < f_j^{k-1}$ **then**
      p_{bestj}^k is set to the current location
    end if
    $gbest^k$=best particle in the whole swarm
    for $j : 1 \to N$ **do**
      Compute $v_j^{K+1}, w_{Lj}^{k+1}, w_{Aj}^{k+1}, w_{Vj}^{k+1}$
      Compute y_j^{K+1}
    end for
  end for
end while

5.5 Multi-Sensorial QoE Model Estimated Response

The parameter estimation of the proposed model was performed by implementing PSO algorithm, using MATLAB and run on a computer with a processing unit of 2.50 GHz Intel (R) Core i5 with 8 GB of RAM. To ensure the algorithm achieve convergence preliminary tests on the PSO had been run, the final convergence of the model presented in Figure 5-6. It can be seen from the figure that the convergence process results in smooth curves, with a rapid decrease at the beginning and then gradually slows down and the algorithm converges to the minimum value of the Mean Squared Error (MSE) within 20 iterations. The resulting PSO parameter setting is summarized in Table 8.

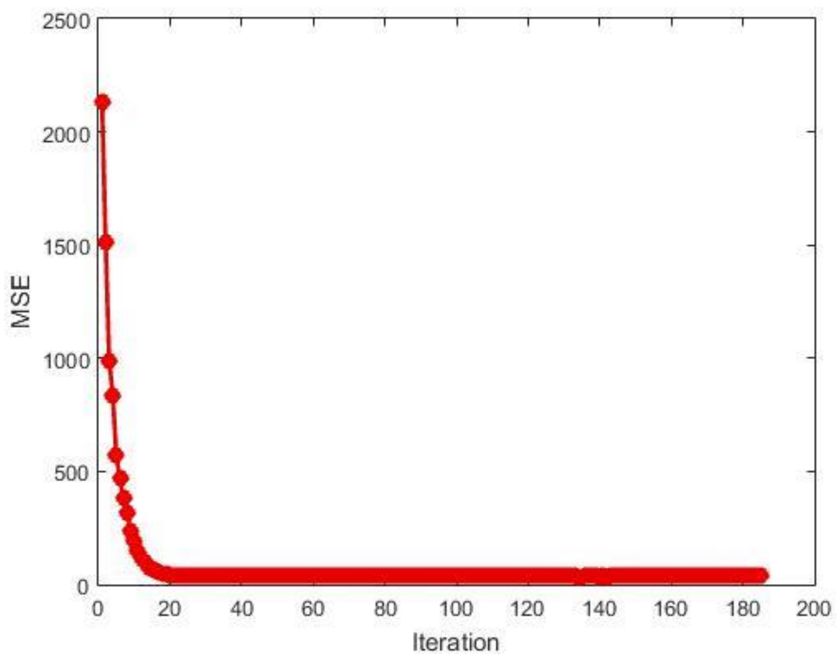


Figure 5-6: The convergence of MSE for the PSO algorithm

Table 8: PSO parameters setting for the experiment

PSO Parameters	Setting
Swarm size, p	15, after running preliminary tests and based on trial and error approach
Maximum number of iterations, t_max	300, after running preliminary tests and based on trial and error approach
Self-confidence factor, c_1 and swarm-confidence factor, c_2	2, as suggested by [95] and [96]
The inertia weight, I	0.02, was selected by running preliminary tests on the selected use case

The estimated parameters value (δ and the weights w_i) for the proposed model are shown in (4).

$$\mathrm{QoE_{eff}} = \mathrm{QoE_{av}} * 1.212 + b_L * \mathrm{QoE_{av}^{0.404}} + b_A * \mathrm{QoE_{av}^{0.418}} + b_V * \mathrm{QoE_{av}^{0.742}} \quad \textbf{(4)}$$

In the subjective assessment, the QoE_{av} represents the quality of the video sequence without any sensory effects, in productive systems, the QoE_{av} can be assessed from the existing QoS models [70], [74], [75] this is done by estimating the QoE_{av} from QoS parameters [97].

5.5.1 Model Estimated Response for the First Empirical Dataset

The subjective assessment data used to conduct the proposed model with the model responses for the three sequences (2012, Pastranas, and Berrecloth) are shown in Figure 5-7, Figure 5-8, and Figure 5-9. The majority of the model responses are inside the Confidence Intervals (CI) (95%) of the subjective quality responses. The model response is also close to the average of the MOS, which show that the proposed model can provide satisfactory estimation accuracy.

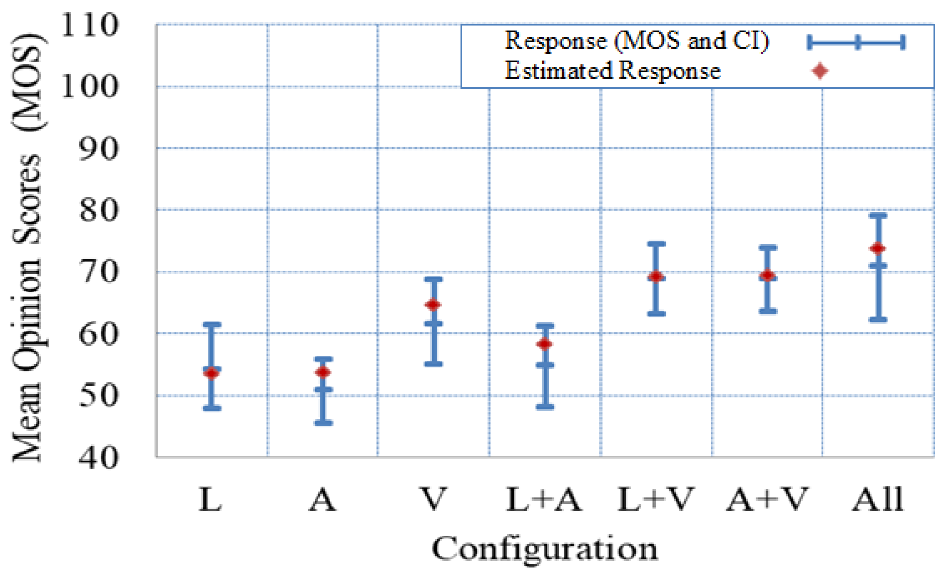


Figure 5-7: The estimated response by the proposed model compared to the MOS for the action sequence 2012

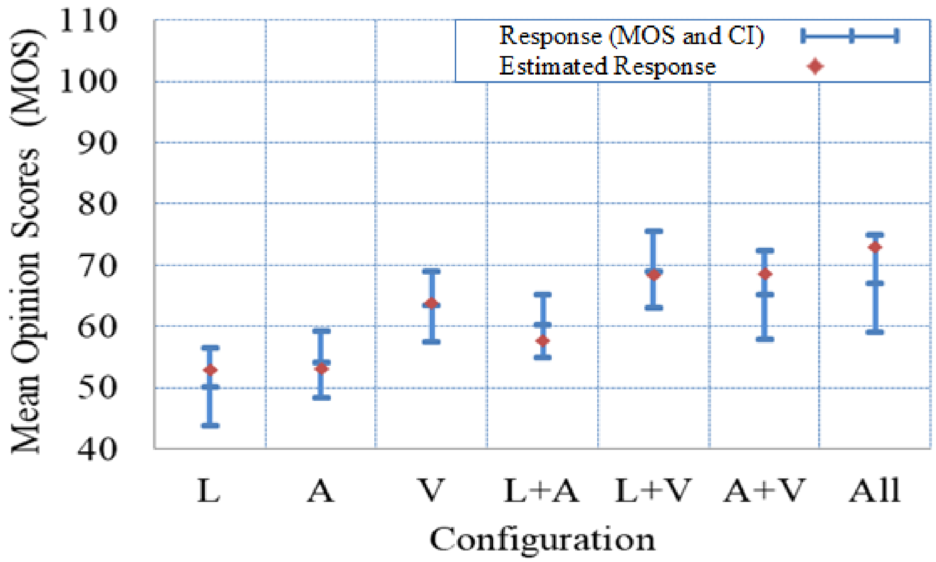


Figure 5-8: The estimated response by the proposed model compared to the MOS for the sport sequence Pastranas

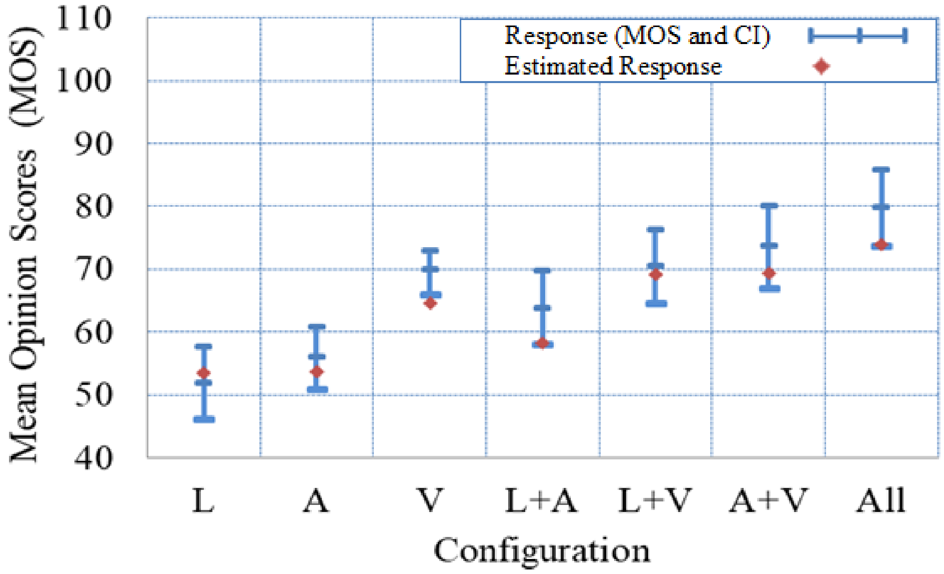


Figure 5-9: The estimated response by the proposed model compared to the MOS for the sport sequence Berrecloth

In order to show the improvement of the proposed model a comparison has been made a with the linear sensory experience model presented in [77]. The performance comparison is in term of MSE which measures the average difference between the model response and the MOS as illustrated in Figure 5-10.

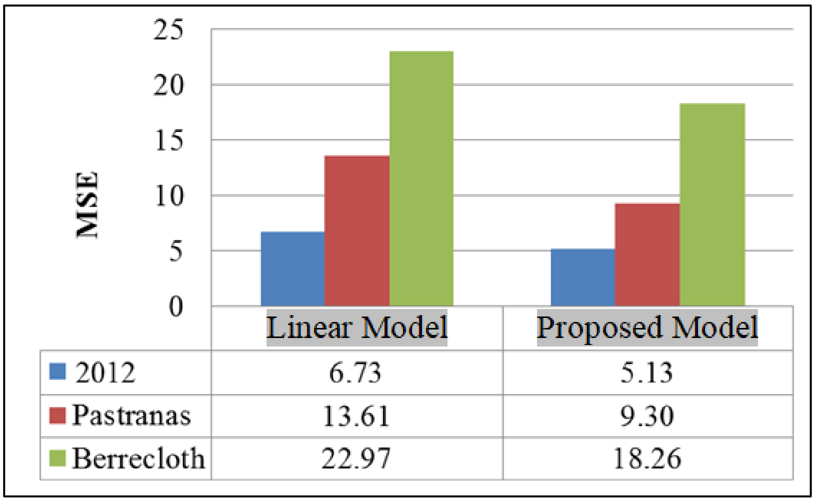


Figure 5-10: Comparison between the linear and the proposed model in term of MSE

According to the results shown in Figure 5-10, it can be concluded that the proposed model give improved results compared to the linear model for the three test sequences. The proposed model allows obtaining an improvement of 11.27% with respect to the linear model presented in [77].

Furthermore, the models prediction accuracies are compared by using the value of the square of multiple correlation coefficients (R^2), where R^2 is the correlation between the actual values and the predicted values. As shown in Table 9, the achieved value of R^2 for the proposed model is higher than the value of the linear model presented in [77]. Therefore, evaluation using the proposed model is more accurate and enhances the estimation accuracy.

Table 9: Comparison between the linear and the proposed model in term of R^2

Model	Linear model [77]	proposed model
R^2	0.782	0.836

5.5.2 Model Estimated Response for the Second Empirical Dataset

The subjective test results (i.e., MOS) for the multi-sensory sequences, with seven different configurations are shown in Figure 5-11, Figure 5-12, Figure 5-13, Figure 5-14, and Figure 5-15. The impact of sensory effects on the MOS differs, depending on the category and contents of the sequence. For the action category, the vibration effect was the most appreciated effect by users. For documentary and commercial containing scenes of nature and landscape, the light was the most preferable. For the sport category the vibration effect was the one with the highest impact, for the sequence Berrecloth, which contain scenes of bicycling down the rock cliffs. On the other hand, for the other sport sequence Pastranas, the airflow effect was preferable. The impact of the combinations of sensory effects (L+A, L+V, V+W) is either lower than or equal to the impact of an individual sensory effect, though the combination of all effects together has the highest impact on the MOS.

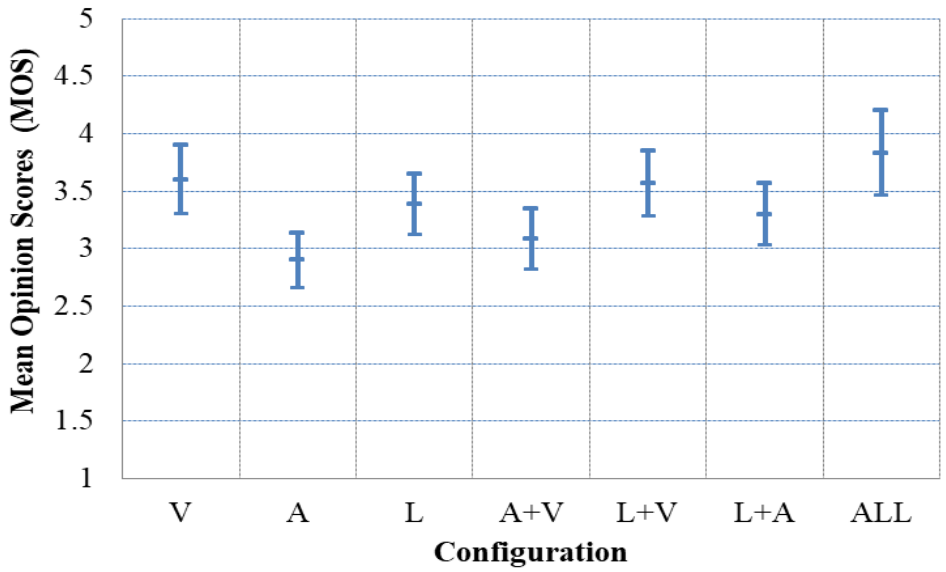


Figure 5-11: MOS and confidence intervals (95%) for the action sequence 2012

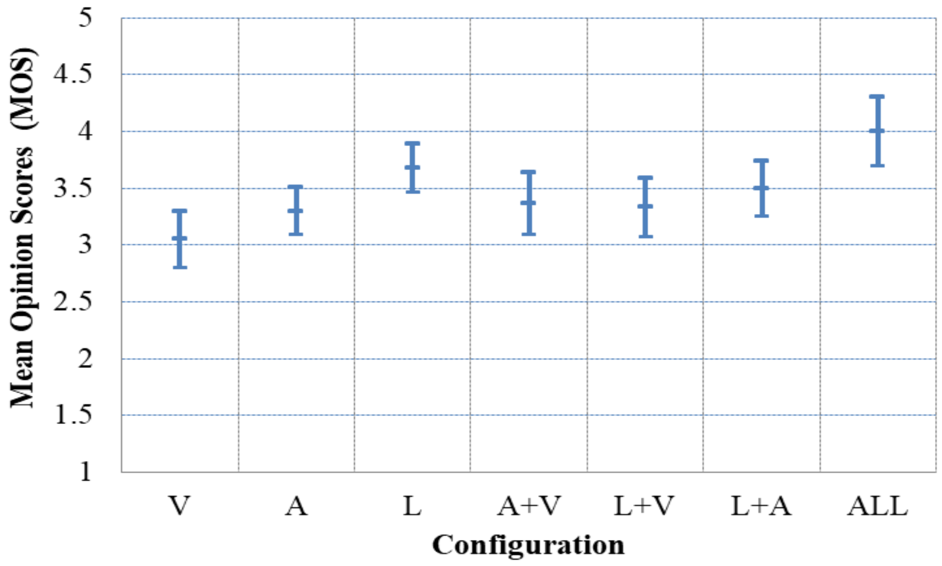


Figure 5-12: MOS and confidence intervals (95%) for the documentary sequence Earth

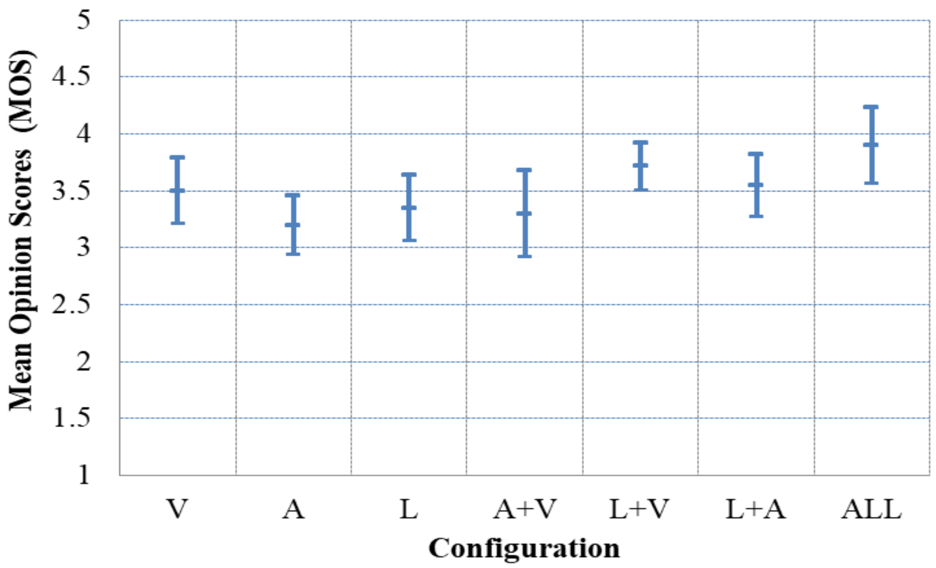


Figure 5-13: MOS and confidence intervals (95%) for the sport sequence Berrecloth

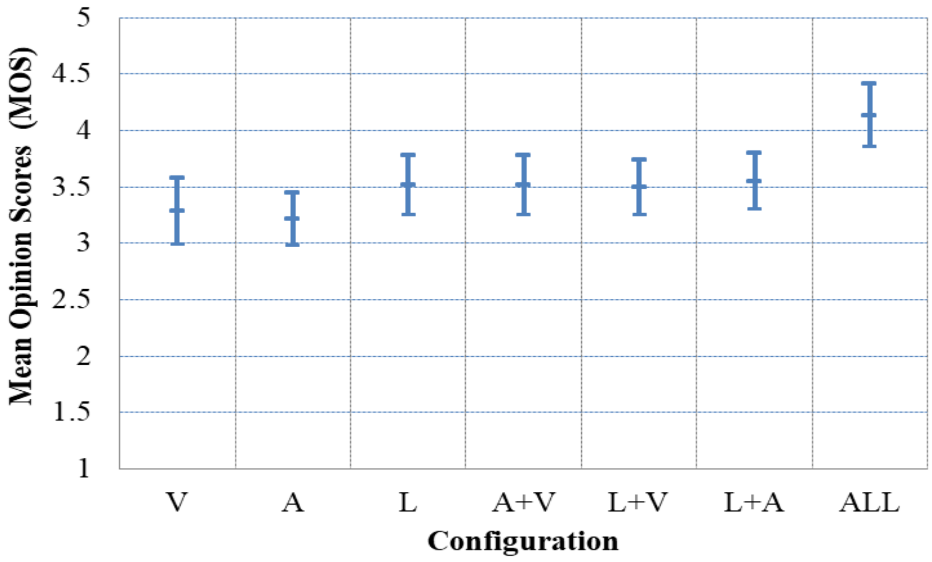


Figure 5-14: MOS and confidence intervals (95%) for the commercial sequence Bridgestone

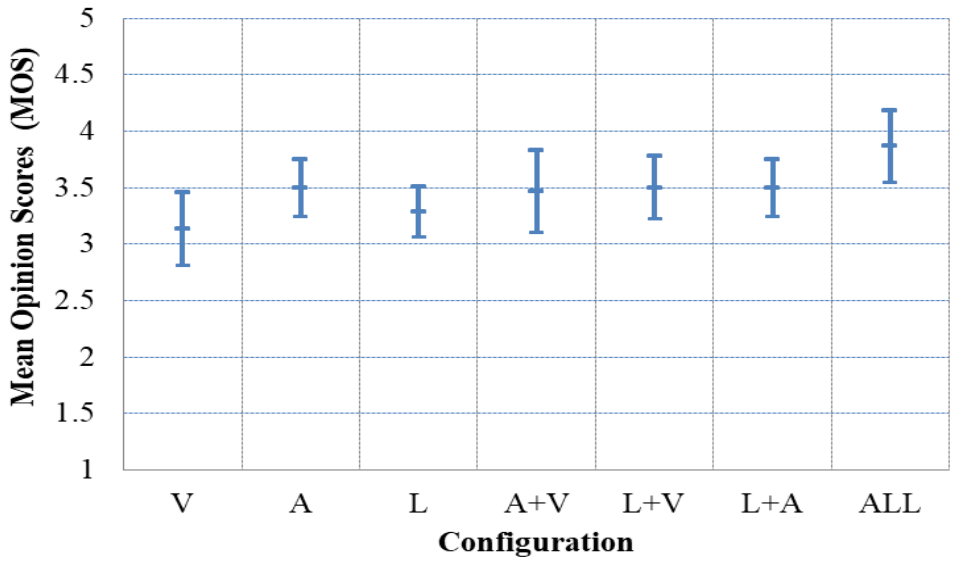


Figure 5-15: MOS and confidence intervals (95%) for the sport sequence Pastranas

The results of the subjective assessment allowed us to validate the proposed model for the prediction of the QoE of multi-sensory media for TV applications on smart home scenario. The subjective assessment results applied to audiovisual sequences selected from the category action, sport, documentary, and commercial which will reduce the need to conduct further subjective quality assessments to assess how the model performs in other category.

The estimated responses by the model for all the sequences are shown in Figure 5-16. The majority of the responses of the proposed model are inside the CI (95%) of the subjective quality responses. The model response is also close to the average of the MOS.

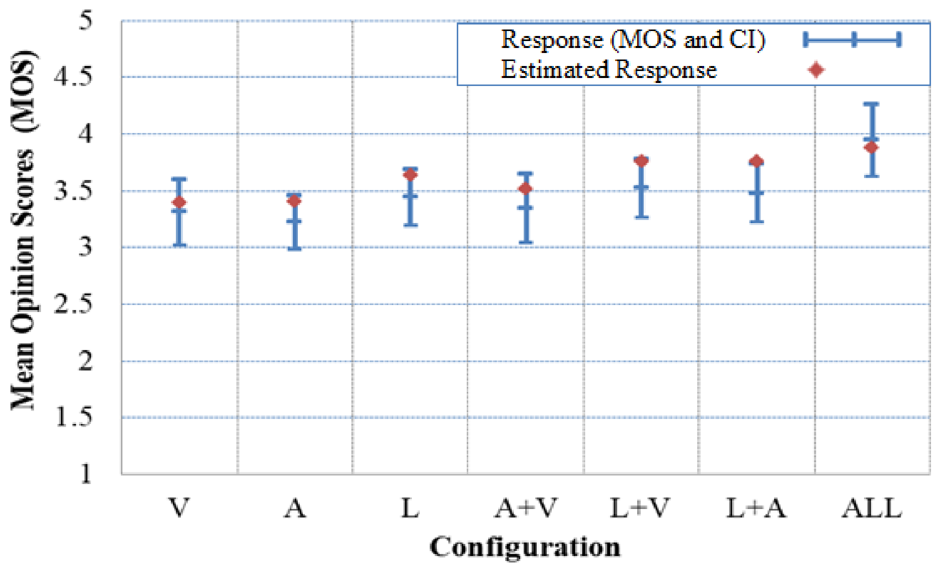


Figure 5-16: The overall estimated response by the proposed model for the second empirical dataset

The accuracy of the model in terms of MSE is very encouraging. The accuracy of the proposed model indicating the relationship between the quality of audio and video contents and user perceived QoE with sensory effects qualifies this work as a contribution in modelling user QoE for multi-sensory media applications.

Chapter 6
Conclusions and Suggestions for Future Works

6.1 Conclusions

In this work, the evaluation of the Quality of Experience (QoE), i.e., “the degree of delight or annoyance of the user of an application or service” for multi-sensorial media has been investigated.

In Chapter 2, the state of the art and current research activities focusing on the role of multi-sensory media in enhancing the quality of experience are discussed, and overviewed some of the existing sensory effect authoring tools, devices, dataset, and the impact of synchronism between sensory effects and multimedia content are discussed. In addition, some subjective studies, different approved subjective quality assessment methods and a variety of assessment parameters are presented. Based on these studies, one can conclude that the QoE can be enhanced by multi-sensory media. The perceived video quality can be enhanced by adding sensory effects, and the sensory effects can partly mask the video quality decrease. In addition, sensory effects can affect the strength of the emotions.

In Chapter 3, the feasibility of an Internet of Things (IoT) based approach to reproduce multi-sensorial media sequences on a real smart home television scenario is described. A cloud IoT architecture has been designed and implemented based on home customary devices by respecting the synchronization constraints. A quality of experience assessment campaign has been performed based on subjective test.

The obtained result showed the feasibility of the proposed approach in terms of synchronization constraint, increase of the sense of reality and general overall satisfaction of the users. Our evaluation was conducted with participants that are between 22 and 50 years old. Due to this, the assessment results valid for the given age group, a general conclusion can be drawn that is sensory effects enhance the viewing experience for the participants by stimulating also other senses than vision and audition.

The IoT solution allows implementing the system in a real smart home scenario without the need to deploy dedicated specific hardware. Furthermore, the IoT approach allows scalability and also the possibility to add customized features to the overall system. In a smart home scenario, the user preferences can be saved by the IoT architecture and the setting of the devices can be adjusted accordingly. This opens new outlooks in the broadcasting area, since it is possible to forecast new services that could be provided to consumers tailored on their experience preferences.

In Chapter 4, the impact of light, vibration, and airflow effects on user experience analyzed in terms of enhance the sense of reality, impact of annoyance, user response to the intensity of the effects, user preference of sensory effects, sequences duration and user overall enjoyment.

Majority of the participants 80% enjoyed the multi-sensorial experience with the three effects. Furthermore most of the assessors 85% consider that sensory effects enhance the sense of reality, and 85% agreed that sensory effects enhance the audiovisual content. In contrast, the majority of the participants (i.e., 95% for light effect, 65% for airflow, and 45% for vibration) did not experience any annoyance as a result of the sensory effects, and 67.5% did not distract by the sensory effects. In addition, results reveal that users assess the intensity of the sensory effect during the assessment as fine (i.e., 72.5% for light effect, 45% for airflow, and 32.5% for vibration). Also 70% of the participants felt their emotions increased with the multi-sensorial media. Most of the participants 77.5% prefer a longer multi-sensorial sequence to better appreciate the impact of sensory effects. In our assessment 67.5% of the participants select the light effect as their preferred sensory effect. The assessment results show that most of the participants were interested in experience additional sensory effects (i.e., 77.5% for thermal effect and 75% for scent effect).

The conclusion on the impact of sensory effects on the user experience is not the same for each video category. The sequences have been selected from the category action, sport, documentary, and commercial. The news category was not included in this assessment, since according to the literature sensory effects have low influence on news category. Additionally, the assessment results indicate that the contents of the video sequence play important role for enriching user viewing experience by sensory effects, sequences with short shots and a lot of transitions affect the voting behavior of the participants.

In chapter 5, a parametric QoE model for multi-sensory media TV in smart home scenario has been instantiated and validated in a real smart home scenario, through Mean Opinion Score (MOS) assessment campaign.

The results of the subjective assessment allow us to derive this model for the quality of sensory experience. The proposed model takes into account the number of effects and their impact on the QoE. Furthermore, this model can be combined with models that allow an estimation of the QoE for audiovisual content. The instantiation of the proposed model based on empirical data and fits with literature user studies. The validity of the proposed model based on the subjective assessment results applied to audiovisual sequences selected from the category action, sport, documentary, and commercial, this will reduce the need to conduct further subjective quality assessment to assess how the model performs in other category.

The proposed model has been instantiated with three sensory effects light, airflow, and vibration, but other sensory effects like scent and temperature can be added. Due to the nonlinearity of the problem, as to the model parameters estimation, a meta-heuristic approach, based on Particle Swarm Optimization (PSO) algorithm, has been implemented. In order to show the effectiveness of the proposed nonlinear model a comparison has been made with a linear multi-sensorial quality of experience model presented in literature based on the same MOS dataset from the subjective test experiments used to validate the linear model. Results show that the quality of experience estimated by the proposed model is more accurate and therefore the proposed model can enhance the estimation accuracy, and can provide satisfactory estimation of the QoE of multi-sensory media applications. The accuracy of the proposed model indicating the relationship between the quality of audiovisual contents and user perceived QoE with sensory effects qualifies this work as a contribution in modelling user QoE for multi-sensory media applications.

6.2 Suggestions for Future Works

The subjective assessment results highlighted some issues that need to be further investigated in future works. The main critical issue derived from the individual comments of the participants, is that vibration effect delivered by the smartphones can be sometime annoying depending on the smartphones placement. Participants either held the smartphone in their hand, or placed it in their pockets, or leaning beside them in the sofa, the vibration effect should be used carefully in order to eliminate the risk of presenting an annoying experience to the participants. Moreover, some of the participants complained about the limitation in impact of the RGB lights led due to the distance set between the sofa and the TV monitor. It seems that the viewing distance as specified by the ITU-T Recommendation P.911 does not match with the need of the users in case of light enhancement effect which need to be studied further in the future. On the other hands the airflow effect strongly depends on the distance between the air conditioner fan and the sofa. It appears evident that different room geometries and more tests have to be performed in order to find the right tradeoff and standardizing a quality assessment procedure with multi-sensory effects that is agreed upon by the community and used widely.

Furthermore providing a new high resolution multi-sensorial sequences dataset suitable for performing evaluations of multi-sensorial media, which will reduce the time consuming procedure of generating the multi-sensorial metadata descriptions, is future works subject. Another future investigation task is intensively study the synchronization between sensory effects and multimedia content for the IoT architecture, which designed and implemented respecting the synchronization constraints.

Finally, we hope if opportunely embedded on the content stream or in the broadcasting transmission technology, multi-sensorial metadata could be interpreted by a smart TV and used in a smart home environment to provide users with an enriched TV experience, via the proposed IoT architecture as a starting point.

References

[1] M. Fadda, M. Murroni and V. Popescu, "An unlicensed indoor HDTV multi-vision system in the DTT Bands," *IEEE Transactions on Broadcasting*, vol. 58, no. 3, pp. 338–346, 2012.

[2] D. Giusto, M. Murroni and G. Soro "Slow motion replay of video sequences using fractal zooming," *IEEE Transactions on Consumer Electronics,* vol. 51, no. 1, pp. 103–111, February 2005.

[3] P. Le Callet, S. Moller and A. Perkis, "Qualinet white paper on definitions of quality of experience," *European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003*), Lausanne, Switzerland, Version 1.2, 2013.

[4] *ITU-T Rec.E.800*, "Definitions of terms related to quality of service," 2008.

[5] F. Kuipers, R. Kooij, D. De, and K. Brunnström, "Techniques for measuring quality of experience," *in Proc. of the International Conference on Wired/Wireless Internet Connections*, 2010

[6] *ITU-T Rec.G.107*, "The E-Model a Computational Model for Use in Transmission Planning," 2003.

[7] A. Mellouk, S. Hoceini and H. Tran, "Quality of Experience for Multimedia: Application to Content Delivery Network Architecture," *Wiley-ISTE*, pp. 235, Nov 2013. ISBN: 978-1-848-21563-4.

[8] G. Ghinea, C. timmerer, W. Lin, and SR. Gulliver, "Mulsemedia: State of the Art, Perspectives, and Challenges," *ACM Transactions on Multimedia Computing, Communications and Applications (TOMM),* vol. 11, no. 17, 2104.

[9] Z. L. Wu and N. Saito, "The Smart Home [Scanning the Issue]," *Proceedings of the IEEE*, vol. 101, no. 11, pp. 2319–2321, Nov. 2013. doi: 10.1109/JPROC.2013.2282668.

[10] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, 2010.

[11] M. Fadda, M. Murroni, and V. Popescu "A cognitive radio indoor HDTV multi-vision system in the TV white spaces," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 2, pp. 302–310, 2012

[12] C. Alippi: Intelligence for Embedded Systems. *Springer Verlag*, 283pp, 2014. ISBN 978-3-319-05278-6.

[13] "Cyber-physical systems," Program Announcements & Information. *The National Science Foundation*, 4201 Wilson Boulevard, Arlington, Virginia 22230, USA. 2008-09-30. Retrieved 2009-07-21

[14] B. Choi, E. Lee, and K.Yoon, "Streaming media with sensory effect," *in Proc. of the International conference on Information Science and Applications (ICISA)*, pp. 1–6, 2011.

[15] M. Waltl, C. Timmerer, B. Rainer, and H. Hellwagner, "Sensory effects for ambient experiences in the World Wide Web," *Multimedia tools and Applications*, vol. 70, no. 2, pp. 1141–60, 2014.

[16] J. J. Kaye, "Making Scents: Aromatic Output for HCI," *Interactions*, vol. 11, no. 1, pp 48–61, 2004.

[17] G. Ghinea and O. Ademoye, "User Perception of Media Content Association in Olfaction-Enhanced Multimedia," *ACM Transactions on Multimedia Computing, Communications and Applications (TOMM)*, vol. 8, no. 52, 2012.

[18] M. Waltl, C. Timmerer, and H. Hellwagner, "Increasing the User Experience of Multimedia Presentations with Sensory Effects," *in Proc. of the 11th International Workshop on Image Analysis for Multimedia Interactive Services*, pp. 1–4, 2010.

[19] Z. Yuan, S. Chen, G. Ghinea, and G. Muntean, "User quality of experience of mulsemedia applications," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 11, no. 1, pp. 15:1–15:19, Sep. 2014.

[20] B. Rainer, M. Waltl, E. Cheng, M. Shujau, C. Timmerer, S. Davis, I. Burnett, C. Ritz and H. Hellwagner, "Investigating the impact of sensory effects on the Quality of Experience and emotional response in web videos,"

in Proc. of the 4th International Workshop on Quality of Multimedia Experience, QoMEX 2012, pp. 278–283, 2012.

[21] M. Waltl, B. Rainer, C. Timmerer, and H. Hellwagner, “An End to-End Tool Chain for Sensory Experience based on MPEG-V,” *Signal Processing: Image Communication*, vol. 28, no. 2, pp. 136–150, 2013.

[22] S. Shin, K. Ha, H. Yun and Y. Nam, “Realistic Media Authoring Tool Based on MPEG-V Proceedings of the 10th International Standard,” *in Proc. of the 8th International Conference on Ubiquitous and Future Networks (ICUFN)*, pp. 730–732, 2016.

[23] E. Saleme, and C. Santos, “PlaySEM: a Platform for Rendering MulSeMedia Compatible with MPEG-V,” *in Proc. of the WebMedia’15*, Manaus, Brazil, pp.145–148, 2015.

[24] M. Waltl, B. Rainer, C. Timmerer, and H. Hellwagner, “Sensory Experience for Videos on the Web,” *in Proc. of the IEEE Workshop Multimedia Web*, pp. 49–51, 2011.

[25] M. Waltl, B. Rainer, C.Timmerer, and H. Hellwagner, “Enhancing the User Experience with the Sensory Effect Media Player and AmbientLib,” *in Proc. of the Advances in Multimedia Modeling-18th International Conference, MMM 2012*, pp. 624–626, 2012.

[26] Sensory Experience Lab. http://selab.itec.aau.at.

[27] B. de Ruyter and E. Aarts, “Ambient Intelligence: Visualizing the Future,” *in Proc. of the Working Conference on Advanced Visual Interfaces ACM*, New York, NY, USA, AVI '04, pp. 203–208, 2004.

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

[29] Philips amBX PC Gaming Accessories. http://bit.ly/jilnyg

[30] Cyborg Gaming Lights. http://www.cyborggaming.com.

[31] M. Waltl, “The Impact of Sensory Effects on the Quality of Multimedia Experience,” *PhD thesis, Alpen-Adria-Universität Klagenfurt*, pp. 234, 2013.

[32] M. J. Smith, and P. J. Kiger, “The Lingering Reek of Smell-O-Vision,” *Article in Los Angeles times*, 2006.

[33] O. Ademoye and G. Ghinea, “Information Recall Task Impact in Olfaction-Enhanced Multimedia,” *ACM Transactions on Multimedia Computing, Communications and Applications (TOMM)*, vol. 9, no. 17, 2013.

[34] G. Ghinea and O. Ademoye, “The Sweet Smell of Success: Enhancing Multimedia Applications with Olfaction,” *ACM Transactions on Multimedia Computing, Communications and Applications (TOMM)*, vol. 8, no. 2, 2012.

[35] G. Ghinea and O. Ademoye, “Olfaction-Enhanced Multimedia: Perspectives and Challenges,” *Multimedia Tools and Applications*, vol. 55, no.3, pp. 601–626, 2011.

[36] Dale Air. http://www.daleair.com.

[37] M. Waltl, “Enriching Multimedia with Sensory Effects, Annotation and Simulation Tools for the Representation of Sensory Effects,” *VDM Verlag Saarbruucken*, Saarbruucken, Germany, 2010.

[38] Y. Kim, J. Cha, J. Ryu, and I. Oakley, “A Tactile Glove Design and Authoring System for Immersive Multimedia,” *IEEE Multimedia*, vol. 17, no.3, pp. 34–45, 2010.

[39] P. Lemmens, F. Crompvoets, D. Brokken, J. V. D. Eerenbeemd, and G.-j. D. Vries, “A body – conforming tactile jacket to enrich movie viewing,” *in Proc. of the Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pp. 7–12, 2009.

[40] S. Palan, R. Wang, N. Naukam, L. Edward, and K. J. Kuchenbecker, “Tactile Gaming Vest (TGV),” 2010. http://iroboticist.com/2010/03/26/tgv

[41] S. ur Rehman, “Turn Your Mobile Into the Ball: Rendering Live Football Game Using Vibration,” *IEEE Transactions on Multimedia*, vol. 10, pp. 1022–1033, Oct. 2008

[42] E. Dijk, A. Weffers-Albu, and T. De Zeeuw, “A tactile actuation blanket to intensify movie experiences with personalised tactile effects,” *in Proc. of*

the International Conference on Intelligent Technologies for Interactive Entertainment, pp. 2–3, 2010.

[43] A. Israr and I. Poupyrev, "Tactile Brush: Drawing on Skin with a Tactile Grid Display," *in Proc. of the Computer Human Interaction*, pp. 2019–2028, 2011.

[44] J. K. Yun, J. H. Jang, K. R. Park and D.W. Han, "Real-Sense Media Representation Technology Using Multiple Devices Synchronization," *in Proc. of the IFIP International Federation for Information Processing*, pp. 343–353, 2009.

[45] N. Murray, Y. Qiao, B. Lee, A. K. Karunakar, and G. Muntean, "Subjective evaluation of olfactory and visual media synchronization," *in Proc. of the 4th ACM Multimedia Systems Conference*, pp. 162–171, 2013.

[46] N. Murray, Y. Qiao, B. Lee, A. K. Karunakar, and G. Muntean, "Multiple-scent enhanced multimedia synchronization," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 11, no.12, 2014.

[47] Y. Ishibashi, T. Kanbara, and S. Tasaka, "Inter-stream synchronization between haptic media and voice in collaborative virtual environments," *in Proc. of the 12th Annu. ACM Int. Conf. Multimedia ACM*, New York, NY, USA, pp. 604–611, 2004.

[48] Q. Zeng, Y. Ishibashi, N. Fukushima, S. Sugawara, and K. E. Psannis, "Influences of inter-stream synchronization errors among haptic media sound and video on quality of experience in networked ensenmble," *in Proc. of the IEEE 2nd Global Conference on Consumer Electronics (GCCE)*, pp. 466–470, 2013.

[49] Z. Yuan, T. Bi, G. Muntean, and G. Ghinea, "Perceived synchronization of mulsemedia services," *IEEE Transactions on Multimedia*, vol. 17, no.7, pp. 957–966, 2015.

[50] M. Waltl, C. Timmerer, B. Rainer, and H. Hellwagner, "Sensory Effect Dataset and Test Setup," *in Proc. of the IEEE 4th Int. Workshop Quality Multimedia Experience*, pp. 115–120, 2012.

[51] *ITU-R Rec. BT.500-13* "Methodology for the subjective assessment of the quality of television pictures," 2012.

[52] R. L. Storms, and M. J. Zyda, "Interactions in Perceived Quality of Auditory-Visual Displays," *Presence*, vol.9, no. 6, pp.557–580, 2000.

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

[54] P. Corriveau "Video Quality Testing," *In H. R. Wu and K. R. Rao, editors, Digital Video Image Quality and Perceptual Coding*, chapter 4, pp.125–153.CRC Press, November 2005.

[55] S. Ishihara, "Ishihara's Tests for Colour-Blindness," *Kanehara & Co., Tokyo*, Japan, 1985.

[56] H. Snellen "Probebuchstaben zur Bestimmung der Sehscharfe," *P. W. van de Weijer*, Utrecht, 1892.

[57] M. Waltl, C. Timmerer, and H. Hellwagner, "Improving the quality of multimedia experience through sensory effects," *in Proc. of the Second International Workshop on Quality of Multimedia Experience (QoMEX'10)*, pp. 124–129, 2010.

[58] Z. Yuan, G. Ghinea, and G. Muntean, "Quality of Experience Study for Multiple Sensorial Media Delivery," *in Proc. of the International Wireless Communications and Mobile Computing Conference (IWCMC)*, pp: 142–1146, 2014.

[59] K. Yoon, B. Choi, ES. Lee, and TB Lim, "4-D Broadcasting with MPEG-V," *in Proc. of the IEEE Int. Workshop Multimedia Signal Process*, pp. 257–262, 2010.

[60] J. K. Yun, J. H. Jang, and K. D. Moon, "Development of the Real-Sense Media Broadcasting Service System based on the SMMD," *in Proc. of the IEEE International Conference on Consumer Electronics (ICCE)*, pp. 435–436, 2011.

[61] E. Ko, M. Han, S. Park, and J. Jang, "Design and implementation of the sensory effect Aggregator for providing the real-sense 4D Broadcasting

Services," *in Proc. of the International Conference on ICT Convergence (ICTC)*, pp. 177–178, 2012.

[62] "The State of Traditional TV: Updated With Q2 2017 Data," https://www.marketingcharts.com/featured-24817.

[63] R Girau, S Martis, and L Atzori "Lysis: A platform for IoT distributed applications over socially connected objects," *IEEE Internet of Things Journal*, vol. 4, no. 1, pp. 40–51, 2017

[64] M. Nitti, V. Pilloni, G. Colistra, and L. Atzori, "The virtual object as a major element of the internet of things: a survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1228–1240, 2015.

[65] Haier inverter technology. http://www.haier.net/en/.

[66] Philips hue personal wireless lighting. https://www2.meethue.com/en-us.

[67] TV 60" UHD 4K Flat Smart Serie 9 JU6800. http://www.samsung.com/it/tvs/uhd-ju6800/UE60JU6800KXZT/.

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

[69] *U.S. Department of Commerce /National bureau of Standards*. "Size of Letters Required for Visibility as a Function of Viewing Distance and Observer Visual Acuity,"

[70] D. S. Hands "A basic multimedia quality model," *IEEE Transactions on Multimedia*, vol. 6, no. 6, pp. 806–816, 2004.

[71] T. Zinner, O. Abboid, O. Hohlfeld, T. Hossfeld, and P. Tran-Gia "Towards QoE Management for Scalable Video Streaming," *in Proc. of the 21st ITC Specialist Seminar on Multimedia Application - Traffic, Performance and QoE*, pp. 64–69. 2010.

[72] T. Wang, A .Pervez, and H. Zou "VQM-based QoS/QoE Mapping for Streaming Video," *in Proc. of the 3rd IEEE International Conference on Broadband Network and Multimedia Technology (IC-BNMT)*, pp. 807–811, 2010.

[73] W. Cherif, A. Ksentini, D. Negru, and M. Sidibe "A_PSQA: Efficient real-time video streaming QoE tool in a future media internet context," *in Proc. of the IEEE International Conference on Multimedia and Expo (ICME)*, Barcelona, Spain, pp. 1–6, 2011.

[74] M. Fiedler, T. Hossfeld, and P. Tran-Gia "A Generic Quantitative Relationship between Quality of Experience and Quality of Service," *IEEE Network*, vol. 24, no. 2, pp. 36–41, 2010.

[75] F. Pereira "A Triple User Characterization Model for Video Adaptation and Quality of Experience Evaluation," *in Proc. of the 7th Workshop on Multimedia Signal Processing (MMSP)*, Shanghai, China, pp.1–4, 2005.

[76] D. De Vera, P. Rodriguez-Bocca, and G. Rubino "Automatic Quality of Experience Measuring on Video Delivering Networks," *SIGMETRICS Perform. Eval. Rev.*, vol. 36, no. 2, pp. 79–82, August 2008.

[77] B. Rainer, and C. Timmerer "A Generic Utility Model Representing the Quality of Sensory Experience," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, ACM, New York, USA, vol. 11, no. 1, pp. 14:1–14:17, 2014.

[78] N. Murray, G. Muntean, Y. Qiao, S. Brennan, and B. Lee, "Modeling User Quality of Experience of Olfaction-Enhanced Multimedia," *IEEE Transactions on Broadcasting*, , vol. 64, no. 2, pp. 539–552, 2018.

[79] GAF. Seber, and CJ. Wild "Nonlinear Regression," *Wiley Series in Probability and Mathematical Statistics*, Wiley, New York, 1989.

[80] V. Özsoy, and H. Örkçü "Estimating the Parameters of Nonlinear Regression Models Through Particle Swarm Optimization," *Gazi University Journal of Science*, vol. 29, no. 1, pp. 187–199, 2016.

[81] R. Eberhart and J. Kennedy, "A New Optimizer Using Particle Swarm Theory," *in Proc. of the 6th International Symposium on Micro Machine and Human Science*, Japan. IEEE Service Center Piscataway NJ, pp. 39–43, 1995.

[82] J. Kennedy, and R. Eberhart “Particle swarm optimization,” *in Proc. of the IEEE International Conference on Neural Networks*, Australia, pp. 1942–1948, 1995.

[83] Y.Zhao and W. Zu, “Real-time obstacle avoidance method for mobile robots based on a modified particle swarm optimization,” *in Proc. of the International Joint Conference on Computational Sciences and Optimization*, pp.269–272, 2009.

[84] L.D. Jalal, “Three-dimensional Off-Line Path Planning For Unmanned Aerial Vehicle Using Modified Particle Swarm Optimization,” *World Academy of Science, Engineering and Technology, International Journal of Mechanical, Aerospace, Industrial, Mechatronic and Manufacturing Engineering*, vol. 9, pp. 1529–1533, 2015.

[85] N.A. Shiltagh and L.D. Jalal, “Optimal Path Planning for Intelligent Mobile Robot Navigation using Modified Particle Swarm Optimization,” *International Journal of Engineering Technology*, vol. 2, pp. 260–267, 2013.

[86] N. A. Shiltagh, L. D. Jalal, “A comparative study: Modified particle swarm optimization and modified genetic algorithm for global mobile robot navigation,” *Int. J. Comput. Appl.*, vol. 89, no. 9, pp. 32–46, 2014.

[87] C. Yang, C. Chuang, J. Jeng and C. Tao, “Constructing the linear regression models for the symbolic interval-values data using PSO algorithm,” *in Proc. of the International Conference on System Science and Engineering (ICSSE)*, pp. 177–181, 2011.

[88] M. Murroni, “Performance analysis of modulation with unequal power allocations over fading channels: A genetic algorithm approach,” *in Proc. of the 14th European Wireless Conference, Electronic Proceedings*, Prague; Czech Republic; 2008.

[89] M. Lixia, V. Popescu, and M. Murroni, “Performance analysis of modulation with unequal power genetic algorithms,” *in Proc. of the Proceedings of the International Conference on Optimisation of Electrical and Electronic Equipment, OPTIM*, Brasov; Romania; 2010.

[90] N. A. Shiltagh and L. D. Jalal, “Path planning of intelligent mobile robot using modified genetic algorithm,” *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 3, pp. 31–36, 2013.

[91] M. Hosseini, S.A. Movahedi, A.A. Dehghani and Y. Khaledian, “Estimation of soil mechanical resistance parameter by using particle swarm optimization, genetic algorithm and multiple regression methods,” *journal of Soil and Tillage Research*, vol. 157, p. 32–42, 2016.

[92] M. Anedda, P. Angueira, J. Morgade and M. Murroni, “Heuristic performance evaluation for DVB-T/T2 SFN network,” *Telecommunication Systems*, vol.63, no.1, pp. 55–63, 2016.

[93] C. Perra, “A low computational complexity blockiness estimation based on spatial analysis,” *in Proc. of the 22nd Telecommunications Forum Telfor (TELFOR)*, Belgrade, 2014.

[94] W. Jatmiko, K. Sekiyama, and T. Fukuda “Modified particle swarm robotic for odor source localization in dynamic environment,” *International Journal of Intelligent Control and Systems: Special Issue on Swarm Robotic*, pp.176–184, 2006.

[95] R. Eberhart, and Y. Shi “Particle swarm optimization: developments, applications and resources,” *in Proc. of the IEEE International Congress on Evolutionary Computation*, pp.81–86, 2001.

[96] Y. Shi, and R. Eberhart “A modified particle swarm optimizer,” *in Proc. of the International Conference on Evolutionary Computation, IEEE World Congress on Computational Intelligence*, pp. 69–73, 1998.

[97] C. Timmerer, B. Rainer, and M. Waltl, “A utility model for sensory experience,” *in Proc. of the Quality of Multimedia Experience (QoMEX), 2013 Fifth International Workshop* ,pp. 224–229, 2013.