

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2019.DOI

A DASH-based Adaptive Multiple Sensorial Content Delivery Solution for Improved User Quality of Experience

LONGHAO ZOU¹, TING BI², AND GABRIEL-MIRO MUNTEAN², (Senior Member, IEEE)

¹College of Computer Science & Software Engineering, Shenzhen University, Shenzhen, China (e-mail: zoulonghao@gmail.com)

²Performance Engineering Laboratory, School of Electronic Engineering, Dublin City University, Ireland (ting.bi2@mail.dcu.ie, gabriel.muntean@dcu.ie)

Corresponding author: Ting Bi (e-mail: ting.bi2@mail.dcu.ie).

This work was supported by the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement no. 688503 for the NEWTON project (<http://newtonproject.eu>) and China Postdoctoral Science Foundation (No. 2017M622777).

ABSTRACT Increasing number of researchers are focusing on the emerging communication technologies which enrich user perceived quality of experience by involving vision, auditory, tactile, olfaction, gustatory and other senses. However, there are multiple challenges related to using multiple sensorial media (i.e. mulsemmedia), including synchronization with the traditional multimedia content and delivery over diverse network environments. This paper proposes MulseDASH, a novel multiple sensorial media content delivery solution based on the Dynamic Adaptive Streaming over HTTP standard (DASH). MulseDASH is described and evaluated in a real test-bed in terms of effectiveness of its adaptive streaming and synchronization mechanisms. Extensive testing involving both network emulation and subjective assessment experiments shows how MulseDASH performs excellent real-time streaming adjustment to match network conditions and improves user quality of experience.

INDEX TERMS Mulsemedia, MPEG DASH, Quality of Experience.

I. INTRODUCTION

CURRENT interactive rich media technologies (e.g. on-line multimedia streaming, social media tools, Virtual Reality (VR) and Augmented Reality (AR) applications, etc.) have narrowed significantly the distances between people, reformed the way people communicate, and provided a more immersive environment people experience. According to the most recent Cisco Internet traffic statistics report [1], the global Internet video traffic over IP which accounted for 75% of all IP data traffic in 2017 is expected to reach 82% in 2022. Moreover, the potential value of VR/AR technology is no longer a secret since numerous industry analyst reports have forecasted enormous growth for the rich media content exchange market, as they fundamentally enhance the way humans interact with digital and physical real worlds [2]. The Cisco data analytics report also indicates that the VR/AR traffic will increase more than 18-fold between 2017 and 2022 [1], and VR/AR headset devices will grow from around 20 million in 2017 to nearly 100 million in terms of numbers by 2022 [3].

Increasing rich media interaction involves more than just audio and video content. The multimedia specialist and in-



FIGURE 1: Multi-sensorial Media Interaction

ventor, Morton L. Heilig produced Sensorama Simulator, the first VR/AR machine, which provided support for offering users a multiple sensorial experience back in 1961 [4]. After several decades, 26 multimedia scientists gathered at the ACM SIGMM Conference in 2005 discussed publicly future directions in multimedia research, and made highly challenging proposals. They focused on making "interactions with remote people and environments nearly the same as

interactions with local people and environments" and explore other media content types alongside audio and video [5]. Figure 1, illustrates haptic, olfaction (smell), airflow (wind) and other potential sensorial inputs which could be used by researchers to enhance the way people interact remotely with equipment, machines, computers and other humans [5].

Since the term *mulsemmedia*, derived from multiple sensorial media, was first introduced in 2010 [6], a high number of pioneering research activities have involved mulsemmedia and user interaction with multi-sensorial content. For instance research works have focused on user experience optimization by employing olfaction (i.e. odor, smell) [7] [8] [9], airflow (i.e. wind effect) [10] [11], tactile interaction (i.e. kinesthetic, haptic, vibration, etc.) [12] [13] [14] and even gustatory stimuli (i.e. targeting taste) [15] [16]. However, most of the research already mentioned is off-line and has involved people interacting with local applications only. Additionally, most of these works are related to a single novel sensorial stimulus apart from audio and video, instead of designing solutions for interaction with multiple stimuli rich media content. Finally, even fewer researchers have focused on mulsemmedia content delivery-related aspects [17].

At the same time, diverse studies indicate a clear trend of user preference and industry push towards rich media content (e.g. ultra high definition video, VR/AR, omni-directional video, etc.), including delivery solutions [1]. Such content has high resolution which impacts positively user experience levels, but also it has high bitrate and low latency delivery requirements [18]. Unfortunately, despite the efforts put in terms of network advancements, including Fifth Generation (5G) technologies in general and 5G *Tactile Internet* research and development in particular [19], there is still need for innovative delivery solutions. Such solutions should support rich media exchange over the existing networks in order to support high user quality of experience (QoE). Related to mulsemmedia delivery, challenging is to balance the need for inclusion of multi-sensorial media components and consequent higher bitrate with latency and network bandwidth delivery-related requirements [17]. Enabling synchronisation between diverse mulsemmedia components and video during their delivery is also challenging [20] [21]. Adaptive solutions which adjust content delivery characteristics and ultimately transmitted bitrate to match network delivery conditions or device properties have had highly positive results, especially in terms of increasing user QoE [22] [23]. Among these solutions very successful are those based on the latest standard which supports multimedia delivery adaptation, the Dynamic Adaptive Streaming over HTTP (DASH) [24].

This paper introduces **MulseDASH**, a novel DASH-based adaptive delivery solution for mulsemmedia content which increases user QoE. In its dynamic adjustment of multiple sensorial content characteristics, MulseDASH performs an innovative trade-off between video quality and presence of diverse sensorial components. This trade off relies on the fact that the presence of other sensorial components has a masking effect on potential video quality variations,

as noted for audio [25]. The paper describes the adaptive mulsemmedia delivery architecture, presents the principles of MulseDASH and introduces its design. MulseDASH evaluation was performed by involving a real-life implementation of MulseDASH and a real test-bed. Testing results show how user perceived QoE increases when using MulseDASH in comparison with classic approaches.

This paper is organized as follows. Section II discusses studies related to multiple sensorial media and rich media content delivery. Section III introduces MulseDASH principle and framework, whereas section IV presents the implementation of MulseDASH and its deployment on a real test-bed. The MulseDASH real-life evaluation of user experience is also described. Section V analyzes the results and highlights the outcome of MulseDASH evaluation. Section VI draws conclusions and presents future work avenues.

II. RELATED WORKS

A. MULSEMEDIA: STATE OF THE ART

There have been many recent advances related to digital content beyond the classic multimedia format, including introduction of high and ultra-high resolution video, omni-directional video content and interactive multimedia. Among these highly diverse rich media content types, multiple sensorial media (*mulsemmedia*) has inspired academic and industrial researchers and developers, especially in relation to its potential to increase perceived QoE levels by improving user sense of reality. Researchers and designers were mostly focused on finding ways to overcome many existing challenges in acquiring, storing, displaying and exchanging mulsemmedia content and propose solutions to address them.

Although the number of recent proposals involving multiple sensorial technologies is much lower than those targeting visual and auditory human senses only, many mulsemmedia solutions have been proposed and have even been deployed, reaching the wide public. For instance haptic wearable devices were introduced and are used for health recovery¹, haptic gloves are part of VR gaming kits², haptic control is employed in robotics³, gas sensors are deployed for environment monitoring⁴, smell displays are used for art design⁵ and aroma diffusers are employed for interior decorations⁶. Next most important research outputs related to the human senses⁷ other than sight and hearing are discussed.

¹Haptic Health Recovery from Manchester Metropolitan University: <https://www.haptic.ro/development-new-haptic-device-aid-walking-rehabilitation/>

²Haptic Gloves: <https://haptx.com/>

³Disney's Remote Control Robot Hand: <https://likethefuture.com/disneys-remote-control-robots-move-just-like-people/>

⁴AMS MEMS Sensors: <http://ams.com/eng/Products/Environmental-Sensors/Gas-Sensors>

⁵Smell Expert Sissel Tolass: <https://www.theguardian.com/lifeandstyle/2016/mar/18/smell-expert-sissel-tolaas-breathes-deep-and-then-follows-her-nose-in-melbourne>

⁶Muji Aroma Diffuser: <http://muji.us/store/ultrasonic-aroma-diffuser.html>

⁷The five human senses are sight (visual), hearing (auditive), smell (olfaction), taste (gustatory) and touch.

1) Taste

Currently, the research related to the gustatory sense is lagging behind, as overcoming the issues related to taste including its perception, interpretation, description, replication, etc. are more complex. However, recently a super-family of G-protein-coupled receptors were found to be responsible to most of human tastes (e.g. sweet, sour, salt, bitter, and savory) [26]. Additionally, an artificial lipid membrane based technical taste sensor (i.e. electronic tongue) was developed to detect the same taste similarly to the human tongue⁸. Different from the costly taste sensors, some cheaper interactive taste actuating devices were implemented to stimulate the human tongue and provide the specific taste experience, which are based on changes in terms of vibration, electric current and temperature [27] [28].

2) Touch

Haptic technologies have been well studied, especially in the context of interaction between humans and machines, and real and virtual worlds. Haptic solutions support user immersion in VR both as input and feedback, to and from the virtual environment. By employing them could benefit various applications in diverse deployment areas such as medicine, entertainment, education, industry, arts and so on. In general, the overall machine haptic sensory-motor loop contains three major components: sensors, controller and actuators. The controller deploys the strategies or algorithms designed to process the sensing information collected by sensors and make actuators to perform actions i.e. provide users with a response or feedback [29]. Haptic sensors can be divided into two primary types: *tactile sensors* for cutaneous perceptual measurement and *kinesthetic sensors* for modeling force and position. Differing in terms of their piezoresistivity, capacitance, piezoelectricity, temperature or humidity transduction, diverse tactile sensors are currently deployed on wearable human-machine interfaces (e.g. haptic gloves), skin prosthesis, strain sensors, blood flow monitors and so on [30]. Regarding kinesthetic sensors, the magneto-resistive angle, optical (e.g. Microsoft Kinect⁹), acoustic and inertial (e.g. gyroscopes) sensors are employed to measure the force, velocity and relative placement information, respectively [31]. Compared to haptic sensors, haptic actuators are more complex and ingenious, providing a tactile-interactive interface between real and virtual worlds, machine and human users. Currently, the haptic actuators working with haptic sensors are based on mechanical structures that offer force or vibration feedback. For example the iPhone 7 *taptic engine* simulates a "3D" tactile haptic feedback when users touch the screen. The engine uses a Linear Resonant Actuator (LRA) to generate vibrations [32]. In general, the mechanical structure of force actuators is composed of a power source, a motor and a component for force transmission based on

⁸Insent Taste Sensors: http://www.insent.co.jp/en/products/taste_sensor_index.html

⁹Kinect: <https://www.xbox.com/en-US/xbox-one/accessories/kinect>

gears, pulleys/belts, oil/air pressure or capstans. For instance, actuator examples include the multiple degree of freedom manipulator Phantom Omni, which is based on a capstan drive [33] and the haptic glove Dexmo, which employs belt and pulley force transfer [34].

3) Olfaction

Olfaction plays an important role in human daily lives, and involves stimulation by odors. Various odors influence human affective states and moods, and facilitate memory retrieval i.e. recollection of events associated with particular odors [35]. Odor detection is one of the main applications of machine olfaction technologies which are based on classic chemical gas sensors, optical sensors, chromatography and other spectrometers (e.g. ion, infrared, mass, etc.) [36]. Recently, a novel Artificial Neural Networks (ANN)-based pattern recognition system was developed based on optical sensing results, which chemically learns about any changes in the surrounding environment [37]. Most machine olfaction sensing technologies (e.g. electronic noses) are employed in environment monitoring, industrial manufacturing, disease diagnostic and so on. Conversely, olfaction actuators working with VR applications have been widely employed in the fields of entertainment (i.e. gaming), environment decoration and education. For instance SBi4 v2, a olfaction diffuser produced by Exhalia¹⁰ has been used to assist in relevant olfaction experiments in academic education. However, it is not easy to control the diffusion direction and intensity. A 3D-printed bespoke was produced to adapt the direction of SBi4 diffusion in [9], and a Surface Acoustic Wave (SAW) device was designed to control the diffusion intensity [38].

4) Air Flow

In terms of sensorial input media, air flow refers to a stream of air perceived by humans. However, it has also become an important actuator, as it makes people have special feelings when the air in motion touches human skin. Often the air flow is also associated with other sensorial inputs, as for instance it carries and diffuses odors. Stand alone or in conjunction with other sensorial media, air flow enhances the immersive experience of users. Matsukura et al. in [39] developed a two-dimension multisensorial field display device which conducts the spatial airflow velocity, controls odor concentration and synchronizes the output with the images in the computer screen to improve user experience. An important challenge noted in relation to air flow is controlling the wind display direction. Kulkarni et al. [40] and Nakano et al. [11] have developed a professional large wind display device called Treadport Active Wind Tunnel (TPAWT) and a portable wind display device based on a matrix of CPU fans, respectively. The former could provide users any air flow distribution with infinite number of degrees of freedom (DoF) in a big room, and the latter could create a controllable air flow experience at a short distance. However, most related works have used air

¹⁰Exhalia: <http://www.exhalia.com>

flow in conjunction with other media types and have focused mainly on its delivery to users.

B. MULSEMEDIA DELIVERY

Diverse streaming solutions have been proposed in order to support delivery of audio-visual content to users anytime and anywhere across the Internet. Adaptive solutions are among the most successful methods to support high delivery performance in dynamic and resource limited network conditions. Muntean et al. have proposed the Quality-Oriented Adaptive Scheme (QOAS) which performed transmission of segments from different video clips, pre-encoded at different quality levels depending on the real-time feedback about network delivery conditions. The authors showed how by controlling the content quality adjustment, random loss decreased and overall user perceived Quality of Experience (QoE) increased. Recently, increasing number of solutions have focused on mobile and wireless multimedia delivery scenarios [41]. This is due to the growing number of devices and users and their increasing demand for mobile rich media networked applications. The adaptation solutions have also extended their focus, concentrating not only on network delivery, but also on other aspects such as those related to energy efficiency [42], wireless access networks [43], device screen resolutions [44], overall user QoE [45] and service cost [46].

The principles behind diverse research-proposed adaptive delivery methods have been deployed as part of mainstream practical protocols and solutions, including Microsoft Smooth Streaming (MSS)¹¹, Apple HTTP Live Streaming (HLS)¹² and Adobe Real Time Message Protocols¹³. More recently the MPEG Dynamic Adaptive Streaming over HTTP (MPEG-DASH)¹⁴ was standardized and has become the most popular technology for low-cost on-demand and live adaptive video streaming over the current network infrastructure. MPEG-DASH, which is also compatible to MSS and HLS, enables client-side selection and request from the server of video segments with different quality levels. The result is a smooth delivery of a video which adapts its bitrate according to dynamically changing network bandwidth.

Nowadays, multiple sensorial content is being used to complement classic audio-visual material. This is fueled by the existing rich support of various display devices, user interest in exploring novel technical infotainment avenues and studies which have shown how by employing multiple sensorial content, user QoE increases [17]. In terms of haptics, an important research avenue primarily focuses on human-machine teleoperation technologies. One of the proposed solution has transmitted haptic information (i.e. force and vibration) over IP networks at distance. Experiments studying haptic remote collaboration over a network connection between USA and UK [47] have shown that the

latency leads to instability of the remote haptic interaction in a Shared Virtual Environment (SVE) [48]. Other works [49] [50] [51] studying SVE collaboration found that the latency tolerance of haptic communication is under 60ms and the jitter is under 10ms.

Olfaction is one of the most popular sensorial media components. A new study in neuroscience found that a VR system which incorporated smell could influence human behaviors in navigation [52]. This finding is similar to those of several other olfaction media works which are focused on user experience of immersive applications (e.g. VR and AR). Richard et al. [53] and Zou et al. [54] have employed odor diffusion to enhance user QoE during immersive learning. 'Smelling Screen', an olfaction display machine was developed and used by Matsukura et al. to present corresponding odor distribution while a user is watching an image sitting in the front of a monitor at a distance of 0.5m [55]. The difficulty in terms of interactive olfaction-enhanced media is to maintain low any potential difference between display times of olfaction and visual content [20]. A series of studies by Murray et al. found that user QoE is impacted by several factors, including skew between video and olfaction media, delivery jitter, number of odors, and user profile. The authors recommend the potential skew between olfaction and video media component should remain within -5s to +10s [21].

Impressive research and standardisation effort has been put in order to enable good synchronisation between different multi-sensorial media components and bridge the gap between virtual and real worlds, including when there is networked interaction with and between remote users. The Virtual Reality Markup Language (VRML) based on Extensible Markup Language (XML), used also as part of the MPEG-4 Binary Format for Scenes (BIFS), was designed to describe some haptic content (e.g. depth, stiffness, friction or any texture of a scene/object) associated with 3D or 2D objects within video content. VRML, popular for development of Web-based 3D or 2D multimedia content, was superseded by X3D developed by Web3D¹⁵. However, latest 3D or VR content development has increasingly relied on Unity3D¹⁶, Unreal¹⁷ or other commercial programming platform recently, so VRML and X3D are less used in the market [56]. The ISO/IEC 23005 MPEG-V standard¹⁸ formulates, describes and organizes sensorial effects in a multimedia content based on the XML format [57]. In particular MPEG-V standardizes a unified format for interaction information between real and virtual worlds, including haptic messages, vibration pattern, thermal effects and so on. In general, the MPEG-V file, formatting sensorial content and corresponding audio-visual content is multiplexed into a MPEG-2 TS container, and then transmitted to users [58]. However, MPEG-V integrated with MPEG-2 TS does not support adaptive streaming over the real-time dynamic changing networks, or varied user profiles

¹¹MSS: <https://www.iis.net/media/experiencesmoothstreaming>

¹²HLS: <https://developer.apple.com/streaming/>

¹³RTMP: <https://www.adobe.com/devnet/rtmp.html>

¹⁴DASH: <https://mpeg.chiariglione.org/standards/mpeg-dash>

¹⁵x3d Web3D: <http://www.web3d.org/x3d/what-x3d>

¹⁶Unity3D: <https://unity3d.com/>

¹⁷Unreal: <https://www.unrealengine.com/>

¹⁸MPEG-V: <https://mpeg.chiariglione.org/standards/mpeg-v>

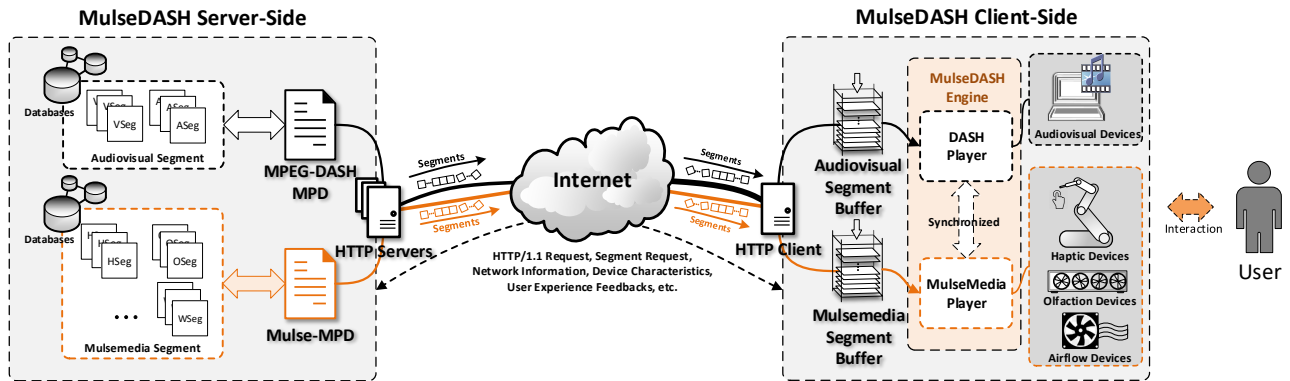


FIGURE 2: The Principle of MulseDASH: Adaptive Mulsemedia Streaming Scheme over HTTP-DASH

and device characteristics, affecting the user QoE during multi-sensorial effect rendering.

The ADAPtive MulSemedia delivery solution (ADAMS), proposed by Yuan et. al [17], is a method for performing adaptive multi-sensorial media delivery. In ADAMS, the metadata set annotating the different sensorial content associated with the video is organized and described using XML based on the MPEG-7 standard¹⁹. Different from MPEG-V, ADAMS adapts the specific sensorial media segments combined with the video packets according to the predicted network bandwidth variation and user profiles.

In this context, the existing mulsemmedia description and delivery solutions organize the sensorial media-related information into an extended XML file associated to the audiovisual content. Additionally, some of them convey the multimedia and mulsemmedia content together to the user-side immediately. They consider or not adaptation of the different levels of the media segments to specific user operational conditions (e.g. network bandwidth, user profile, device, etc). Unfortunately, it is not enough if the adaptive mulsemmedia delivery solutions consider user profile-related feedback only, due to the fact that the existing sensors and actuators are highly complex and heterogeneous. Therefore, network conditions (e.g. buffer-based network information measurement) and sensory-oriented characteristics (e.g. type/model, predefined priority, effect synchronization delay, etc.) should be taken into account in conjunction in order to best enhance user perceived QoE levels. Additionally, the extreme low latency requirements of mulsemmedia interactive services are a great challenge for the current network architecture. Moreover, the current XML-based mulsemmedia description standards (e.g. MPEG-V, MPEG-7, MPEG-4 BIFS²⁰, etc.) are unable to address the requirements of mulsemmedia delivery in the context of the heterogeneous 5G network architecture and services. Therefore, this paper introduces MulseDASH, a

novel adaptive mulsemmedia streaming solution over MPEG-DASH which supports the following advanced features beyond those offered by existing related solutions:

- *Hierarchical MulseMedia* Presentation Description
- *Multi-sensorial Content Encapsulation using JSON*²¹
- *Receiver Buffer-based* and *Multi-sensory-oriented* Adaptive Mulsemedia Streaming Scheme

Additionally, alongside MulseDASH design, this paper presents its deployment and testing in a real experimental test-bed. This demonstrates the benefit of using an adaptive mulsemmedia streaming scheme with existing sensors, actuators and equipment.

III. MULSEDASH FRAMEWORK DESIGN

Inspired by the adaptive multimedia streaming standard DASH, the proposed mulsemmedia streaming framework MulseDASH inherits its advantages in terms of media information organization and extends it by integrating new features. These features include support for multiple sensorial media components and integration of an adaptive delivery scheme that maintains smooth mulsemmedia streaming and sensory synchronization.

The MulseDASH framework design is illustrated in Figure 2 and involves two main components which intercommunicate via the Internet: MulseDASH Server and MulseDASH Client, which are introduced next.

A. MULSEDASH SERVER

1) HTTP Server and Data Storage

As shown in Figure 2, the MulseDASH Server facilitates the HTTP responses to the multimedia/mulsemmedia segment requests received from the client-side, integration of mulsemmedia and multimedia, handling of specific feedback from clients (e.g. network information, device characteristics, user experience feedback) and distribution management of multimedia/mulsemmedia data. The MulseDASH Server provides

¹⁹MPEG-7: <https://mpeg.chiariglione.org/standards/mpeg-7>

²⁰MPEG-4 Part 11: <https://mpeg.chiariglione.org/standards/mpeg-4/scene-description-and-application-engine>

²¹ECMA 404 Std., Dec. 2017, <http://www.ecma-international.org/publications/files/ECMA-ST/ECMA-404.pdf>

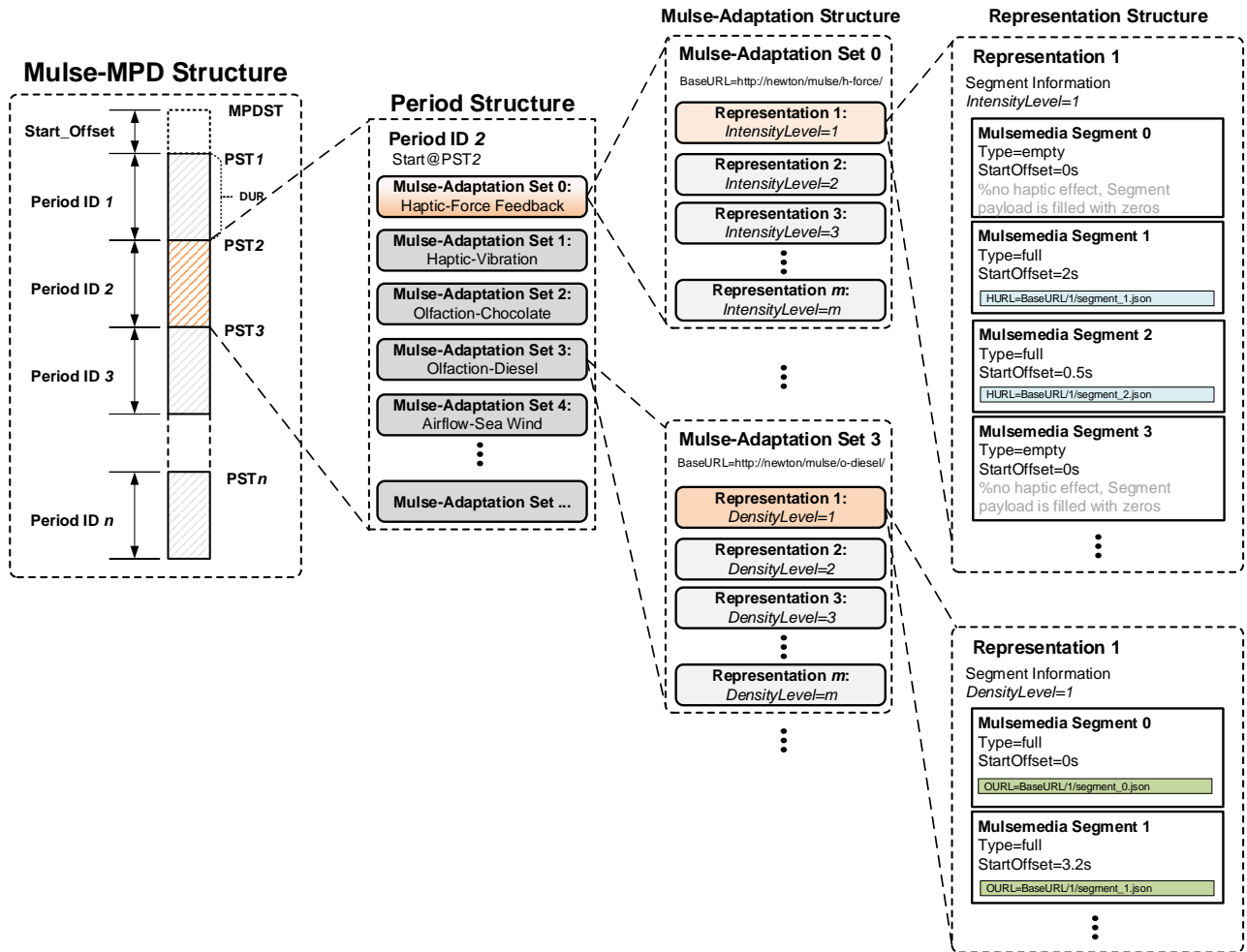


FIGURE 3: MulseMPD Structure

the interface between the mulsemedia/multimedia content and the MulseDASH clients over the HTTP-based transmission, which contains both MPEG-DASH metadata in Media Presentation Description (MPD) format and mulsemedia metadata, labeled Mulse-MPD, which will be described separately. After MulseDASH clients retrieve the MPEG-DASH MPD and Mulse-MPD from the server-side, based on their content and following the proposed MulseDASH algorithm, the clients will access the appropriate media segments from their distributed storage locations.

2) Hierarchical Mulse-MPD and Mulsemedia Segments

MulseDASH employs MPEG-DASH standard MPD structure for its audio-visual components. However, in order to accommodate its multi-sensorial media components, Mulse-MPD was introduced. Mulse-MPD extends MPEG-DASH and inherits the XML-based hierarchical architecture, providing a flexible and reliable content organization for different sensory effects, quality/intensity levels, adaptation sets and

play periods. However, different from the classic XML-based description of sensorial media segment information in MPEG-V and ADAMS/MPEG-7, Mulse-MPD employs a JSON-based encapsulation for mulsemedia segments. This is as comparative performance studies of JSON and classic XML [59], are in favour of JSON when large number of objects encoded in JSON and XML are transmitted. The hierarchical architecture of Mulse-MPD consists of **Periods**, **Mulse-Adaptation Sets**, **Representations** and **Mulsemedia Segments**, as illustrated in Figure 3. The Mulse-MPD structure includes a sequence of Periods, where a **Period** contains the top-level description of a sensory element including start time (i.e. **PST**) and duration (i.e. **DUR**). As the current multi-sensorial devices follow diverse production standards, sensory media codecs and inter-communication protocols, Mulse-MPD also accommodates this rich variability. Therefore, each **Period** contains several **Mulse-Adaptation Sets** that are associated with different adaptation groups defined by users or systems depending on the mulsemedia effect

types, device characteristics, user preferences, scenarios and so on.

For example, two different types of haptic effects are going to be rendered via two different haptic devices, respectively. One haptic effect is played on a vibrating mouse, whereas the other one is rendered via a haptic gaming vest produced by a different company, so two **Mulse-Adaptation Sets** are needed in Mulse-MPD. Also another example is shown in Figure 3 presenting how different types of odors (i.e. chocolate and diesel) are decoded by different **Mulse-Adaptation Sets**, respectively. Additionally, different numbers of **Mulse-Adaptation Sets** can be grouped and delivered based on different user preferences and scenarios designed in the audiovisual content. For instance, three different Mulse-Adaptation Sets with three different types of odors are configured in the **Mulse-MPD** for a movie scene designed with three types of odors. Additionally, a haptic **Mulse-Adaptation Set** is configured in **Mulse-MPD** if the user prefers the haptic effect or if the haptic device is available (i.e. some devices may be "offline" or with low battery levels).

In each **Mulse-Adaptation Set**, the same sensory content can be encoded and shown in different **Mulse-MPD Representations**. Different from the classic MPD Representation which is dependent on the bitrates or other video quality features, the **Mulse-MPD Representation** considers specific features of the sensory content. For instance, the haptic effect can have different **Mulse-MPD Representations** which differ in terms of their intensity levels. Similarly, the olfaction effect can have different intensity levels, as shown in Figure 3. Additionally, the different **Mulse-MPD Representations** can have different start offsets and play durations depending on user preferences. The **Representation** can enable adaptation of different sensory effects based on the *network conditions, device characteristics, pricing strategies* and so on. Considering a delay-sensitive case for example, the higher level **Mulse-MPD Representation** with high intensity of vibration effects can be downgraded to a lower **Mulse-MPD Representation** with lower intensity when the network delay becomes longer. Notably, the sensory content organization and configuration in **Mulse-MPD Representation** is much more flexible than those designed based on MPEG-V and MPEG-7. **Mulsemmedia Segments** contain the actual sensory media information and describe the type of sensory effect, start time, duration, effect intensity and other content-related data, stored in JSON format. Figure 3 also includes **WURL**, **HURL** and **OURL**, which are used to indicate the URL address of Wind effect, Haptic effect and Olfaction effect segments, respectively. Note that other sensorial effects can also be considered.

B. MULSEDASH CLIENT

Extending the classic MPEG-DASH client, the MulseDASH client retrieves both MPD and Mulse-MPD files from the MulseDASH server. An innovative MulseDASH Adaptive Streaming Algorithm is introduced which governs the manner in which audio-visual and multi-sensorial segments are

requested and presented to the viewers in order to achieve increased user QoE levels.

The MulseDASH client addresses two major concerns: adaptive mulsemmedia streaming and synchronisation between multimedia and mulsemmedia content. They are introduced next.

1) Adaptive Mulsemmedia Streaming Algorithm

Different from the conventional multimedia streaming, mulsemmedia streaming is not always continuous. Depending on the audio-visual scene design, at any moment in time, there may or may not be required a multi-sensorial effect playout. Therefore these effects are distributed discretely along the continuous audio-visual content timeline. According to the Mulse-MPD structure, the discretely distributed mulsemmedia content will be divided into segments and only some of them involve mulsemmedia content. This is indicated by their **Type** set to "full", or "empty" (i.e. indicating there is no mulsemmedia effect during this time slot and zero padding is included (see Figure 3). At design and as shown in Figure 4, the temporal length of mulsemmedia segments τ , expressed in milliseconds, is the same as that of audio-visual segments. Additionally, the **Mulse-MPD Segment** information contains the start time offset of mulsemmedia playout in each segment, namely **StartOffset** or $\Delta\tau^{start}$. Mulsemmedia streaming also considers the delay tolerances of different mulsemmedia effects. For instance, a user may prefer that the haptic effect is more important than other effects, and would benefit in their experience from haptic higher intensity or may want to make sure the odor diffuses earlier and thus increase the level of scene immersiveness.

A) Buffered-based Quantized Rate Adaptation Scheme (BQRAS): Consider that the types of different mulsemmedia content rendering on different mulsemmedia devices are denoted by $\mathcal{I} := \{1, 2, 3, \dots, i, \dots, I\}$ and $|\mathcal{I}| = I$. Each mulsemmedia streaming content is composed of segments, namely $\mathcal{N} := \{1, 2, 3, \dots, n, \dots, N\}$ and $|\mathcal{N}| = N$ shown in Figure 4. The downloading bitrate of the n_{th} segment of the i_{th} type of mulsemmedia effect is represented as $r_i(n) \in \mathcal{R}_i$ and the size in bytes and length in milliseconds of the segment are $s_i(n)$ and $\tau_i(n)$ (i.e. $r_i(n) = \frac{s_i(n)}{\tau_i(n)}$, where $\tilde{\tau}_i(n) \leq \tau_i(n)$ is the actual length of mulsemmedia effect playout), respectively. In general, the MulseDASH client initiates a HTTP-based request to the MulseDASH server for the n_{th} segment of the i_{th} type of mulsemmedia effect with the bitrate $r_i(n)$, and then the downloading starts immediately. Let $T_i^d(i)$ be the download duration. Then the next segment of the i_{th} type of mulsemmedia effect will start to be downloaded after time $T_i(n)$:

$$T_i(n) = \max \{T_i^d(n) \cdot a_i(n), T_i^p(n-1) \cdot b_i(n-1)\} \quad (1)$$

where $T_i^p(n)$ is the target duration to playback a segment, and the segment request indicator denoted by $a_i(n)$ which is 1 if the download request of the n_{th} segment is made and 0 otherwise. $b_i(n)$ is 1 if the next requested segment is "Full" and 0 otherwise. After its arrival, n_{th} segment is

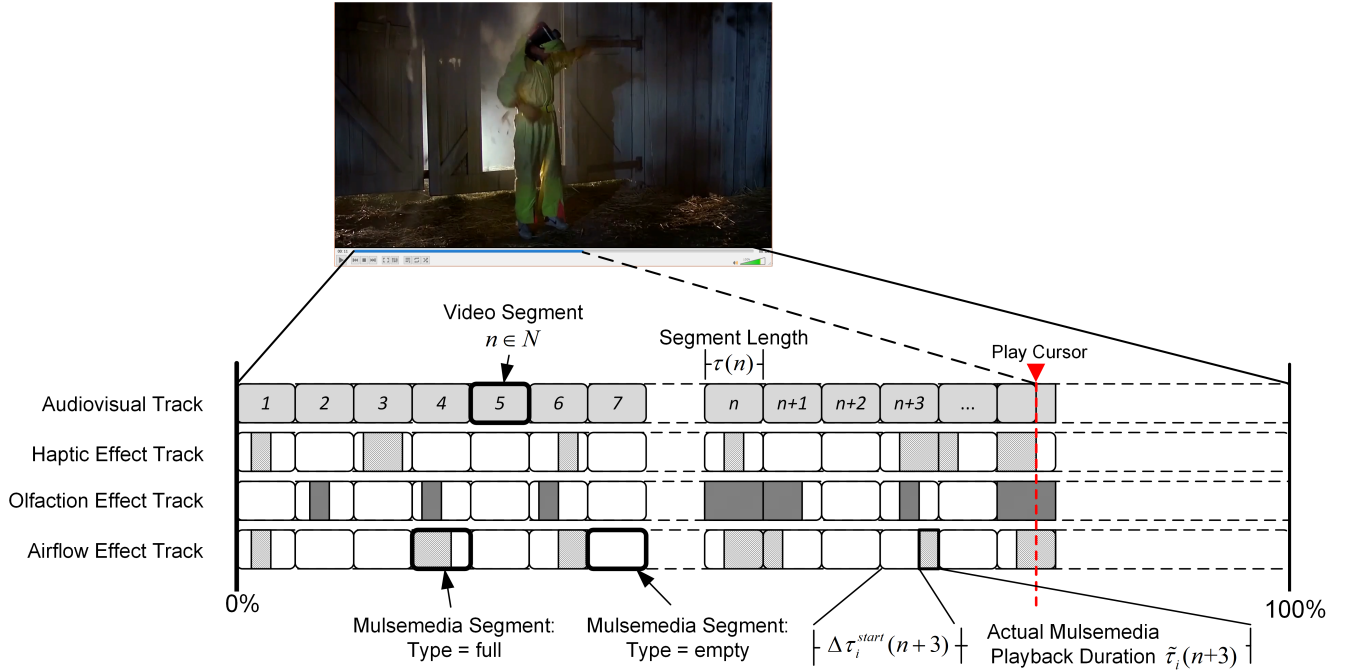


FIGURE 4: Timeline of Diverse Multimedia and Mulsemedia Segments

stored into the client-side playback buffer from where it is consumed by the mulsemmedia player. The consumption rate is the same rate as that of one mulsemmedia or audio-visual segment, hence the instantaneous buffer which stores all the mulsemmedia segments measured in video time (milliseconds) can be expressed as in (2):

$$B(n) = \max \left\{ B(n-1) - \sum_{i=1}^I T_i(n), 0 \right\} + \sum_{i=1}^I \tau_i(n) a_i(n) \quad (2)$$

The playback buffer level is strongly affected by the network conditions. Then the buffer level reading is noise-filtered to yield the smoothed buffer level $\bar{B}_i(n)$ depending on the moving average of the historical buffer levels, with the aim of removing the abnormal volatility of transmission delay measurement. The smoothed buffer level can be calculated as in (3):

$$\bar{B}(n) = (1 - \beta_n) \cdot B(n) + \beta_n \cdot \bar{B}(n-1) \quad (3)$$

where the coefficient weight β_n is exponentially denoted by:

$$\beta_n = 1 - \exp \left(- \left| \frac{\bar{B}(n-1) - \bar{B}(n-k-1)}{\bar{B}(n-k-1) - \bar{B}(n-2k-1)} \right| \right) \quad (4)$$

The β_n is set based on the difference between the past buffer level moving averages which are sampled at every k segments. If the difference is high, a larger coefficient weight is given to the current buffer level reading, otherwise a larger weight is associated with the older buffer level reading.

In this paper, a quantized method is proposed to control mulsemmedia adaptation. A different number q of mulsemmedia

effect deliveries will be adjusted based on the quantized buffer level $\bar{B}_i(n)$, according to the expression shown in (5):

$$q_{n+1} := Q_t(z) \quad (5)$$

where $z = \bar{B}_i(n) - \frac{1}{2} B_{max}$, and B_{max} is the maximum buffer level customized by the mulsemmedia player. The quantization method $Q_t(\cdot)$ uses the logistic function $Q_t(x) = \frac{1}{1 + \exp(-\gamma x)}$, $\gamma > 0$ to activate the shifting up or down of the number q of the adapted mulsemmedia effects:

$$q_{n+1} = \begin{cases} q_n + 1 & \text{if } Q_t(z) > Q_t(\frac{1}{2} B_{max}) \\ & \& 0 \leq q_n < I \\ q_n - 1 & \text{if } Q_t(z) < Q_t(B_{min}) \\ & \& 0 < q_n \leq I \\ q_n & \text{otherwise} \end{cases} \quad (6)$$

When $Q_t(z) > Q_t(\frac{1}{2} B_{max})$, the number of mulsemmedia effects affected will increase and q will decrease when the buffer level is lower than the minimum B_{min} . Otherwise q will remain the same.

B) Priority-aware Request Scheduling Scheme (PRSS):

Starting from the same number of mulsemmedia effects result of BQRAS adaptation, there are multiple combinations of mulsemmedia effects for possible selection. In Figure 2, the mulsemmedia player also cooperates with multiple mulsemmedia devices which require different levels of tolerance to the network conditions. In order to solve both effect selection and network tolerance issues, a priority-aware request scheduling scheme (PRSS) is used to determine the best adapted combination set of the mulsemmedia effects triggered by the different mulsemmedia effects based on network performance, pre-

defined priorities, segment types and effect of mulsemmedia payout *StartOffset* times.

In this context, PRSS selects the adapted combination subset \mathcal{I}_{sub}^* of mulsemmedia effects with the highest priority (i.e. resulted from summation of all the next requested mulsemmedia segments belonging to the set) as in (7):

$$\mathcal{I}_{sub}^* = \arg \max_{\substack{\mathcal{I}_{sub}^{q_{n+1}} \subseteq [\mathcal{I}]^{q_{n+1}} \\ \text{s.t. } i \in \mathcal{I}_{sub}^{q_{n+1}}}} \sum_i p_i a_i(n+1) b_i(n+1) \hat{T} h_i \Delta \hat{\tau}_i^{start} \quad (7)$$

where $[\mathcal{I}]^{q_{n+1}}$ denotes the collection of all subsets of the mulsemmedia effect type set \mathcal{I} of size q_{n+1} , and the $\mathcal{I}_{sub}^{q_{n+1}}$ is the subset of the collection. $p_i \in [0, 1]$ is the pre-defined priority for the users. The normalized network throughput is $\hat{T} h_i = 2 \arctan(\frac{r_i(n) * \hat{\tau}_i(n)}{T_i^d(n)}) / \pi$, $\hat{T} h_i \in [0, 1]$, and the normalized **StartOffset** value is $\Delta \hat{\tau}_i^{start} = \frac{\tau_i(n+1) - \Delta \tau_i^{start}(n+1)}{\tau_i(n+1)}$, $\Delta \hat{\tau}_i^{start} \in [0, 1]$.

Considering the two stages and the proposed BQRAS and PRSS, the proposed MulseDASH adaptation algorithm is also composed of two parts which are described in Algorithm 1.

Algorithm 1: Mulsemmedia Adaptation and Scheduling Algorithm

- 1 initialization: Mulsemmedia Effect Set \mathcal{I} , and \mathcal{N}
Segments of each mulsemmedia effect;
 - 2 //Part I: Buffered-based Quantized
 - 3 //Rate Adaptation Scheme
 - 4 Calculate $T_i(n)$ by using (1);
 - 5 Calculate $B(n)$ by using (2);
 - 6 Update $\bar{B}(n)$ by using (3) and (4);
 - 7 Calculate $Q_i(z)$ and select the next adapted quantized number q of mulsemmedia effects by using (5) and (6);
 - 8 //Part II: Priority-aware Request
 - 9 //Scheduling Scheme
 - 10 **foreach** $\mathcal{I}_{sub}^{q_{n+1}} \subseteq \mathcal{I}_{sub}^{q_{n+1}}$ **do**
 - 11 **foreach** $i \in \mathcal{I}_{sub}^{q_{n+1}}$ **do**
 - 12 Calculate the priority of each mulsemmedia effect for next requested segment based on (7);
 - 13 **end**
 - 14 **end**
 - 15 The next requested mulsemmedia segments within the combination subset \mathcal{I}_{sub}^* with the highest priority are scheduled to be downloaded based on (7);
-

2) Synchronization between Multimedia and Mulsemmedia Content

There is evidence [60] that the response time of tactile stimuli for individuals was 28% and 34% shorter than those for auditory and visual stimuli, respectively. Moreover, our previous research [61] and [20], on mulsemmedia synchronization has shown that the acceptable skews for tactile/haptic effects (i.e. $[0, 1]$ seconds) were much “narrow” than those of airflow (i.e.

$[-5, 3]$ seconds) and olfaction (i.e. $[-7.5, 10]$ seconds), respectively. This suggests the “inter-stream” synchronization involving haptic/tactile effect is much more sensitive than that of other effects in terms of user QoE. As a matter of fact, real mulsemmedia and multimedia devices are affected by many encoding, decoding and playback issues which may cause varying “inter-stream” playback delay and lower viewer QoE. This paper focuses on introducing a novel multimedia-mulsemmedia content synchronization algorithm to offer “smooth” user experience when availing remotely from MulseDASH content. The algorithm for mulsemmedia synchronization is presented in Algorithm 2.

Algorithm 2: Mulsemmedia Synchronization Algorithm

- 1 initialization: T_{c_base} , $T_i^{c_diff}$, $T_i^{c_start}$, d_i ;
 - 2 Get the base clock time: $T_{c_base}(n) = T_i^c(n)$ // the current playback timestamp of the n_{th} segment of the i_{th} type of mulsemmedia effect ;
 - 3 **foreach** $j \in \mathcal{I} \setminus i$ **do**
 - 4 $T_j^{c_diff}(n) = T_j^c(n) - T_{c_base}(n)$;
 - 5 $d_j(n) = T_j^{c_start}(n) - T_j^{c_start}(n-1)$;
 - 6 $d_{thres} = (d_j(n) > \delta ? \delta : d_j(n))$;
 - 7 **if** $|T_j^{c_diff}(n)| < \tau_j$ **then**
 - 8 **if** $T_j^{c_diff}(n) \leq -d_{thres}$ **then**
 - 9 $d_j(n) = 0$
 - 10 **else**
 - 11 **if** $T_j^{c_diff}(n) \geq d_{thres}$ **then**
 - 12 $d_j(n) = \lambda \cdot d_j(n)$
 - 13 **end**
 - 14 **end**
 - 15 **end**
 - 16 $T_j^{c_start}(n+1) = T_j^{c_start}(n) + d_j(n)$;
 - 17 Update next segment playback start time $T_j^{c_start}(n+1)$;
 - 18 **end**
-

Assuming that the i_{th} effect is selected as the base clock and the other effect $j_{th} \in \mathcal{I} \setminus i$ is about to synchronize. Due to its higher sensitivity, the haptic effect playing time is selected as the reference time clock (i.e. base clock). The time difference $T_j^{c_diff}(n)$ between the base clock and other effect’s playback clock is used to adjust the new play timestamp $T_j^{c_start}(n+1)$ of that effect. When $T_j^{c_diff}(n)$ is smaller than the length of a segment, if the $T_j^{c_diff}(n)$ is lower than the threshold d_{thres} (δ is the target threshold value defined by the system), it means that the payout rate of the effect is slower than that of the haptic effect and suggests the next segment be played immediately; otherwise the payout rate is faster than that of the haptic effect and the next segment will be played after a sleep moment $\lambda \cdot d_j(n)$ (i.e. λ is a synchronization factor which can be tuned). Following this solution, after some necessary adjustments, the mulsemmedia effect payout times will be synchronized.

In general, the video and audio components are already

synchronized. If the mulsemmedia is about to synchronize with the multimedia content, then next mulsemmedia and multimedia tracks can be synchronized following (8):

$$T^{c_start}(n+1) = T^{c_start}(n) + \max \{d_{mulse}(n), d_{av}(n)\} \quad (8)$$

where d_{mulse} is the final synchronized delay of mulsemmedia tracks calculated based on Algorithm 2, and d_{av} is calculated based on the audiovisual media synchronization mechanism which defined by general audiovisual player (i.e. it is not the point of this paper). The next played segments of mulsemmedia and multimedia will start after the same delay of $\max\{d_{mulse}, d_{av}\}$.

IV. MULSEDASH DEPLOYMENT AND EVALUATION

MulseDASH was deployed in a real-life system and was employed for delivery of multi-sensorial media including airflow, haptic and olfactive stimuli alongside the audiovisual components. This section describes the performance evaluation and subjective testing setup in a real network environment.

Due to the lack of mature mulsemmedia hardware support, the devices used for MulseDASH testing have been modified to fit the purpose.

A. HARDWARE SETUP

Due to the lack of mature mulsemmedia hardware support, the devices used for MulseDASH testing have been modified to fit the purpose. The following devices were modified and used, as illustrated in Figure 5:

- **Airflow Generator** consists of a Pulse-width Modulation fan and an Arduino board-based circuitry which was designed to enable control of the fan. A specific Arduino program code was written, compiled and run on the Arduino board to control the fan ON, OFF and its speed.
- **Haptic Mouse** is a SteelSeries Rival 700 professional gaming mouse which provides vibration effects during user interaction²². Rival 700 mouse was modified to generate tactile/vibration stimuli of different intensity, duration and frequency for users during the multimedia content playback.
- **Olfaction Diffuser Sbi4** was produced by Exhalia²³. The diffuser has 4 fans and when equipped with 4 aromatic cubes can distribute 4 different scents at different times. The support SDK toolkit help to control the diffusion in terms of scent type, delay, strength and density.

B. SOFTWARE SETUP

The MulseDASH player was developed as a Web-based mulsemmedia player application using JavaScript to enable deployment of MulseDASH and work with the modified

TABLE 1: Network Emulation Configuration

	$X \sim Poisson(\mu)$	$X \sim Exp(\lambda^{-1})$
Scenario 1	$\mu = 7$	$\lambda^{-1} = 2$
Scenario 2	$\mu = 5$	$\lambda^{-1} = 5$
Scenario 3	$\mu = 3$	$\lambda^{-1} = 8$

devices. The player follows the MulseDASH architecture illustrated in Figure 2, extends the MPEG DASH player developed part of the *dash.js* project²⁴ and integrates a dashboard-based management module. The mulsemmedia player and its dashboard-based management module support Mulse-MPD request and retrieval, network connection and receiver buffer information management, multimedia display, mulsemmedia device connectivity, multi-sensorial effect rendering and control and mulsemmedia and multimedia synchronization. A screenshot is shown in Figure 5.

C. MULSEDASH PERFORMANCE EVALUATION SETUP

In order to evaluate the performance of MulseDASH and especially its mulsemmedia and multimedia synchronization algorithm proposed in this paper, a real network environment was setup and mulsemmedia and multimedia delivery experiments were run. The network evaluation framework illustrated in Figure 6 consists of a Network Emulator, a Linux-based HTTP server and a HTTP client. MPEG-DASH and MulseDASH content was stored at the HTTP server, and *Node.js* was deployed. The proposed MulseDASH player was installed at the client and was working with the mulsemmedia devices. The Linux Traffic Control (TC) utility supported by Network Emulator (NetEm)²⁵ was deployed between server and client to emulate desired changes of the real network conditions. TC invokes the Linux kernel packet scheduler to control packet delay and loss and limits the outbound bottleneck at the server to simulate network load. Poisson and Exponential distributions (i.e. $X \sim Poisson(\mu)$ and $X \sim Exp(\lambda^{-1})$, μ and λ^{-1} are the means of the distributions, respectively) are utilized to model the outbound bandwidth limitation and inter-arrival time of the concurrent events in our network emulation, respectively. The configuration of the network emulation is illustrated in Table 1.

In the network emulation experiments, the server outbound total bandwidth is limited to 10 Mbps and three different scenarios are designed and run using Python scripts:

- **Scenario 1:** A Poisson distribution with $\mu = 7$ was used to model a high number of users who access the streaming service at the same time, and an Exponential distribution with $\lambda^{-1} = 2$ was employed to simulate the high inter-arrival frequency of the concurrent events;
- **Scenario 2:** A Poisson distribution with $\mu = 5$ and an Exponential distribution with $\lambda^{-1} = 5$ were used to configure a medium level of concurrent user numbers and inter-arrival frequency, respectively;

²²SteelSeries Rival 700: <https://steelseries.com/gaming-mice/rival-710>

²³Exhalia: <https://www.exhalia.com/us/produits/espaces-olfactifs/>

²⁴dash.js: <https://github.com/Dash-Industry-Forum/dash.js/wiki>

²⁵Linux NetEm: <https://wiki.linuxfoundation.org/networking/netem>

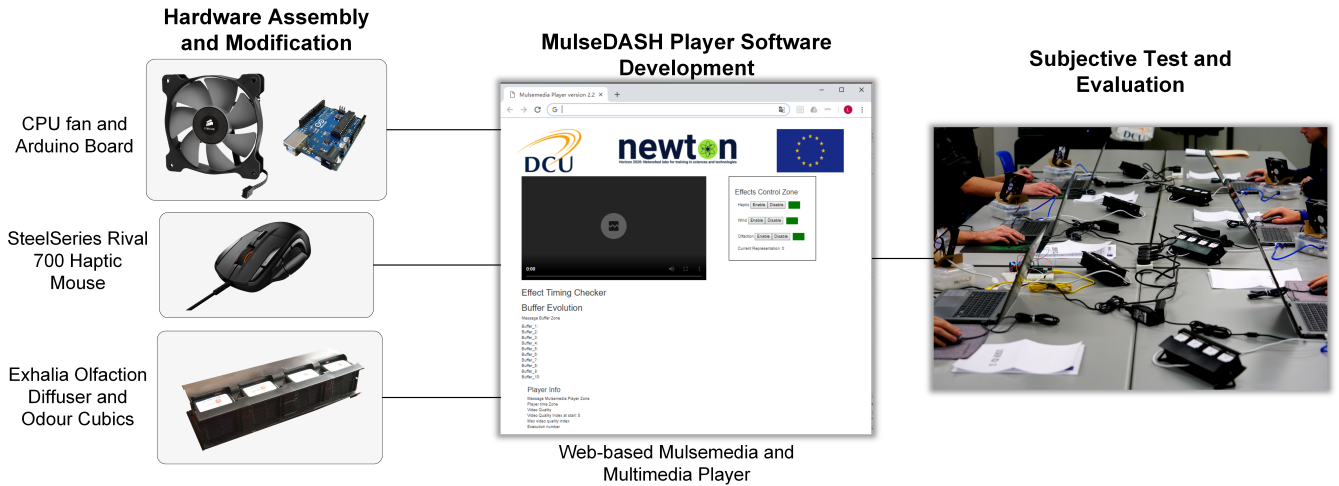


FIGURE 5: MulseDASH Implementations and Evaluations

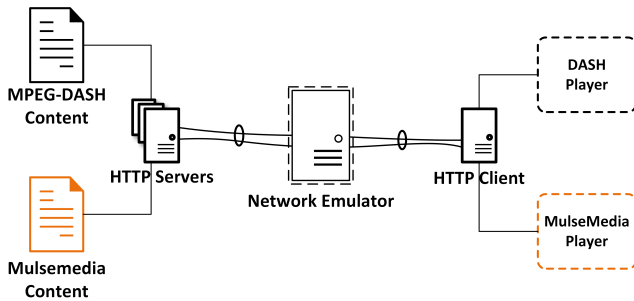


FIGURE 6: MulseDASH Network Emulation Setup

- **Scenario 3:** The means of the distributions, $\mu = 3$ and $\lambda = 8$ are used to model a low concurrent user number and low event inter-arrival frequency, respectively.

A high-quality video clip (i.e. 1080P, 3840kbps, 30fps) was cropped from the *Big Buck Bunny* animation movie (i.e. from 2:10 to 7:30) and was encoded with the 3 types of mulsemmedia effects in order to test the synchronization mechanism. The indexes i of the effects are as follows: haptic $i = 1$, olfaction $i = 2$ and airflow $i = 3$, and the priorities p_i of the effects in the adaptation mechanism are all set to 1. The 320-second long video clip contains content with high temporal and spatial encoding complexity [44] which causes high data rate variance of the adaptive multimedia streaming segments (see Figure 7c).

TABLE 2: Multimedia Content Configuration

Big Buck Bunny, 596 seconds				
Quality Level	Resolution	Bitrate	Frame rate	Format and Codec
Low	480P	2000kbps	24fps	MPEG-4 AVC/H.264, x264
Medium	720P	3000kbps	30fps	
High	1080P	3840kbps	30fps	

TABLE 3: Mulsemedia Content Configuration

Index	number of mulsemmedia effects	combination subsets of mulsemmedia effects
0	$ Z_{sub}^0 = 1$	no effect
1	$ Z_{sub}^1 = 3$	haptic
2		olfaction
3		wind
4	$ Z_{sub}^2 = 3$	haptic, Olfaction
5		haptic, wind
6		olfaction, wind
7	$ Z_{sub}^3 = 1$	haptic, olfaction, wind

D. MULSEDASH SUBJECTIVE TESTING SETUP

Subjective testing experiments are performed to assess the user perceived quality when employing MulseDASH-based adaptation. The video clips used in the subjective tests are pre-encoded from the animation movie, *Big Buck Bunny*, using different quality levels from high to low, as described in Table 2. Eight 30-seconds video clips with a wide range of temporal and spatial complexity are selected from the 596-second long movie. The eight video clips encoded at 3 different quality levels (i.e. generating 24 video clips) are compiled by MP4Box²⁶ in MPEG-DASH-formatted MPD files with 2-second long video segments.

According to the proposed MulseDASH adaptive mulsemmedia streaming algorithm introduced in Section III, different combination subsets of mulsemmedia effects are created for each video clip. Table 3 shows how each video clip is associated with 0, 1, 2 or 3 types of mulsemmedia effects, generating 8 different possible combinations. Considering there are 24 different quality video clips, there are 192 mulsemmedia and multimedia combination samples for the subjective testing.

²⁶MP4Box: <https://gpac.wp.imt.fr/mp4box/>

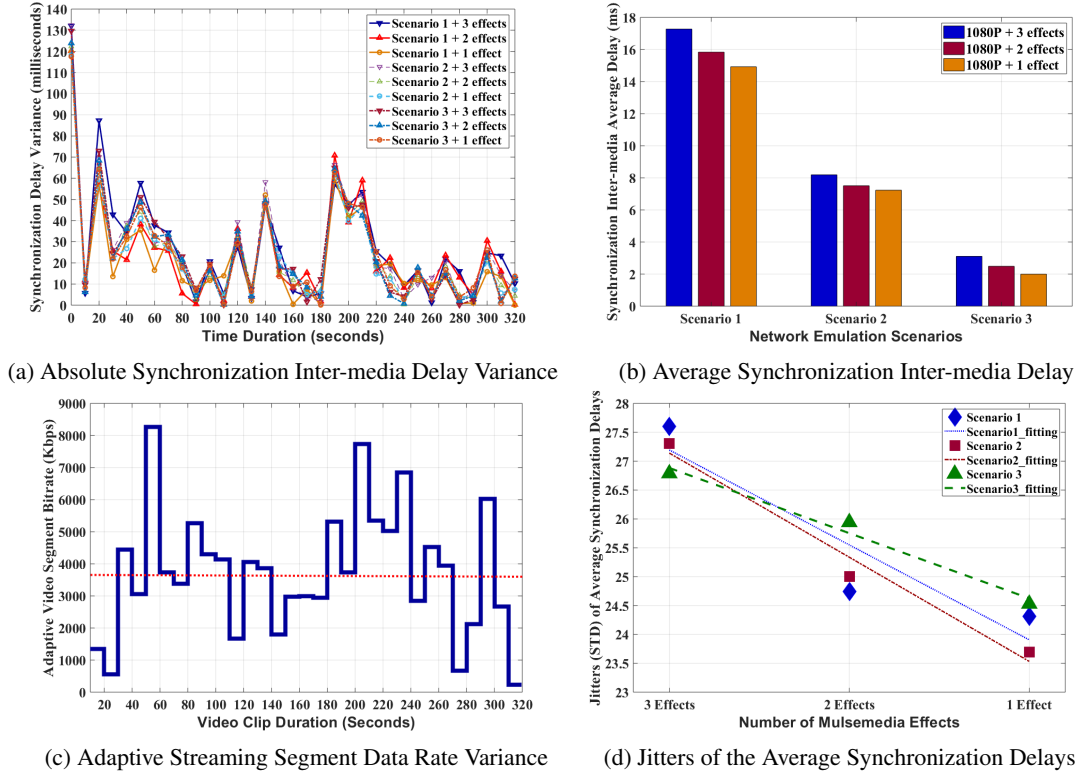


FIGURE 7: Results of Network Emulation for the Synchronization Mechanism

V. MULSEDASH TESTING - RESULT ANALYSIS

A. MULSEDASH SYNCHRONIZATION PERFORMANCE ANALYSIS

1) Impact of Adaptive Streaming on the Synchronization Inter-media Delay

From the "up-and-down" variances illustrated in Figure 7a and Figure 7c, both inter-media delays and adaptive segment bitrate levels are recorded. Notably, each change of the inter-media delay in a current time slot is triggered by a change of the adaptive bitrate level in the past slot, and a high level bitrate causes larger download time. At the beginning of the video clip (i.e. from 0s to 80s), the adaptive bitrate levels start from low to high, and then the inter-media delay varies from a very high level to a low level based on the mechanism proposed in (8). This mechanism helps reduce the gap of the inter-media delay between the mulsemmedia and multimedia download time. Another case of impact of adaptive multimedia bitrate levels on the inter-media delay is shown from 150s to 200s. In this period, the inter-media delay is boosted from a "flat" level (i.e. from 150s to 180s) to a "steep" variation (i.e. from 180s to 200s) due to the sudden change of adaptive segment bitrate. Moreover, the proposed synchronization mechanism helps to smoothen the changes of inter-media delay after 2 or 3 segments after 200s. Similar cases also happen from 240s to 320s as shown in Figure 7a and Figure 7c.

2) Impact of Mulsemedia Effects on the Synchronization Inter-media Delay

MulseDASH is tested with different numbers of mulsemmedia effects in the three different load network emulation scenarios. The time-varying inter-media delays in each scenario shown in Figure 7a indicate that the 3 mulsemmedia effects in the heaviest traffic environment cause high inter-media delay variation in comparison with those in the other tested scenarios. The average results of the inter-media delay suggest that the adaptive streaming with highest number of mulsemmedia effects causes higher delays than those with lower number of mulsemmedia effects. For example in Figure 7b, the average inter-media delay of the 3 mulsemmedia effects in Scenario 1 is higher by 15.7% than that for 1 effect in Scenario 1. The same comparisons reveals the delay increasing by 13.3% and 55.8% in Scenario 2 and Scenario 3, respectively. Additionally, the inter-media delay of the same numbers of mulsemmedia effects streaming tested in the lowest load scenario (i.e. Scenario 3) is on average reduced by 84.3% compared to those in the highest load scenario (i.e. Scenario 1). Moreover Figure 7d, which presents the results of average jitters in different scenarios, shows how the proposed synchronization mechanism for the mulsemmedia effects is faster to reduce the jitter when the number of mulsemmedia effects decreases in the high traffic environment (e.g. Scenarios 1 and 2). In general, the results of the whole network and synchronization experiments in Figure 7 demonstrate good performance for the inter-media delays and jitters (i.e. lower

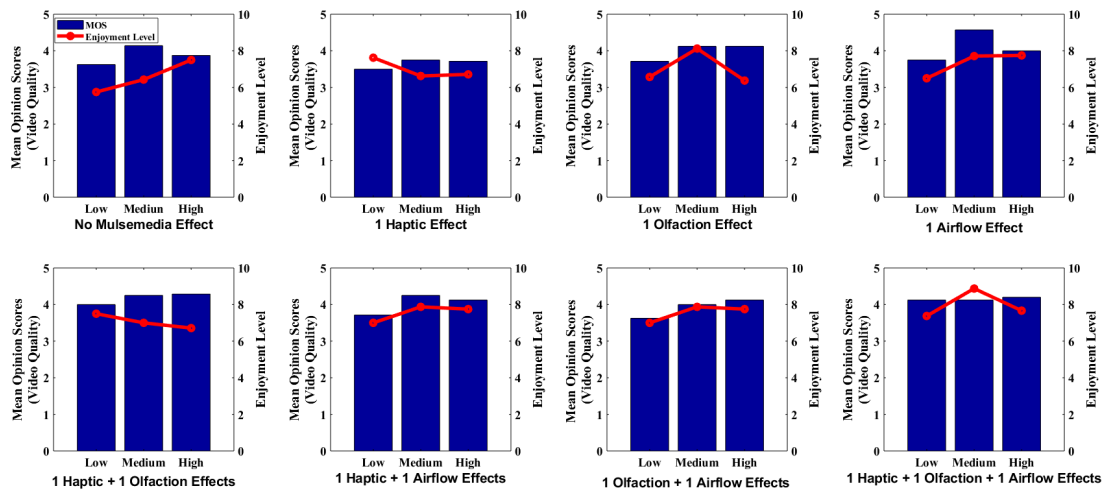


FIGURE 8: User Perceived Experience for Different Types of Mulsemedia Effects

than 20ms) due to the proposed MulseDASH synchronization mechanism.

B. MULSEDASH: USER PERCEIVED QUALITY ANALYSIS

During the subjective evaluation of MulseDASH, 24 participants were invited to experience MulseDASH adaptive streaming. Each testing section has involved 64 combinations of random quality video clips with random numbers of mulsemedia effects selected from the 192 mulsemedia and multimedia combination samples defined in Section IV. The subjective test results expressed in terms of Mean Opinion Score (MOS from 1 to 5) and predefined user enjoyment level (i.e. from 1 to 10) are retrieved and presented in Figure 8, Figure 9 and Figure 10, respectively. In this paper, the MOS is used to measure user perceived QoE of the video in the presence of mulsemedia effects, whereas the predefined user enjoyment level is supposed to grade user overall experience when subject to multimedia and mulsemedia effects.

1) Impact of Different Mulsemedia Effects on User QoE

Figure 8 presents the user experience quality grading for one case with no mulsemedia effects and 7 cases of different mulsemedia effect combinations.

1) No Mulsemedia Effect: In the experiments without mulsemedia effects, the enjoyment level grades have increased gradually with the increase in multimedia quality (i.e. from low to high). However, MOS has not followed a similar increase pattern. For example, the MOS for high quality videos is lower than that for medium quality clips. A possible reason is the fact that MOS is averaged from different video clips with different temporal and spatial complexity content which probably affects the subjective grading of MOS. A second reason can be MOS scale's limited granularity.

2) Haptic Effect: From Figure 8, it can be noted how the MOS results have increased with video quality levels, and the lowest video quality is graded higher with addition of the haptic effect, enhancing the enjoyment level for the whole user experience. Yet, the enhancement is limited for the medium and high quality level videos, which probably is caused by the simple haptic effects (e.g. vibrations) generated by the mouse.

3) Olfaction Effect: The MOS grades show that the olfaction effect improves the user perceived quality level of multimedia. In terms of enjoyment level, the grades vary most likely due to the fact that some users are not happy with olfaction effects during video playing. However, the average scores are higher by 2% compared to the tests without effects.

4) Airflow Effect: the airflow provides the best experience quality for the participants in the tests. Compared to the haptic effect, the MOS and enjoyment level of airflow are increased by 12.3% and 4.7%, respectively. Similarly, the airflow effect also increases with 2.9% the subjective video quality and provides with 4.2% higher user experience than those due to the olfaction effect.

2) Impact of Different Number of Mulsemedia Effects on User Experience

Depending on the analysis for the impact of different types of mulsemedia effects, it has been confirmed that the mulsemedia effects influence the user experience during video clip playout, even increase the perceived video quality level when watching a lower quality video clip. Moreover, the results of the experiments with multiple effect combinations, shown in Figure 8, provide more favorable evidences to demonstrate that the number of effects also influences users' grading. For example, the combination of one type of haptic and one type of airflow effects increases the user enjoyment level by 7.8% compared to the case with only one haptic effect is

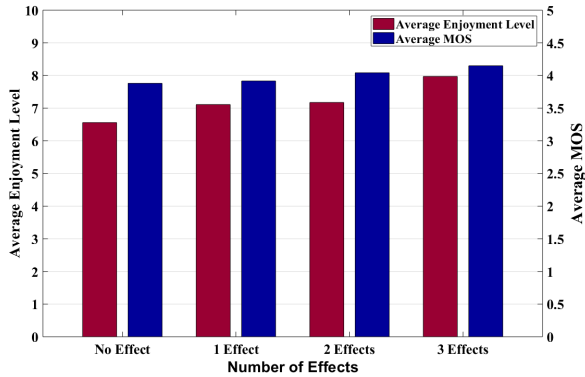


FIGURE 9: User Perceived Experience for Different Numbers of Mulsemmedia Effects

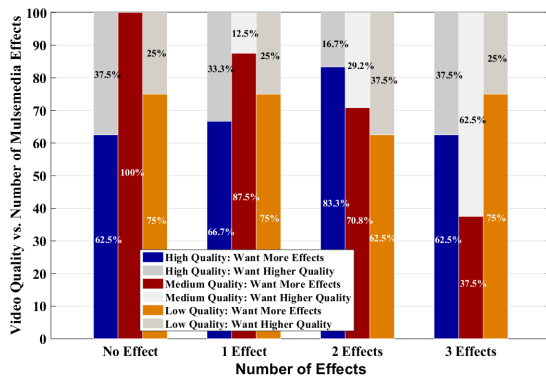


FIGURE 10: More Mulsemmedia Effects VS. Higher Video Quality Level

employed. The combination of haptic and olfaction effects also enhances the user perceived quality (i.e. MOS) and the whole enjoyment level by 9.6% and 1% compared to those of any one of them, respectively. However, the combination of olfaction and airflow does not have any improvement in terms of user experience. The potential reason is that the airflow would boost the diffusion and concentration of the odours which might irritate some of the participants who do not like the smell. Due to the random types and numbers of mulsemmedia effects that would be generated in the real world, the average results of MOS and enjoyment level for the different numbers of effects are presented in Figure 9. Both user perceived quality and enjoyment level are increased gradually with the number of effects growing from 0 to 3. For instance, the average increasing rate of MOS with the number of effects is 2.26%, and the average increasing rate of enjoyment level score is 6.80%. In total, all of the results shown in Figure 8 and Figure 9 indicate that the participants have mostly enjoyed when they are watching a video clip with 3 types of mulsemmedia effects, for which both MOS and enjoyment level score have reached the highest level.

3) Study of Trade-off between Mulsemmedia Effects and Video Quality

There is an interesting dilemma: what do the participants actually prefer to access: more mulsemmedia effects or higher video quality level? The trade-offs between employing mulsemmedia effects and improving video quality are displayed in Figure 10. Between 60% to 80% of the participants who are watching high quality and low quality videos, prefer adding more mulsemmedia effects. More mulsemmedia effects help those participants to improve their enjoyment level. The results show that the participants who are watching medium quality video clips have a good balance between the number of mulsemmedia effect and video quality. Those of them who prefer increasing the video quality level gradually when the number of mulsemmedia effects is also increasing.

VI. CONCLUSION AND FUTURE WORKS

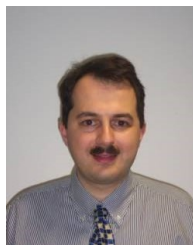
This paper has proposed MulseDASH, an innovative adaptive mulsemmedia streaming solution which is designed to improve user QoE levels. MulseDASH was tested and the network performance evaluation results present much lower inter-media delay results (i.e the average value $< 18ms$) against the skew requirement between different media provided in [21] [49] [50] [51]. The subjective tests for the different combinations of mulsemmedia effects show that the MulseDASH improves the user enjoyment level and perceived quality during the whole video playing. Moreover, more users prefer adding more effects while watching the video in a lower quality level. Future work, should include MulseDASH improvement and evaluation in wireless network environments which contain more complex heterogeneous conditions and higher latency requirement applied in the future Internet of Things or 5G scenarios.

REFERENCES

- [1] Cisco, "The Zettabyte Era: Trends and Analysis," Tech. Rep., Jun. 2018.
- [2] PTC, "The State of Industrial Augmented Reality 2017," Tech. Rep., 2018.
- [3] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, from 2017 to 2022," Tech. Rep., Feb. 2019.
- [4] M. L. Heilig, "Sensorama simulator," Patent US3 050 870A, Jan., 1961.
- [5] L. A. Rowe and R. Jain, "ACM SIGMM retreat report on future directions in multimedia research," ACM Transactions on Multimedia Computing, Communications, and Applications, vol. 1, no. 1, pp. 3–13, Feb. 2005.
- [6] G. Ghinea and O. Ademoye, "A user perspective of olfaction-enhanced mulsemmedia," in Proceedings of the International Conference on Management of Emergent Digital EcoSystems, ser. MEDES '10. New York, NY, USA: ACM, 2010, pp. 277–280.
- [7] G. Ghinea and O. A. Ademoye, "Olfaction-enhanced multimedia: perspectives and challenges," Multimedia Tools and Applications, vol. 55, no. 3, pp. 601–626, Dec 2011.
- [8] H. Matsukura, T. Yoneda, and H. Ishida, "Smelling screen: Development and evaluation of an olfactory display system for presenting a virtual odor source," IEEE Transactions on Visualization & Computer Graphics, vol. 19, no. 4, pp. 606–615, April 2013.
- [9] N. Murray, Y. Qiao, B. Lee, and G.-M. Muntean, "User-profile-based perceived olfactory and visual media synchronization," ACM Trans. Multimedia Comput. Commun. Appl., vol. 10, no. 1s, pp. 11:1–11:24, Jan. 2014.
- [10] H. Matsukura, T. Nihei, and H. Ishida, "Multi-sensorial field display: Presenting spatial distribution of airflow and odor," in 2011 IEEE Virtual Reality Conference, March 2011, pp. 119–122.

- [11] T. Nakano, Y. Yoshioka, and Y. Yanagida, "Effects of wind source configuration of wind displays on property of wind direction perception width of wind velocity distribution and accuracy of wind source alignment," 2014.
- [12] J. Cha, M. Eid, and A. E. Saddik, "Touchable 3d video system," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 5, no. 4, pp. 29:1–29:25, Nov. 2009.
- [13] M. Eid, J. Cha, and A. E. Saddik, "Hugme: A haptic videoconferencing system for interpersonal communication," in 2008 IEEE Conference on Virtual Environments, Human-Computer Interfaces and Measurement Systems, July 2008, pp. 5–9.
- [14] A. Marzo, A. Ghobrial, L. Cox, M. Caleap, A. Croxford, and B. W. Drinkwater, "Realization of compact tractor beams using acoustic delay-lines," *Applied Physics Letters*, vol. 110, no. 1, p. 014102, Jan. 2017.
- [15] T. Narumi, T. Kajinami, S. Nishizaka, T. Tanikawa, and M. Hirose, "Pseudo-gustatory display system based on cross-modal integration of vision, olfaction and gustation," in 2011 IEEE Virtual Reality Conference, March 2011, pp. 127–130.
- [16] M. Murer, I. Aslan, and M. Tscheligi, "Lollo: Exploring taste as playful modality," in Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction, ser. TEI '13. New York, NY, USA: ACM, 2013, pp. 299–302.
- [17] Z. Yuan, G. Ghinea, and G. M. Muntean, "Beyond multimedia adaptation: Quality of experience-aware multi-sensorial media delivery," *IEEE Transactions on Multimedia*, vol. 17, no. 1, pp. 104–117, Jan. 2015.
- [18] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis, "5g-enabled tactile internet," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 3, pp. 460–473, March 2016.
- [19] G. P. Fettweis, "The tactile internet: Applications and challenges," *IEEE Vehicular Technology Magazine*, vol. 9, no. 1, pp. 64–70, March 2014.
- [20] Z. Yuan, T. Bi, G. M. Muntean, and G. Ghinea, "Perceived synchronization of mulsemmedia services," *IEEE Transactions on Multimedia*, vol. 17, no. 7, pp. 957–966, July 2015.
- [21] N. Murray, B. Lee, Y. Qiao, and G.-M. Muntean, "Multiple-scent enhanced multimedia synchronization," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 11, no. 1s, pp. 12:1–12:28, Oct. 2014.
- [22] G.-M. Muntean, P. Perry, and L. Murphy, "Objective and subjective evaluation of qoas video streaming over broadband networks," *IEEE Transactions on Network and Service Management*, vol. 2, no. 1, pp. 19–28, Nov. 2005.
- [23] B. Ciubotaru, G. Ghinea, and G. Muntean, "Subjective assessment of region of interest-aware adaptive multimedia streaming quality," *IEEE Transactions on Broadcasting*, vol. 60, no. 1, pp. 50–60, March 2014.
- [24] Z. Jiang, C. Xu, J. Guan, Y. Liu, and G. Muntean, "Stochastic analysis of dash-based video service in high-speed railway networks," *IEEE Transactions on Multimedia*, pp. 1–1, 2018.
- [25] O. A. Ademoye, N. Murray, G.-M. Muntean, and G. Ghinea, "Audio masking effect on inter-component skews in olfaction-enhanced multimedia presentations," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 12, no. 4, pp. 51:1–51:14, Aug. 2016.
- [26] R. Fredriksson, M. C. Lagerstrom, L. Lundin, and H. B. Schioth, "The g-protein-coupled receptors in the human genome form five main families. phylogenetic analysis, paralogon groups, and fingerprints." *Molecular Pharmacology*, vol. 63, no. 6, pp. 1256–1272, 2003.
- [27] N. Ranasinghe, A. D. Cheok, O. N. N. Fernando, H. Nii, and G. Ponampalam, "Electronic taste stimulation," in Proceedings of the 13th International Conference on Ubiquitous Computing, ser. UbiComp '11. New York, NY, USA: ACM, 2011, pp. 561–562.
- [28] M. Murer, I. Aslan, and M. Tscheligi, "Lollo: Exploring taste as playful modality," in Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction, ser. TEI '13. New York, NY, USA: ACM, 2013, pp. 299–302.
- [29] M. A. Srinivasan and C. Basdogan, "Haptics in virtual environments: Taxonomy, research status, and challenges," *Computers & Graphics*, vol. 21, no. 4, pp. 393–404, 1997.
- [30] T. Yang, D. Xie, Z. Li, and H. Zhu, "Recent advances in wearable tactile sensors: Materials, sensing mechanisms, and device performance," *Materials Science and Engineering: R: Reports*, vol. 115, pp. 1–37, 2017.
- [31] Z. F. Quek, S. B. Schorr, I. Nisky, W. R. Provancher, and A. M. Okamura, "Sensory substitution and augmentation using 3-degree-of-freedom skin deformation feedback," *IEEE Transactions on Haptics*, vol. 8, no. 2, pp. 209–221, 2015.
- [32] E. E. Hoggan, S. A. Brewster, and J. Johnston, "Investigating the effectiveness of tactile feedback for mobile touchscreens," pp. 1573–1582, 2008.
- [33] A. J. Silva, O. A. D. Ramirez, V. P. Vega, and J. P. O. Oliver, "Phantom omni haptic device: Kinematic and manipulability," in 2009 Electronics, Robotics and Automotive Mechanics Conference (CERMA), Sept 2009, pp. 193–198.
- [34] Z. Ma and P. Bentzvi, "Design and optimization of a five-finger haptic glove mechanism," *Journal of Mechanisms and Robotics*, vol. 7, no. 4, p. 041008, 2015.
- [35] O. Baus and S. Bouchard, "Exposure to an unpleasant odour increases the sense of presence in virtual reality," *Virtual Reality*, vol. 21, no. 2, pp. 59–74, Jun 2017.
- [36] F. Rock, N. Barsan, and U. Weimar, "Electronic nose: current status and future trends," *Chemical Reviews*, vol. 108, no. 2, pp. 705–725, 2008.
- [37] S. K. Ryman, N. D. Bruce, and M. S. Freund, "Temporal responses of chemically diverse sensor arrays for machine olfaction using artificial intelligence," *Sensors and Actuators B: Chemical*, vol. 231, pp. 666–674, 2016.
- [38] K. Hashimoto and T. Nakamoto, "Tiny olfactory display using surface acoustic wave device and micropumps for wearable applications," *IEEE Sensors Journal*, vol. 16, no. 12, pp. 4974–4980, June 2016.
- [39] H. Matsukura, T. Nihei, and H. Ishida, "Multi-sensorial field display: Presenting spatial distribution of airflow and odor," in 2011 IEEE Virtual Reality Conference, March 2011, pp. 119–122.
- [40] S. D. Kulkarni, M. A. Minor, M. W. Deaver, and E. R. Pardyjak, "Output feedback control of wind display in a virtual environment," in Proceedings 2007 IEEE International Conference on Robotics and Automation, April 2007, pp. 832–839.
- [41] G.-M. Muntean and N. Cranley, "Resource efficient quality-oriented wireless broadcasting of adaptive multimedia content," *IEEE Transactions on Broadcasting*, vol. 53, no. 1, pp. 362–368, March 2007.
- [42] M. Kennedy, H. Venkataraman, and G. M. Muntean, "Dynamic stream control for energy efficient video streaming," in 2011 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), June 2011, pp. 1–6.
- [43] L. Zou, R. Trestian, and G. M. Muntean, "A utility-based priority scheduling scheme for multimedia delivery over lte networks," in 2013 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), June 2013, pp. 1–7.
- [44] —, "E3doas: Balancing qoe and energy-saving for multi-device adaptation in future mobile wireless video delivery," *IEEE Transactions on Broadcasting*, vol. 64, no. 1, pp. 26–40, March 2018.
- [45] A. Moldovan and C. H. Muntean, "Subjective assessment of bitdetect - a mechanism for energy-aware multimedia content adaptation," *IEEE Transactions on Broadcasting*, vol. 58, no. 3, pp. 480–492, Sep. 2012.
- [46] G. Gao, Y. Wen, and J. Cai, "vcache: Supporting cost-efficient adaptive bitrate streaming," *IEEE MultiMedia*, vol. 24, no. 3, pp. 19–27, 2017.
- [47] J. Kim, H. Kim, B. K. Tay, M. Muniyandi, M. A. Srinivasan, J. Jordan, J. Mortensen, M. Oliveira, and M. Slater, "Transatlantic touch: A study of haptic collaboration over long distance," *Presence*, vol. 13, no. 3, pp. 328–337, June 2004.
- [48] C. Basdogan, C.-H. Ho, M. A. Srinivasan, and M. Slater, "An experimental study on the role of touch in shared virtual environments," *ACM Trans. Comput.-Hum. Interact.*, vol. 7, no. 4, pp. 443–460, Dec. 2000.
- [49] R. Iglesias, S. Casado, T. Gutiérrez, A. García-Alonso, W. Yu, and A. Marshall, "Simultaneous remote haptic collaboration for assembling tasks," *Multimedia Syst.*, vol. 13, no. 4, pp. 263–274, Jan. 2008.
- [50] T. Watanabe, Y. Ishibashi, N. Fukushima, and S. Sugawara, "Dynamic switching control of haptic transmission direction in remote control system," in 2010 IEEE Haptics Symposium, March 2010, pp. 207–213.
- [51] M. Eid, J. Cha, and A. E. Saddik, "Admux: An adaptive multiplexer for haptic-audio-visual data communication," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 1, pp. 21–31, Jan 2011.
- [52] B. A. Radvansky and D. A. Dombeck, "An olfactory virtual reality system for mice," *Nature Communications*, vol. 9, no. 1, Dec. 2018.
- [53] E. Richard, A. Tijou, P. Richard, and J.-L. Ferrier, "Multi-modal virtual environments for education with haptic and olfactory feedback," *Virtual Reality*, vol. 10, no. 3, pp. 207–225, Dec 2006.
- [54] L. Zou, I. Tal, A. Covaci, E. Ibarrola, G. Ghinea, and G.-M. Muntean, "Can multisensorial media improve learner experience?" in Proceedings of the 8th ACM on Multimedia Systems Conference, ser. MMSys'17. New York, NY, USA: ACM, 2017, pp. 315–320.
- [55] H. Matsukura, T. Yoneda, and H. Ishida, "Smelling screen: Development and evaluation of an olfactory display system for presenting a virtual odor source," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 4, pp. 606–615, 2013.

- [56] F. Danieau, A. Lecuyer, P. Guillotel, J. Fleureau, N. Mollet, and M. Christie, "Enhancing audiovisual experience with haptic feedback: A survey on hav," *IEEE Transactions on Haptics*, vol. 6, no. 2, pp. 193–205, April–June 2013.
- [57] ISO/MPEG, "ISO/IEC 23005 Information technology—Media context and control Part 1-5," Tech. Rep., Feb. 2009.
- [58] K. Yoon, B. Choi, E. S. Lee, and T. B. Lim, "4-d broadcasting with mpeg-v," in 2010 IEEE International Workshop on Multimedia Signal Processing, Oct 2010, pp. 257–262.
- [59] N. Nurseitov, M. Paulson, R. Reynolds, and C. Izurieta, "Comparison of JSON and XML Data Interchange Formats: A Case Study," *Computer Applications in Industry and Engineering*, pp. 157–162, 2009.
- [60] A. Chan and A. Ng, "Finger response times to visual, auditory and tactile modality stimuli," in *Lecture Notes in Engineering and Computer Science*, vol. 2196. Newswood Limited, 2012, pp. 1449–1454.
- [61] N. Murray, Y. Qiao, B. Lee, A. K. Karunakar, and G.-M. Muntean, "Subjective evaluation of olfactory and visual media synchronization," in *Proceedings of the 4th ACM Multimedia Systems Conference*, ser. MMSys '13. New York, NY, USA: ACM, 2013, pp. 162–171.



GABRIEL-MIRO MUNTEAN (S'02-M'04-SM'17) received the B.Eng. and M.Sc. degrees from Politehnica University of Timisoara, Romania in 1996 and 1997, respectively, and the Ph.D. degree from the School of Electronic Engineering, Dublin City University (DCU), Ireland, in 2003 for his research on quality-oriented adaptive multimedia streaming over wired networks. He is currently an Associate Professor with the School of Electronic Engineering DCU and co-Director of the Performance Engineering Laboratory DCU. He has published over 300 papers in prestigious international journals and conferences, has authored three books and 16 book chapters, and has edited five other books. His research interests include quality-oriented and performance related issues of adaptive multimedia delivery, performance of wired and wireless communications, energy-aware networking, and personalized technology-enhanced learning. Dr. Muntean is an Associate Editor of the *IEEE Transactions on Broadcasting*, the *Multimedia Communications Area Editor* of the *IEEE Communication Surveys and Tutorials*, and a reviewer for other important international journals, conferences, and funding agencies. He is a senior member of IEEE and IEEE Broadcast Technology Society and coordinates the EU Horizon 2020-funded NEWTON project (<http://newtonproject.eu>).

...



LONGHAO ZOU (S'12) received the B.Eng and Ph.D degrees from Beijing University of Posts and Telecommunications (BUPT), Beijing, China and Dublin City University (DCU), Ireland in 2011 and 2016, respectively. He was a postdoctoral researcher with the EU Horizon 2020 NEWTON Project at DCU. Now he is with College of Computer Science & Software Engineering, Shenzhen University, Shenzhen, China. His research interests include mobile and wireless communications,

adaptive multimedia and mulsemmedia streaming, resource allocation and user quality of experience.



TING BI (S'12-M'18) received the B.Eng. degree in Software Engineering from Wuhan University, China in 2010, and received the M.Eng. and PhD degrees in Telecommunications from Dublin City University, Ireland in 2011 and 2017, respectively. He is currently a Postdoctoral Researcher with the Performance Engineering Laboratory and the Insight Centre for Data Analytics in the School of Electronic Engineering at Dublin City University. He has published in prestigious international

conferences and journals. His research interests include mobile and wireless communications, multimedia and multi-sensory media streaming over wireless access networks, user quality of experience, handover and network selection strategies, and energy saving for mobile devices. He is a member of IEEE, ACM, RDA, IEEE Young Professionals, IEEE Communications Society and IEEE Broadcast Technology Society.