A DASH-based Mulsemedia Adaptive Delivery Solution

Ting Bi
Dublin City University
Dublin, Ireland
ting.bi2@mail.dcu.ie

Antoine Pichon
Dublin City University
Dublin, Ireland
antoine.pichon2@mail.dcu.ie

Longhao Zou
Shenzhen University
Shenzhen, China
zoulonghao@gmail.com

Shengyang Chen
Dublin City University
Dublin, Ireland
shengyang.chen5@mail.dcu.ie

Gheorghita Ghinea
Brunel University
London, United Kingdom
george.ghinea@brunel.ac.uk

Gabriel-Miro Muntean
Dublin City University
Dublin, Ireland
gabriel.muntean@dcu.ie

ABSTRACT

The Dynamic Adaptive Streaming over HTTP (DASH) is a standard designed for adaptive multimedia delivery. This paper describes an adaptive multimedia and mulsemedia delivery system which employs an extension of DASH to support multiple sensorial media. Testing using a real-life test-bed shows how user satisfaction increases when employing the proposed solution which exposes users to multi-sensorial content. The motivation behind this work is the creation the first solution for remote adaptive multi-sensory media (mulsemedia) content delivery in order to enhance user satisfaction and increase their quality of experience. A DASH-based mulsemedia delivery system (DASH-MS) was built and tested. This work focuses on a fundamental issue for remote multi-sensory media systems: synchronization and shows how the proposed DASH-MS addresses it. Results of real life subjective testing indicate that the average levels of user satisfaction when exposed to content with the mulsemedia effects are better or at least equal to those when no such effects were included. Additionally, the highest value of the level of satisfaction corresponds to the case when these effects were correctly synchronized. Finally, users seem to have preferred effects in advance than late effects.

CCS CONCEPTS

• **Human-centered computing → Graphical user interfaces**; **User interface design**;

KEYWORDS

DASH, Mulsemedia, Adaptation, Synchronization

ACM Reference Format:

Ting Bi, Antoine Pichon, Longhao Zou, Shengyang Chen, Gheorghita Ghinea, and Gabriel-Miro Muntean. 2018. A DASH-based Mulsemedia Adaptive Delivery Solution. In *MMVE'18: 10th International Workshop on Immersive Mixed and Virtual Environment Systems, June 12–15, 2018, Amsterdam, Netherlands.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3210438.3210443

MMVE'18, June 12–15, 2018, Amsterdam, Netherlands

ACM ISBN 978-1-4503-5771-5/18/06…$15.00
https://doi.org/10.1145/3210438.3210443

1 INTRODUCTION

Multimedia applications are omnipresent in our day-to-day activities. However, in terms of its components, multimedia is mostly composed of audio and video, and very seldom text, targeting two human senses only. Mulsemedia refers to the combination of multimedia objects (video and audio) and components targeting other human senses such as touch, smell, and taste. Mulsemedia delivery refers to multi-sensory media distribution and its related aspects such as synchronization between multimedia content and multisensorial effects which are essential for the success of the future mulsemedia communication systems. [16].

Currently mulsemedia distribution is performed mostly locally. The few networked-enabled solutions proposed employ a one-solution-fits-all approach or require specific protocols for content delivery and do not support the widely used HTTP protocol. At the same time, there are important works which focus on remote delivery of adaptive multimedia content. Some of these employ MPEG-DASH, a recent standard for dynamic adaptive multimedia delivery over HTTP. The latest browsers (e.g. Chrome, Firefox, Safari, Edge) also support DASH-based content delivery. However, their design targets multimedia, and mulsemedia content delivery is not supported.

In this context, in order to create an immersive multimedia environment that stimulates user experience with multiple media elements engaging three or more human senses and allows content adjustment in existing delivery networks, a novel DASH-based mulsemedia delivery solution (DASH-MS) is proposed.

This paper describes the proposed DASH-MS and presents the design and testing of a DASH-based multimedia and mulsemedia delivery system which deploys DASH-MS. The article also focuses on a fundamental issue for remote multi-sensory media solutions: synchronization and shows how the proposed DASH-MS addresses it. As very few studies have investigated the user perceived satisfaction associated with the use of mulsemedia, this paper also studies user perceived satisfaction in relation to multiple media components synchronization.

This paper is organized as follows. Section II introduces the state-of-the-art related works in mulsemedia communications. Section III presents the proposed DASH-based mulsemedia delivery system.

Section IV describes the multi-sensory effect synchronization mechanism and section V analyzes the subjective test results. Finally, Section VI concludes the paper.

2 RELATED WORKS

The term *mulsemedia* was first coined in 2010 [4] and refers to media elements engaging three or more human senses. Current mulsemedia applications can use two standards designed by the Moving Picture Experts Group (MPEG): MPEG-7 [8] and MPEG-V [15]. MPEG-7 is designed to describe multimedia content data, whereas MPEG-V is designed to interface with virtual worlds.

Adaptation techniques have been widely exploited in content delivery systems, especially for the video delivery over the Internet. Moreover, various industry and academic research teams proposed innovative solutions to address different problems which occur during video delivery, while also satisfying diverse requirements in terms of transmission environment and users' quality of experience. In this context, the solutions proposed were known as adaptive video delivery schemes and, depending on the different requirements for video streaming, they are classified in the literature into five categories [1, 3, 12, 14, 18, 20]: 1) Content-aware adaptive solutions relate to content-aware encoding and content classification algorithms; 2) Quality-oriented/Quality of Service (QoS)-aware solutions consider the transmission channel conditions and QoS requirements in the adaptation process; 3) User Experience-aware solutions mainly consider the users' Quality of Experience (QoE) levels when adapting the quality of the delivered video stream based on clients' feedback; 4) Energy-efficiency/saving solutions adopt green algorithms for saving power at the sender or at the client during video delivery; 5) Other schemes that employ other solutions for video adaptation.

Dynamic Adaptive Streaming over HTTP (DASH), referred to as MPEG-DASH, is a standard that enables adaptive bitrate video delivery over the Internet. The media content is partitioned into one or more segments and is delivered from conventional HTTP web servers to the clients using HTTP [13]. However MPEG-DASH does not indicate the actual adaptation mechanism that is employed during the media content delivery. This adaptation mechanism is designed separately and many solutions have already been proposed, including Quality-Oriented Adaptive Scheme (QOAS) [10, 11], Prioritized Adaptive Mechanism (PAM) [19], Region of Interests Adaptive Scheme (ROIAS) [2] and Enhanced Energy-aware Device Oriented Adaptive Scheme (E2DOAS) [21]. Additionally MPEG-DASH is designed for adaptive multimedia delivery, and does not consider adaptive mulsemedia content delivery.

The proposed adaptive multimedia delivery methods can also be classified into different categories, depending on the object to be adapted. Indeed, in order to maintain high QoE levels, some methods use users' preferences or device energy/power levels. Other methods adapt the quality of the video/audio component to network conditions. Similarly, adaptive mulsemedia solutions could use some network delivery-related characteristics, user profile information or multi-sensory device-related data. However, as far as the authors of this paper are aware, they are the only researchers who have proposed an adaptive mulsemedia delivery solution to date [17].

The ADAptive MulSemedia delivery solution (ADAMS) proposed in [17] employs an algorithm that extends the video quality adjustment process described in QOAS [9] for multimedia to suit mulsemedia data. The server receives feedback regarding both network conditions and user preferences on sensory effects from the client and makes an adaptive decision. However, ADAMS is not based on the DASH standard, cannot be employed in conjunction with existing DASH videos and therefore cannot be easily embedded in future streaming systems.

It is particularly relevant to design and build a DASH-based adaptive application capable to deliver mulsemedia content. This is based on an extension of the DASH protocol, which was designed for multimedia content only. The proposed extension to the existing DASH protocol is intended to support adaptive mulsemedia content delivery.

3 SYSTEM ARCHITECTURE

The architecture of the proposed DASH-based adaptive multimedia and mulsemedia delivery system is illustrated in Fig. 1. The system is deployed at both client and server sides. At the client side, a mulseplayer renders various sense segments into video/audio streaming and multi-sensorial effects (haptic, olfaction, etc.) which will be deployed via different devices. At the server side, the platform annotates different sense information and encodes all information into an adaptive multimedia and multi-sensory stream and associates it with two Media Presentation Description (MPD) files.

3.1 Encoding

Fig. 2 illustrates the encoding component of the proposed DASH-based Adaptive Multimedia and Multisensorial Delivery System architecture. Multimedia content delivery is very well developed. In this system, the open source tools "mp4box"[1] and "bento4"[2] are used to encode the original video/audio content in MPEG DASH. For all other senses' content/information, an annotation tool is proposed to generate the *MPD*.

3.2 Delivery

After encoding, the sender delivers the adaptive multimedia and multi-sensory stream from the sender buffer via the network based on the network conditions and user preferences. The stream is received at the receiver buffer at the client side. Then, after decoding and rendering, a mulsemedia player on the client side will enable the video/audio content playout, as well as the control of the multisensory devices to generate the multi-sensory user experience in synch with the multimedia stream.

4 SYNCHRONIZATION MECHANISM

4.1 Mulsemedia Player

The block-level architecture of the mulsemedia player is illustrated in Fig. 3. The player contains two components: the dash multimedia player "multi.js" and the mulsemedia effect synchronizer "mulse.js". Both "multi.js" and "mulse.js" are JavaScript applications which are fetched from the server and executed by the client web browser.

[1] mp4box: https://gpac.wp.imt.fr/mp4box/
[2] bento4: https://www.bento4.com/developers/dash/

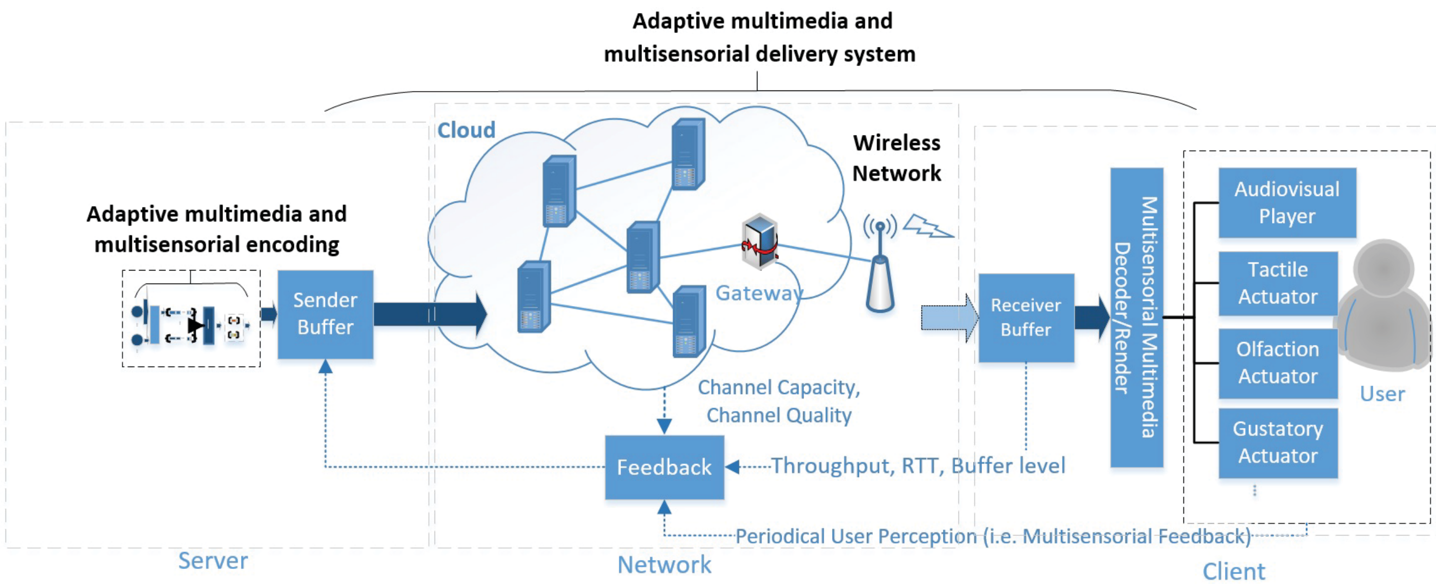


Figure 1: DASH-based Adaptive Multimedia and Mulsemedia Delivery System

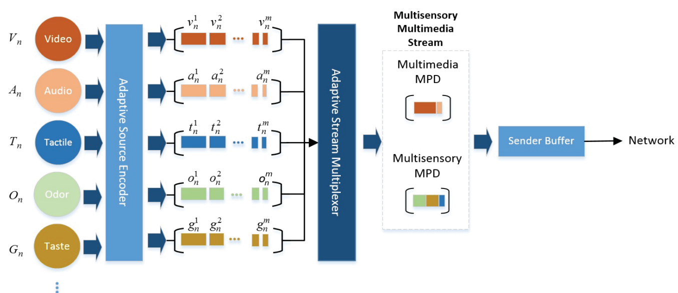


Figure 2: Encoding Architecture

"multi.js" is a DASH multimedia player which extends "dash.js"[3] with a user profile. When the web page is loaded, "multi.js" downloads the video components from the server and plays them on web page. Mulsemedia segment files are progressively downloaded by the "mulse.js" and is synchronized with the multimedia DASH player "multi.js".

Folowing DASH principle, "mulse.js" application can download diverse segment files. In this case, there are three different types of segment files: haptic, olfaction and wind. However in principle, there could be more. Each segment file corresponds to a duration of 10 seconds in the video. During the duration of a single segment there can be different types of effects. All mulsemedia segment files are using the JavaScript Object Notation (JSON) format.

Once the mulseplayer downloads the mulsemedia segment files, the mulsemedia player fills the mulsemedia buffers, which is in fact a table of mulsemedia segments. Once a segment is played, another one is downloaded (or an "empty" segment is inserted) and occupies the position in the buffer of the segment that has just been played. Then, the segment situated in the next position in the buffer

[3]dash.js: https://github.com/Dash-Industry-Forum/dash.js

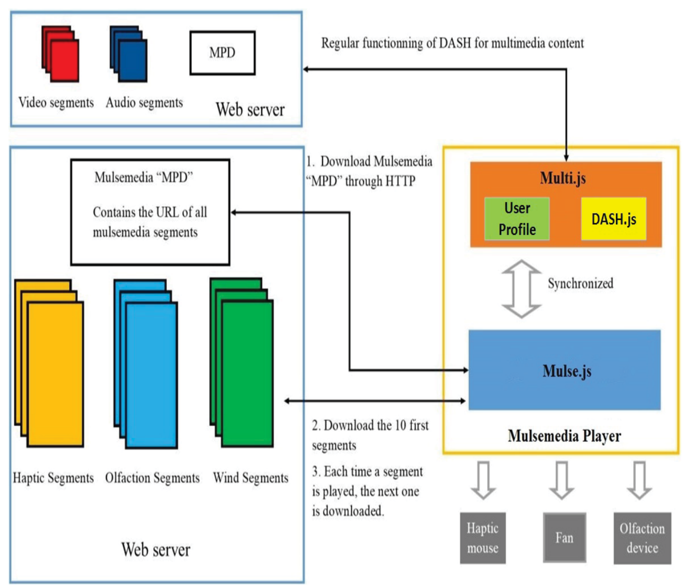


Figure 3: Block level architecture of the mulsemedia player

is played. Therefore, for instance assuming a buffer length of 10 and a segment duration of 10 seconds, the segment that has just been downloaded will be played 90 seconds later. In case the user goes to a specific moment into the video, the buffer is automatically filled again according to the new time. In the example given, the buffer always contains 9 segments (plus the one which is played). If the video content stops at 100s and the mulseplayer time is at 80s, the buffer will contain the segment that starts at 80s, the one that starts at 90s and 8 "empty" segments, which correspond to segments starting after 100s.

4.2 Playout and Synchronization

Several functions are designed and are used to play the multi-sensory effect segments in the receive buffer: "play-mulsemedia", "synchronize-and-play" and "play-segment".

The "play-mulsemedia" function starts by playing the segment situated in position 0 in the buffer by calling the function "synchronize-and-play". This function returns the current delay between the multi.js player actual time and the starting time of the current mulsemedia segment. This is the skew between the time at which the multi-sensorial effects occur and that of the display of the relevant multimedia content. This delay can be either positive or negative. Then, "play-mulsemedia" waits for 10 seconds minus the synchronization delay (the result is inferior to 10 seconds if the delay was positive, and superior to 10 seconds if the delay was negative), and plays the segment contained in buffer 1. It also replaces the segment contained in buffer 0 (the one that has just been played), by the one that will be played in 9*10 = 90 seconds later. "play-mulsemedia" is a recursive function. After playing segment contained in buffer 10, it plays segment in buffer 0.

The function "synchronize-and-play" is the one that handles synchronization with the multimedia player "multi.js", by computing the delay between the time of "multi.js" and the starting time of a mulsemedia segment. In case the delay (called "synchro-ms" for "synchronization in milliseconds") is positive, it means that compared to multi.js, the mulsemedia effects are late. So the segment is played with a starting time equal to "synchro-ms". A negative delay refers to the effect being in advance compared to those in "multi.js", so the program waits for a time of "synchro-ms" before playing the segment with a starting point at 0.

The function "play-segment" is called by "synchronize-and-play". It is the function that plays a segment via calling "bind-game-event", "send-game-event", "play-olfaction" and "play-wind" (these are the functions used to play a mulsemedia effect. A "real" segment is composed of 3 different tables containing all of the effect descriptions ("haptic-effects", "olfaction-effects", and "wind-effects"). The function "play-segment" launches a timer for every effect contained in these tables, and plays them when their corresponding timer expires. The timer depends on the starting point of the effect inside the segment and on the starting point of the segment. Indeed, "play-segment" is able to start to play the segment at different starting time. If the segment is started at 3 seconds, the effects scheduled before second 3 are not played and an effect to be played at 5 seconds would start after 2 seconds. In case of haptic effects, there is a delay of 250 ms between the binding time and the sending time. A haptic effect is played if the segment starts before the binding time of the effect. In case the segment starts between the binding time and the sending time, the effect is not played.

5 EXPERIMENTAL SETUP

Satisfactory synchronization is essential to guarantee a temporal ordering of multiple sensorial effects in a mulsemedia delivery system [16]. In this section, the perceived experience of haptic, olfaction and air-flow effects with low synchronization delay in the proposed DASH-based adaptive multimedia and mulsemedia delivery system are investigated. In order to measure the user's satisfaction with DASH-MS, a subjective experiment is performed with multiple individuals, as the human perception varies depending on individual user preference. Also, the synchronization delay is measured.

37 users from different backgrounds (e.g. students, staff, researchers, engineers) participated in the subjective tests. The study was promoted via institutional email and through a specially created "doodle". The on-line scheduling tool was used to arrange and fill a pre-set scheduling table. All the participants accepted body haptic effects and were screened against anosmia and color blindness. It took around 10 minutes for each participant to complete the whole test.

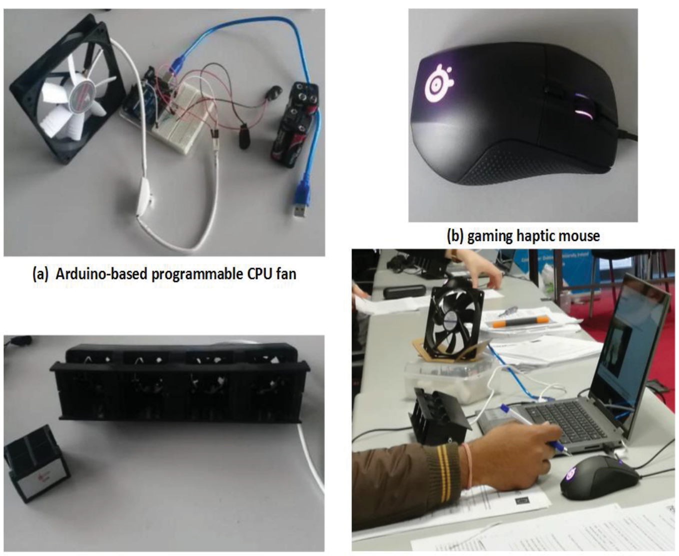
Four video clips were selected from the video "Big Buck Bunny". The video clips were synchronized with haptic, air-flow and olfaction effects. They had the same characteristics: 2 minute long, resolution of 1920x1080 pixels, and 30 fps frame rate. The ***video clip 1*** was presented with no additional sensorial effects. Multi-sensory effects were correctly synchronised with the ***video clip 2***. Haptic effects had a synchronization delay of +1 second, whereas wind and olfaction effects had synchronization delays of +5 seconds in ***video clip 3***. Haptic effects had a synchronization delay of -1 second, whereas wind and olfaction effects had synchronization delays of -5 seconds in ***video clip 4***.

Compared to the subjective testing of the traditional multimedia content, additional multi-sensory actuator equipment is required to present the mulsemedia effects to the end users with the mulseplayer in DASH-MS. A gaming haptic mouse, an Exhalia scent diffuser and an Arduino-based programmable CPU fan were used to produce haptic, olfaction and air-flow effects alongside the multimedia content, respectively. According to the content scenario, the multi-sensory effects were manually synchronized with the corresponding sensorial content in the multimedia clips by setting the start and end timestamps, then the "mulsemedia segment file" is generated by a multi-sensory data annotation tool.

Fig. 4 shows the interface of the mulseplayer developed for DASH-MS. On the top of the page there are the DCU, EU Horizon 2020 project NEWTON and European Union logos. In the video play area, mulseplayer have control buttons for play/pause, volume control and full screen functions. The Effects Control Zone is a manual control approach for the user to select the mulsemedia effect of their preference. The mulseplayer will detect the user profile in "multi.js" and enable/disable the multi-sensory effect accordingly. The default settings enable all multi-sensory effects and the block shows the green color; once the button pressed, the color will change to red and the corresponding effect is disabled (i.e. and will not occur anymore). "Current Representation" corresponds to the "mulsemedia representation" currently used by the mulseplayer (default setting is "0").

Fig. 5 illustrates the equipment and the test-bed employed in the tests. Fig. 5(a) shows an Arduino-based programmable controlled fan that provides the air-flow effect. The fan's on/off state and strength was controlled by a C++ program using the Arduino board. The gaming haptic mouse in Fig. 5(b) was available from Rival and supports full control of the haptic effect in terms of frequency and duration. Fig. 5(c) presents the Exhalia scent diffuser that enables to diffuse scents from each of its four small fans. The subjective test-bed shown in Fig. 5(d) was built in a separate room in the Performance Engineering Lab at Dublin City University, Ireland.

Figure 4: Interface of the mulseplayer



(a) Arduino-based programmable CPU fan

(b) gaming haptic mouse

(c) Exhalia scent diffuser

(d) subjective test

Figure 5: Mulsemedia Equipment and Test-bed

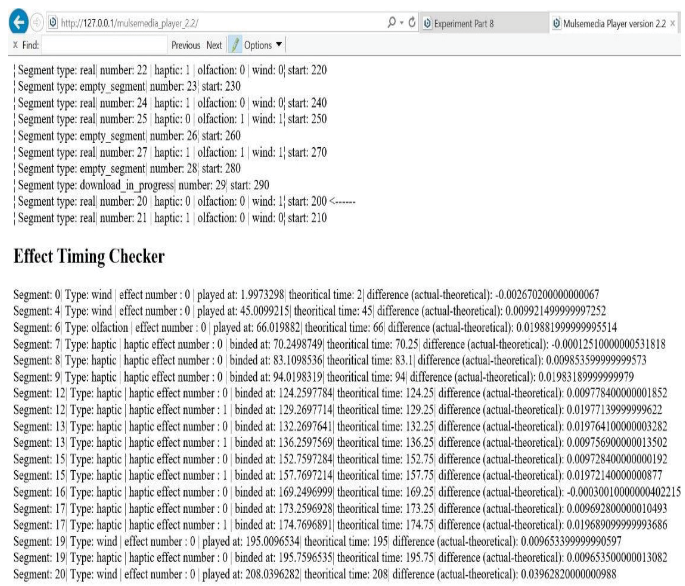


Figure 6: Effect Timing Checker Result

In order to avoid any potential disturbances during the tests, the testing environment was set up in the way shown above, obeying all the recommendations of ITU-T R.P.910 [6], ITU-T R.P.911 [5] and ITU-T R.P.913 [7].

6 TEST RESULT AND ANALYSIS

6.1 Synchronization Delay

The way effects are synchronized with the video is described in section 4. A function is designed and called "time-checker" in "mulse.js" which measures the difference between the moment an effect is played and the moment it theoretically should have been played. A screenshot of results given by this function is provided in fig. 6. The evolution of the delay between mulsemedia effects and the video for a 5-min clip with 33 effects (without any pause or seek operation) is presented in fig. 7. The synchronization mechanism of the mulseplayer is working on mulsemedia segments of 10 seconds duration. Consequently, the effects contained in the same segments have close synchronization delays.

The synchronization mechanism is able to adapt the starting point of mulsemedia segments to the actual time of the "multi.js" player. Therefore, delays rarely exceed ±0.07 seconds. The importance of synchronization between effects and video is shown in [16]. The delay for a haptic effect should not exceed 1 second, but wind effects can have delays up to -5 seconds or +3 seconds.

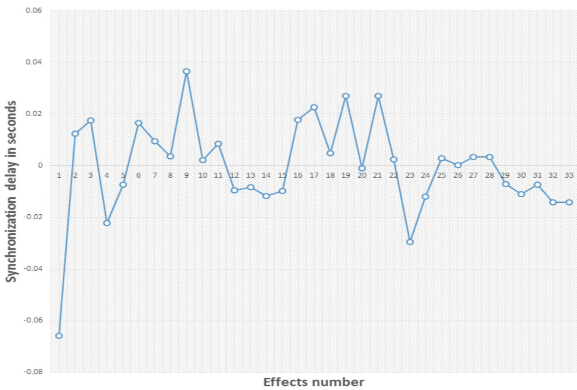


Figure 7: Evolution of the delay between effects and video for a 5-min clip without pause or seek operation

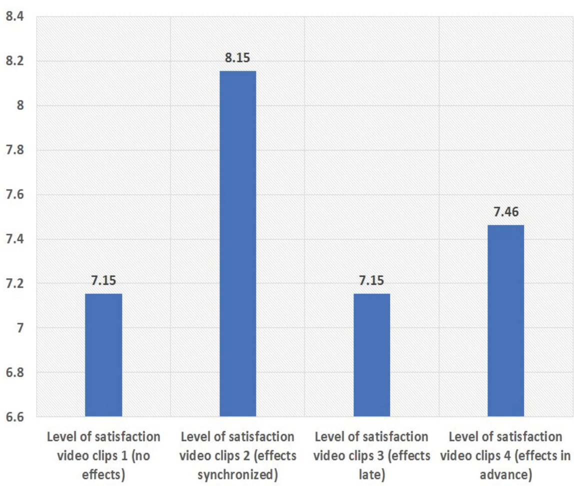


Figure 8: Average Levels of Satisfaction

6.2 User Satisfaction

Each participant was asked to answer a set of questions about their perceptual experience of the video in presence of a subset of these multi-sensorial stimuli (i.e. give marks between 1 to 10, where 1 represents "bad" and 10 represents "excellent" satisfaction levels). Fig. 8 illustrates the results in terms of average levels of satisfaction for different synchronization delays. It is interesting to note that the average levels of satisfaction with the mulsemedia effects are always superior or equal to the average level of satisfaction with no effects. As expected, the highest value of the level of satisfaction corresponds to video clip 2 in which the effects were correctly synchronized. Additionally, users seem to have preferred effects in advance than late effects.

7 CONCLUSIONS AND FUTURE WORK

This paper describes, illustrates and tests an adaptive mulsemedia delivery system, and a multi-sensorial player (mulseplayer). The mulseplayer deploys some important concepts such as the decomposition of effects in segments of different types and a metadata file named mulsemedia MPD. These new concepts can be used in future adaptive mulsemedia streaming applications.

ACKNOWLEDGMENTS

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

REFERENCES

[1] Velibor Adzic, Hari Kalva, and Borko Furht. 2012. Content aware video encoding for adaptive HTTP streaming. In *Consumer Electronics (ICCE), 2012 IEEE International Conference on*. IEEE, 92–93.

[2] B. Ciubotaru, G. Ghinea, and G. M. Muntean. 2014. Subjective Assessment of Region of Interest-Aware Adaptive Multimedia Streaming Quality. *IEEE Transactions on Broadcasting* 60, 1 (March 2014), 50–60. https://doi.org/10.1109/TBC.2013.2290238

[3] Jen-Wen Ding, Der-Jiunn Deng, Tin-Yu Wu, and Hsiao-Hwa Chen. 2010. Quality-aware bandwidth allocation for scalable on-demand streaming in wireless networks. *IEEE Journal on Selected Areas in Communications* 28, 3 (2010).

[4] Gheorghita Ghinea and Oluwakemi Ademoye. 2010. A user perspective of olfaction-enhanced mulsemedia. In *Proceedings of the International Conference on Management of Emergent Digital EcoSystems*. ACM, 277–280.

[5] P911 ITU-T RECOMMENDATION. 1998. Subjective audiovisual quality assessment methods for multimedia applications. (1998).

[6] P910 ITU-T RECOMMENDATION. 1999. Subjective video quality assessment methods for multimedia applications. (1999).

[7] P913 ITU-T RECOMMENDATION. 2014. Methods for the Subjective Assessment of Video Quality Audio Quality and Audiovisual Quality of Internet Video and Distribution Quality Television in Any Environment. (2014).

[8] Sylvie Jeannin, Leszek Cieplinski, Jens Rainer Ohm, and Munchurl Kim. 2001. Mpeg-7 visual part of experimentation model version 9.0. *ISO/IEC JTC1/SC29/WG11 N* 3914 (2001).

[9] G-M Muntean, Philip Perry, and Liam Murphy. 2004. A new adaptive multimedia streaming system for all-IP multi-service networks. *IEEE Transactions on Broadcasting* 50, 1 (2004), 1–10.

[10] G. M. Muntean, P. Perry, and L. Murphy. 2005. Objective and subjective evaluation of QOAS video streaming over broadband networks. *IEEE Transactions on Network and Service Management* 2, 1 (Nov 2005), 19–28.

[11] G. M. Muntean, P. Perry, and L. Murphy. 2007. A Comparison-Based Study of Quality-Oriented Video on Demand. *IEEE Transactions on Broadcasting* 53, 1 (March 2007), 92–102. https://doi.org/10.1109/TBC.2006.886451

[12] Vishwanath Ramamurthi and Ozgur Oyman. 2014. Video-QoE aware radio resource allocation for HTTP adaptive streaming. In *Communications (ICC), 2014 IEEE International Conference on*. IEEE, 1076–1081.

[13] Thomas Stockhammer. 2011. Dynamic adaptive streaming over HTTP–: standards and design principles. In *Proceedings of the second annual ACM conference on Multimedia systems*. ACM, 133–144.

[14] Hrishikesh Venkataraman, Ting Bi, Ting Wu, Ioana Ghergulescu, and Gabriel-Miro Muntean. [n. d.]. DE-BAR: Device Energy-Centric Backlight and Adaptive Region of Interest Mechanism for Wireless Mobile Devices. *Wireless Personal Communications* ([n. d.]), 1–27.

[15] Kyoungro Yoon, Sang-Kyun Kim, Jae Joon Han, Seungju Han, and Marius Preda. 2015. *MPEG-V: bridging the virtual and real world*. Academic Press.

[16] Zhenhui Yuan, Ting Bi, Gabriel-Miro Muntean, and Gheorghita Ghinea. 2015. Perceived synchronization of mulsemedia services. *IEEE Transactions on Multimedia* 17, 7 (2015), 957–966.

[17] Zhenhui Yuan, Gheorghita Ghinea, and Gabriel-Miro Muntean. 2015. Beyond multimedia adaptation: Quality of experience-aware multi-sensorial media delivery. *IEEE Transactions on Multimedia* 17, 1 (2015), 104–117.

[18] Zhenhui Yuan and Gabriel-Miro Muntean. 2013. A prioritized adaptive scheme for multimedia services over IEEE 802.11 WLANs. *IEEE Transactions on Network and Service Management* 10, 4 (2013), 340–355.

[19] Z. Yuan and G. M. Muntean. 2013. A Prioritized Adaptive Scheme for Multimedia Services over IEEE 802.11 WLANs. *IEEE Transactions on Network and Service Management* 10, 4 (December 2013), 340–355. https://doi.org/10.1109/TNSM.2013.110513.130490

[20] Longhao Zou, Ramona Trestian, and Gabriel-Miro Muntean. 2015. E 2 DOAS: user experience meets energy saving for multi-device adaptive video delivery. In *Computer Communications Workshops (INFOCOM WKSHPS), 2015 IEEE Conference on*. IEEE, 444–449.

[21] L. Zou, R. Trestian, and G.-M.. Muntean. 2015. E2DOAS: user experience meets energy saving for multi-device adaptive video delivery. In *Computer Communications Workshops (INFOCOM WKSHPS), 2015 IEEE Conference on*.