IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 4 July 2022, accepted 24 July 2022, date of publication 27 July 2022, date of current version 2 August 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3194170

RESEARCH ARTICLE

Integration of Multisensorial Effects in Synchronised Immersive Hybrid TV Scenarios

DANIEL MARFIL[1], FERNANDO BORONAT[1], (Senior Member, IEEE), JUAN GONZÁLEZ[1], AND ALMANZOR SAPENA[2]

[1]Department of Communications, Universitat Politècnica de València, Gandia Campus, 46730 Grao de Gandia, Spain
[2]Department of Mathematics, Universitat Politècnica de València, Gandia Campus, 46730 Grao de Gandia, Spain

Corresponding author: Fernando Boronat (fboronat@dcom.upv.es)

This work was supported in part by the "Vicerrectorado de Investigación de la Universitat Politècnica de València" under Project PAID-11-21 and Project PAID-12-21.

ABSTRACT Traditionally, TV media content has exclusively involved 2D or 3D audiovisual streams consumed by using a simple TV device. However, in order to generate more immersive media consumption experiences, other new types of content (e.g., omnidirectional video), consumption devices (e.g., Head Mounted Displays or HMD) and solutions to stimulate other senses beyond the traditional ones of sight and hearing, can be used. Multi-sensorial media content (a.k.a. mulsemedia) facilitates additional sensory effects that stimulate other senses during the media consumption, with the aim of providing the consumers with a more immersive and realistic experience. They provide the users with a greater degree of realism and immersion, but can also provide greater social integration (e.g., people with AV deficiencies or attention span problems) and even contribute to creating better educational programs (e.g., for learning through the senses in educational content or scientific divulgation). Examples of sensory effects that can be used are olfactory effects (scents), tactile effects (e.g., vibration, wind or pressure effects), and ambient effects (e.g., temperature or lighting). In this paper, a solution for providing multi-sensorial and immersive hybrid (broadcast/broadband) TV content consumption experiences, including omnidirectional video and sensory effects, is presented. It has been designed, implemented, and subjectively evaluated (by 32 participants) in an end-to-end platform for hybrid content generation, delivery and synchronised consumption. The satisfactory results which were obtained regarding the perception of fine synchronisation between sensory effects and multimedia content, and regarding the users' perceived QoE, are summarised and discussed.

INDEX TERMS Advanced TV, digital TV, HbbTV standards, hybrid TV, immersive TV, mulsemedia, multi-sensorial media, multisensory integration, sensory effects, TV.

I. INTRODUCTION

Traditionally, the consumed media contents have exclusively involved audiovisual (AV) streams. However, in order to generate more immersive media consumption experiences, other new types of content (e.g., omnidirectional video), consumption devices (e.g., Head Mounted Displays or HMD) and solutions to stimulate other senses beyond the traditional ones of sight and hearing, can be used. Multi-sensorial media content (a.k.a. mulsemedia [1]) allows providing additional sensory effects that stimulate other senses during the media consumption, with the aim of providing the consumers with a more immersive and realistic consumption experience (*"Seeing is believing, but feeling is the truth." —T. Fuller* [2]). They provide a greater degree of realism and immersion during media consumption experiences, but also can provide

The associate editor coordinating the review of this manuscript and approving it for publication was Yue Zhang.

greater social integration (e.g., people with AV deficiencies or attention span problems) and can even contribute to creating better educational programs (e.g., for learning through the senses in educational content or scientific divulgation).

Examples of sensory effects that can be used are olfactory effects (scents), tactile effects (e.g., vibration, wind or pressure effects), and ambient effects (e.g., temperature or lighting). Fig. 1 shows an example of a scenario for mulsemedia content consumption, including ambient lighting, smell, temperature, vibration and wind effects.

Mulsemedia has been a widely researched area for many years [1]. Back in 1999, a subjective evaluation involving more than 300 participants was carried out in [3]. In that paper, the way scents affected the participants were investigated to assess wether this type of enhanced media has any influence on improving the sensation of presence in virtual reality (VR) scenarios. Moreover, it was also studied how those sensory effects influence the memory regarding the virtual objects in the virtual environment. The obtained results suggest that those additional sensory effects increase the feeling of presence and help to better remember the objects represented in the virtual environment.

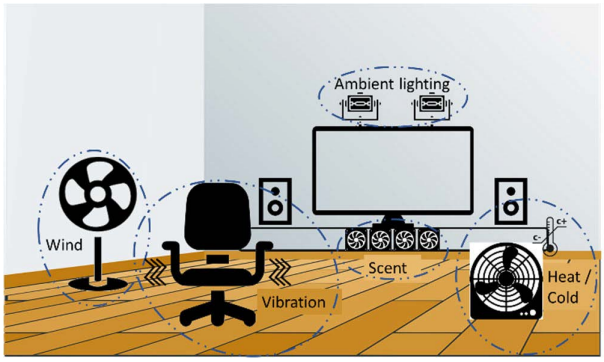

FIGURE 1. Example of a mulsemedia content consumption scenario.

As consumption-related technologies evolve and improve the user experience, new types of content, consumption devices and ways of consuming the content appear, and, consequently, new associated challenges must be faced. For instance, the 360 mulsemedia concept is introduced in [4]. It is proposed to enhance the omnidirectional video content with additional sensory effects, to be triggered when the user is watching specific regions of the content. In particular, the variation of the intensity of the olfaction effect depending on the user's field of view (hereafter, FoV) is proposed.

As it can be assumed, playing out and/or triggering in a synchronous way all the involved streams and sensory effects in a multi-sensorial media consumption experience is a crucial requirement in providing a satisfactory Quality of Experience (QoE) to users. Media synchronisation (abbreviated as sync, hereafter) is one of the many challenges that still need to be faced in current hybrid consumption environments. These environments are the ones in which the different streams involved in the experience are delivered through different (hybrid) technologies (e.g., broadcast and/or broadband). In them, the concurrent sync of the media playout of those involved (hybrid) streams is referred to as *hybrid sync*.

Current commercially deployed sync mechanisms still need to be enhanced, and new hybrid sync mechanisms are required to successfully deploy consumption experiences in mulsemedia scenarios involving more than one delivery technology and several devices. As an example, the 2018 Super Bowl event was provided through both broadcast and broadband delivery networks. The IP-based (broadband) content arrived from 28 to 47 seconds behind real-time [5]. In such a case, if the content is received only through the broadband network, it may not imply any inconvenience to the users' perceived QoE. However, if the broadband content is complementary to any other content received through another delivery technology (e.g., Digital Video Broadcasting or DVB) with less latency, this can imply a deterioration of the perceived QoE. Therefore, sync mechanisms must be implemented and performed for that type of scenario.

The Hybrid Broadcast Broadband TV standard (HbbTV [6]) provides mechanisms for harmonising the delivery and consumption of broadcast and broadband (hybrid) TV-related contents through connected TVs (main devices) and companion or secondary devices, such as smartphones or tablets. It also provides mechanisms to provide Inter-Device Media Sync (IDES). Nevertheless, a survey carried out in [7] shows that, on the one hand, 35.8% of respondents think that the deployment of HbbTV-based services is still behind expectations. On the other hand, despite the pessimistic vision of the HbbTV deployment, most of the respondents think that HbbTV is a very interesting (23.2%) or moderately interesting (45.2%) technology of advanced TV services.

Despite the massive increase in content availability and consumption via broadband networks, the primary source of content consumption is still the traditional broadcast TV [8], [9]. As an example, regarding multimedia content accessed through the Internet, there is still a significant 36% of Spanish citizens over 16 years old (approx. 40M people) that do not consume digital content, and, from that percentage, a third does not even know about the existence of the complementary services provided by connected TVs [10]. However, in the same study, it is stated that 71% of users that consume multimedia content through the Internet use a connected TV to do so. In particular, 5.6% only use connected TVs and 65.4% use connected TVs and other devices, such as smartphones, set-top boxes, etc.

Therefore, IP delivery already has a real impact on the TV industry, offering viewers unrivalled choice and flexibility. Nevertheless, the less noticed and discussed trend is that, in Europe, DVB-based Digital Terrestrial Television (DTT) has continued to grow its overall number of TV households over the years. It has gone on to evolve and perform very well in the market.[1] Currently, with the aid of the hybrid TV model, DTT is now very well placed to underpin the

[1] https://dtg.org.uk/2021/05/13/state-of-the-nation-2021-full-report/

market, considering as an IP-only scenario is still a very long way off. Instead, the combination of DTT and IP into a hybrid environment is a model set to remain relevant for many years to come as it meets the needs of present-day consumers.

In this paper, a solution for providing multi-sensorial and immersive hybrid (broadcast/broadband) TV content consumption experiences, including omnidirectional video and sensory effects, is presented. Sensory effects can be added to the consumption experience of both the main broadcast AV content delivered through a DVB-based network, and the related complementary (omnidirectional or traditional) content delivered through a broadband network. The proposed solution has been implemented through an enhanced version of a refined end-to-end platform for hybrid content generation, delivery and consumption, previously presented in [11] and [12], and compatible with the HbbTV standard [6]. New features have been added to it in order to make it also compatible with the MPEG-V standard [13] and to handle mulsemedia streams correctly. The implementation of the proposed solution has been subjectively evaluated (by 32 participants) with satisfactory results.

The main contributions of this work regarding the authors' previous related works are the following:

- The design of an enriched end-to-end hybrid delivery and consumption solution for the hybrid TV use case, including traditional multimedia content, mulsemedia content and immersive (omnidirectional video) content.
- A mechanism to signal available mulsemedia content within the broadcast TV media streams.
- Design and implementation of a device to control the triggering of sensory effects, so-called mulsemedia controller, which provides a RESTful service to enable communications between it and the user's consumption (main or secondary) device. This controller manages and acts as a communication gateway between the main or secondary device and the hardware in charge of generating or presenting the sensory effects.
- Implementation of an IDES sync mechanism between the main (or secondary) device and the mulsemedia controller.
- Enhanced testbed to subjectively assess the proposed solution, involving hybrid TV functionalities, and omnidirectional and mulsemedia content.
- Subjective assessment (N = 32) of the proposed solution obtaining very satisfactory results in terms of users' immersion and perceived QoE.

The paper is structured as follows. In Section II, previous works regarding mulsemedia content and (hybrid) sync mechanisms are summarised. In Section III, the requirements and involved processes to generate, handle and provide mulsemedia content are described. In Section IV, the implemented testbed, the evaluation scenario and the followed evaluation methodology are presented. The evaluation results are also summarised and discussed in this section. Finally, in Section V, some conclusions are drawn and some future projects are addressed.

II. PREVIOUS WORKS

A. MULSEMEDIA

Mulsemedia has historically been a research area (e.g., [1], [2], [12], [14]–[17]). Many applications have been developed in the past, from adopting this type of content for AV and VR applications or video games to studying how the inclusion of sensory effects affects the users' perceived QoE. Actually, in [2] it is stated that there are still open issues and challenges to face, such as the need for defining authoring tools for this type of content. In [12], a survey is presented to understand the role that mulsemedia can play in the QoE enhancement in future broadcasting services. In that work, it is concluded that mulsemedia can: 1) improve the perceived video quality; 2) partially mask eventual video quality decrease; and 3) impact the strength of the emotions. However, it is also stated that further research is still pending in order to face other issues, such as the (semi-)automatic generation of sensory effects or the tolerable asynchrony between the AV content and the sensory effects.

In [18], it is stated that there is a lack of formal and explicit representation of what mulsemedia is, which may result in a failure to understand it adequately. It is proposed to establish a common conceptualization in regard to mulsemedia systems through a reference ontology, named MulseOnto, covering their main notions and, therefore, providing a wider picture of mulsemedia systems. For instance, the authors of that research try to answer relevant questions such as what mulsemedia computer/software systems are, or what the user interface for a mulsemedia computer system is.

A detailed study on how sensory effects affect the users' perceived QoE is presented in [19]. That work aims to identify, classify and quantify the users' preference regarding individual sensorial components of mulsemedia streams. In that work it is stated that sync between the AV content and the multi-sensorial effects is a key factor in providing a positive impact on the users' QoE. Besides, the research in [20] reflects the existing difficulties for assessing multimedia QoE, which can be due to many reasons, as QoE is user-centric, individual, multi-dimensional and multi-sensorial.

An extensive compilation of mulsemedia related work is presented in [14]. Additionally, the authors in that work also list the beneficial areas where mulsemedia can enrich the users' QoE in human-computer interaction (HCI), such as telecommunications, education, e-health, e-commerce, advertising, entertainment or tourism [21]. In [15], some recommendations to design and evaluate mulsemedia applications, focusing on olfactory effects integration are proposed. Moreover, authors in that paper list as key factors challenges such as: olfaction based mulsemedia integration, sync, standardisation, olfactory sensor and display development, intensity and duration of effects, applicability in different areas (e.g., health, education, tourism…) and remote delivery. In that paper, some interesting instructions are provided, regarding: 1) mulsemedia quality evaluation in terms of laboratory design; 2) assessor preparation; 3) experimental design; and 4) different olfaction characteristics that should

be taken into consideration, such as advice to assessors (e.g., avoiding drinking coffee before the assessment), mechanisms to remove lingering scents, etc. Additionally, regarding olfactory content, authors recommend not using more than ten different scents during an evaluation session by following the defined specifications in [22]. In [16], a survey of mulsemedia devices is also presented. In that study, a guide for users to build their own mulsemedia environment, both on desktop and in immersive 360 scenarios, is provided. In particular, regarding 360 environments, the authors in that study state that there is still limited research involving olfactory media.

There have been many efforts to provide mulsemedia experiences through the development of frameworks and/or platforms in order both improve the users' perceived QoE and/or perform subjective studies (e.g., [23]–[31]). The research in [23] describes an open-source end-to-end tool (based on MPEG-V Part 3: Sensory Information [13]) to generate and experience mulsemedia content. In that research, a subjective evaluation involving three different experiments was carried out. Some interesting conclusions were drawn, such as: 1) mulsemedia content can prove to be even annoying, depending on the type of consumed content (e.g., news); 2) sensory effects have an impact on the perceived video quality (higher mean opinion score -MOS- in average); or 3) emotions (e.g., fun, fear, etc.) are stronger if sensory effects are included.

In order to signal and insert the required information for multi-sensorial content, a.k.a. Sensory Effect Metadata (SEM), a framework is proposed in [24], including a Single Media Multiple Device (SMMD) controller. This controller is in charge of receiving the mulsemedia content (including AV content) and sending each content to the corresponding device (e.g., AV to a TV, scents to a scent generator, heat or cold information to an air conditioner unit, etc.). All this information is signalled by adopting the MPEG-7 standard [32], which specifies a metadata system for describing multimedia content. Additionally, a GUI based authoring tool is also presented to allow non-expert users to signal mulsemedia information into AV content.

In [25], the MPEG-7 standard is adopted in order to design an adaptive mulsemedia framework (called ADAMS). It includes both coarse- and fine-grained adaptation modules (i.e., mulsemedia flow adaptation and packet priority scheduling) on the server side. Those modules have been developed based on the results from a performed subjective evaluation and the existing network conditions. Specifically, the subjective evaluation is performed in order to study the users' QoE in loaded network delivery conditions and how the proposed framework is capable of minimising the impact on the perceived QoE.

The requirements and associated challenges to overcome the delivery of mulsemedia content are stated in [26], in order to provide a compliant and interoperable sensory effects delivery, such as: 1) multifunctionality and reusability; 2) reactivity and timeliness; and 3) manageability and configurability. Additionally, a framework to encode, transmit and consume this type of media is also presented in that work, in order to provide truly immersive media experiences. Moreover, the same authors present in [27], [28] a detailed description of the framework and its capabilities. The first version of that framework is called PlaySEM Sensory Effects Renderer (SER) [27] and PlaySEM SER 2 [28] is the evolved version. PlaySEM SER 2 supports multi-communication and multi-connectivity protocols, multi-standards (e.g., MPEG-V standard [13]) and facilitates the accommodation of new technology. In order to do this, it is essential that its set of architectural and design patterns are applied successfully. As a result, that evolution led towards an interoperable mulsemedia framework for delivering different sensory effects such as scent, vibration, light or wind. In that study, different case studies with different profiles of PlaySEM SER 2 are described in order to prove its reusability and adaptability to different possible scenarios (e.g., a video clip enriched with external light, smell, vibration and wind, or a smell-intensive system). Furthermore, in [29], a complete toolset, including the player of the PlaySEM SER framework, is presented. In particular, an authoring tool called *Sensory Effect Video Annotation* (SEVino, Fig. 2) is described. This tool enables the generation of mulsemedia metadata for multimedia content and stores it in an XML file by following the MPEG-V standard [13]. Additionally, another different tool is also presented in that work, called Sensory Effect Simulator (SESim), which allows users to simulate and playout any mulsemedia content within a computer, providing graphical emulations of the involved devices (e.g., a fan picture with an overlaid intensity number from 0-100 for a certain moment in which a wind effect is required).

Similarly, in [30] an authoring tool based on MPEG-V [13] is presented. It is divided in two different parts: the generation of the SEM tool and the SEM authoring tool. In particular, the first tool generates a SEM file in an XML format, which follows the Sensory Effect Description Language (SEDL) schema, defined in [33]. The second tool consists of a 3D virtual environment in which the AV content and the sensory effects are simulated.

Another approach for representing sensory effects as first-class entities in multimedia models and authoring languages is presented in [34], enabling multimedia applications to synchronise sensorial media to interactive AV content in a high-level specification. A simulator and an application example are provided. The application example is specified with another similar multimedia authoring language, NCL (Nested Context Language), and Lua.[2] On the one hand, NCL also uses a declarative approach based on XML. On the other hand, Lua components are used for translating sensory effect high-level attributes to MPEG-V SEM files.

The solution presented in this paper allows for the use of different authoring languages for declarative approaches

[2]Lua is the auxiliary language used by NCL when an author needs to perform some kind of task that is out of the scope of NCL.

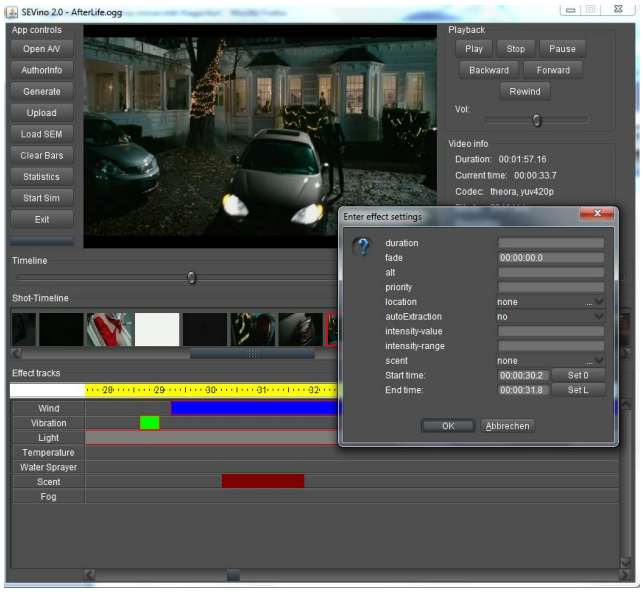

FIGURE 2. SEVino 2.0 tool [29].

based on XML, but in our testbed, we have used the SEVino (Sensory Effect Video Annotation) tool that makes use of SEDL.

In [31], the concept of *Five Sense Media Playback* is presented. It includes the consumption of AV content, among other sensory effects, such as wind, light or vibration. The included SEM is registered in an XML file. Moreover, a sync mechanism is proposed consisting of a calculated activation time when the sensory effects should be triggered, by taking into account the execution time of the device and the transmission time of the data.

Olfactory can be considered the third most stimulated sense regarding mulsemedia content. There is a huge variety of previous studies where this sense is stimulated to enhance the users' perceived QoE (e.g., [35]–[45]).

As an example, in [35]–[38] the influence of the olfactory sense in 360 scenarios is studied. In [39], Sensory Experience is defined as the QoE for multi-sensorial media, i.e., not only related to the traditional AV content. Additionally, after an empirical study, authors in that study conclude that a better users' perceived QoE is achieved if the olfaction effects are delivered to the user after the video, as opposed to presenting them before the video. They also conclude that more studies are needed to better understand the impact that audio has on the perception of scent.

In [40] early considerations are presented in relation to the existing olfaction-related studies and devices. It is stated that this type of media can improve visual data in VR training situations, in which scent can complement critical cues or alarms.

The olfactory sense is used in [41] to classify and tag photos, similarly to using traditional text tags (both techniques are compared). The main aim in that study is to test if the olfactory sense can also be used to memorise and recall photos. Performed tests over some participants confirm that text-based tags are useful and accurate and, alternatively, scent-based tags, although achieving lower accuracy, are also significatively useful and can be used to aid photo recall successfully.

In [42], a study regarding the scent delivery is carried out. A comparison between a real scenario (a classroom) and a virtual scenario (the same classroom modelled in 3D), is implemented. In both virtual and real scenarios, the experiments consisted of evaluating the time delay it takes from the start of the scent generation until it is detected by participants. Result shows similar "scent-arrival" time in both scenarios. However, it is concluded that the sense of smell is strongly dependent on each individual, regarding how and when the scent is detected, thus this area is still challenging. The same authors in [43] present a study in which it is observed how long it takes for a person to adapt to (i.e., fail to notice) the smell of lemon in a virtual environment. Participants viewed the virtual environment twice, with two different qualities, the first one being high quality and the second one variable (decreasing) quality. They were separated in three groups: a first group with no smell effects; a second group with smell effects during the viewing; and a third group with smell effects before and during the viewing. Participants had to indicate when they thought that the second viewing had decreased its quality compared to the first viewing. The first and third groups of participants coincided in the time boundary. The adaptation happens sometime after 150 seconds, and after 199 seconds participants might ignore any scent presence. Moreover, in [44] the same authors state that it is possible to decrease the quality of the rendering of specific regions in a virtual scenario by attracting users with the olfactory sense to a different region. So, the users do not realise this decrease in quality and that the rendering of time can be significantly reduced without any impact on the users' perceived QoE.

In [45], a report comparing the results of two separate research studies, which explored the users' QoE of olfaction-enhanced multimedia when the audio component is presented. Results show that when no audio component is presented together with the video content, out-of-sync situations between the video content and multi-sensorial media are detected with lower values of asynchrony than when the audio component is present. It is concluded that audio can provide a masking effect to sync issues between the video content and multi-sensorial media.

There are also other previous works that focus on other senses (e.g., [46]–[51]). In [46] a prototype of a mulsemedia environment is presented, including wind and scent effects. A detailed description of the designed hardware devices for the prototype is provided. It includes a wind device (a pc-controlled fan), an olfactory device (using vaporising scents and compressed air injection) and a pneumatic device (to emulate an impactful sense of wind).

In [47] a mulsemedia delivery system is described, including scent, air effects and haptic devices to enhance the traditional AV consumption experience. A subjective assessment

is performed with AV content with average and high quality (853 × 480 and 1280 × 720, respectively) together with the aforementioned sensory effects. It is concluded 1) that the most preferred sensory effect is the haptic, which is delivered to the user though a special vest; and 2) that sensory effects enhance and improve the user's enjoyment and sense of reality.

In [48], subjective assessment of the users' perceived QoE in a multi-sensorial consumption web-based environment is performed. In particular, how the light effects can influence the consumption experience and the impact of different settings for the colour calculation is studied. As in other works, the SEM formats are also used to signal the involved sensory effects. Results indicate that the inclusion of light effects enhances the perceived QoE, although authors state that including other additional sensory effects would enrich the experience even more, as they concluded previously in [52]. Additionally, a plugin for the Firefox web browser capable of reading and interpreting SEM data is also described.

A research study on how to design art experiences whilst considering all the senses is presented in [49]. The Tate Sensorium exhibition at the Tate Britain art gallery in London, UK, is studied focusing on the touch sense and how it is integrated with sound. An air haptic device is configured to enhance the experience of observing a painting with three different patterns. Authors in that study conclude that those types of interdisciplinary works (i.e., the technical team and the artistic non-technical team) can improve the museum attendees' experience and, moreover. They show a first glance of how to create, conduct and evaluate multi-sensorial experiences in a museum. In [50], in order to emulate food and drinks in a virtual environment, electronic devices are presented. In [51], how users experience multi-sensorial content is explored taking into account cross modal correspondence principles, as described in [53]. An alternative exploration method to traditional studies on QoE is proposed by considering other parameters such as the heart rate or the gaze behaviour during evaluations. It is suggested that when cross-modal principles are considered in the design of mulsemedia systems, the content is perceived differently from when using semantic congruence. However, regarding the reported QoE there is not a significant impact.

Regarding the consumption of mulsemedia content delivered through broadcast technology there is not so much research. How the sensorial media is signalled, embedded or delivered is an additional challenge that still needs to be dealt with to be able to provide the content via this type of technology.

In [54] a broadcasting system including sensory effects information is proposed. For that purpose, the definition and signalling of the mulsemedia requirements is defined in an XML, which is embedded during the encoding of the MPEG2-Transport Stream (MPEG2-TS). Similarly, in [55] a framework to broadcast mulsemedia content, adopting the standard MPEG-V Part 3: Sensory Information, is presented. However, no evaluations are provided.

In [56] and [57] a testbed of a real smart home scenario enhanced with sensory effects (with a Smart TV and IoT devices capable of generating the effects) is presented and implemented. Common home devices such as air conditioning devices or lights are used. Subjective tests based on MOS to obtain the users' perceived QoE are carried out. Participants had to watch 10 sequences of 30-40s interleaved with 5s of grey screen. Satisfactory results are obtained, proving the feasibility of the multisensorial TV services.

An IPTV-based mulsemedia scenario is described in [58] but with little detail. It is proposed to register the SEM in an XML file embedded (but it is not explained how) in the MPEG2-TS that is delivered to the users' home receivers. The sensory effects generation devices should be capable of decoding and analysing the mulsemedia data.

B. NEED FOR SYNC

Through the end-to-end media consumption chain, there are many different causes that can add undesired latency. The variable end-to-end delays can cause media contents to arrive at different instances although they were generated to be consumed simultaneously. Thus, those contents could not be played out as planned (i.e., at the same time or in-sync). The different sources of latencies in the chain are studied in [59] and [60], and existing delays and added latencies in current delivery networks are studied in [60], [61] and [62]. On the one hand, regarding broadcast technologies, delay differences can be up to 5s in national scenarios and up to 6s in international scenarios [60]. On the other hand, regarding broadband technologies, typical values of delay and jitter, from 20ms to 500ms and from 0ms to 500ms, respectively, are reported in the International Telecommunications Union (ITU-T) G.1050 standard [63].

As in traditional media, in mulsemedia scenarios there is also a need for achieving a synchronised state between the involved streams, in order to include mulsemedia content into traditional multimedia (i.e., AV) content. In previous studies, such as [64]–[66] and [67], this specific challenge has been studied.

In particular, a sync algorithm for multi-device scenarios to provide synchronised mulsemedia experiences is presented in [64]. In that study, the mulsemedia content is separated into different streams, depending on the involved sense or effect (e.g., heat, wind, scent, light or vibration). The proposed algorithm is based on the RTP/RTCP protocol [68] and consisted on a Master/Slave scheme on which the device in charge of presenting the AV content adopts the master role. An evaluation is performed by comparing the asynchrony between the AV content playout and the sensory effects generation when the proposed mechanism is enabled and when it is not. Asynchrony values are kept below 30ms if the proposed mechanism is enabled and conversely, asynchrony values can exceed 100ms if there is no sync mechanism. So, the results show the need to adopt sync mechanisms in mulsemedia scenarios.

Regarding the need for sync between the sensory effect generation and the main AV content playout, in [65] the temporal relationship between the AV content playout and the olfaction-based sensory effects is studied. A subjective evaluation is performed in order to analyse the user's ability to detect asynchrony and thus, their perceived QoE. An asynchrony interval of [−5, 10] seconds is defined as valid to deliver olfaction-based sensory effects and to be perceived as synchronised with the AV content; and out-of-sync if this effect is provided beyond the defined limits.

A study of the users' perceived QoE in which two olfactory effects are synchronised with AV media is carried out in [66]. This study is performed in order to determine if sync boundaries are affected by using two olfactory effects and the impact that jitter (in particular, the mixing of scents) has on the users' QoE. It is concluded that olfaction effects can be considered *in-sync* if they are presented with asynchrony values within the interval [−10, +15] seconds regarding the AV content playout, and values beyond that boundary can be considered *out-of-sync*.

Additionally, in [67], another subjective evaluation is performed in order to study the acceptable sync boundaries regarding sensory effects such as haptic and air-flow. It is concluded that the accepted asynchrony interval to consider those effects *in-sync* is [0, +1] seconds for haptic effects and [−5, +3] seconds for air-flow effects with regard to the main AV content.

In relation to the need for hybrid sync mechanisms involving traditional multimedia content (i.e., only text, audio and video streams) displayed in one or many devices (i.e., IDES), consumed at one or many destinations (i.e., IDMS or Inter Destinations Media Synchronisation [69]), and transmitted through one or many delivery technologies, authors have already extensively reviewed previous works in [69], [70], [71], and [72]. From those works, the adoption of a hybrid sync mechanism called *Timing and External Media Information* (TEMI) [73] can be highlighted. It has been proposed by MPEG and DVB, as an amendment to ISO/IEC13818-1 [74] and provides (in summary) the following features:

- Insertion of an extrinsic, absolute and stable timeline within the MPEG2-TS within the so-called *temi timeline* descriptor.
- Insertion of URLs within the so-called *temi location* descriptor.

With the use of both types of descriptors, broadband available content can be: i) signalled inside the broadcasted content (by inserting its location in the *temi location* descriptor), thus, making it available at the user's home; and ii) played synchronously with the main content, as this will have an inserted global timeline (inside the *temi timeline* descriptor) which can be used as a reference. This mechanism has been used in preliminary tests on BBC UK public TV to allow for personalised broadcast services, based on the use of HbbTV, (by having some of the content replaced with IP-delivered content). In particular, *an MPEG TEMI timeline is read from the broadcast and used to time when the switch from broadcast to an IP delivered DASH stream takes place*.[3]

III. SOLUTION FOR THE INTEGRATION OF MULTI-SENSORIAL EFFECTS IN HYBRID TV CONSUMPTION EXPERIENCES

In this Section, a solution for the integration of multisensorial effects (i.e., mulsemedia content) to enhance traditional TV consumption experience is presented. The mulsemedia generation process, the signalling of the sensory effects into the broadcasted multimedia main content, and the required functionalities in order to satisfactorily consume the mulsemedia streams beyond audio and video are described. Specifically, the proposed solution involves both the content provider part which includes the content generation and transmission and the consumer's home part (including the sensory effects management, and generation in sync with the TV main content playout). Fig. 3 shows an overview of the proposed end-to-end solution. On the one hand, on the Content Provider side (left side in Fig. 3) the generation and signalling of the mulsemedia content is carried out. In particular, regarding the mulsemedia content, a Sensory Effect Metadata (SEM) file is generated to link specific sensory effects to different instants of the main TV content (explained in subsections III.A and III.B). On the other hand, in order to consume the mulsemedia content, there is a need for the management and triggering of the available sensory effects generation when indicated in the SEM file. For that purpose, a Mulsemedia Controller (abbreviated as MC, hereafter) has been designed in order to receive the mulsemedia-related information from the main or secondary device and send the corresponding orders to the effect generation devices to trigger the appropriate effects (explained in subsection III.C).

A. GENERATION OF MULSEMEDIA CONTENT

In order to generate the mulsemedia content, a SEM file is created. The SEM file follows the SEDL schema, defined in [33] and included in the MPEG V standard [13]. It uses tags to provide the required information to link a specific sensory effect and its parameters (e.g., type of effect, intensity or other sensory effect-specific attributes) to a specific instant of the main TV content. Specifically, the generated SEM file, links a specific Presentation Timestamp (hereafter, PTS) instant value from the main TV content to one or several sensory effects. PTS values providing intrinsic relative timelines for synchronising the information within the same MPEG2-TS, but they have no significance outside the media included in it [71]. Moreover, these temporal relationships can be overwritten by networking equipment throughout the end-to-end delivery chain, without being aware of this transformation. Therefore, when the content is received by consumers, the sync process between the main TV content playout and the generation of the associated sensory effects can fail. To avoid

[3]https://www.bbc.co.uk/rd/blog/2018-09-content-substitution-personalised-broadcasting-hbbtv (last access june 2022)

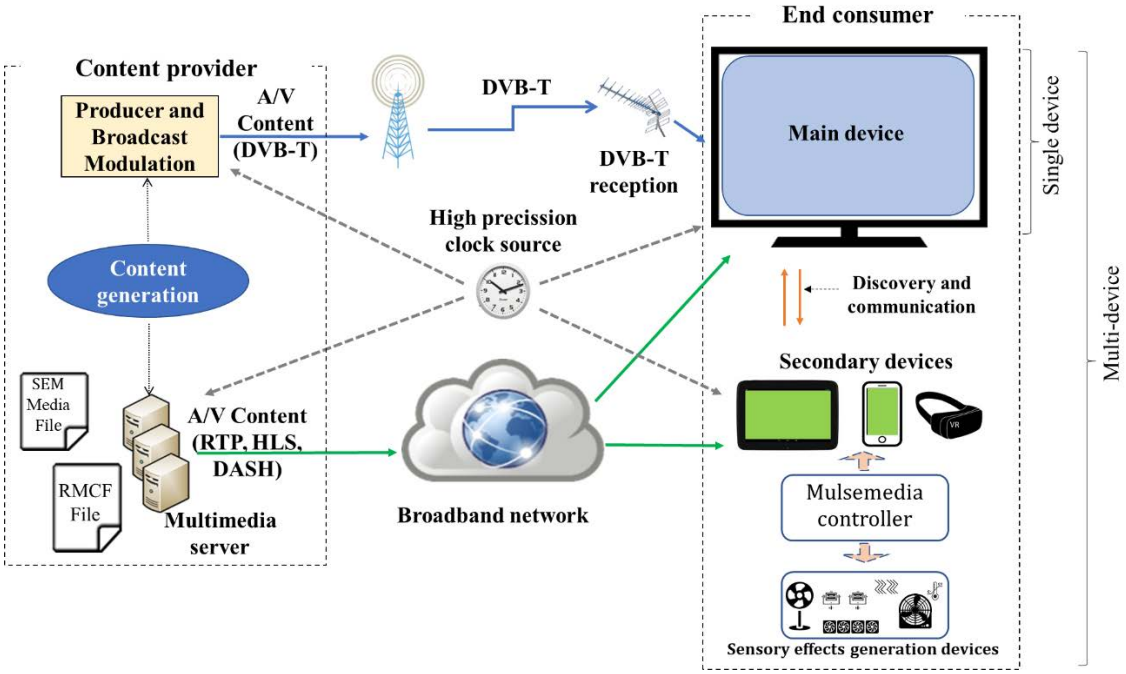

FIGURE 3. Overview of the proposed end-to-end solution to include multi-sensorial effects in synchronized immersive hybrid TV scenarios.

```
<sedl:Effect xsi:type="sev:LightType" autoExtraction="both" activate="true"
   si:anchorElement="true" si:pts="0" global-timestamp:"3763721890965118886"/>

<sedl:GroupOfEffects si:anchorElement="true" si:pts="180000"
   global-timestamp:"3763721890965118886">

  <sedl:Effect xsi:type="sev:ScentType" intensity-value="100.0"
     intensity-range="0.0 100.0" activate="true" sev:scent="smoke"/>

  <sedl:Effect xsi:type="sev:TemperatureType" intensity-value="40.0"
     intensity-range="0.0 40.0" activate="true"/>

</sedl:GroupOfEffects>
```

FIGURE 4. Example of SEM.

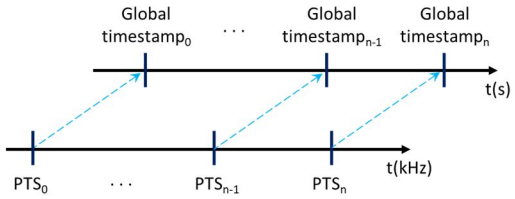

FIGURE 5. Matching process between intrinsic PTS timestamps and global timestamps.

this issue, a global time reference is used and an additional property for each effect is provided including the equivalent global timestamp corresponding with the original PTS instant. In particular, an NTP[4]-formatted timestamp under the property name ''global-timestamp'' is attached to every effect tag where a PTS value is included. Fig. 4 shows an example of the SEM file with this property included. For that, the very first NTP-formatted timestamp embedded into the multimedia MPEG2-TS *temi timeline* descriptor is associated with the PTS = 0 value. The next PTS values are increased consequently to keep matching the same instant of the multimedia content, by converting Hz into seconds and then, adding this value to the initial NTP-formatted timestamp (Fig. 5). This way, both the main TV content and the additional mulsemedia contents involved in the consumption experience, share the same global timeline and, thus, it is possible to achieve an accurate sync level at the consumer's home.

[4]NTP: Network Time Protocol, defined in RFC 5905.

B. SIGNALING OF MULSEMEDIA CONTENT

Once the mulsemedia content is generated, it is necessary to include some notification mechanisms regarding the availability of this type of content once compatible players have received the main TV broadcast content. For that purpose, also a tag-based file, so-called *Related Media Content File* (hereafter, RMCF) is designed in order to signal all the available related or complementary broadband content to the broadcast main TV content. The RMCF has been previously defined and used in other studies about Hybrid TV, for similar purposes (e.g., [72]). In this study, in order to signal available mulsemedia content, some existing tags (MEDIA and SOURCE tags, see Fig. 6 and Table 1) have been updated to provide the required information to handle this type of content. For such a case, the essential information to enable the sensory effects integration into the hybrid TV consumption experience is the location of the SEM file (e.g., its Uniform Resource Identifier or URI). Thus, the end-user's consumption devices can access the SEM file and manage the corresponding effects.

```
<?xml version="1.0" encoding="UTF-8"?>
<Related Media Contents File>
    <MEDIA id="media0" media_type="360AV" media_format="h264/aac" metadata="360_cam"
    temi_init="3763721899965118886">
        <source protocol="http/dash" projection="ERP" tiled ="false"
        uri="http://server.com/360/erp/stream.mpd"/>
    </MEDIA>
    <MEDIA id="media1" media_type="AV" media_format="h264/aac" metadata="TV_cam/front view"
    temi_init="3763721899965118886">
        <source protocol="http/dash" uri="http://server.com/front/stream.mpd"/>
    </MEDIA>
    <MEDIA id="mulsemedia0" media_type="SEM" media_format="xml" metadata="mulsemedia"
    temi_init="3763721899965118886">
        <source protocol="http" uri=http://server.com/mulsemedia/mmedia_effects.xml/>
    <MEDIA/>
</Related Media Contents File >
```

FIGURE 6. Example of RMCF including the signaling of available mulsemedia content related to the main TV content.

TABLE 1. Tags from [71] updated to support mulsemedia content.

Tag	Property	Description
MEDIA		It specifies the necessary metadata for any available related or complementary AV or mulsemedia content
	id	Unique identifier for the element
	media_type	Media content type. Value for mulsemedia content: 'SEM'.
	media_format	Format or encoding information
	metadata	Brief description
	temi_init	Absolute global time when the content generation started
SOURCE		It allows to specify alternative origins for the same content
	uri	Uniform resource identifier

C. MULSEMEDIA CONSUMPTION FUNCTIONALITIES

Multi-sensorial effects can involve different senses, depending on the provided main TV content and, finally, on the consumer's choice. For that reason, a Mulsemedia Controller (MC) has been designed, which is connected to the main (TV) or secondary device and in charge of handling the sensory effects with the available generation devices at the user's home. The MC can be implemented in a low-powered microcontroller, as it does not require any powerful technical specification requirement. It is designed to be a simple gateway between the main or secondary devices and the effect generation devices, for instance, to activate some relays, such as to switch on or off effect generation devices. Also, to output low voltage signals, as soon as the main (or secondary) device sends any order involving any sensory effect generation. Current microcontrollers (e.g., Arduino, Adafruit) can implement IP wired/wireless network connectivity. On the one hand, this functionality can be used to adopt the DIAL protocol [75] in order to automatically discover main and/or secondary devices in the network (Fig. 7). That protocol is recommended by the HbbTV specification (Section 14 in [76]) to discover the involved devices and to establish a communication channel between them. On the other hand, wireless connectivity can provide a more comfortable environment for users, because including sensory effect generation devices may involve an overly ''wired'' environment that can negatively affect the users' perceived QoE. Despite having a wireless connection may add additional latency, common latency for wireless local area network values should not exceed 60ms [77]. This is an insignificant value for the most restrictive asynchrony intervals (see Section II.B), in order to consider as in-sync the perception of the sensory effects.

Once the main (or secondary) device is aware of the local IP address of the MC device, any sensory effect-related request can be sent through a Web service that follows the Representational State Transfer (REST) software architecture, which is implemented in the MC (Fig. 7). Specifically, a RESTful Application Program Interface (API) allows the requesting systems (in this work, the main or secondary devices) to access and manipulate textual representations of Web resources by using a uniform and predefined set of stateless operations. In particular, HTTP methods such as GET, POST or DELETE can be used by those devices to interact with the MC device through the RESTful API and, consequently, with the sensory effect generation devices.

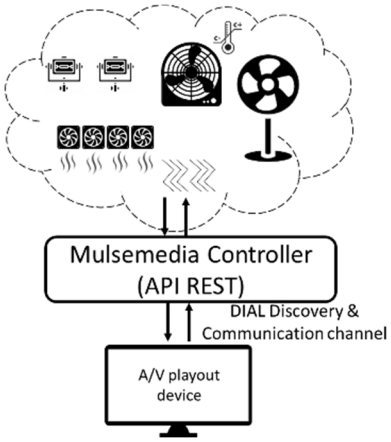

FIGURE 7. Discovery of the A/V playout device with DIAL [75] and management of the involved sensory effects through a REST API.

Through this API the main (or secondary) device can be aware of the available sensory effect generation devices which can be triggered during the media consumption experience.

1) MULSEMEDIA REST API

The adopted RESTful API has been designed in order to manage the involved sensory effect generation devices in a multi-sensorial hybrid TV consumption experience. As a RESTful system, it follows the next constraints:

- Client-server architecture: to allow multi platform communication and improve scalability.
- Statelessness: each client request is self-contained, i.e., all the information necessary to service the request is included in the request itself.
- Cacheability: the server must notify whether it is cacheable or not, to improve the performance.
- Layered system: clients are not aware if they are directly communicating with the server or with intermediates,

and this does not affect the involved communications client-server.
- Code on demand: client can transfer executable code to customise functionalities.
- Uniform interface: to simplify and decouple the architecture.

In the proposed solution, the most basic interaction between the main or secondary device (client) and the MC device (server) is the use of the HTTP GET and POST requests. The REST API for this specific use has been designed for two different types of functions: obtaining the list of the available effects and setting up a specific sensory effect from the available list. In this case, additional parameters are also included into the request, such as: the selected effect, a digital 0 or 1 (i.e., on or off states) or an 8-bit value if the selected effect supports different intensities (see Table 2).

TABLE 2. Implemented REST requests.

Request type	URL structure
GET	http://IP_MC/list_available_effects
POST	http://IP_MC/set/*effect*/{ 0 | 1 | intensity} (e.g., http://192.168.10.10/set/mint_scent/255)

On the one hand, the HTTP GET request is used in order to obtain the information of all the available sensory effect generation devices. This way, the main or secondary device is aware of the sensory effects that can be triggered during the media consumption experience. On the other hand, the HTTP POST request is used to trigger and/or stop the generation of sensory effects. As it can be observed in Fig. 8, the main or secondary device performs a specific HTTP POST request when it needs to trigger a particular sensory effect. As an example, the POST request to *http://{MC_local_IP}/set/15/1* means to enable (setting to '1') a relay which is connected to the output pin number 15 of the MC, where a fan is connected (to emulate the wind effect). If the request is successful, an HTTP "200 OK" message is responded. Contrarily, if an error occurs (e.g., if there is no device attached to the indicated output pin number or there is no device for that specific sensory effect) an HTTP "400 Bad request" message is responded. For that same example, in order to stop the wind effect, a POST request to *http://{MC_local_IP}/set/15/0* should be performed.

2) MANAGEMENT OF THE MULSEMEDIA SEM FILE BY THE MAIN OR SECONDARY DEVICES

Once the main device receives the broadcast TV content, it simultaneously starts the playout process and also analyses the MPEG2-TS to get the information included within the *temi location* and *timeline* descriptors (a more detailed description of this process can be found in [71]). The RMCF file (previously explained in subsection III.B) contains the required information to obtain, through the broadband network, the SEM file associated with the AV content being received and played. Therefore, the device can request this file and next, inspect and interpret the different sensory effects that can be triggered for that specific content. Then, the device dynamically creates a table including: i) the associated timestamp of every sensory effect; ii) the type of sensory effect to handle; and iii) other necessary information, such as the intensity, the scent type, or any other required information, separated by, e.g., semicolons. It can also share this information with the secondary devices (if needed). An example of how this table could look is shown in Table 3.

TABLE 3. Example of a list of sensory effects stored in the main or secondary device.

Associated Timestamp (ns)	Effect Type	Additional Data
3763721890965118886	Scent	type=forest; intensity=100/100
3763721892965118886	Wind	intensity=50/100
3763721896965118886	Vibration	intensity=10/100
[...]	[...]	[...]

During the playout process of the hybrid TV content, the main or secondary device must be aware of the instant being played, in order to decide whether it is the most suitable instant to send a request to the MC device, in order to trigger any involved sensory effect. Therefore, the device must periodically check the global timestamp associated with the current instant being played.

As the device knows the current instant of the AV content playout and the instant when the mulsemedia content needs to be triggered, it only has to compare both global timestamps (as they share the same global timeline) and, as soon as the sensory effect global timestamp fits within the configured acceptable boundaries, the device can send the MC the request to trigger the involved sensory effect. Fig. 9 summarises the different tasks that the main device performs in order to handle the involved sensory effects, once it has stored the involved sensory effects in the aforementioned table and is able to establish a communication channel with the MC.

3) MANAGEMENT OF THE SENSORY EFFECT GENERATION DEVICES BY THE MC

As previously defined, the MC can be implemented in a device with low performance, so it is intended not to be overwhelmed with a high workload. Considering this fact, it has been designed only to implement two different tasks: i) the previously explained REST service, in order to be capable of establishing a communication with the users' main or secondary devices; and ii) the transmission of electrical signals through its output pins. Regarding the latter, the MC can use its pins to transmit digital '1' or '0' values, e.g., to activate a relay in which a fan is connected; or to transmit analog electrical signals, e.g., to set the intensity from a range of possible values for a vibration effect generation device. The connection of any sensory effect generation device with the MC is independent/agnostic from the users' main or secondary devices. It is not relevant for the latter to know how the connection between the MC and the sensory effects generation devices is carried out.

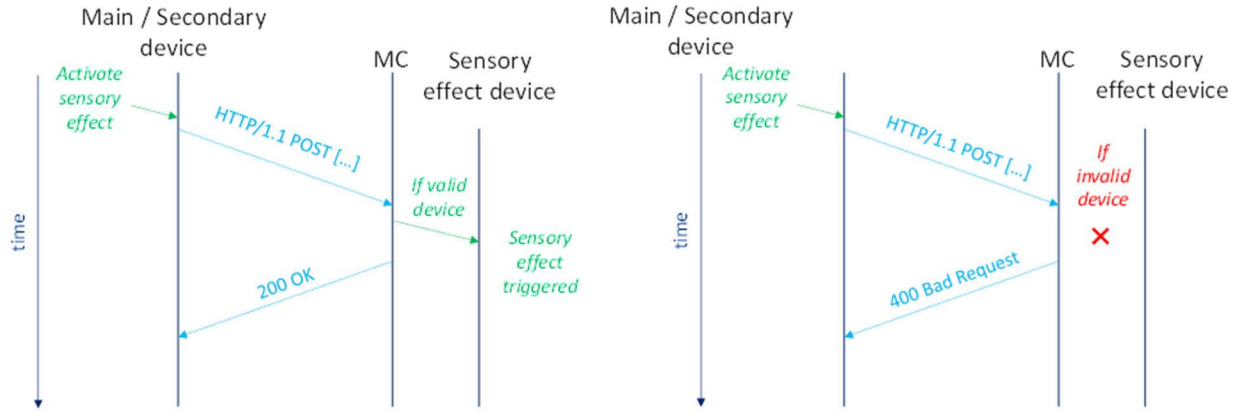

FIGURE 8. Example of a successful (left) and an unsuccessful (right) request between the main (or secondary) device and the MC.

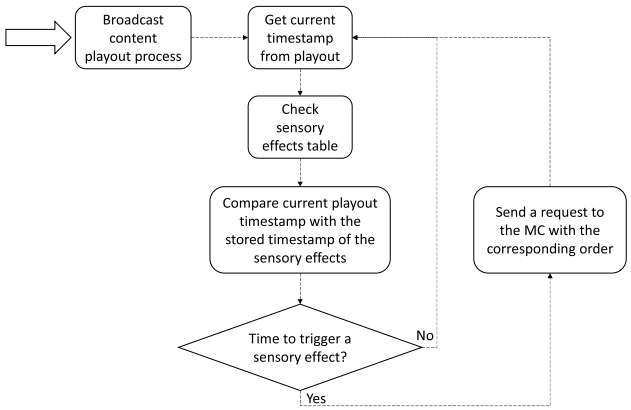

FIGURE 9. General overview of the involved routines of the main device to manage the activation of sensory effects.

IV. EVALUATION AND RESULTS

In this Section the conducted subjective evaluation process is described. Firstly, the implemented testbed is defined, including the generation of mulsemedia content at the content provider side (left side of Figure 3) and the implemented functionalities to trigger the sensory effects at the consumer side (right side of Figure 3). Secondly, the description of the evaluation process is detailed, and the obtained results are summarised and discussed.

The main goal of the conducted subjective evaluation has been to obtain and quantify the QoE perceived by the involved participants in the implemented hybrid TV scenario including multi-sensorial effects and omnidirectional AV content in order to provide a more immersive TV consumption experience. For that purpose, a PC-emulated TV is used to watch the broadcast TV main AV content (main device) and, additionally, a Head Mounted Display or HMD (secondary device) is used to watch the complementary broadband omnidirectional AV content.

A. USED TESTBED

In this study, the end-to-end platform for hybrid content generation, delivery and consumption has used some of the authors' previous research ([71] or [72]) has been enhanced to create a hybrid TV testbed supporting mulsemedia content generation and consumption. The multimedia devices of the testbed are presented in Table 4.

TABLE 4. List of involved devices in the subjective evaluation.

Device	Role
Intel Core i7-6700 @ 3.40GHz CPU, 8GB RAM, W10 with DTA-2111 modulator	Broadcaster
Intel Xeon E5420, 8GB RAM, Ubuntu 14.04, Fast Ethernet 100	Broadband Media Server
PC: Celeron N3050 @ 1.60GHz CPU, 4GB RAM, Ubuntu 14.04 with DVB USB rtl2832, connected to an LG 32LF592U 32" Smart TV (MS) via HDMI	End-user TV
Smartphone Samsung Note 9 embedded into an HMD	End-user HMD

As in our previous study in [72], on the one hand, regarding the broadcast transmission, the *DekTec StreamXpress* software configuration tool, provided with the DekTec DTA-2111 modulator PCI card, has been used to force packet losses in the broadcast connection. Each of the MPEG2-TS transport packets are usually extended by a short-ended Reed-Solomon error protection code, leading to a DVB MPEG2-TS packet with a length of 204 bytes. In combination with convolution coding and appropriate modulation schemes, a so-called quasi-error free (QEF) transport of DVB services can be guaranteed, which means that, on average, only one non-correctable error occurs within one hour of program presentation (equivalent to a BER of 1×10^{-11}).[5] Therefore, we have evaluated the scenario with a symbolic packet loss probability of 0,05% (much larger than the QEF average) in DVB transmission. Meanwhile, regarding the broadband network between Broadband Media Server and the end-user secondary device, (i.e, HMD) this has been conditioned with *netem* tool,[6] a network emulation functionality able to

[5] As stated in ETSI EN 300 744 V.1.6.2 (2015-10) Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for digital terrestrial television: "Quasi Error Free (QEF) means less than one uncorrected error event per hour, corresponding to BER = 10-11 at the input of the MPEG-2 demultiplexer".

[6] https://wiki.linuxfoundation.org/networking/netem (last access june 2022)

emulate delay, loss, duplication, and re-ordering of transmission packets. In particular, the broadband network has been conditioned with a delay of 60ms ± 20ms by following a normal distribution, which corresponds to what can be observed within long-distance fixed line connections or reasonable mobile networks and, thus, is representative for a broad range of application scenarios.

1) MULSEMEDIA CONTENT GENERATION

In order to generate the mulsemedia content, the tools described in [29] have been adopted. First, the so-called SEVino2 tool (Fig. 2) has been used to generate the sensory effects for the multimedia content into a SEM file. Next, the SESim2 tool has been used to check that the generated SEM file with the description of the involved sensory effects from SEVino2 has been generated correctly. Once the generated SEM file is validated, the MPEG2-TS stream including the embedded signalling information can be generated in order to be transmitted through the broadcast technology. For that purpose, the GPAC [78] tool has been used to generate the MPEG2-TS, including the *temi timeline* descriptor (with a global NTP-formatted time instant value) and the *temi location* descriptor (including the URL of the RMCF, which, in turn, includes the URI of the generated SEM file, as explained in section III). Then, the very first NPT-formatted timestamp inserted in the *temi timeline* descriptor is obtained and used to modify the SEM file by adding the proposed ''global-timestamp'' property in every tag where a sensory effect is involved (also explained in section III).

For scent and wind generation the NTP-timestamps have been manually modified in order to trigger the activation of their generation in advance with enough time to let the user experience the effect in sync with the AV content. Several tests have been conducted in the testbed by the authors to calculate the triggering advancement. In the testbed, the configured time has been manually set within the multimedia content player, which requests the activation of the corresponding sensory effect. This time value mainly depends on the distance between the sensory effect device, the end user and the features of that device (e.g., the diameter and the rounds per minute or RPM of a fan). Therefore, in real broadcasting scenarios it should be configured by the user depending on the devices used in their own scenario and their experience and perception. How to do this is out of the scope of this paper and is left for further work.

In order to exemplify the adopted sensory effects and their integration into the AV content, Figure 10 illustrates the timeline and the triggering instants of a reduced[7] list of sensory effects to be triggered during the AV content playout.

2) ADOPTED MULSEMEDIA CONSUMPTION MODULES

In order to manage the involved sensory effects involved in a mulsemedia content, the adopted device as MC in the

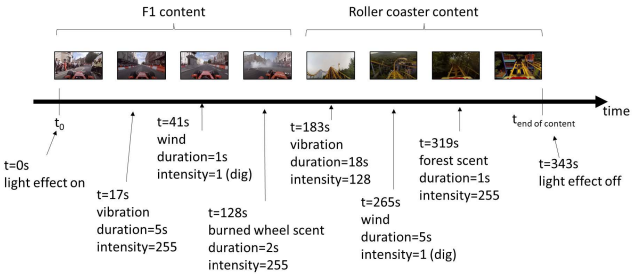

FIGURE 10. Generated mulsemedia content.

testbed has been the NodeMCU v3 device.[8] It is based on the ESP8266 WiFi System on Chip (SoC), and it can be programmed with the Arduino Integrated Development Environment (IDE). It can output both digital and analog electrical signals, besides providing 802.11 connectivity functionalities. Therefore, it can support the defined requirements (i.e., RESTful service and communicating with the connected sensory effect generation devices) to play a role as an MC. It is wirelessly connected to the user's main and/or secondary devices, and, at the same time, it is wired up to devices generating the following sensory effects: scent, heat, wind, vibration and ambient lighting.

The involved hardware of each generation device (i.e., sensory effect actuator) is the following:

- Heat and Wind: a common heater and fan, respectively (Fig. 12), connected to the MC through electronic relays.
- Scent: a modified version of the SBIv4 device from Exhalia,[9] in order to trigger the involved fans from the analog pin outputs (to provide different scent intensities) of the MC instead of managing this device from a web browser (default configuration).
- Vibration: Monacor BR-50 Bass rocker,[10] a low frequency speaker (30Hz to 300Hz) capable of vibrating, which produces a structure-borne sound and was mounted to a sofa in the testbed. The MC also controls this device through its analog pin outputs, in order to provide different vibration intensities.
- Ambient lighting: a Raspberry Pi 3 running Kodi OS (i.e., a media centre) with Hyperion functionalities (i.e., an opensource Ambilight[11]-like implementation). The MC is connected to this device through a relay, in order to turn the light system on/off. This hardware enables the automatic generation of light colours similar to the ones being presented on the TV screen boundaries, in order to ''extend'' the colour of the multimedia content over the back wall.[12]

[7] In order to make the figure clearer, only a reduced number of effects are represented in it.

[8] https://nodemcu.readthedocs.io/en/master/ (last access: june 2022)

[9] https://www.exhalia.com/us/ (last access: june 2022)

[10] https://www.monacor.de/produkte/components/lautsprechertechnik/hi-fi-chassis-/br-50/ (last access: june 2022)

[11] https://www.philips.co.uk/c-m-so/tv/p/ambilight (last access: june 2022)

[12] https://forums.raspberrypi.com/viewtopic.php?t = 128492 (last access: june 2022)

The description of the electronic circuits regarding how the sensory effect generation devices are implemented and connected to the MC or powered is out of the scope of this paper.

The flow diagram of the behaviour of the MC is shown in Fig. 11.

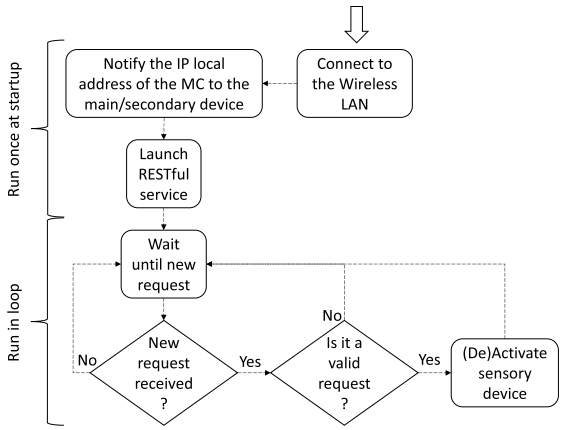

FIGURE 11. Behavior of the implemented MC.

As it can be observed, its functionalities have been designed to be as simple as possible, in order to use its technical resources in the most efficient manner.

3) TESTBED SCENARIO

A small space in a research lab has been conditioned to emulate a living room.

The subjective evaluation has been carried out in a properly conditioned 9m^2 space (Fig. 12), including a PC-emulated smart TV, an HMD, a sofa with a vibration effect generation device, with additional scent, wind, and ambient lighting generation devices. An additional low-noise fan has also been used in it for removing lingering scents before each evaluation session with different users (not visible in Fig. 12).

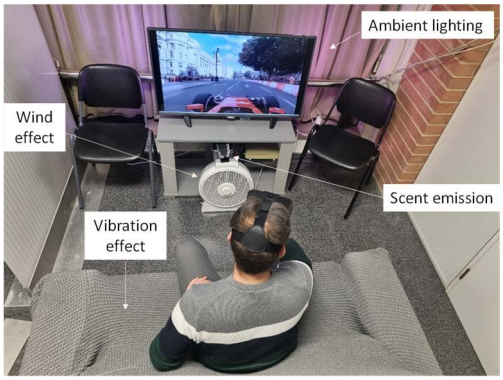

FIGURE 12. Emulated living room within the research lab.

B. AVAILABLE CONTENT DURING THE EVALUATION

Regarding the main (broadcast via DVB-T) TV content used during the evaluation, it has been generated by cropping the related (broadband via Internet) omnidirectional content (the expected frontal or main view of the content). Tables 5 and 6 show the main features and source of the used content and the included sensory effects in the testbed, respectively. The resolution of the broadcast content (MPEG2-TS) is 1920×1080px, while the broadband 360 content resolutions in DASH, with 3-second of video chunks/segments, are 7680×3840px; 3840×1920px; 1920×960px; 960×480px; and 480×240px.

C. EVALUATION PROCESS

For the subjective assessment, the Absolute Category Rating (ACR) method has been followed (as in many other related works [14]), and the evaluation scenario has been set in a controlled environment, following the Recommendation ITU-T P.913 [79].

The ACR method consists of presenting test stimuli to the participants one at a time and rating them independently on a category scale. In particular, ACR is a single stimulus method and, once the participants observe one stimulus, they have to rate it in a five-level score scale, ranging from 5 (excellent) to 1 (bad).

The evaluation was carried out in pairs of participants, and it was divided into three stages, following the suggestions and recommendations in [80], in order to prepare the environment in a home-like setting and not lasting more than 30 minutes. The first stage consisted of a brief introduction to the evaluation scenario, as well as the fulfilment of an initial questionnaire to collect participants' profiles. The second stage consisted of a guided use of the available AV consumption devices (and HMD) and apps to familiarise the participants with the scenario and its main features. Finally, the third stage consisted of a free use of the AV consumption devices, with no guidance or monitoring. After this stage, a final questionnaire was completed by the participants in order to evaluate their perceived QoE and the usability of the developed testbed (by adopting the SUS or *System Usability Scale* [81]).

Additionally, both the QoE and the perceived synchronisation level between AV content and sensory effects has also been evaluated with the adoption of a MOS metric (Mean Opinion Score [82]) by using a five level Likert scale.

D. PARTICIPANTS

A total of 32 users participated in the assessment and have consented to the publishing of the obtained results. Table 7 summarises their personal data obtained during the first stage of the assessment.

53% of the participants have a technical profile regarding their sphere of work. It also has to be stated that none of the participants had any AV impairment or had any post-Covid symptoms (e.g., lack of smell or taste) that could have an affect on the subjective evaluation.

Regarding smoking habits, 15% of participants smoke at least a cigarette per day, 9% occasionally smoke (less than 1 cigarette per day), 3% are passive smokers and the

TABLE 5. Generated content for the evaluation process.

Title	Duration (s)	Encoding	Available 2D views	Available omnidirectional views	Omnidirectional projection type	Source
F1 & Roller Coaster	360	H.264 + AAC	3	1	CMP	https://www.youtube.com/watch?v=fQoVFraBOnc https://www.youtube.com/watch?v=8lsB-P8nGSMs

TABLE 6. Generated sensory effects for the evaluation process.

Type of effect	Comments
Ambient lighting	Light effect which amplifies colors of the content from the TV, expanding them to the back wall.
Scent	Two different scents have been used, one of burned-like tire for the F1 content part and a forest scent during the roller-coaster content part.
Wind	This effect has been activated in specific moments to provide a feeling of fast motion (speed).
Vibration	This effect has been activated in specific moments to provide a feeling of fast (speed) or sudden motion.

TABLE 7. Profile of participants.

Genre		Age		Level of Studies	
Male	Female	<21	3%	Secondary	13%
55%	45%	21-25	19%	Professional training	25%
		26-30	13%	Degree	28%
		31-35	13%	Master	19%
		36-40	0%	PhD	15%
		41-45	3%		
		46-50	28%		
		>50	21%		

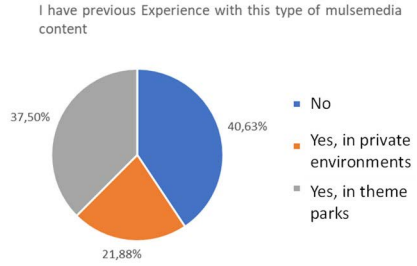

FIGURE 13. Participants' previous experience with mulsemedia content.

remaining 73% do not actively or passively smoke. By following recommendations presented in [15], participants did not smoke, drink coffee or eat in the previous 2h to the evaluation.

Fig. 13 summarises the participants' previous experience with the consumption of mulsemedia content. 41% of them did not have any similar experience before, while 38% had a previous similar experience, mainly in theme parks.

E. OBTAINED RESULTS

1) USER's PERCEIVED QoE AND IMMERSIVENESS PERCEPTION

Regarding the participants' previous opinion (asked prior to the evaluation process) about the statement "*sensory effects improve the QoE*", 75% of participants totally agreed, 22% partially agreed and only 3% partially disagreed with the statement. Nevertheless, it improved after the evaluation. When asked if *sensory effects improved their perceived QoE during the consumption sessions*, 91% of the participants totally agreed, while the remaining 9% neither agreed nor disagreed.

Additionally, participants evaluated their *perceived QoE when using both the TV and the HMD devices*, obtaining MOS scores of 4.03 $\pm$ 0.27 (C.I.95%) and 4.19 $\pm$ 0.26 (C.I.95%), respectively. Participants were asked about their perception of immersiveness in both situations.

When using only the TV, results show that 76% of participants totally agreed with the statement "*During the content consumption with TV and sensory effects, I have felt immersed into the experience*", while 16% partially agreed and the remaining 8% were neutral. When using the HMD, 88% of participants totally agreed with the statement "*During the content consumption with HMD and sensory effects, I have felt immersed into the experience*", while 6% partially agreed and the remaining 6% were neutral. Additionally, participants also were asked if at *some moment they felt like they had lost the notion of time*. When using only the TV, nobody lost the notion of time, while, when using HMD, 27% of participants totally agreed, 40% partially agreed, 27% were neutral and the last 7% totally disagreed. The duration of the video was quite short (360s) and, despite being immersed, most of the participants did not have time to lose the notion of time, especially when using only the TV device.

2) SENSORY EFFECTS SYNCHRONISATION

To assess whether in the implemented testbed the generated effects were correctly triggered during the consumption experience and perceived by users in the correct instants during the AV content playout, some questions/statements were included in the questionnaire. It was desirable to know whether out of sync situations could appear in our testbed during the assessment, since they would negatively affect

the consumption experience and, therefore, the results of our tests. On the one hand, during the immersive TV consumption experience using only the TV, 97% of participants (totally or partially) agreed with the statement "*sensory effects were in-sync with the AV content*" and the remaining 3% was not sure at all (Figure 14).

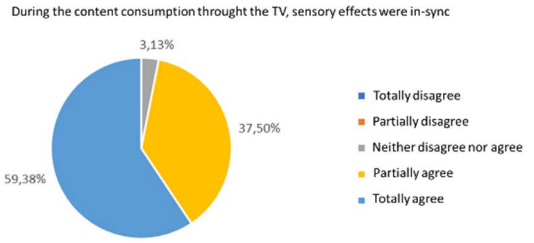

FIGURE 14. Perception of synchronization between TV and sensory effects.

However, on the other hand, during the immersive hybrid TV consumption experience using the HMD, all of the participants (totally or partially) agreed with that statement (Figure 15). As shown in other works, the immersion provided when using an HMD enhances the user's tolerance to small asynchronies that could be annoying when they are not used.

Regarding the inter-device (TV and HMD) synchronization (IDES) perception during the interaction between participants, the ones consuming the content on the TV were asked whether they considered that their colleague with the HMD was consuming the same content (i.e., they were in-sync) as him/her. 75% of participants agreed while 25% of participants did not agree or disagree.

So, we could conclude that during the tests, the effects were triggered in sync with the AV content playout and both devices were playing hybrid content (MPEG2-TS broadcast content on TV, and MPEG DASH broadband content on the HMD) in a synchronised way.

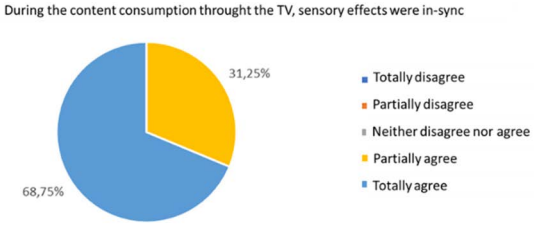

FIGURE 15. Perception of synchronization between HMD and sensory effects.

3) SENSORY EFFECT PREFERENCES

When asked about the participant's *preference regarding a specific sensory effect*, vibration is the most preferred effect, followed by wind, various scents and ambient lighting effects, these results are shown in Figure 16. This could be explained by the different degree of novelty of the effects. For example, the vibration effect is not commonly used in current multimedia systems, while the lighting effect is currently

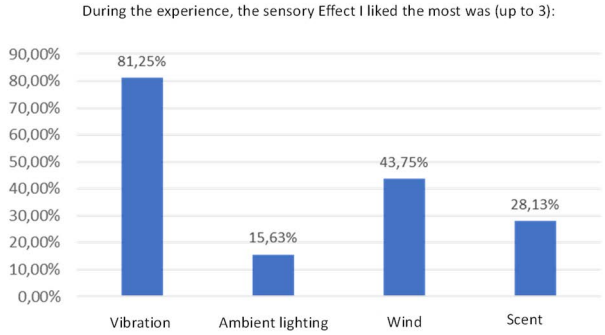

FIGURE 16. Most preferred sensory effects by participants.

a commercially available effect (e.g., in many Philips TV models including the *Ambilight* system).

Figure 17 shows *the most annoying sensory effects* for the participants. 16% of them did not like the wind effect, followed by vibration (6.25%) and scents (3.13%). Some of them stated that the noise of some of the generators (relays, fan, engines…) was annoying and some others stated that they were cold because of the wind effect. 75% of participants stated that no sensory effect was annoying at all.

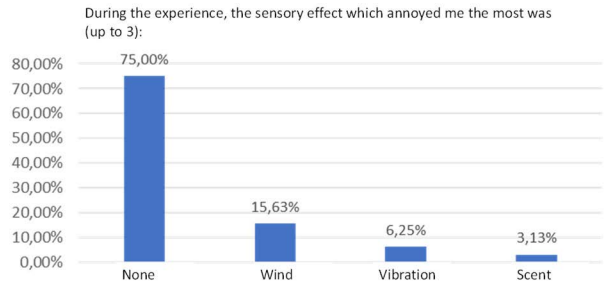

FIGURE 17. Most annoying sensory effects for participants.

4) PREFERENCES AND EXPECTATIONS

Regarding the *preferred device to be used during the consumption experience*, 72% prefered the use of an HMD for this type of experience, while 28% prefered the use of the TV. Participants were also asked if they would *prefer a multisensorial environment at home with only a TV or, with both, a TV and an HMD*. The obtained results are shown in Figures 18 and 19. Participants mostly prefered having both devices at home.

Additionally, with respect to the *participants' interest in going to events or facilities with this type of content* (Figure 20), 18.75% would always be interested in going, while 25% would go very often. The most selected answer was "from time to time" by 34.38% of the participants, while 21.88% of them would hardly ever go if given the opportunity.

Finally, when asked about their *expectations to have this type of environment at home*, 53% of the participants consider this could be a reality in a 2-5 years time, while 25% expect to have this environment in less than 2 years, 16% in 5-10 year time and 6% in more than 10 years.

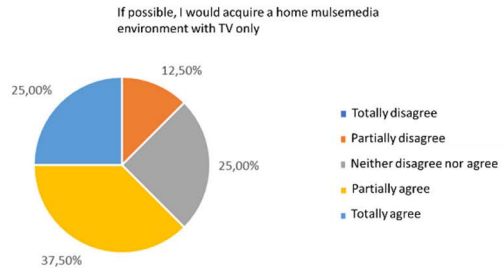

FIGURE 18. Will of participants to have a mulsemedia environment (only TV).

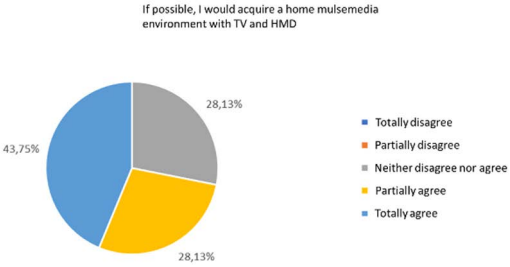

FIGURE 19. Will of participants to have a mulsemedia environment (TV and HMD).

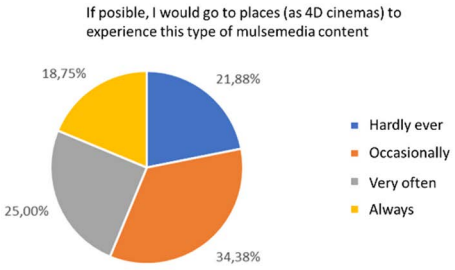

FIGURE 20. Will of participants to go to places or events where these types of contents are provided.

TABLE 8. SUS questionnaire.

Q1	I think that I would like to use this system frequently
Q2	I found the system unnecessarily complex
Q3	I thought the system was easy to use
Q4	I think that I would need the support of a technical person to be able to use this system
Q5	I found the various functions in this system were well integrated
Q6	I thought there was too much inconsistency in this system
Q7	I would imagine that most people would learn to use this system very quickly
Q8	I found the system very cumbersome to use
Q9	I felt very confident using the system things before I could get going
Q10	I needed to learn a lot of with this system

5) USABILITY

Regarding the usability of the tools and features of the developed testbed, as previously stated, the SUS questionnaire has been used. Table 8 lists the 10 questions included in it. The result provides a score from 0 to 100 points, being 0 points the worst imaginable usability score and 100 points the best imaginable one.

The obtained results are presented in Fig. 21. According to [83], they provide a score of 94.83 with a standard deviation of ± 10.19. According to [84], with this score, the overall usability of the implemented solution can be considered as *excellent*.

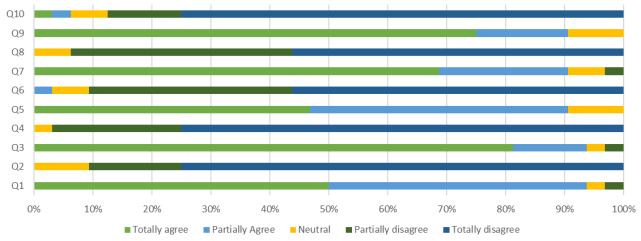

FIGURE 21. Obtained responses for the SUS questionaire.

F. LIMITATIONS OF THE STUDY

The conducted study has some limitations to take into account before extracting general conclusions. Principaly, few users have participated in it, and it was carried out in a polytechnical university environment. Considering that most of the participants were students or technical workers, it would be necessary to expand this research to involve more participants. Ideally, with more heterogeneous profiles, in order to obtain a more general view of the perception and acceptability of this new type of immersive hybrid TV content consumption. Nevertheless, previous studies point out that expected results should not be significatively different based on users' expectations [70]. Furthermore, the study has focused on conducting tests based on the use of a limited testbed with a specific AV content and just four sensory effects, generated by low-cost devices with some drawbacks (e.g., annoying noises). Tests with more AV contents and with more effects (and more scents) and with better and more expensive effect generators should be conducted. This aspect will be left for further work.

However, it has served to test the proposed end-to-end solution and to obtain valuable insights from the participants.

V. CONCLUSION AND FUTURE WORK

In this paper, an end-to-end solution for providing multisensorial and immersive hybrid TV consumption experiences has been described. It has been designed, included and subjectively evaluated in a testbed implemented in an enhanced end-to-end platform for hybrid TV content generation, delivery and consumption. That platform has been updated and enhanced to support the MPEG-V standard, including all the necessary steps from the generation by the content provider to the consumption on the consumer's devices of this type of content. In addition, a mechanism for signalling this type of content embedded in the broadcast main AV content (in the MPEG2-TS) has been proposed and implemented in the testbed, as well as a control module to manage the interpretation and sinchronised generation of the sensory effects. The usability of the implemented testbed has been subjectively

evaluated by 32 users as excellent, according to the SUS questionnaire. In the testbed the sync level between the playout of the AV content and the generation of the sensory effects perceived by those users is more than acceptable and it is even better when an HMD device is adopted as the consumer device. As in other studies, the obtained results also show that the use of that device helps to increase the participant's tolerance to small asynchrony situations, indirectly affecting their perceived QoE. Consequently, HMD device is the preferred consumption device for these types of experiences, but also because it provides a full immersion (not provided by other devices, such as TV, tablets…). In fact, most users would prefer a multi-sensorial environment at home with both devices, a TV and an HMD.

From the obtained results in the subjective assessment of the implemented testbed, it can be concluded that the inclusion of sensory effects improves the end-users' feeling of immersion considerably and, consequently, their QoE during the hybrid TV content consumption experience.

Regarding the users' preferences in relation to the included sensory effects in the testbed, there is a clear preference for the vibration effect and a significantly lower interest in ambient lighting effect, which, as explained, could be due to the different degree of novelty of both effects in current multimedia systems. Notice that some participants did not like the wind effect, and some of them have even considered it as annoying (some because of the noise of the fan but others because they were cold as a result of the wind effect).

The authors would like to emphasise that the solution proposed in this paper can work in real broadcast scenarios, since the adopted protocols, mechanisms and delivery networks are already being used in them and in current home environments. Nevertheless, despite this and the increasing interest in these types of services, like most of the participants do, authors also expect that they will become a reality in a mid- or long-term period. It is still necessary to delve a bit more into the integration of these types of effects on multimedia systems to be able to reach a compromise in assuring the users' perception of them, without negatively affecting their consumption experience. Low-noise devices should be designed and used to generate the desired sensorial effects in the least intrusive way, in order not to annoy the users during the experiences, especially when several sensory effects are combined (left for further work). More tests with different genres of AV contents and with more effects and scents should be conducted (left for further work). Additionally, the users should be capable of configuring some parameters (e.g., delay from the instant a scent generation is triggered to the instant that scent is detected by the user's nose) and the effects to be noticed, by minimising or deactivating those that can be annoying, depending on the situation. For example, the use of a wind effect with high intensity can be very annoying during the winter. For this purpose, friendly configuration interfaces should be provided (left for further work in our testbed).

ACKNOWLEDGMENT

Funding for open access charge: Universitat Politècnica de València.

REFERENCES

[1] C. Velasco and M. Obrist, *Multisensory Experiences: Where the Senses Meet Techonology*. New York, NY, USA: Oxford, 2020, p. 112, doi: 10.1093/oso/9780198849629.001.0001.

[2] G. Ghinea, C. Timmerer, W. Lin, and S. R. Gulliver, "Mulsemedia: State of the art, perspectives, and challenges," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 11, no. 1s, pp. 1–23, Oct. 2014, doi: 10.1145/2617994.

[3] H. Q. Dinh, N. Walker, L. F. Hodges, C. Song, and A. Kobayashi, "Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments," in *Proc. IEEE Virtual Reality*, Mar. 2003, pp. 222–228, doi: 10.1109/vr.1999.756955.

[4] I.-S. Comsa, R. Trestian, and G. Ghinea, "360° mulsemedia experience over next generation wireless networks—A reinforcement learning approach," in *Proc. 10th Int. Conf. Quality Multimedia Exper. (QoMEX)*, May 2018, pp. 1–6, doi: 10.1109/QoMEX.2018.8463409.

[5] Advanced Television. (2019). *Study: Super Bowl Streams 47s Behind Real-Time*. [Online]. Available: https://advanced-television.com/2019/02/06/study-super-bowl-streams-47s-behind-real-time/

[6] *ETSI TS 102 796 V1.4.1. HbbTV 2.0.2 Specification*, HbbTV Association, Geneva, Switzerland, 2018.

[7] Digital TV Europe. (2019). *Digital TV Europe Industry Survey 2019*. [Online]. Available: https://www.digitaltveurope.com/magazine/digital-tv-europe-industry-survey-2019/

[8] Advanced Television. (2019). *Study: TV Still the Dominant Screen for Youngsters*. [Online]. Available: https://advanced-television.com/2019/02/19/study-more-screen-time-for-young-children/

[9] *Mind the Gap—A Closer Look at Video Advertising Reach in the Age of Increasing Media Fragmentation*, Ebiquity, London, U.K., 2020.

[10] IAB Spain. (2019). *Estudio de Televisión Conectada 2019*. [Online]. Available: https://iabspain.es/estudio/estudio-de-television-conectada-2019/

[11] D. Marfil, F. Boronat, A. Sapena, and A. Vidal, "Synchronization mechanisms for multi-user and multi-device hybrid broadcast and broadband distributed scenarios," *IEEE Access*, vol. 7, pp. 605–624, 2019, doi: 10.1109/ACCESS.2018.2885580.

[12] L. Jalal and M. Murroni, "Enhancing TV broadcasting services: A survey on mulsemedia quality of experience," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast. (BMSB)*, Jun. 2017, pp. 1–7, doi: 10.1109/BMSB.2017.7986192.

[13] *MPEG-V. Media Context and Control*, document MPEG, ISO/IEC 23005, 2010.

[14] Y. Sulema, "Mulsemedia vs. multimedia: State of the art and future trends," in *Proc. Int. Conf. Syst., Signals Image Process. (IWSSIP)*, May 2016, pp. 1–5, doi: 10.1109/IWSSIP.2016.7502696.

[15] N. Murray, O. A. Ademoye, G. Ghinea, and G.-M. Muntean, "A tutorial for olfaction-based multisensorial media application design and evaluation," *ACM Comput. Surv.*, vol. 50, no. 5, pp. 1–30, Sep. 2018, doi: 10.1145/3108243.

[16] E. B. Saleme, A. Covaci, G. Mesfin, C. A. S. Santos, and G. Ghinea, "Mulsemedia DIY: A survey of devices and a tutorial for building your own mulsemedia environment," *ACM Comput. Surv.*, vol. 52, no. 3, pp. 1–29, May 2020, doi: 10.1145/3319853.

[17] A. Covaci, R. Trestian, E. B. Saleme, I.-S. Comsa, G. Assres, C. A. S. Santos, and G. Ghinea, "360° mulsemedia: A way to improve subjective QoE in 360° videos," in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 2378–2386, doi: 10.1145/3343031.3350954.

[18] E. B. Saleme, C. A. S. Santos, R. A. Falbo, G. Ghinea, and F. Andres, "Towards a reference ontology on mulsemedia systems," in *Proc. 10th Int. Conf. Manage. Digit. EcoSyst.*, Sep. 2018, pp. 23–30, doi: 10.1145/3281375.3281378.

[19] Z. Yuan, S. Chen, G. Ghinea, and G.-M. Muntean, "User quality of experience of mulsemedia applications," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 11, no. 1, pp. 1–19, 2014, doi: 10.1145/2661329.

[20] Z. Akhtar, K. Siddique, A. Rattani, S. L. Lutfi, and T. H. Falk, "Why is multimedia quality of experience assessment a challenging problem?" *IEEE Access*, vol. 7, pp. 117897–117915, 2019, doi: 10.1109/ACCESS.2019.2936470.

[21] M. Melo, H. Coelho, G. Gonçalves, N. Losada, F. Jorge, M. S. Teixeira, and M. Bessa, "Immersive multisensory virtual reality technologies for virtual tourism: A study of the user's sense of presence, satisfaction, emotions, and attitudes," *Multimedia Syst.*, vol. 28, no. 3, pp. 1027–1037, Jun. 2022, doi: 10.1007/s00530-022-00898-7.

[22] *Sensory Analysis Methodology Initiation and Training of Assessors in the Detection and Recognition of Odours*, document ISO 5496:2006, International Organization for Standardization, 2006.

[23] M. Waltl, B. Rainer, C. Timmerer, and H. Hellwagner, "An end-to-end tool chain for sensory experience based on MPEG-V," *Signal Process., Image Commun.*, vol. 28, no. 2, pp. 136–150, Feb. 2013, doi: 10.1016/J.IMAGE.2012.10.009.

[24] B. S. Choi, S. H. Joo, and H. Y. Lee, "Sensory effect metadata for SMMD media service," in *Proc. 2009 4th Int. Conf. Internet Web Appl. Services, (ICIW)*, vol. 2009, 2009, pp. 649–654, doi: 10.1109/ICIW.2009.104.

[25] Z. Yuan, G. Ghinea, and G. M. Muntean, "Beyond multimedia adaptation: Quality of experience-aware multi-sensorial media delivery," *IEEE Trans. Multimedia*, vol. 17, no. 1, pp. 104–117, Jan. 2015, doi: 10.1109/TMM.2014.2371240.

[26] E. B. Saleme, C. A. S. Santos, and G. Ghinea, "Coping with the challenges of delivering multiple sensorial media," *IEEE Multimedia Mag.*, vol. 26, no. 2, pp. 66–75, Apr. 2019.

[27] E. B. Saleme and C. A. S. Santos, "PlaySEM: A platform for rendering MulSeMedia compatible with MPEG-V," in *Proc. 21st Brazilian Symp. Multimedia Web*, Oct. 2015, pp. 145–148, doi: 10.1145/2820426.2820450.

[28] E. B. Saleme, C. A. S. Santos, and G. Ghinea, "A mulsemedia framework for delivering sensory efects to heterogeneous systems," *Multimedia Syst.*, vol. 25, pp. 421–447, May 2019, doi: 10.1007/s00530-019-00618-8.

[29] M. Waltl, B. Rainer, C. Timmerer, and H. Hellwagner, "A toolset for the authoring, simulation, and rendering of sensory experiences," in *Proc. 20th ACM Int. Conf. Multimedia (MM)*, 2012, pp. 1469–1472, doi: 10.1145/2393347.2396522.

[30] S.-H. Shin, K.-S. Ha, H.-O. Yun, and Y.-S. Nam, "Realistic media authoring tool based on MPEG-V international standard," in *Proc. 8th Int. Conf. Ubiquitous Future Netw. (ICUFN)*, Jul. 2016, pp. 730–732, doi: 10.1109/ICUFN.2016.7537133.

[31] J.-K. Yun, J.-H. Jang, and K.-D. Moon, "Five sense media playback technology using multile devices synchronization," in *Proc. 5th Int. Conf. Comput. Sci. Converg. Inf. Technol.*, Nov. 2010, pp. 73–75, doi: 10.1109/ICCIT.2010.5711032.

[32] *MPEG, MPEG-7 Part 5: MDS*, document ISO/IEC 15938-5, 2003.

[33] M. Waltl, C. Timmerer, and H. Hellwagner, "A test-bed for quality of multimedia experience evaluation of sensory effects," in *Proc. Int. Workshop Quality Multimedia Exper.*, Jul. 2009, pp. 145–150, doi: 10.1109/QOMEX.2009.5246962.

[34] M. Josué, R. Abreu, F. Barreto, D. Mattos, G. Amorim, J. dos Santos, and D. Muchaluat-Saade, "Modeling sensory effects as first-class entities in multimedia applications," in *Proc. 9th ACM Multimedia Syst. Conf.*, Jun. 2018, pp. 225–236, doi: 10.1145/3204949.3204967.

[35] J. P. Sexton, A. A. Simiscuka, K. Mcguinness, and G.-M. Muntean, "Automatic CNN-based enhancement of 360° video experience with multisensorial effects," *IEEE Access*, vol. 9, pp. 133156–133169, 2021, doi: 10.1109/ACCESS.2021.3115701.

[36] I. S. Comsa, E. B. Saleme, A. Covaci, G. M. Assres, R. Trestian, C. A. Santos, and G. Ghinea, "Do I smell coffee? The tale of a 360° mulsemedia experience," *IEEE Multimedia*, vol. 27, no. 1, pp. 27–36, Mar. 2020, doi: 10.1109/MMUL.2019.2954405.

[37] G. Brianza, P. Cornelio, E. Maggioni, and M. Obrist, "Sniff before you act: Exploration of scent-feature associations for designing future interactions," in *Proc. IFIP Conf. Hum.-Comput. Interact.* in Lecture Notes in Computer Science, vol. 12933, 2021, pp. 281–301, doi: 10.1007/978-3-030-85616-8_17.

[38] D. Narciso, M. Melo, J. Vasconcelos-Raposo, and M. Bessa, "The impact of olfactory and wind stimuli on 360 videos using head-mounted displays," *ACM Trans. Appl. Perception*, vol. 17, no. 1, pp. 1–13, Jan. 2020, doi: 10.1145/3380903.

[39] C. Timmerer, M. Waltl, B. Rainer, and N. Murray, *Sensory Experience: Quality of Experience Beyond Audio-Visual*. Cham, Switzerland: Springer, 2014, pp. 351–365.

[40] D. A. Washburn and L. M. Jones, "Could olfactory displays improve data visualization?" *Comput. Sci. Eng.*, vol. 6, no. 6, pp. 80–83, Nov. 2004, doi: 10.1109/MCSE.2004.66.

[41] S. Brewster, D. McGookin, and C. Miller, "Olfoto: Designing a smell-based interaction," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Apr. 2006, p. 653, doi: 10.1145/1124772.1124869.

[42] B. Ramic-Brkic and A. Chalmers, "Virtual smell: Authentic smell diffusion in virtual environments," in *Proc. 7th Int. Conf. Comput. Graph., Virtual Reality, Visualisation Interact. Afr. (AFRIGRAPH)*, 2010, p. 45, doi: 10.1145/1811158.1811166.

[43] B. Ramic-Brkic and A. Chalmers, "Olfactory adaptation in virtual environments," *ACM Trans. Appl. Perception*, vol. 11, no. 2, pp. 1–16, Jul. 2014, doi: 10.1145/2617917.

[44] B. Ramic, A. Chalmers, J. Hasic, and S. Rizvic, "Selective rendering in a multi-modal environment: Scent and graphics," in *Proc. 23rd Spring Conf. Comput. Graph. (SCCG)*, 2007, pp. 147–151, doi: 10.1145/2614348.2614369.

[45] 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. 1–14, Aug. 2016, doi: 10.1145/2957753.

[46] K. Hirota, S. Ebisawa, T. Amemiya, and Y. Ikei, "A theater for viewing and editing multi-sensory content," in *Proc. IEEE Int. Symp. VR Innov.*, Mar. 2011, pp. 239–244, doi: 10.1109/ISVRI.2011.5759643.

[47] Z. Yuan, G. Ghinea, and G.-M. Muntean, "Quality of experience study for multiple sensorial media delivery," in *Proc. Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Aug. 2014, pp. 1142–1146, doi: 10.1109/IWCMC.2014.6906515.

[48] M. Waltl, C. Timmerer, B. Rainer, and H. Hellwagner, "Sensory effects for ambient experiences in the world wide Web," *Multimedia Tools Appl.*, vol. 70, no. 2, pp. 1141–1160, May 2014, doi: 10.1007/s11042-012-1099-8.

[49] V. C. Thanh, D. Ablart, E. Gatti, C. Velasco, and M. Obrist, "Not just seeing, but also feeling art: Mid-air haptic experiences integrated in a multisensory art exhibition," *Int. J. Hum.-Comput. Stud.*, vol. 108, pp. 1–14, Dec. 2017, doi: 10.1016/j.ijhcs.2017.06.004.

[50] N. Ranasinghe, K.-Y. Lee, G. Suthokumar, and E. Y.-L. Do, "Virtual ingredients for food and beverages to create immersive taste experiences," *Multimedia Tools Appl.*, vol. 75, no. 20, pp. 12291–12309, Oct. 2016, doi: 10.1007/s11042-015-3162-8.

[51] A. Covaci, E. B. Saleme, G. A. Mesfin, N. Hussain, E. Kani-Zabihi, and G. Ghinea, "How do we experience crossmodal correspondent mulsemedia content?" *IEEE Trans. Multimedia*, vol. 22, no. 5, pp. 1249–1258, May 2020, doi: 10.1109/TMM.2019.2941274.

[52] M. Waltl, C. Timmerer, and H. Hellwagner, "Improving the quality of multimedia experience through sensory effects," in *Proc. 2nd Int. Workshop Quality Multimedia Exper. (QoMEX)*, Jun. 2010, pp. 124–129, doi: 10.1109/QOMEX.2010.5517704.

[53] C. Spence, "Crossmodal correspondences: A tutorial review," *Attention, Perception, Psychophys.*, vol. 73, no. 4, pp. 971–995, May 2011, doi: 10.3758/s13414-010-0073-7.

[54] J. K. Yun, M. G. Kim, J. H. Jang, and K. R. Park, "Media/playback device synchronization for the 4D broadcasting service system," in *Proc. IEEE Int. Conf. Consum. Electron.*, Jan. 2012, pp. 329–330, doi: 10.1109/ICCE.2012.6161891.

[55] K. Yoon, B. Choi, E.-S. Lee, and T.-B. Lim, "4-D broadcasting with MPEG-V," in *Proc. IEEE Int. Workshop Multimedia Signal Process.*, Oct. 2010, pp. 257–262, doi: 10.1109/MMSP.2010.5662029.

[56] L. Jalal, M. Anedda, V. Popescu, and M. Murroni, "QoE assessment for broadcasting multi sensorial media in smart home scenario," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast. (BMSB)*, Jun. 2018, pp. 1–5, doi: 10.1109/BMSB.2018.8436875.

[57] L. Jalal, M. Anedda, V. Popescu, and M. Murroni, "Internet of Things for enabling multi sensorial TV in smart home," in *Proc. IEEE Broadcast Symp. (BTS)*, Oct. 2018, pp. 1–5, doi: 10.1109/BTS.2018.8550959.

[58] J. K. Yun, J. H. Jang, and K. D. Moon, "Development of the real-sense media broadcasting service system based on the SMMD," in *Proc. IEEE Int. Conf. Consum. Electron.*, Jan. 2011, pp. 435–436, doi: 10.1109/ICCE.2011.5722669.

[59] M. Montagud, F. Boronat, H. Stokking, and R. van Brandenburg, "Inter-destination multimedia synchronization: Schemes, use cases and standardization," *Multimedia Syst.*, vol. 18, no. 6, pp. 459–482, 2012, doi: 10.1007/s00530-012-0278-9.

[60] W. J. Kooij, H. M. Stokking, R. van Brandenburg, and P.-T. de Boer, "Playout delay of TV signals: Measurement system design, validation and results," in *Proc. ACM Int. Conf. Interact. Exper. TV Online Video (TVX)*, 2014, pp. 23–30, doi: 10.1145/2602299.2602310.

[61] M. O. van Deventer, H. M. Stokking, O. A. Niamut, F. A. Walraven, and V. B. Klos, "Advanced interactive television services require content synchronization," in *Proc. 15th Int. Conf. Syst., Signals Image Process.*, Jun. 2008, pp. 109–112, doi: 10.1109/IWSSIP.2008.4604379.

[62] R. N. Mekuria. (2011). *Inter-Destination Media Synchronization for TV Broadcasts?: TU Delft Institutional Repository*. [Online]. Available: https://repository.tudelft.nl/islandora/object/uuid:61907582-3fe2-4e1f-86af-47c0eacbfb04

[63] *Network Model for Evaluating Multimedia Transmission Performance Over Internet Protocol*, document G.1050, International Telecommunication Union (ITU), 2016.

[64] J. K. Yun, J. H. Jang, K. R. Park, and D. W. Han, "Real-sense media representation technology using multiple devices synchronization," in *Proc. IFIP Int. Workshop Softw. Technolgies Embedded Ubiquitous Syst.* in Lecture Notes in Computer Science, vol. 5860, 2009, pp. 343–353, doi: 10.1007/978-3-642-10265-3_31.

[65] N. Murray, G. M. Muntean, Y. Qiao, and B. Lee, "Olfaction-enhanced multimedia synchronization," in *MediaSync: Handbook Multimedia Synchronization*. Cham, Switzerland: Springer, 2018, pp. 319–356.

[66] N. Murray, Y. Qiao, G.-M. Muntean, and B. Lee, "Multiple-scent enhanced multimedia synchronization," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 11, no. 1, pp. 1–28, 2014, doi: 10.1145/2637293.

[67] Z. Yuan, T. Bi, G. M. Muntean, and G. Ghinea, "Perceived synchronization of mulsemedia services," *IEEE Trans. Multimedia*, vol. 17, no. 7, pp. 957–966, Jul. 2015, doi: 10.1109/TMM.2015.2431915.

[68] R. Frederick, *RTP?: A Transport Protocol for Real-Time Applications*, document Internet Soc. RFC 3550, 2003, pp. 1–89.

[69] D. Marfil, F. Boronat, M. Montagud, and A. Sapena, "IDMS solution for hybrid broadcast broadband delivery within the context of HbbTV standard," *IEEE Trans. Broadcast.*, vol. 65, no. 4, pp. 645–663, Dec. 2019, doi: 10.1109/TBC.2018.2878285.

[70] F. Boronat, M. Montagud, D. Marfil, and C. Luzón, "Hybrid broadcast/broadband TV services and media synchronization: Demands, preferences and expectations of Spanish consumers," *IEEE Trans. Broadcast.*, vol. 64, no. 1, pp. 52–69, Mar. 2018, doi: 10.1109/TBC.2017.2737819.

[71] F. Boronat, D. Marfil, M. Montagud, and J. Pastor, "HbbTV-compliant platform for hybrid media delivery and synchronization on single- and multi-device scenarios," *IEEE Trans. Broadcast.*, vol. 64, no. 3, pp. 721–746, Sep. 2018, doi: 10.1109/TBC.2017.2781124.

[72] D. Marfil, F. Boronat, J. Lopez, and A. Vidal, "Enhancing the broadcasted TV consumption experience with broadband omnidirectional video content," *IEEE Access*, vol. 7, pp. 171864–171883, 2019, doi: 10.1109/ACCESS.2019.2956084.

[73] *Delivery of Timeline for External Data*, document ISO/IEC, ISO/IEC 13818-1:2013/PDAM 6, 2013.

[74] *Information Technology Generic Coding of Moving Pictures and Associated Audio Information*, document ISO/IEC, ISO/IEC 13818-1, 2018.

[75] *DIAL. Discovery and Launch Protocol Specificacion. Version 1.7.2*, Netflix, Scotts Valley, CA, USA, 2015.

[76] *ETSI TS 102 796 V1.6.1. HbbTV 2.0.3 Specification*, HbbTV Association, Grand-Saconnex, Switzerland, 2020.

[77] I. Grigorik, *High-Performance Browser Networking*. Sebastopol, CA, USA: O'Reilly Media, 2013.

[78] GPAC. *GPAC Framework*. Accessed: Jul. 2022. [Online]. Available: https://gpac.wp.imt.fr/

[79] *Methods for the Subjective Assessment of Video Quality, Audio Quality and Audiovisual Quality of Internet Video and Distribution Quality Television in any Environment*, document ITU-T Recommendation P.913, Recomm. ITU-T P.913, 2016.

[80] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document ITU-R BT.500-13, ITU, no. BT.500-13, 2012, pp. 1–48, vol. 211.

[81] J. Brooke, "SUS-A quick and dirty usability scale," *Usability Eval. Ind.*, vol. 189, no. 194, pp. 4–7, 1996.

[82] *Methods for Objective and Subjective Assessment of Quality*, document ITU-T P.800, International Telecommunication Union, 1998.

[83] J. Brooke, "SUS—A quick and dirty usability scale industrial usability evaluation," in *Usability Evaluation in Industry*, I. Jordan, P. W. Thomas, B. A. Weerdmeester and McClelland, Ed. London, U.K.: Taylor & Francis, 2011, pp. 189–194.

[84] A. Bangor, P. Kortum, and J. Miller, "Determining what individual SUS scores mean: Adding an adjective rating scale," *J. Usability Stud.*, vol. 4, no. 3, pp. 114–123, May 2009.

DANIEL MARFIL was born in Gandia, Spain. He received the B.Sc. degree in informatics technical engineering, the B.Sc. degree in telecommunications, and the M.Sc. degree in telecommunication technologies, systems and networks from the Universitat Politècnica de València (UPV), Spain, in 2011, 2015, and 2016, respectively, where he is currently pursuing the Ph.D. degree with the Immersive Interactive Media Research and Development Group. He is an Assistant Researcher and a Developer with the Immersive Interactive Media Research and Development Group, UPV. He has authored one book chapter and several research and conference papers. His main research interests include communication networks, code developing, and media synchronization.

FERNANDO BORONAT (Senior Member, IEEE) was born in Gandia, Spain. He received the bachelor's degree in telecommunications engineering from the Universitat Politècnica de València (UPV), Spain, and the M.E. and Ph.D. degrees in telecommunication engineering from UPV, in 1994 and 2004, respectively. After working for several Spanish telecommunication companies, he moved back to UPV, in 1996. He is currently an Assistant Professor with the Department of Communications. He is also the Head of the Immersive Interactive Media Research and Development Group (http://iim.webs.upv.es), UPV, Gandia Campus. He has authored two books, several book chapters, an IETF RFC, and more than 100 research articles. He is involved in several IPCs of national and international journals and conferences. His main research interests include communication networks, multimedia systems, multimedia protocols, and media synchronization. He is a member of ACM. He is an Editor of *MediaSync: Handbook on Multimedia Synchronization* (Springer, 2018).

JUAN GONZÁLEZ was born in Simat de la Valldigna, Spain. He received the B.Sc. degree in telecommunications from the Universitat Politècnica de València (UPV), in 2016, and the M.Sc. degree in bioinformatics and biostatistics from the Universitat Oberta de Catalunya i la Universitat de Barcelona. He is currently pursuing the Ph.D. degree with the Immersive Interactive Media Research and Development Group, UPV. He is an Assistant Researcher and a Developer with the Immersive Interactive Media Research and Development Group, UPV. His main research interests include web and mobile application development, virtual reality, augmented reality, and machine learning techniques.

ALMANZOR SAPENA received the B.S. degree in mathematics from the University of Valencia and the Ph.D. degree from the Universitat Politècnica de València (UPV), Spain, in 2002, with a focus of topological properties in fuzzy metric spaces. He is currently an Associate Professor with the Department of Applied Mathematics, UPV, and is doing research on fuzzy topology and noise reduction in digital images and on adaptive media playout techniques. He has published some papers on these topics in international journals and conferences. He is a member of the Immersive Interactive Media Research and Development Group, UPV, Gandia Campus.