IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received November 23, 2019, accepted November 23, 2019, date of publication November 26, 2019, date of current version December 12, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2956084

Enhancing the Broadcasted TV Consumption Experience With Broadband Omnidirectional Video Content

DANI MARFIL[1], FERNANDO BORONAT[1], (Senior Member, IEEE), JAIR LÓPEZ[1], AND ANNA VIDAL[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: Dani Marfil (damarre@dcom.upv.es)

This work was supported by the Generalitat Valenciana, Investigación Competitiva Proyectos, through the Research and Development Program Grants for Research Groups to be Consolidated, under Grant AICO/2017/059 and Grant AICO/2017.

ABSTRACT The current wide range of heterogeneous consumption devices and delivery technologies, offers the opportunity to provide related contents in order to enhance and enrich the TV consumption experience. This paper describes a solution to handle the delivery and synchronous consumption of traditional broadcast TV content and related broadband omnidirectional video content. The solution is intended to support both hybrid (broadcast/broadband) delivery technologies and has been designed to be compatible with the Hybrid Broadcast Broadband TV (HbbTV) standard. In particular, some specifications of HbbTV, such as the use of global timestamps or discovery mechanisms, have been adopted. However, additional functionalities have been designed to achieve accurate synchronization and to support the playout of omnidirectional video content in current consumption devices. In order to prove that commercial hybrid environments could be immediately enhanced with this type of content, the proposed solution has been included in a testbed, and objectively and subjectively evaluated. Regarding the omnidirectional video content, the two most common types of projections are supported: equirectangular and cube map. The results of the objective assessment show that the playout of broadband delivered omnidirectional video content in companion devices can be accurately synchronized with the playout on TV of traditional broadcast 2D content. The results of the subjective assessment show the high interest of users in this type of new enriched and immersive experience that contributes to enhance their Quality of Experience (QoE) and engagement.

INDEX TERMS 360 video, broadband, broadcast, DASH, DVB, HTTP-based adaptive streaming, hybrid TV, IDES, inter-device synchronization, media synchronization, omnidirectional video.

I. INTRODUCTION

Nowadays, the use of hybrid technologies (i.e., broadcast and broadband) for the delivery and consumption of TV related content is a reality. Moreover, the wide range of heterogeneous consumption devices (e.g., from traditional TVs to smartphones), offers the opportunity to broadcasters and network operators to deliver different related content in order to enrich the TV consumption experience. This combination of delivery technologies (and also, consumption devices) enable new scenarios, in which traditional TV broadcast content can be complemented with additional related contents delivered via broadband networks. Those contents can be consumed synchronously in the same device (a.k.a. hybrid device) or in several devices (e.g., [1]). An evidence of the efforts in this direction is the Hybrid Broadcast Broadband TV (HbbTV) standard [2], which provides mechanisms for harmonizing the delivery and consumption of interactive broadcast and broadband TV-related contents through connected TVs and companion devices.

Moreover, there is a current trend that also evidences that content consumption through the Internet (by using different devices) is increasing year after year. For instance, according to [3], on the one hand, Netflix has been the 9th most

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

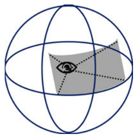

FIGURE 1. User's Field of View.

downloaded application for smartphones in 2018 and, on the other hand, 92% of Internet users claim that they watch videos online.

During the last few years, omnidirectional video (a.k.a. 360 video) content generation and consumption has been also experiencing a significant increase. Moreover, as stated in [4], 11% of US users claim to own virtual reality (VR) hardware (i.e., Head Mounted Displays or HMD) as of November 2018, compared to August 2017, when roughly a 7% owned these devices. According to these data, authors expect that HMD devices will be a more common device to consume broadband omnidirectional content related to a main broadcast TV content in the future. This will provide more immersive TV consumption experiences. Actually, Spanish public broadcaster, RTVE, has already experimentally introduced VR and interactivity in TV series 'The Ministry of Time' and 'The Case'. Moreover, a proof of the growing demand of omnidirectional content is the many efforts to efficiently encode and deliver this type of content (as explained in Section II). Delivery of omnidirectional content is a key challenge. End users only watch a small percentage of the whole 360 view (a.k.a. Field of View or FoV), as shown in Figure 1. If all the available 360 content is delivered to the user, it will imply an inefficient use of the bandwidth, due to the fact that the number of pixels being watched by the user compared to the total number of delivered pixels, is small.

Regarding the simultaneous hybrid content consumption in several devices (or even in the same device), there are still many challenges to face, and media synchronization (sync, hereafter) is one of them. The term *hybrid sync* is commonly used when different related streams, delivered via hybrid broadcast/broadband technologies, need to be played out in a synchronous way. New hybrid sync mechanisms are needed, and the existing ones still need to be improved. For instance, in the last Super Bowl event in 2018, in which live IP (i.e., broadband) streams were requested by 20% of viewers, some latency measurements were conducted. "Live streams" were anywhere from 28 to 47 seconds behind real-time [5]. Those latency values can be considered as tolerable when that source is the only one from which the content comes from. Nevertheless, when the available stream content is considered as an extension or complement of the broadcast TV content, sync mechanisms must be implemented. Moreover, related to the aforementioned HbbTV standard [2], a survey in [6] shows that, on the one hand, 35.8% of respondents think that the deployment of these services is still behind expectations. On the other hand, despite of the pessimistic vision of the HbbTV deployment, most of respondents think that HbbTV is a very attractive (23.2%) or moderately attractive (45.2%) technology of advanced TV services.

Notwithstanding the increase of content consumption through broadband technologies, traditional broadcast TV consumption is still the main source of entertainment for young people [7]. The DTG Summit 2018[1], held in London in May, explored the changes affecting the broadcasting world, envisaging how TV is set to evolve over the next 10 to 20 years. Although IP delivery already have a real impact on the TV industry, offering viewers unrivalled choice and flexibility, the less noticed and discussed trend is that, in Europe, DVB-based Digital Terrestrial Television (DTT) continues to grow its overall number of TV households over the years, and to evolve and perform very well in the market (see FEH Media Insight blog[2] for more details). DTT has served people well for so long and remain the mainstay of modern TV for many segments of viewers. The evolution to a Hybrid TV service (DTT and IP) has meant that DTT is now very well placed to underpin the market as the market moves towards a pay-lite model. Viewers can receive a free base of content delivered via DTT and overlay an IP-delivered offer or catch-up and Subscription Video on Demand (SVoD) services from a range of providers. The success and appetite from viewers for the hybrid model suggest that an IP-only future is still a very long way off. Instead, continuing to blend the best of IP and DTT into a workable hybrid offer that meets the needs of nowadays consumer is a model which is set to remain relevant for many years to come.

Thus, this is the main reason why we propose in this paper a hybrid solution in which complementary (omnidirectional) video content is delivered via broadband technology (IP-based networks), but keeping the content being delivered through the conventional broadcast TV technology (e.g., DVB) as the main content. In this manner, the solution is also backwards compatible. Moreover, to provide a realistic and close-to-commercial hybrid environment, the proposed solution has been designed to be compatible with the HbbTV standard [2].

The proposed solution in this paper has been designed to improve a previous solution presented by the authors in [1], with more and better functionalities. The main contributions of this paper regarding that previous work are the following:

- An enhanced end-to-end hybrid delivery and consumption platform, including omnidirectional video content as broadband content.
- A new mechanism to signal available omnidirectional video content within the broadcast TV media streams.
- Design of a multi-platform, responsive and adaptive web-based immersive omnidirectional scenario for the consumption of the multiple available video content through traditional (laptops, smartphones and tablets) and VR devices, such as standalone Head Mounted Displays (HMD) or smartphones to be embedded in HMDs.

[1] https://dtg.org.uk/the-dtg-summit-2018/

[2] http://www.feh-mi.com/blog/the-dtt-platform-succeeding-in-the-face-of-the-svod-challenge/

- Design and implementation of an adaptive omnidirectional video player supporting the two most common types of projections: equirectangular and cube map.
- Implementation of an inter-device sync mechanism between Main Screen and Companion Screens when playing related 360 video content, following the specifications of HbbTV standard.
- Evolved testbed to assess in close-to-commercial scenarios the proposed hybrid environment involving synchronized playout of traditional broadcast TV content and omnidirectional video content, by adopting the designed functionalities (e.g., the signalling mechanisms or the omnidirectional player) in this work.
- Objective evaluation of the sync performance between the playout of the broadcast TV content and the one of the broadband omnidirectional video content, under non-ideal network conditions, obtaining satisfactory results.
- Subjective assessment of the designed immersive scenario, the overall usability of the proposed solution and the perceived level of achieved sync, obtaining very optimistic results and very good levels of the users' perceived Quality of Experience (QoE).

The paper is structured as follows. In Section II, previous works regarding omnidirectional video content and hybrid sync are described. In Section III, the proposed solution and its functionalities are presented. In Section IV, the designed virtual environment for the immersive consumption of the available broadband related content is described. In Section V, a developed testbed implementing the proposed solution is described. In Section VI, the conducted objective and subjective evaluations and the obtained results are presented and discussed. Finally, the paper ends with some conclusions and future work in Section VII.

II. RELATED WORK

In this section, some issues and related works regarding omnidirectional video content and the associated challenges in order to achieve hybrid sync are summarized.

A. OMNIDIRECTIONAL VIDEO CONTENT

Currently, video generation has evolved from traditional 2D video content to a more immersive video content, such as the omnidirectional video (360) or even volumetric (i.e., holograms) video content [8]. This new wave of video content requires more bandwidth and the support of technologies which allow higher quality encoding. The MPEG-I standard [9] specifies these emergent types of video content. Its main aim is the standardization of immersive visual media. Features of the immersive visual media, distinguishing it from the classical 2D media are described in that standard. Some of these features are: the generation of 360 or stereoscopic content (and the many possibilities and mechanisms to create it), the adoption of new encoding techniques for high resolution content (such as High Efficiency Video Coding, HEVC [10]), the usage of specific displays such as the HMDs, etc.

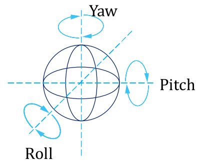

FIGURE 2. 3 DoF representation: yaw, pitch and roll.

Regarding the ability to consume 360 video content, there are different levels of freedom. A three Degrees of Freedom (3DoF) video content allows the user to watch a 360 video with an HMD. Actually, the user chooses what portion of the content to watch (FoV, Figure 1) with the head's rotation. Yaw, pitch and roll movements allow the user to watch the different angles of the 360 video content (see Figure 2). Although this is the most common type of omnidirectional video content, it is not enough to provide a full immersive experience. In fact, a so-called 3DoF+ concept has been described as an evolution from the 3DoF feature [11]. It allows limited movements along the yaw, pitch and roll axis, which requires the content to have additional depth and texture information. Finally, the most immersive feature is 6DoF, where the user can ''freely'' move through the VR scene without limitations.

1) VIDEO PROJECTIONS

When generating content with the use of 360 cameras, the way the resulting images are projected to flat planes is not evident [12]. During the generation process, some drawbacks such as distortion may happen. For instance, multiple local projections with multiple viewpoints to create more realistic projections are proposed in [12]. Moreover, in [13], an extensive review of the different types of projection and quality assessment methodologies is presented. In that work, it is stated that video projections can be classified in two categories: viewport-independent (such as the Equirectangular Projection -ERP-, the Cylindrical Equal-Area Projection -EAP- or the Cube Map Projection -CMP-); and viewport-dependent, which only takes into account the user's region of interest (i.e., the FoV), such as the pyramid projection proposed by Facebook in [14]. Examples of both types of projections are shown in Figure 3.

2) STREAMING OF OMNIDIRECTIONAL VIDEO CONTENT

As it can be seen, omnidirectional content requires higher bandwidth demands than traditional video content. This is due, as stated in the previous section, to the fact that the FoV only involves small regions of the whole omnidirectional video content. Therefore, a solution called *tiled-streaming* has been proposed and, actually, it is included in standards such as HEVC [10]. This solution consists in dividing the content in tiles, so the users can demand only specific regions of the content and thus, save bandwidth resources. Actually, in [15], HEVC is adopted to create two tiled versions of Virtual Reality (VR) content, one with high bitrate and

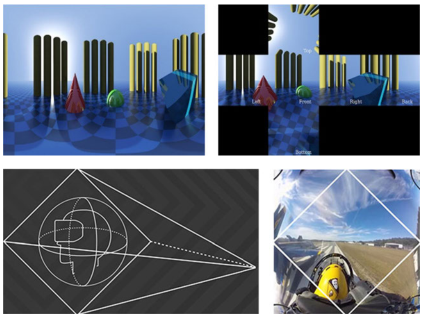

FIGURE 3. Viewport-independent[3] ERP (top left), CMP (top right) projections and viewport-dependent pyramid projection (bottom).

another one with lower bitrate. Results presented in that work show that up to 30-40% of the bitrate is saved when providing two qualities compared to when only the high bitrate version of the content is provided. Authors in that work use an HMD device to perform the evaluation.

A different technique to save bandwidth costs can be found in [16], where an open-source Dynamic Adaptive Streaming over HTTP (DASH, [17]) video player is presented, although the device used for the experimental assessment is not specified. It can manage tiled-based content in order to acquire the highest possible quality for the user's FoV while efficiently managing the rest of tiles outside the FoV. For this, it is proposed to include the MPEG-DASH's Spatial Representation Description feature (DASH-SRD [18]). This feature is defined in the second amendment of MPEG-DASH standard part 1, 23009-1:2014. There, an approach for streaming only specific tiles of a content is defined. It is achieved by extending the Media Presentation Description (MPD) file with a description of the spatial relationships between tiles. Additionally, in [19], DASH-SRD is also adopted to provide a more bandwidth efficient solution for streaming 360 content. In that work, an experimental assessment is carried out with the Samsung Gear VR with a Samsung S7 smartphone and the player in that work is not implemented for browsers. In that work, it is concluded that up to 72% of the bandwidth can be saved by using the SRD feature defined in [18].

In [20], three mechanisms of tile-based streams delivery are described (from least to most bandwidth efficient): 1) *full delivery basic*, in which all the involved tiles are delivered in the highest feasible quality; 2) *full delivery advanced*: in which all the tiled-content is delivered, but tiles outside the user's FoV are requested in a lower quality than the ones inside the user's FoV; and 3) *partial delivery*, in which only the tiles inside the user's FoV are delivered in the highest possible quality and the other tiles are not even requested. In that work, an HMD device is used for the assessment, although the specific model is not specified and the involved player does not run in browsers.

However, as stated in [21], the number of requests made by clients to get the tiled-based video content, when HTTP-based Adaptive Streaming (HAS) technologies, such as DASH, are adopted, is very high. As an example, authors in that work describe that tiling the video introduces an overhead of requests compared to the non-tiled version, which varies between 6% and 22%. To solve this issue, the use of the server-push functionality of the HTTP/2 protocol is proposed. That way, a client only requests the content once and afterwards, all the involved tiles are pushed from the server automatically. In that work, tiles are also sent bandwidth-efficiently, i.e., only tiles inside the user's FoV are sent at the highest possible quality.

While tile-based streaming requires minimal changes on the server side, it implies an important load on consumption devices, which need to determine the tiles to download, decode and merge for rendering and display. The prediction of FoV considering the user's movements or interaction require complex algorithms. Moreover, considering that the decoding capability of mobile devices is limited by their hardware video decoders, tile-based 360 video streaming requires very powerful mobile devices to be able to play the videos smoothly. Edge computing-based solutions can bring both processing and content close to end users, but it is an area still under research and development. In [22], a multi-viewpoint (MVP) 360 content acquisition and navigation framework is presented. In particular, it enables the navigation through some specific viewpoints (VP) in the 360 content, enabling six DoF (6DoF) for the users. In order to allow fast VP switching, authors in that work state that the translation movements in the 6DoF scene should be anticipated. For this, two different client strategies, a proactive one and a reactive one, are considered. In order to provide fast VP switching, both take into account three variables: 1) the distortion and the bit rate of each encoded video segment; 2) the user trajectories inside the MVP video; and 3) the average download bandwidth available during each downloaded chunk. Tiled-content streaming is shown to improve the overall performance of the framework and proactive strategies are less bandwidth-efficient than reactive ones.

In [23], a commercial platform to consume omnidirectional video content in a live video sharing website is presented. During a live streaming session, the system was used by 10% of the usual consumers. Although adopting tiled-streaming technologies, in which only the FoV is requested at the highest possible quality, in that work authors found that segments with higher bitrates arrived later due to their larger size.

Therefore, all the previous works presented in this subsection illustrate the relevance of optimizing content transmission in order to reduce cost in terms of bandwidth usage.

B. DELAY VARIABILITY, GLOBAL TIMELINES AND INTER-DEVICE SYNCHRONIZATION

Extra delays can be added throughout the end-to-end delivery chain due to numerous reasons. In [24] and [25] the sources of these delays and their effect for both broadband and broadcast technologies are presented, respectively. Existing delay differences between destinations in current delivery

[3] Source of the images: https://docs.unity3d.com/Manual/VideoPanoramic.html

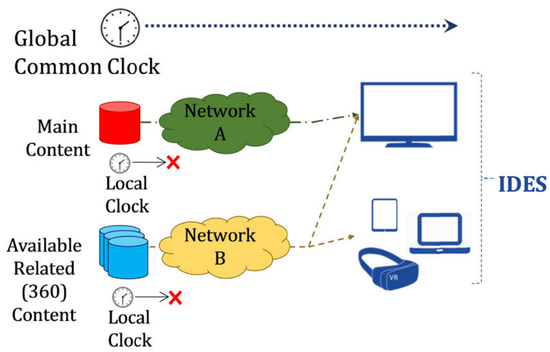

FIGURE 4. Global clock protocols can provide a common timeline throughout all the delivery chain.

networks have already been identified in previous works (e.g., [25]–[27]). First, in [25], it was stated that differences in end-to-end delays between receivers in an IPTV scenario can be larger than 6s. Second, the study in [26] showed that the delay differences when delivering the same media content via different broadcast variants and in different formats can accumulate up to 6s, and that HAS solutions can be more than 1 minute slower than broadcast technologies. In that work, it was also pointed that significant delay differences between receivers occur, even when the receivers use exactly the same TV delivery technology, setup combination (e.g., subscription type/quality) and equipment. However, no numbers were provided due to the lack of sufficient measurements from multiple geographically distributed sites. In [27] it was stated that depending on the resolution of the video content (i.e., Standard or High Definition -SD or HD-), delay differences can reach 8s. Although there are no data about delays when delivering omnidirectional content, it is expected to have even higher delay differences.

All the above issues reflect the need for adaptive and accurate hybrid media sync solutions to compensate for that end-to-end delay variability. When using just one device (a.k.a. hybrid terminal) an inter- stream sync mechanism is needed. When using several devices, an inter-device sync (IDES, hereafter) mechanism should be adopted. Otherwise, these contents will not be played out at the same time in the involved devices [28].

Typically, media delivery technologies rely on the insertion of *intrinsic* and relative timelines into the media streams, which are commonly obtained from local clocks. However, these timelines have no signification outside the media included in these streams. Thus, these types of timelines are not valid for IDES, as media content can be generated by different sources, and even networking equipment can override *intrinsic* timelines throughout the delivery chain. To solve this, absolute and *extrinsic* timelines can be obtained from the same or traceable clock sources (e.g., Network Time Protocol -NTP- [29]) and its reference keeps unmodified throughout all the delivery chain (see Figure 4). For example, this is the approach that follows RTP/RTCP streaming, as well as other mechanism proposed by the European Telecommunications Standards Institute (ETSI) in [30]. The latter is based on inserting an external and absolute timeline into MPEG2-TS streams. This mechanism is used in [31] to achieve IDES in second screen TV experiences. In a similar way, by also using [30], [32] combines a tiled streaming solution via HTTP Live Streaming (HLS) for ultra-high resolution media content while receiving a DVB stream through the main TV. In [33] and [34], implementations of the DVB-Companion Screens and Streams (DVB-CSS) specification [35] (also adopted by HbbTV standard) are presented. Although no conclusive results are stated, in [34] asynchrony delays from -500ms to 1000ms are shown to be unnoticeable for the participants in the conducted subjective tests.

Additionally, another mechanism called Timeline and External Media Information (a.k.a. TEMI) has been proposed by MPEG and DVB as an amendment to ISO/IEC 13818-1 [36]. Specifically, the TEMI mechanism has been adopted for the solution proposed in this work, as it is considered in the HbbTV specification as a valid mechanism to synchronize media.

Regarding IDES in hybrid TV consumption scenarios, the need for this type of sync has become more popular due to the spread of heterogeneous consumption devices, as addressed in [1] and [28]. In particular, the HbbTV standard [2] provides functionalities and mechanisms to synchronously consume hybrid content in the main TV device (Main Screen, hereafter MS) and, optionally, in one (or more) secondary devices (Companion Screen, hereafter CS), such as smartphones, Head Mounted Displays (HMD), tablets or laptops. Regarding these *hybrid media sync* mechanisms, works in [31], [33] or [32] describe different solutions to achieve accurate hybrid sync between devices. In particular, in [31] and [32], an external global timeline is inserted into the MPEG2-TS; and in [33] the DVB-CSS [35] protocol is adopted. In [1], authors present a solution based on the use of the TEMI mechanism (which is also the one used in this work) and, additionally, a more extended compilation and description of the existing *hybrid media sync* mechanisms and the existing (related) hybrid platforms including them are provided.

III. HYBRID END TO END DELIVERY AND CONSUMPTION SOLUTION INCLUDING OMNIDIRECTIONAL VIDEO CONTENT

In this section the proposed solution is presented. It is based on the authors' previous work described in [1]. It has been designed to support omnidirectional video content processing, broadband delivery and consumption in a synchronous way together with the consumption of broadcast TV content. First, the solution is briefly described, and then the new added functionalities and mechanisms along the end-to-end delivery and consumption ecosystem are summarized.

A. HYBRID DELIVERY AND CONSUMPTION SOLUTION

In this work, the solution described in [1] and enhanced in [37] has been extended to include preparation, delivery and consumption of omnidirectional video content as

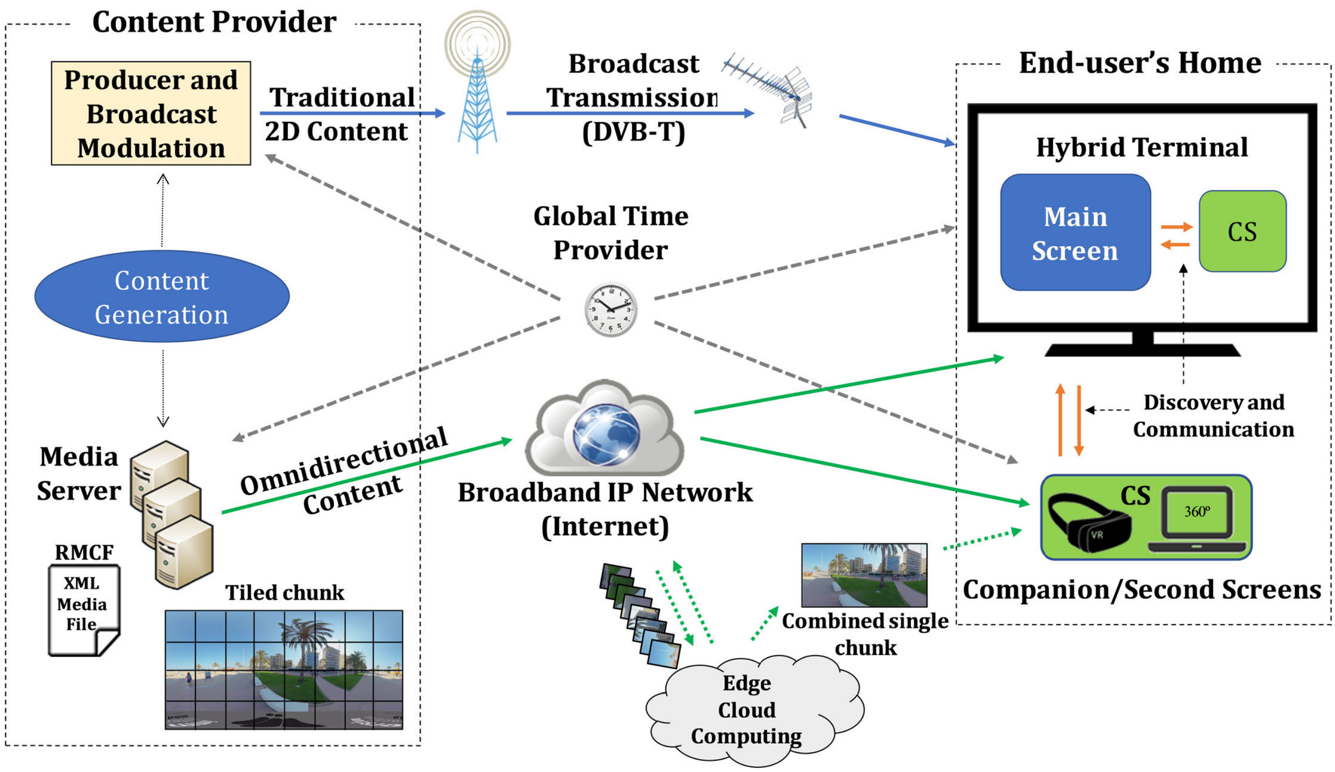

FIGURE 5. **Proposed solution's overall architecture, updated from [1].**

complementary to broadcast TV content. Figure 5 presents an overview. The architecture of the solution can be divided in two parts: 1) Content provider's side and 2) End user's side. On the one hand, inside the Content provider's side, the tasks of generating, encoding and transmitting the available contents are carried out. It also includes a global time provider. An XML file (called *Related Media Content File*, *RMCF*) is generated in order to signal the complementary broadband-available content (e.g., the omnidirectional video content). For this, by adopting the TEMI mechanism, the URL of the *RMCF* is inserted into the *TEMI location descriptor* [36]. This way, as soon as the broadcast content is received, the hybrid terminal (with the MS, e.g., the TV screen) will be able to extract and analyze the URL of the *RMCF* embedded in the *TEMI location descriptor*. On the other hand, inside the End user's side, the tasks of selecting, requesting, receiving, decoding and presenting the selected content are carried out.

Meanwhile, additional CS devices (e.g., a smartphone or an HMD) can discover the MS and then establish a communication channel with it to exchange useful information, such as the requirements to playout additional complementary content or playout timing information, to achieve accurate sync. For this, timestamps should have previously been inserted into the broadcast content in the *TEMI timeline descriptor* [36] during the encoding process. An extended description of all the involved basic elements and processes in the solution can be found in [1].

B. OMNIDIRECTIONAL VIDEO ENHANCED FUNCTIONALITIES

In this subsection, only the new specific features and functionalities related to omnidirectional video content which have been included in the new solution are described. All the other ones can be found in [1].

1) CONTENT PROVIDER'S SIDE

In this side, as seen in Figure 5, the proposed solution does not require any specific extra functionality but the obvious need for the generation of omnidirectional video content and an update of the already existing elements. Regarding the omnidirectional video content, in the platform different projections can be adopted and supported. Once the content is recorded, the process to store it and make it available via any HAS technology is similar to converting traditional 2D content into this type of streaming technology. Moreover, in order to provide this type of content in a more efficient way, recent technologies such as HEVC [10] or the DASH-SRD extension [18] can be adopted in addition to the already supported technologies, such as H.264.

Additionally, taking into account that the end user's player or device could not support the aforementioned tile-based technologies, and also in order to avoid processing load and latency in those devices (e.g., smartphones used in HMD), an additional module is proposed to be located in the (edge) cloud. Specifically, on the one hand, the *Edge Cloud Computing module* will be in charge of requesting the

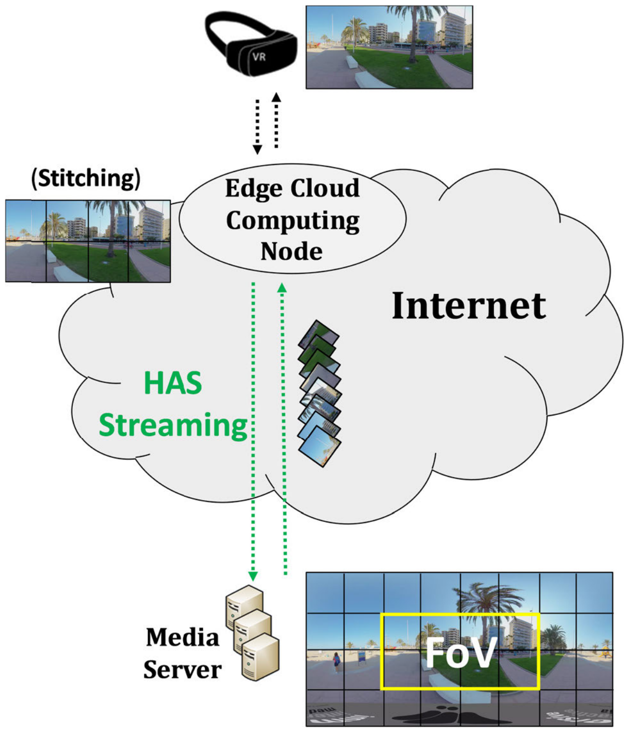

FIGURE 6. Functioning of the Edge Cloud Computing Node.

tiles inside the user's FoV in each chunk from the Media Server and then decode and stitch them together to generate one single video chunk. The combined single chunks are the ones to be locally delivered to the CSs (Figure 5), when they are not capable of directly consuming tiled content (due to, e.g., poor performance, incompatibility issues, etc.) Figure 6 shows the involved steps that this procedure follows: 1) the users requests the video chunks providing information about the current FoV to the Edge Cloud Computing Node; 2) the Edge Cloud Computing Node requests the involved tiles for that specific FoV to the Media Server at the Content Provider's side; 3) the Edge Cloud Computing Node receives the involved tiles, stitches them and generates a single chunk which is sent to the users. On the other hand, the use of this module can provide the use of sophisticated machine learning algorithms for predicting future changes of FoV with high accuracy. Some of those algorithms are very complex and may not run on current mobile devices yet. Therefore, authors want to emphasize that the design and implementation of a tiled-based 360 video streaming solution based on cloud edge computing is out of the scope of this paper and is left for further work.

The signaling mechanism described in [1] has also been updated to notify the existence of complementary omnidirectional video content. The *MEDIA* tag in the *RMCF* file, defined in [1], has been extended and new properties have been amended to support this new type of content. Table 1 shows the updated tag and its properties. In particular, in the proposed solution, the *media_type* property of the *MEDIA* tag currently supports two different types of values: *AV* for traditional 2D content and *360AV* for the omnidirectional video

TABLE 1. Tags from [1] updated to support omnidirectional video content.

Tag	Property	Description
MEDIA		It specifies the necessary metadata for any available AV media content
	id	Unique identifier for the element
	media_type	Media content type. Value for omnidirectional video content: '360AV'.
	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
	projection	Used projection type (e.g., 'ERP', 'CMP'…)
	protocol	Used protocol (e.g., HLS, DASH)
	tiled	Boolean value to indicate if the content format is tiled or not
	uri	Uniform resource identifier

content. Moreover, an additional property called *projection* has been included in the *SOURCE* tag in order to describe the type of projection which has been employed for the content generation. For instance, in this tag, ERP or CMP projections can be notified to the involved players.

Just in case that the end user's device does not support encoding formats involving tiling mechanisms (such as HEVC), when analyzing the *RMCF* the device will look for the boolean value in the *tiled*property inside the *source*tag. If it is *true*, it implies that the stored content is tiled-based. Otherwise, the content is not divided in tiles. If the content is tiled-based, the involved CS will request the content (including information about its FoV in the request) to the Edge Cloud Computing node (contained in the *uri* property), which will be in charge of assuming the performance cost for predicting future changes of FoV, requesting the corresponding tiles inside the FoV, with an acceptable quality, decoding and merging them into a single video chunk, and sending it to the CS. This way, all that process is transparent to the end user's consumption device and ensure that all devices are able to play omnidirectional video content.

2) END USER'S SIDE

In this side, once the MS has received the broadcast content, it can extract the *TEMI location descriptor* and request the *RMCF*. For the correct interpretation of this file and the successful acquisition of information regarding the omnidirectional video content, the playout process of the MS has also been updated. In the screen of MS device, a small pop-up emerges in order to notify the user about the existence of available related omnidirectional and/or traditional 2D video contents. Simultaneously, and transparently for the user, if a CS has established a communication channel with the MS, it will receive the analyzed information regarding available (and playable) content. For the omnidirectional video content, a specific player, capable of presenting this type of content, is required. It is intended that omnidirectional video content can be played in heterogeneous devices, such as specific (e.g., an HMD) or more general devices (e.g., smartphones, tablets or computers/laptops). In particular, HMDs, smartphones or tablets can make use of their sensors to display a particular FoV depending on the orientation axis (Figure 3). However, if the omnidirectional video content is being consumed through a computer, the navigation can be done by clicking and dragging the mouse to the region of interest. Additionally, smartphones or tablets can also be handled this way. After requesting the omnidirectional video content, the MS periodically sends playout timing information to the CSs. The mechanisms and functionalities to discover the MS, to create a communication mechanism between the involved devices (i.e., a MS and at least, one CS) and to achieve IDES are extensively explained in [1].

C. IDES REQUIREMENTS

In the proposed solution, omnidirectional video content is understood as complementary content to the main broadcast TV content, which is being played out in the MS (e.g., a TV or a set-top box). Several users can be consuming the hybrid content in different devices (TV and companion devices) at the same time/place and can interact (e.g., discussing about the content being watched, regardless the type of format). Thus, if one or several end users are consuming the omnidirectional video content (preferably, through an HMD), they might get totally immersed and, therefore, disconnected from the main content playout on TV. In that context, although there is not a strict need to achieve frame-accurate sync, a lip-sync (i.e., ±80ms [38]) level should be required in order to avoid noticeable asynchronies between MS and CS (especially if all the users are only listening to the audio from the MS).

IV. IMMERSIVE SCENARIO FOR BROADBAND MEDIA CONSUMPTION

This Section describes the design proposed for the immersive scenario where the available related video (2D or 360) contents are played out.

A significant design requirement, in order to recreate an immersive scenario, has been to emulate a living room where a virtual TV is available, in which the related 2D contents can be also consumed. Moreover, the scenario should include as many interaction points as needed, in order to notify the end user about the availability of additional and complementary (2D or 360) content and let the end user interact with the scenario (e.g., selecting the content to play). In particular, this design has been developed by taking into account the opinions (compiled in a brainstorm-style meeting) of all the members of the research group, among which there is a professional designer, regarding the user interface and the best way to interact with the available items and signaled content. After the meeting, and some design iterations, the mockup of the scenario shown in Figure 7 was obtained. It summarizes the different functionalities that should be available in the scenario. Starting from that idea, the designer prepared several designs (360 photos) for the virtual living room and, next, one of them was selected by consensus (shown in Figure 8 and in Section V).

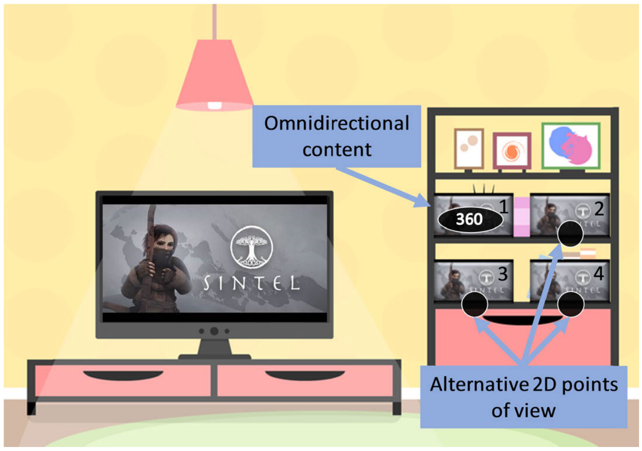

FIGURE 7. Mockup of the immersive virtual scenario to be designed.

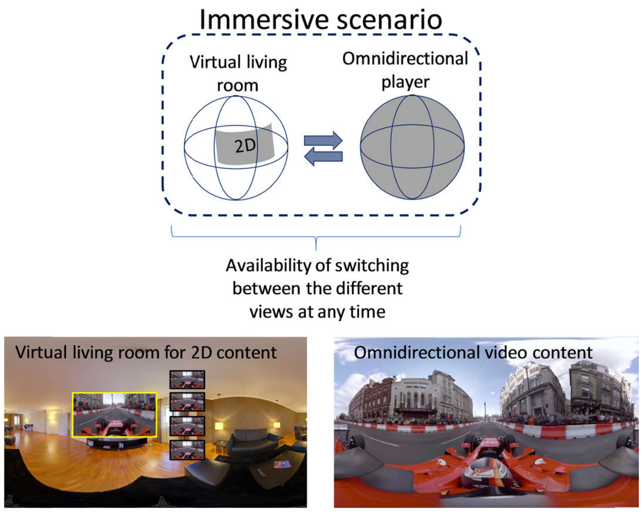

FIGURE 8. Available views within the immersive scenario.

As it can be observed in Figure 7, the virtual TV in the simulated living room can be playing a video (the open-source Sintel movie[4], in the Figure), and the availability of related contents is notified to the end user in a non-intrusive way, by including their associated interaction points. For the example in the Figure, it is assumed that, for that movie, there are four available related contents: three 2D videos with different points of view (as if it were a multi camera movie) and an additional omnidirectional content (with the text "360" overlaid). Depending on the consumption device, the required interaction to trigger the playout of those contents can be performed by clicking it (in a PC or laptop), touching it (in a tablet or smartphone) or by collision functionalities between the point of interest and the user's viewpoint geometrical center (in an HMD).

This design involves implementing an omnidirectional scenario that is comprised of two parts. On the one hand, the first part includes the virtual living room. Although the displayed

[4]https://durian.blender.org/

content in this part (in the virtual TV) is 2D, the living room itself is omnidirectional (a 360 image), in order to recreate a scenario as immersive for the end user as possible. The user point of view has been located in the center of the scenario and it has 3DoF. On the other hand, the second part refers to the embedded omnidirectional video player, which is independent to the first part and should provide 360-degree playout functionalities to let the end user consume any section of the provided omnidirectional content. The switching process between both view modes (virtual living room and omnidirectional) should be available for the end user at any time through interaction points, as described in this Section. Figure 8 summarizes both viewing modes (top) and the examples of the implementation for both cases (bottom).

Regarding the used technologies to implement the proposed design, web-based technologies have been adopted. This manner, it is compatible with modern smartphones (which can be included inside an HMD or used as a simple flat screen in the traditional way), tablets and computers. The implementation is based on the use of *dash.js* [39] and *three.js* [40] JavaScript libraries:

- Dash.js is a library that allows the playout of content based on MPEG-DASH [17] in browsers.
- Three.js is a library that allows the creation and management of computer graphics in browsers. It also can be combined with other HTML elements (e.g., the *canvas* item). In particular, this library has been adopted to implement the immersive scenario of the virtual living room and, moreover, to project the omnidirectional content in a 360° view. Additionally, this library also allows to manage the stereoscopic presentation of the content in both the virtual living room scenario and the omnidirectional player, when an HMD is used. It allows to shift one of the stereoscopic views, so that each eye can distinguish the same scene with some distance one from the other, thus providing an impression of depth.

On the one hand, when implementing the virtual living room scenario, the three.js library handles a mesh composed of a spherical geometry which is centered in the origin of the coordinates and it is covered by a texture which is the 360 image of the selected living room. In order to visualize it, a camera with 3DoF (i.e., from where the user visualizes this scenario) is located in the center of the sphere. Moreover, to emulate a traditional TV inside this scenario, an additional rectangular mesh has been added, in which the 2D video content will be projected (with the use of the dash.js library). To interact with the virtual living room and the signaling of additional content, some sprites have been created (one per additional content), with icons or thumbnails as their textures. Interactions (e.g. by using mouse or touch control) and collisions (e.g., by using *raycasting*) are enabled with these sprites, which trigger associated events, such as, for example, playing out a different camera point of view of the main content.

TABLE 2. Devices involved in the testbed.

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
Celeron N3050 @ 1.60GHz CPU, 4GB RAM, Ubuntu 14.04 with DVB USB rtl2832, connected to an LG 32LF592U 32” Smart TV via HDMI	Hybrid Terminal + MS
Tablet Samsung Galaxy Tab S4	CS #1
Smartphone Samsung Galaxy S5	CS #2
Smartphone Samsung Galaxy S6	CS #3
Smartphone Samsung Note 9	CS #4
Laptop w/ Intel Core i7-7700HQ @ 2.8GHz, 16GB RAM, Windows 10 OS	CS #5
TP-Link AC1900	Wireless Router

TABLE 3. Features of the available content.

Feature	Configured Value
Content Encoding	H.264 (not tiled) +AAC
Content Frame Rate	25fps
Broadcast Content Resolution (MPEG2-TS)	1920x1080px
Broadband 360 Content Resolution (DASH), with 3-second video chunks/segments	7680x3840px 3840x1920px 1920x960px 960x480px 480x240px

On the other hand, in order to play 360 content in the omnidirectional player, the three.js library is also used. Depending on the type of projection of the 360 content, ERP or CMP, a rectangular or spherical mesh is created, respectively. In both cases, the texture covering the mesh is the 360 video, which is played out by the dash.js library.

For both view modes (virtual living room and omnidirectional), there are also interaction sprites which allow the switching between them. When using an HMD, the collision functionality is performed by using a dot-based pointer and *raycasting* by reading the user's head orientation thanks to the existing sensors (e.g., accelerometers) in the device.

V. DEVELOPED TESTBED

In this Section, the implemented testbed following the architecture shown in Figure 5 is presented. As it is still under development, source code is not publicly available yet, but it will be published as soon as it is completely finished.

Firstly, the involved devices and the features of the used contents are defined in this Section. Secondly, the required changes and improvements that have been performed to the presented testbed in [1], [37], [48] in order to include omnidirectional content, and to support its synchronized playout, are described.

Table 2 lists the devices used in the testbed, while Table 3 specifies the main technical features of the available content. Regarding the video encoding format, H.264 has been adopted. At this stage, although it is not network friendly, the implemented player in the CS is based on a FoV-agnostic DASH-based streaming solution that delivers chunks of the

TABLE 4. Prepared content for the evaluations.

Title	Duration (s)	Encoding	Number of 2D views	Number of 3D views	3D Projection Type	Source
F1 & Roller Coaster	360	H.264 + AAC	3	1	CMP	https://www.youtube.com/watch?v=fQoVFraBOnc&t=9s https://www.youtube.com/watch?v=8lsB-P8nGSM&t=89s
NBA Stephen Curry	143	H.264 + AAC	1	1	CMP	https://www.youtube.com/watch?v=Y7TW6XZ_h_0
Paris	271	H.264 + AAC	1	1	CMP	https://www.youtube.com/watch?v=EkshFcLESPU
Flight	684	H.264 + AAC	1	1	CMP	https://www.youtube.com/watch?v=HEEIzZ7UjRg&t=159s
Sharks	198	H.264 + AAC	1	1	CMP	https://www.youtube.com/watch?v=j4ZT_XlhZBQ&t=5s
Shawn Mendes Live 1	303	H.264 + AAC	2	1	CMP	https://www.youtube.com/watch?v=LueM9tEu2wI https://www.youtube.com/watch?v=We7ZKgwqFKU
Shawn Mendes Live 2	240	H.264 + AAC	2	1	CMP	https://www.youtube.com/watch?v=ttByPyz3XFQ https://www.youtube.com/watch?v=rWx_m1VAp_E

```
<RMCF>
 …
 <MEDIA id="1" media_type="360AV" media_format="h264/aac“ metadata="360 video"
            temi_init="3763721899965118886">
  <SOURCE protocol="http/dash“ projection=“CMP” tiled=“false”
            uri="http://media_server.edu/360/stream.mpd"/>
  …
 </MEDIA>
 …
</RMCF>
```

FIGURE 9. Example of a ***RMFC*** file.

entire panoramic view. The reason is that, although more modern formats like HEVC natively adopt tiled streaming mechanisms, this type of technologies are not globally supported by mobile browsers yet [41]. However, H.264 is still the most compatible video format by most browsers [42]. Moreover, tiled-based encoding requires more computation on mobile devices to decode and process multiple tiles. Additionally, only the support for the CMP and ERP projections has been implemented, as both are the most common projections. As mentioned before, the implementation and, therefore, the assessment of the proposal including tiled-based encoding and edge cloud computing node features is left for further work.

In particular, the generated broadcast content has been extracted from 360 online video contents (explained in the next subsection). Regarding the genre of the used media contents, sports, documentary and music content have been selected. The generation and signalling of the broadcast and broadband content processes are explained in the next subsection. Moreover, further information, such as the type of 3D projection or the source of the content can be found in Table 4.

A. CONTENT PROVIDER'S SIDE

In order to provide end users with different media content, different types of content have been generated and prepared. In particular, online available 8K omnidirectional video content has been acquired (e.g., from YouTube). That content has been prepared an encoded in DASH format with the use of the tool described in [43], to deliver different qualities in chunks/segments of 3 seconds, depending on the end user's available bandwidth and, additionally for each content, the most significant region (for us) of the omnidirectional video content has been extracted and cropped (in 1920x1080 px resolution by using the FFmpeg [44] and GPAC [45] tools) in order to prepare, encode and deliver it through the broadcast channel as the main TV content. During that process, *TEMI location* and *timeline descriptors* have been inserted in the broadcast content to signal the available related broadband media. To enable the IDES mechanism, an NTP-based global time provider located in this side is in charge of providing the common timeline necessary to achieve sync between the involved devices. Finally, the *RMCF* has been generated with all the required tags and information to successfully provide all the needed metadata about the additional broadband media content to the end users (i.e., by using the MEDIA and *SOURCE* tags explained in Section 3). Figure 9 shows an example of an extract of the *RMCF* file including the needed information about the 360 video content.

B. END USER'S SIDE

In this subsection, the end user's side, and more specifically, the CS implementing the proposed immersive scenario, is explained. Figure 10 shows the main modules in both the MS and CS devices. The MS periodically sends playout timing

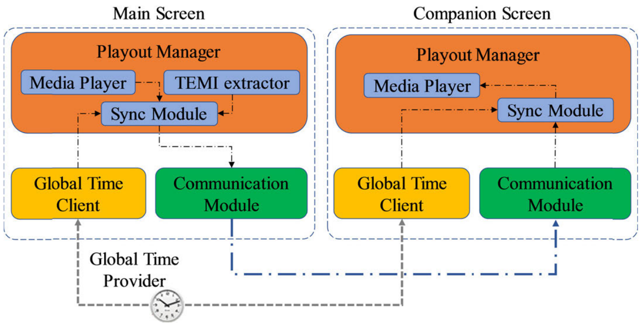

FIGURE 10. Involved modules in the MS and CS devices (End-users home in Figure 5).

FIGURE 11. Immersive scenario seen through a flat display (top) and through and a stereoscopic display or HMD (bottom).

information through the communication module by taking into account the extracted timestamp from the content and the current global instant. In the same way, the CS receives this information and decides (after some calculations) if an adjustment is required (explained later in the following subsections). Further information regarding the communication between MS and CS, and the synchronization of the CS's playout process can be consulted in [1].

1) IMPLEMENTED OMNIDIRECTIONAL SCENARIO

The implemented living room immersive scenario is shown in Figure 11. It includes an omnidirectional living room with a virtual TV, and, in the right part, additional complementary contents of the received main broadcast content being consumed in the MS (e.g., the real TV). The availability of these extra contents is notified by inserting a thumbnail of a frame of each content. This way, the end user is aware of their nature (i.e., additional 2D viewpoints or 360 video versions of the main content).

The thumbnails of the available omnidirectional video contents have an overlay with the "360" text. If the end user selects the playout of this type of content, the omnidirectional video player is launched in full screen mode or in stereoscopic display mode, depending on the device in use. This can be observed in Figure 12, where a screenshot of the developed player is shown, when a flat screen of, e.g., a PC or tablet is used (top side) or when an HMD is used (bottom). As it can be observed, several 3D buttons have been included: a button for muting the audio of the player (in order to listen to the audio of the real TV), and another button to let the user close the omnidirectional player and switch back to the living room immersive scenario.

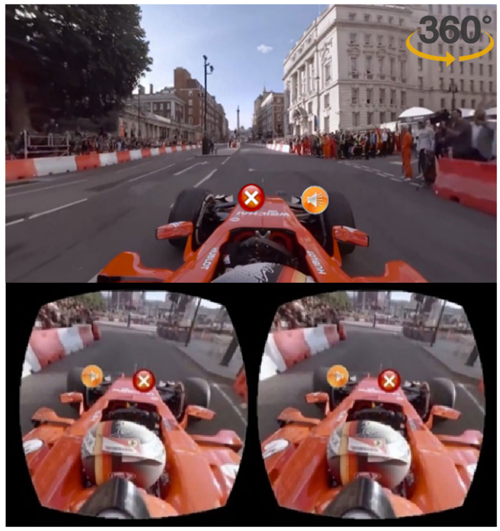

FIGURE 12. Omnidirectional video content played out in a flat display, e.g., a PC display (top) and a stereoscopic display or or an HMD (bottom).

C. NAVIGATION THROUGH THE IMMERSIVE SCENARIO. UI FUNCTIONALITIES

The main device for which the immersive scenario has been designed is the HMD. As previously explained, hands free functionalities have been designed and implemented. In case the end user is consuming the content through an HMD, the visualization of a small centered solid point is enabled. This point is used in order to collide with the interaction points (3D buttons or video thumbnails). Besides, in order to be also compatible with other devices, such as PC or tablets, the interactions with the points of interest can also be performed by using a mouse (if a PC or laptop is used) or the fingers (if a smartphone or a tablet is used).

1) ADOPTED SYNC MECHANISMS

In order to achieve a synchronized state between the MS and the involved CS devices the IDES mechanisms defined by authors in [1] has been followed. The Master/Slave synchronization control scheme [24] has been adopted, in which the MS acts as the Master and the CSs act as the Slaves. The initial steps to achieve hybrid sync, using the Master/Slave scheme, consist of: i) retrieving the TEMI timelines from the incoming MPEG2-TS, which give information about the generation instants of the video frames; ii) tracking the video frames until the rendering elements (i.e., the audio/video

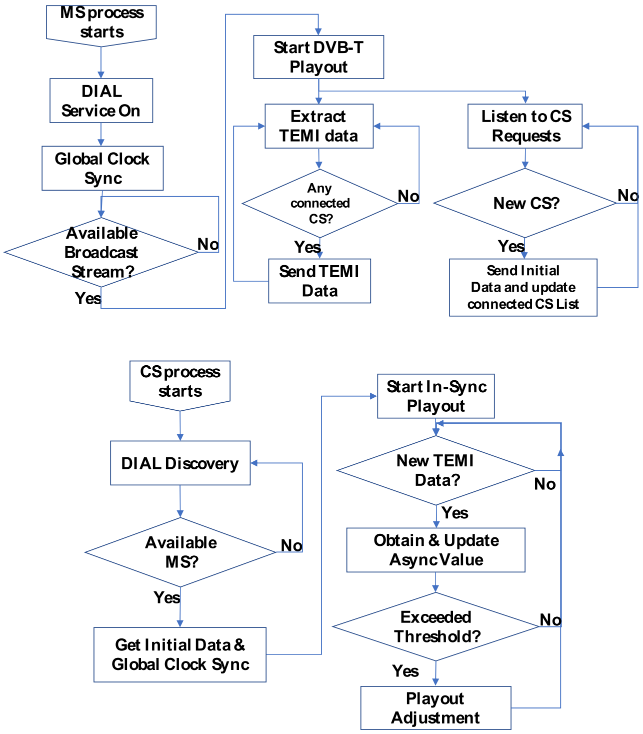

FIGURE 13. MS (top) and CS (bottom) flow diagrams of the processes [1].

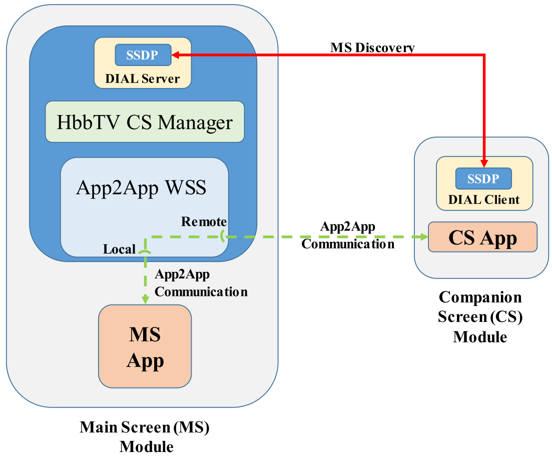

FIGURE 14. Components and communication functionalities involved in the HbbTV communication modules [1].

sinks); and iii) registering the presentation (absolute, NTP-based) timestamps of these video frames. The tuple of NTP-based generation timestamp (TEMI timeline) of the currently processed content and its presentation timestamp will be sent to the involved CS devices (being executed on either the same hybrid terminal and/or on companion devices).

Upon receiving this information, each CS device will be able to compare its own playout timing with the one of the MS and perform playout adjustments (if needed) to be in-sync with the MS. To achieve a coherent notion of time in the session and to compensate for the effect of network delays between the MS and CS Apps, a common clock reference is used. Figure 13 presents the flow diagrams of the involved processes in both MS (top) and CS (bottom) devices. Readers can consult reference [1], in order to have a more detailed description of those processes.

For the explained mechanism, a communication channel between MS and CSs is needed. The implemented testbed follows the specifications of the HbbTV standard and all the steps described in [1]. DIAL mechanisms [46] and, consequently, WebSocket technology [47], have been used In particular, in order to use the specific procedures, defined in the HbbTV standard, the involved communication modules between MS and CSs have been adopted from the public repository from Fraunhofer Institute for Open Communication Systems (FOKUS),[5] in which the HbbTV 2.0 Companion Screen components are fully implemented. This way, MS is aware of the connected CSs and is able to send them the necessary timing information to achieve a tolerable sync level between devices. Figure 14 shows the involved processes and modules related to the HbbTV 2.0 communication functionalities (more details about this Figure can be found in [1]). The MS module runs a DIAL server which can be discovered by CSs. Once the CSs know the existence of an active MS, a bi-directional communication channel (called App2App communication channel) can be established. By using this channel, the MS can send data of complementary broadband content or playout timing information to the CSs.

The broadband available DASH content has been codified in five different qualities and segmented in 3s-length chunks (Table 3). The CS previously downloads the chunks to be played (to be in-sync with the content played on the MS) and, then, according to the timing information received from the MS, smooth playout adjustments in the playout process of the CS can be performed (by using Adaptive Media Playout, AMP, techniques [37]) to achieve the desired sync level between MS and CS devices.

Regarding the enhancements of the solution in order to support omnidirectional video content, on the one hand, the playout process involving the MS is similar to the one in [1], but with the extended capability of interpreting the notification of available omnidirectional video content. On the other hand, the playout process of the CS has been updated by also adding the support for omnidirectional video content playout (with the implemented web-based player). Figure 15 shows the synchronized playout of the main 2D content in one real TV (MS) and the playout of the available omnidirectional content in several companion devices (CSs) running the developed omnidirectional player (but selecting in each a FoV coinciding with the 2D view on TV, approximately). The right performance of the implemented IDES solution can be appreciated.

VI. OBJECTIVE AND SUBJECTIVE EVALUATIONS

In this Section, the methodology and the results of the conducted objective and subjective evaluations are presented and discussed.

[5] https://github.com/fraunhoferfokus/node-hbbtv

FIGURE 15. Synchronized playout in MS (2D main content) and CSs (360 video).

A. OBJECTIVE EVALUATION

In order to assess the achieved sync accuracy in the testbed for the proposed solution, an objective evaluation has been conducted. The asynchrony between the playout processes of MS and CSs (playing broadcast TV content and related broadband omnidirectional video content, respectively) has been evaluated when different devices are used as CSs (a laptop and several Samsung devices, such as Galaxy S5, S6 and Note 9 smartphones, and a Galaxy Tab S4 tablet)

In order to avoid continuous playout adjustments, a maximum tolerable asynchrony threshold value has been defined. If the measured asynchrony between players do not exceed that value, they will not make any playout adjustment. This asynchrony threshold has been set to ±80ms, to achieve lip-sync accuracy [38] (as mentioned, needed when all the users are only listening to the MS or TV audio output). Additionally, in order to evaluate the proposed solution in close-to-real scenarios, additional packet loss probability for the broadcast content and network delay for the broadband content has been added. On the one hand, regarding DVB broadcasting, each of the MPEG2-TS transport packets are usually extended by a shortened 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, in average, only one non-correctable error occurs within one hour of program presentation (equivalent to a BER of 10^{-11}). Taking into account that the available videos used in the evaluation are up to 6 minutes long, we have evaluated the scenario with a symbolic packet loss probability of 0.1% (much larger than the QEF average) in DVB Transmission.[6] On the other hand, regarding broadband delivery, the network delay parameter has been set to 60ms ± 20ms (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.[7] Authors have adopted these values (for the packet loss probability and network delay) before in [37] and [48].

TABLE 5. Obtained mean asynchrony value and 95% confidence interval.

CS Device	Mean asynchrony (s)	95% Confidence Interval (s)
Laptop	0.059	±0.007
Note 9	0.060	±0.002
Galaxy Tab S4	0.062	±0.005
S6	0.066	±0.003
S5	0.067	±0.003

In order to validate the implemented IDES mechanism when omnidirectional video content is being consumed together with broadcast TV content, the asynchrony values measured in 10 sessions of approximately 3 minutes have been registered when using each companion device (i.e. CS). The selected content for the objective evaluation has been the "F1 & Roller Coaster" video content, as it involves both slow and fast motion scenes. Table 5 summarizes the obtained results, including the mean asynchrony value and the 95% confidence interval, in the 10 sessions for each device.

As observed in Table 5, the mean value of the asynchrony is maintained below the configured threshold (±80ms). The best results have been obtained when using the laptop to consume the omnidirectional video content. Nevertheless, the differences between the values obtained for each employed device are minimal. Asynchrony values measured when using newer and more powerful devices, such as Note 9, Galaxy Tab S4 and the laptop, are slightly better than when using the older ones, such as S5 and S6. Figures 16 and 17 show the frequency of repetition of the sync registered values for each device in all the sessions and the cumulative frequency distribution (CFD) of the measured asynchrony values, respectively.

Moreover, an additional 10-minute session involving all the devices being used simultaneously has been conducted in order to register the values of the asynchrony measured in each device (playing the related omnidirectional content) during the session. As an example, Figure 18 presents the values of the asynchrony measured and the magnitude of the performed playout adjustments in the Galaxy S5 device and Figure 18 present the performance of the five involved devices. Only the results of the S5 device are provided in detail because it turned out to be the one providing the worst performance. It can be observed, however, how this device makes the necessary adjustments to keep its sync level within the configured threshold. Additionally, in both Figures 19 and 20 a moving average (10 samples) has been also plotted to make the interpretation of the measured asynchronies easier.

As it can be observed in Figure 19, the measured asynchrony values, in general, do not exceed the maximum tolerable asynchrony values (±80ms, highlighted as black

[6]The broadcast packet loss probability has been forced with the StreamXpress software, which is in charge of modulating and transmitting the broadcast content.

[7]The broadband network delay has been forced with the Netem tool, available in https://wiki.linuxfoundation.org/networking/netem

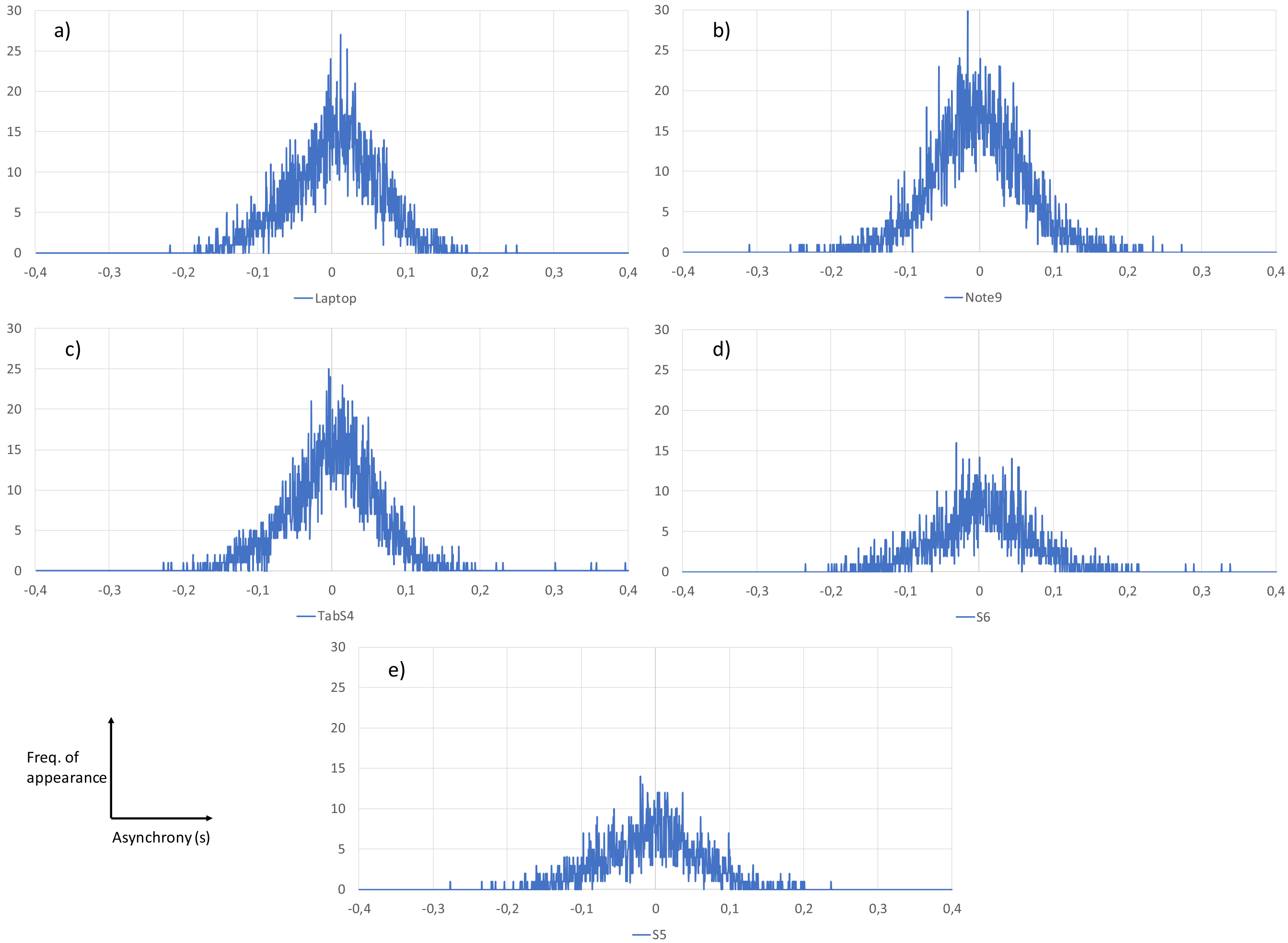

FIGURE 16. Registered asynchrony values in each CS.

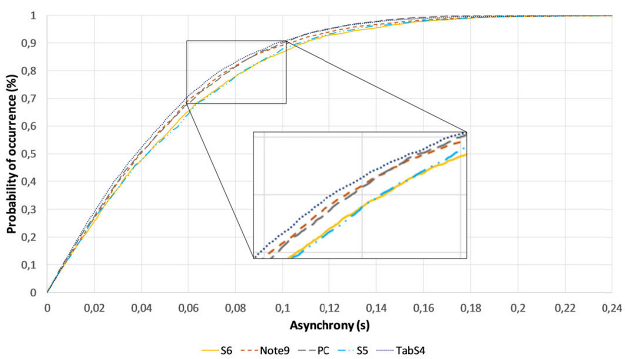

FIGURE 17. Obtained CFD of the measured asynchrony values in each CS device involved in the session.

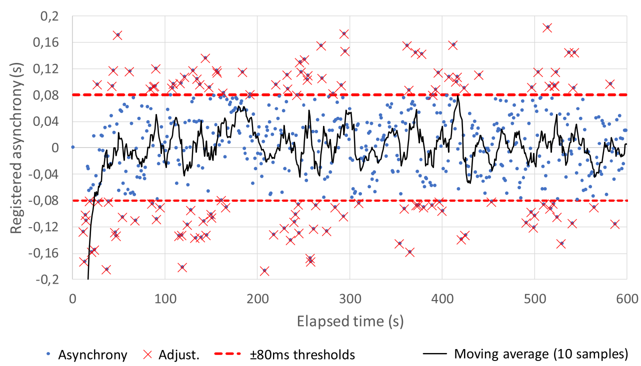

FIGURE 18. Registered asynchrony values by the S5 device during the 10-minute session. Performed adjustments and moving average (10 samples) of the registered asynchrony values.

dashed lines). Moreover, in Figure 18 it can be appreciated that the Laptop, the Note 9 smartphone and the Galaxy Tab S4 provide better performance in terms of sync accuracy than the S6 or the S5 devices. It can be seen that the S6 and S5 devices have less probability of registering an 80ms or lower asynchrony value than the remaining three devices, although, as observed in Table 5, all the mean asynchrony values of all the devices are within the configured threshold. Therefore, taking into account the obtained results, it can be stated that the developed testbed, which follows the proposed hybrid sync solution, enables the a satisfactory sync between the playout of broadband omnidirectional video content in a CS and the playout of the main broadcast content on a real TV (MS).

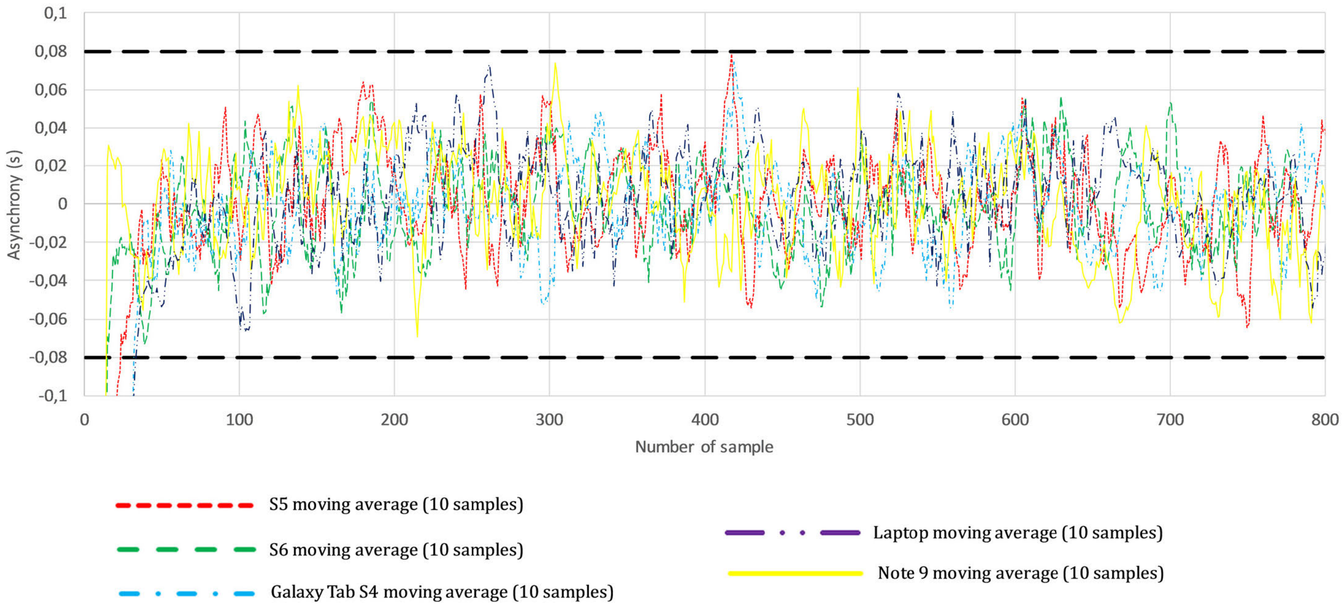

FIGURE 19. **Registered asynchrony values in each CS.**

B. SUBJECTIVE EVALUATION

In this subsection, the subjective evaluation process is described, regarding the content preparation, the methodology, the participant's relevant information and the obtained data.

A total of 32 subjects participated in the study. After analyzing the validity of all their responses, 2 subjects have been discarded due to their contradictory (and incompatible) responses. The results from the remaining 30 subjects have been taken into account. None of the participants presented any audiovisual impairment that could affect the evaluation process.

For this evaluation, only the laptop, the tablet Galaxy Tab S4 and the smartphone Samsung Galaxy Note 9 (inside an HMD) were provided as CSs to the users, as they have been the devices which have provided a better performance in the objective evaluation.

1) EVALUATION PROCESS

At the beginning, all the participants, after signing a consent form, have read a brief description of the testbed, the immersive scenario and embedded omnidirectional player, including its goals and features. Then, in order to guide the participants throughout the process, the evaluation has been divided in four different stages with a total duration of 40 minutes, approximately:

1. Initial Questionnaire (6 minutes): previous to any interaction with the implemented solution, participants have answered a questionnaire regarding their personal information, such as the gender, age range or the professional profile, among other information;

2. Tutorial-like guided use of the developed immersive scenario (12 minutes): to make participants familiar with the available functionalities by using all them in the three devices provided as CSs. At this stage, couples of users (user A and user B) have been forced to use the devices following these 3-minute steps: a) user A watching TV (MS) and user B using the HMD; b) user B watching TV (MS) and user A using the HMD; c) user A using the Tablet and user 2 using the Laptop, both in front of the real TV playing the main broadcast content; d) user B using the Tablet and user A using the Laptop, both also in front of the real TV.

3. Free use of the implemented solution (15 minutes): participants, located in front of the real TV playing the main broadcast content,can freely choose any of the available types of additional related content (2D or 360) and interact with the immersive scenario running on any of the devices provided as CSs;

4. Final questionnaire and interview (7 minutes): to collect the participant's opinion about the overall experience and the usability (by adopting the System Usability Scale -SUS- defined in [49]) of the implemented immersive scenario and the perception of the achieved sync.

During the stages 2 and 3, the users chose the genre of the content to be consumed from the available ones, described in Table 4.

2) PARTICIPANTS' RELEVANT INFORMATION

The results and information presented in this subsection correspond to the first stage of the evaluation process and are presented in Table 6. The answers provided by the participants in this stage were collected before any use of the testbed.

Related to the consumption habits of the participants, on the one hand, they were asked about the time during the day which they spend watching the traditional broadcast TV content. Results show that 50% of the participants consume less than 1h of their daily time, 27% consume

TABLE 6. Participants' profiles.

Gender		Age		Level of studies	
Male	Female	<21 y.o.	0%	High School	10%
57%	43%	21-25 y.o.	37%	Vocational Training	10%
		26-30 y.o.	13%	University Degree	37%
		31-35 y.o.	0%	University Master	13%
		36-40 y.o.	13%	PhD	30%
		41-45 y.o.	13%		
		>45 y.o.	23%		

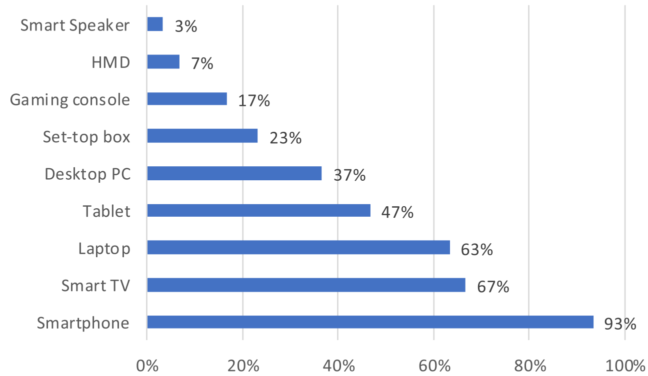

FIGURE 20. Owned devices by participants.

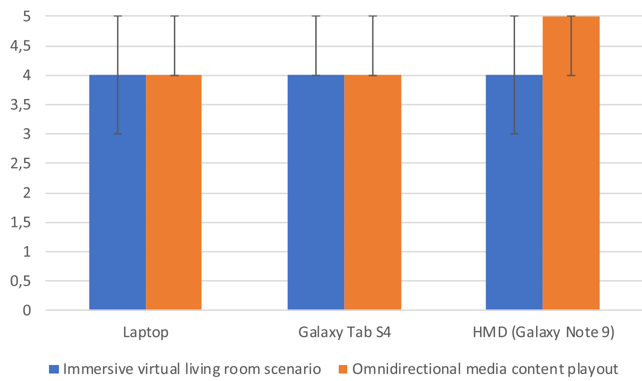

FIGURE 21. MOS results regarding the personal experience when using each device (median, upper and lower quartile).

TABLE 7. Obtained mean values.

	Laptop	Galaxy Tab S4	HMD (Galaxy Note 9)
Immersive virtual living room scenario	4.1	4.23	4.16
Omnidirectional media content playout	4.13	4.26	4.46

between 1 and 2h, a 16% consume between 2 and 3h, and only a 6% consume between 3 and 4h. None of the participants declared to consume more than 4h of daily traditional TV content.

On the other hand, regarding broadband delivered content, (e.g., Video on Demand -VoD- platforms such as Netflix, HBO or Amazon Prime Video), results show that 34% only spend between 1 and 2h of their daily time, a 26% of the participants spend between 2 and 3h, a 10% of the participants spend between 3 and 4h and only a small percentage of the participants consume between 4 and 5h (3%) or more than 5h (3%).

By analyzing these results, there is a significant difference between the traditional media consumption and VoD content consumption. However, it can be stated that both types of content consumption are still used (as confirmed in [3] and [7], explained in Section I).

Regarding the consumption devices owned by the participants, most of them, at least, have a smartphone and a TV. Figure 20 shows the owned devices.

3) OBTAINED RESULTS

This subsection summarizes the results obtained in the stages 2 and 3 of the evaluation. During the second stage the participants had a Tutorial-like guided use of the scenario, in order to show them how to use all the available functionalities in each available device and experience some forced situations. In the third stage, participants could freely choose the device(s) to use and even the content to consume (Table 5). The adopted scale to quantify the participants' experience has been the Mean Opinion Score [50], by using a 5-level Likert-type scale. A 5-point score means the experience has been very good and a 1-point score means that the experience has been very bad. Table 7 and Figure 21 present the obtained results when using each device, and regarding the way they consumed the related contents: (2D views on the virtual TV in the immersive virtual living room scenario or 360 videos in the omnidirectional video player). In particular, Table 7 shows the obtained mean values and Figure 21 shows the median value and the lower and upper quartiles.

The obtained values are pretty similar for every device. However, when using the HMD, there is a clear difference in the participant's QoE when the omnidirectional video content is played out. That is the case in which the HMD is used at its full potential providing a fully immersivity when the user is able to view the whole 360° scene of the video sequence, and, therefore, the score is a bit higher compared to the view of 2D contents on the virtual TV in the immersive virtual living room scenario.

When participants could choose which devices to use (users could demand more than one device), the most demanded ones were the HMD (53%), the tablet (53%) and the laptop (16%). When users were asked about the reasons why they chose that device, the participants who selected HMD claimed that that device provides a more realistic and immersive experience. The participants who selected the tablet claimed that that device provided an easier way of interaction and that it was more comfortable for their eyes. Finally, the participants who chose the laptop claimed to have the sensation of having more control over the environment.

As participants used more than one device, they were also asked about whether they noticed anything strange during the switching between devices. Also, another 5-level Likert-type scale with the values *Totally Disagree, Partially Disagree, Neutral, Partially Agree and Totally Agree* was used to let the participants provide their opinion about the following statement "*When I switched to a different device, I did NOT notice any strange issue*". 73.7% agreed (68.4% completely, and 5.3% partially), 15.8% were neutral, 5.3% (i.e., 1 person) partially disagreed and 5.3% totally disagreed. Participants who partially disagreed, claimed that when they switched to another device, it took a noticeable initial period of time to get in-sync with the TV (i.e., to achieve IDES). This is understandable as the IDES process needs an initial period for MS discovering, exchange initial timing information, and for making the initial the playout adjustment. This out-of-sync period of very few seconds can be perceptible and a bit annoying for users. Regarding the perceived sync, participants did not notice asynchrony issues after that initial period, i.e., during the steady stage of the playout processes.

When using the laptop or the tablet as CS, the participants could consume simultaneously the broadcast content in the MS (real TV) and the broadcast content in the CS. In those cases, they were asked about whether they had watched the MS during the non-guided stage. Only 23% of them claimed to have watched the MS. Some of the provided reasons were that they wanted to consume the main content and have that reference compared to what they were consuming in the CS or to check the achieved sync level between the playout of both devices. When asked whether they wanted to specifically use a single device as the CS, 50% stated that they wanted to use the HMD, 30% would choose the tablet and 13% the laptop. As additional information, from the participants which used the HMD, 23% of them did it while standing up and moving around and the rest preferred to be seated. Regarding the type of content the participants are interested in consuming in this kind of platform, the selected options are shown in Figure 22.

Participants were also asked about the overall experience. Figure 23 summarizes their answers about the statement "*I think omnidirectional consumption can enhance the hybrid TV environment*".

Regarding the perceived level of immersiveness, participants totally agreed (76.6%), partially agreed (16.7%), were neutral (3.3%) and disagreed (3.3%) with the fact of feeling more integrated/immersed in the consumption experience (*During this visualization experience I have felt more integrated/immersed with the content*, Figure 24). Actually, also related to the immersiveness feeling, Figure 25 shows the level of agreement with the statement "*At some moment during the experience with the HMD I have lost the notion of time*".

Finally, participants were also asked about whether they would like to repeat an experience like the experienced one by using the testbed during the evaluation. 87% of them totally agreed and the remaining 13% partially agreed.

Regarding the usability of the developed environment, as previously stated, the SUS questionnaire [49] has been used. Table 8 lists the 10 questions of that questionnaire.

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

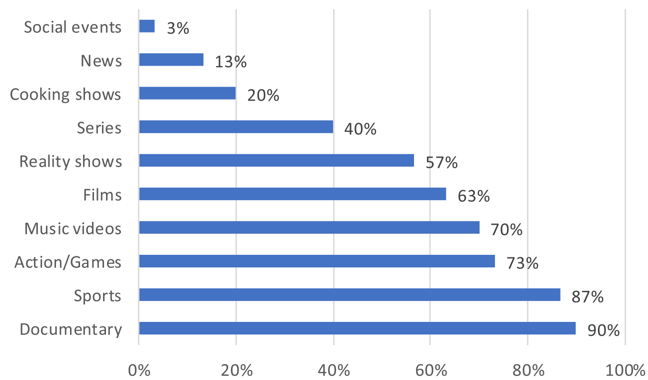

FIGURE 22. Most popular content genres to be consumed in the platform.

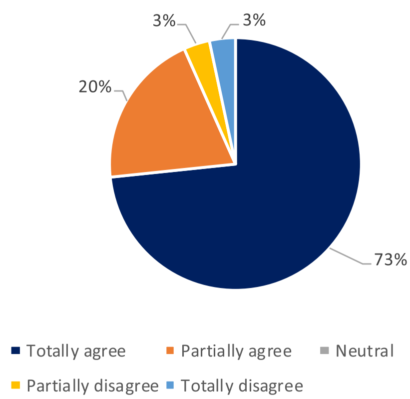

FIGURE 23. Level of agreement with the statement "*I think omnidirectional consumption can enhance the hybrid TV environment*".

The result provides a score from 0 to 100 points, being 0 points the worst imaginable usability score and 100 points the best imaginable usability score, as defined in [51].

The obtained results are presented in Figure 26. According to [49], they provide a score of 92.67 with a standard deviation of ±13.17. According to [51], with this score, the overall usability of the implemented solution can be considered as *excellent*. However, participants were asked for any change or suggestion in order to improve the user experience, and the collected concerns were mainly related to the consumption device, especially when using the HMD device. It induced some dizziness to two participants, who, at some time, preferred to consume the content by using the

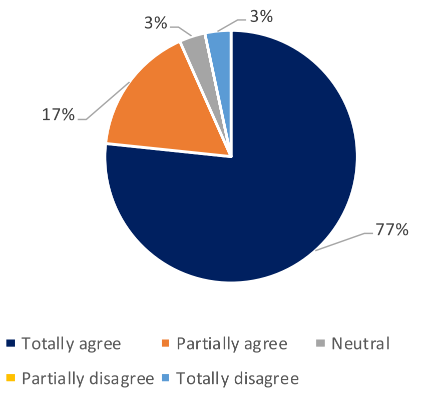

FIGURE 24. Level of agreement with the statement "*During this visualization experience I have felt more integrated/immersed with the content*".

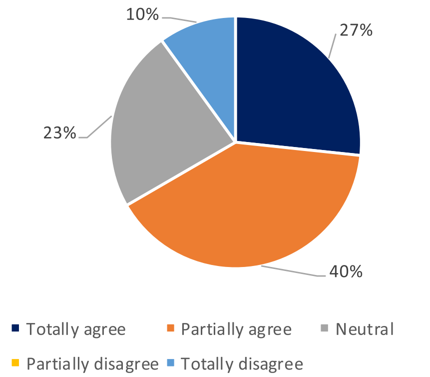

FIGURE 25. Level of agreement with the statement "*At some moment during the experience with the HMD I have lost the notion of time*".

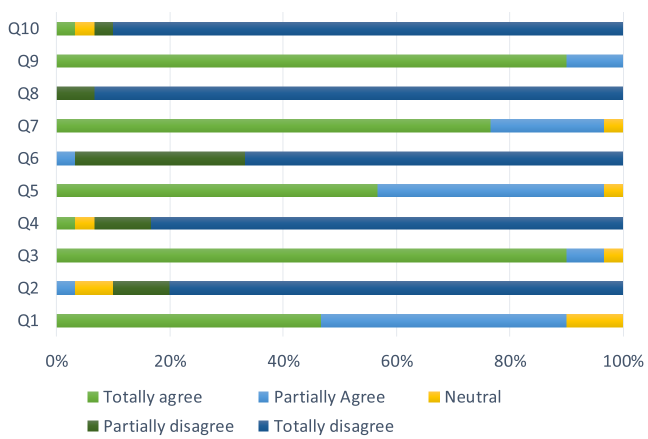

FIGURE 26. Obtained responses for the SUS questionnaire.

tablet or laptop devices because of that reason. Regarding the user interface, participants suggested minor modifications (e.g., bigger size of the virtual TV, or to place the interaction buttons in a less intrusive position in the omnidirectional player (these buttons were initially located in the low-center part of the FoV when using the HMD device and, in some occasions, important content was occluded by them).

It can be seen how both the 360 omnidirectional video consumption along with the provided immersive scenario help to enhance the users' TV media consumption experience when additional complementary omnidirectional video contents are available and can be consumed within accurate sync levels.

VII. CONCLUSION AND FUTURE WORK

In this paper, a possible solution for enabling the delivery and consumption of complementary omnidirectional video content related to the traditional broadcast TV content in a synchronous way has been described. Hybrid broadcast and broadband contents can be generated, encoded, transmitted, received, decoded and presented synchronously. In order to provide the additional broadband content to the end user's home, the required information is signalled inside the broadcast content. The proposed solution could be implemented in current infrastructures and services, since it does not require any additional infrastructure or technology than the already in-use ones. To be able to objectively and subjectively assess the proposed solution, it has been implemented by updating and augmenting an own end-to-end hybrid media delivery testbed. Results of the objective assessment prove that the achieved hybrid sync is good enough to guarantee a good QoE when consuming omnidirectional content related to the main broadcast TV content. Regarding the results obtained in the subjective evaluation, participants have shown a clear interest in this type of new enriched experience. Although most participants stated that they do not own any HMD device, during the evaluation this device has been the most demanded to consume the omnidirectional video content. However, it can be observed that some participants may prefer other devices, such as the tablet or the laptop, as they are more used to them. In conclusion, both objective and subjective results can be considered as very satisfactory and promising.

The employed testbed is still in a first stage of development and many research results associated with 360 video streaming are still pending to be included. Some features are currently under work in order to be implemented in it. Some examples are: the tiled-based encoding support for the adopted projections, the support of new video projections, and the design and implementation of a full FoV adaptive 360 video content delivery solution including the edge cloud computing node, as summarized in Figure 6.

As a proof of evidence, two demo videos showing the capabilities of the developed solution for the laptop and the Note 9 smartphone within an HMD can be watched at https://bit.ly/2o5kNYR and https://bit.ly/2nnJOya, respectively.

REFERENCES

[1] 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.

[2] *HbbTV 2.0.2 Specification*, Standard ETSI TS 102 796 V1.4.1, HbbTV Association, 2018.

[3] Hootsuite. (2019). *Global Digital Report 2019—We are Social*. Accessed: Oct. 2019. [Online]. Available: https://wearesocial.com/global-digital-report-2019

[4] YouGov. *YouGov: VR Adoption Grows to 11% of US Adults—Digital TV Europe*. Accessed: Oct. 2019. [Online]. Available: https://www.digitaltveurope.com/2019/03/22/yougov-vr-adoption-grows-to-11-of-us-adults/

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

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

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

[8] J. Park, P. A. Chou, and J.-N. Hwang, "Volumetric media streaming for augmented reality," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–6.

[9] M. Domański, O. Stankiewicz, K. Wegner, and T. Grajek, "Immersive visual media—MPEG-I: 360 video, virtual navigation and beyond," in *Proc. Int. Conf. Syst., Signals Image Process. (IWSSIP)*, May 2017, pp. 1–9.

[10] *High Efficiency Coding and Media Delivery in Heterogeneous Environment—Part 2: High Efficiency Video Coding*, Standard ISO/IEC 23008-2:2015, International Organization for Standardization, 2015.

[11] J. Jeong, D. Jang, J.-W. Son, and E.-S. Ryu, "Bitrate efficient 3DoF+360 video view synthesis for immersive VR video streaming," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2018, pp. 581–586.

[12] L. Zelnik-Manor, G. Peters, and P. Perona, "Squaring the circle in panoramas," in *Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV)*, vols. 1–2, Oct. 2005, pp. 1292–1299.

[13] Z. Chen, Y. Li, and Y. Zhang, "Recent advances in omnidirectional video coding for virtual reality: Projection and evaluation," *Signal Process.*, vol. 146, pp. 66–78, May 2018.

[14] E. Kuzyakov and D. Pio. (2016). *Next-Generation Video Encoding Techniques for 360 Video and VR*. [Online]. Available: https://code.fb.com/virtual-reality/next-generation-video-encoding-techniques-for-360-video-and-vr/

[15] A. Zare, A. Aminlou, M. M. Hannuksela, and M. Gabbouj, "HEVC-compliant tile-based streaming of panoramic video for virtual reality applications," in *Proc. 24th ACM Int. Conf. Multimedia (MM)*, 2016, pp. 601–605.

[16] J. Le Feuvre and C. Concolato, "Tiled-based adaptive streaming using MPEG-DASH," in *Proc. 7th Int. Conf. Multimedia Syst. (MMSys)*, 2016, Art. no. 41.

[17] *Information Technology—Dynamic Adaptive Streaming Over HTTP (DASH)—Part 1: Media Presentation Description and Segment Formats*, Standard ISO/IEC 23009-1:2014, International Organization for Standardization, 2014.

[18] O. A. Niamut, E. Thomas, L. D'Acunto, C. Concolato, F. Denoual, and S. Y. Lim, "MPEG DASH SRD: Spatial relationship description," in *Proc. 7th Int. Conf. Multimedia Syst. (MMSys)*, 2016, pp. 1–8.

[19] M. Hosseini and V. Swaminathan, "Adaptive 360 VR video streaming: Divide and conquer," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2016, pp. 107–110.

[20] M. Graf, C. Timmerer, and C. Mueller, "Towards bandwidth efficient adaptive streaming of omnidirectional video over HTTP: Design, implementation, and evaluation," in *Proc. 8th ACM Multimedia Syst. Conf. (MMSys)*, 2017, pp. 261–271.

[21] S. Petrangeli, V. Swaminathan, M. Hosseini, and F. De Turck, "An HTTP/2-based adaptive streaming framework for 360° virtual reality videos," in *Proc. ACM Multimedia Conf. (MM)*, 2017, pp. 306–314

[22] X. Corbillon, F. De Simone, G. Simon, and P. Frossard, "Dynamic adaptive streaming for multi-viewpoint omnidirectional videos," in *Proc. 9th ACM Multimedia Syst. Conf. (MMSys)*, 2018, pp. 237–249.

[23] D. Ochi, Y. Kunita, A. Kameda, A. Kojima, and S. Iwaki, "Live streaming system for omnidirectional video," in *Proc. IEEE Virtual Reality (VR)*, Mar. 2015, pp. 349–350.

[24] 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, Nov. 2012.

[25] 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. Experiences TV Online Video (TVX)*, 2014, pp. 23–30.

[26] 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.

[27] R. N. Mekuria, "Inter-destination media synchronization for TV broadcasts," M.S. thesis, Delft Univ. Technol., Delft, The Netherlands, 2011.

[28] 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.

[29] D. Mills, U. Delaware, E. J. Martin, J. Burbank, and W. Kasch, *Network Time Protocol Version 4: Protocol and Algorithms Specification*, document RFC 5905, IETF, 2010.

[30] *Specification for the Carriage of Synchronised Auxiliary Data in DVB Transport Streams*, Standard ETSI TS 102 823 V1.2.1, Digital Video Broadcasting, 2014.

[31] C. Howson, E. Gautier, P. Gilberton, A. Laurent, and Y. Legallais, "Second screen TV synchronization," in *IEEE Int. Conf. Consum. Electron. Dig. Tech. Papers*, Sep. 2011, pp. 361–365.

[32] R. van Brandenburg and A. Veenhuizen, "Immersive second-screen experiences using hybrid media synchronization," in *Proc. Media Synchronization Workshop (MediaSync)*, 2013, pp. 1–7.

[33] V. Vinayagamoorthy, R. Ramdhany, and M. Hammond, "Enabling frame-accurate synchronised companion screen experiences," in *Proc. ACM Int. Conf. Interact. Experiences TV Online Video (TVX)*, 2016, pp. 83–92.

[34] C. Ziegler, C. Keimel, R. Ramdhany, and V. Vinayagamoorthy, "On time or not on time: A user study on delays in a synchronised companion-screen experience," in *Proc. ACM Int. Conf. Interact. Experiences TV Online Video (TVX)*, 2017, pp. 105–114.

[35] *Companion Screens and Streams; Part 2: Content Identification and Media Synchronization*, Standard ETSI TS 106 286-1, Digital Video Broadcasting, 2017.

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

[37] 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.

[38] R. Steinmetz, "Human perception of jitter and media synchronization," *IEEE J. Sel. Areas Commun.*, vol. 14, no. 1, pp. 61–72, Jan. 1996.

[39] DASHIF. *GitHub—Dash-Industry-Forum/Dash.js: A Reference Client Implementation for the Playback of MPEG DASH Via Javascript and Compliant Browsers*. Accessed: Oct. 2019. [Online]. Available: https://github.com/Dash-Industry-Forum/dash.js

[40] Three.js. *Three.js—Javascript 3D Library*. Accessed: Oct. 2019. [Online]. Available: https://threejs.org/

[41] *Can I Use: HEVC*. Accessed: Oct. 2019. [Online]. Available: https://caniuse.com/#search=hevc

[42] *Can I Use: MPEG4*. Accessed: Oct. 2019. [Online]. Available: https://caniuse.com/#search=MPEG-4

[43] D. Gómez, F. Boronat, M. Montagud, and C. Luzón, "End-to-end DASH platform including a network-based and client-based adaptive quality switching module," in *Proc. 7th Int. Conf. Multimedia Syst. (MMSys)*, 2016, pp. 1–4.

[44] *FFmpeg*. Accessed: Oct. 2019. [Online]. Available: https://ffmpeg.org/

[45] GPAC. *GPAC Framework*. Accessed: Oct. 2019. [Online]. Available: https://gpac.wp.minestelecom.fr/

[46] *DIAL Discovery and Launch Protocol Specificacion Version 1.7.2*, Netflix, Los Gatos, CA, USA, 2015.

[47] I. Fette and A. Melnikov, *The WebSocket Protocol*, document RFC 6455, 2011.

[48] 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.*, to be published.

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

[50] *Methods for Subjective Determination of Transmission Quality*, document ITU-T Rec. P.800, 1996.

[51] 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, 2009.

DANI 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. He is also an Assistant Researcher and a Developer with the Immersive Interactive Media R&D Group. He is the author of one book chapter and several research and conference papers. His main topics of interest are communication networks, code developing, and media synchronization.

FERNANDO BORONAT (M'93–SM'11) was born in Gandia, Spain. He received the bachelor's degree in telecommunications engineering and the M.E. and Ph.D. degrees in telecommunication engineering from the Universitat Politècnica de València (UPV), Spain, in 1994 and 2004, respectively. After working for several Spanish telecommunication companies, he moved back to the UPV, in 1996. He is currently an Assistant Professor with the Communications Department. He is also the Head of the Immersive Interactive Media R&D Group, Gandia Campus, UPV. He is the author of two books, several book chapters, an IETF RFC, and more than 100 research articles. His main topics of interest are communication networks, multimedia systems, multimedia protocols, and media synchronization. He is member of ACM. He is an Editor of *MediaSync: Handbook on Multimedia Synchronization* (Springer, 2018). He is involved in several IPCs of national and international journals and conferences.

JAIR LÓPEZ was born in La Paz, Bolivia. He received the B.Sc. degree in telecommunication systems, sound, and image from the Universitat Politècnica de València (UPV), Gandia Campus, in 2019. He is currently pursuing the M.Sc. degree in telecommunications systems and networks. He is also a Developer with the Immersive Interactive Media R&D Group, Gandia Campus, UPV. His main topics of interests are communication networks and multimedia systems.

ANNA VIDAL received the B.Sc. degree in mathematics, in 1986, and the Ph.D. degree focused on fuzzy topology research and its applications, in 1999. In 1990, she became a Full Assistant Professor. In the past, she has participated in several research projects, mainly focused on teaching and educational innovation projects, leading some of them and managing the Educational Innovation and Research Group in Scientific Matters, Universitat Politècnica de València (UPV), Gandia Campus, where she started to work, in 1986. She is currently with the Department of Mathematics and also researching in the area of design, modeling, and analysis of media synchronization techniques. She is a member of the Immersive Interactive Media R&D Group, Gandia Campus, UPV.