

Personalized and Immersive Presentation of Video, Audio and Subtitles in 360° Environments: An Opera Use Case

Isaac Fraile, David Gómez, Juan A. Núñez, Mario Montagud, Sergi Fernández
i2CAT Foundation

C/ Gran Capità 2-4, Edifici Nexus I, Barcelona, Spain

{isaac.fraile, david.gomez, juan.antonio.nunez, mario.montagud, sergi.fernandez}@i2cat.net

ABSTRACT

This paper presents an end-to-end system for a personalized presentation of accessibility and immersive contents in multi-screen scenarios, by focusing on an opera use case. In particular, the system allows experiencing the opera event using the classical audiovisual formats, but it additionally supports a seamless integration of 360° video, spatial audio and the use of Head Mounted Displays (HMDs). The availability of multiple 360° cameras allows experiencing the event from the preferred viewpoint, while the presented audio will match the selected camera position and current user's viewpoint, providing a highly immersive and realistic experience. Finally, a personalized and assistive presentation of subtitles also contributes to a higher accessibility.

Author Keywords

Immersive TV, Multi-Screen Scenarios, Omnidirectional Media, Spatial Audio, Subtitles, Virtual Reality.

ACM Classification Keywords

H.5. Information interfaces and presentation (e.g., HCI).

INTRODUCTION

The TV consumption landscape is rapidly evolving. On the one hand, the combined usage of TVs and companion devices for media consumption is attracting the interest of the research community, industry and consumers (e.g. [1-2]). On the other hand, a proper integration of omnidirectional contents within traditional TV services can bring new fascinating scenarios (e.g. [3]). In this context, ImmersiaTV is a H2020 European project (www.immersiatv.eu/) targeted at overcoming the existing challenges to enable customizable and immersive multi-screen TV experiences.

Apart from audio and video, subtitles also play a key role in TV-related scenarios. Previous works have targeted at proposing solutions for an adaptive and customizable presentation of subtitles, even in multi-screen scenarios (e.g. [4]) and for 360° videos (e.g. [5]). However, many challenges need to be overcome in order to efficiently integrate accessibility services (including subtitling, audio subtitling, audio description and sign language interpreting) in immersive environments. In this context, ImAc is a H2020 European project (www.imac-project.eu) targeted at exploring and adopting proper mechanisms to ensure accessible and inclusive immersive experiences to the whole consumers' spectrum, regardless of their sensorial capacities or other specific impairments.

This paper presents an end-to-end system for a personalized and immersive presentation of video, audio and subtitles in 360° environments, by focusing on an opera use case (Figure 1). In particular, the system allows experiencing an opera event in the classical audiovisual formats, but additionally considering multi-view (or even free viewpoint) and multi-screen scenarios, immersive audio and video formats (e.g. 360° video and spatial audio), and Virtual Reality (VR) consumption devices (e.g. Head Mounted Displays or HMDs). This allows experiencing the opera performance from the preferred viewpoint, without being restricted to the director-controlled scenes, using the preferred media formats and devices, and being also able to dynamically switch between viewpoints and devices. In relation to this, the presented audio will match the selected camera position and current user's viewpoint, thus providing a highly immersive and realistic experience. Beyond the immersive layer, the system allows the presentation of subtitles (i.e., transcripts of the opera) in a customizable and assistive manner, thus also providing an accessibility layer. The system has been set up by combining contributions from the mentioned ImmersiaTV and ImAC projects.

END-TO-END SYSTEM FOR 360° ENVIRONMENTS

Next, the key building blocks of the end-to-end system are briefly introduced.

Content Capture and Stitching

A novel sub-system for the capture and processing of both traditional and omnidirectional contents has been developed. It allows a seamless integration between 360° video camera rigs and conventional ones, while leveraging current practices in camera and computing technology systems.

Content Production and Edition

A custom production workflow supporting the live and offline edition of traditional and omnidirectional contents has been developed in ImmersiaTV. It is based on both new and adapted tools, and includes a new plugin for Adobe Premiere Pro. This workflow is being extended in ImAc in order to properly integrate accessibility contents.

Content Encoding, Packaging, Signaling and Delivery

The system includes new standard-compliant solutions for the adaptive encoding of the immersive and accessibility contents. On the one hand, Dynamic Adaptive Streaming over HTTP (DASH) has been adopted for adaptive media delivery. On the other hand, a proper metadata model for the signaling, discovery, association and orchestration of the

contents has been specified, being also standard-compliant with the DASH signaling formats.

Interactive and Personalized Content Consumption

The immersive and accessibility contents are adaptively presented on the involved consumption devices, based on their resources, on the narrative(s) specified during production and on the specific interactions and preferences of the consumers. Different presentation modes are enabled, such as adding overlays (“portals”) to the omnidirectional videos with additional media assets, as well as adding intuitive and assistive notifications, transitions and effects.

Likewise, proper service and device discovery, association, app launching and inter-device synchronization solutions have been developed to enable coherent multi-screen scenarios. All these solutions are compliant with the latest release of Hybrid Broadcast Broadband TV (HbbTV) standard (version 2.0.1).

Finally, a customizable, adaptive and assistive presentation of subtitles for the 360° videos, to be consumed on HMDs, is also supported. In particular, different colors for each speaker (i.e., singer) are added, and different font sizes and languages can be selected. Subtitles are, by default, presented at the bottom region, as in typical video players. However, it is also possible to present them at the top region of the player (Figure 2). This presentation option is known as super-titles or surtitles, and is typically used and preferred in musical and theatre performances. In order to properly identify the active speaker(s) and main actions while freely exploring the 360° area, different guiding mechanisms are provided, such as adding arrows, a compass, or sided text.

CONTENTS FROM THE “ROMÉO ET JULIETTE” OPERA

A performance of the “Roméo et Juliette” opera was recorded at the *Gran Teatre del Liceu* Opera House (Barcelona). The goal is to create a ground-breaking product that will allow enjoying an opera performance in a highly interactive, immersive, personalized and accessible manner.

Multiple cameras (4 360°, 2 320°, 4 170°, and 5 directive cameras) and more than 80 audio sources (including 3D – Ambisonics– and binaural formats), distributed both on and off the stage, were used in the recording. This not only allows the selection of the preferred viewpoint, but also providing detailed sound landscapes, depending on the selected camera and viewpoint. Finally, a personalized and assistive presentation of subtitles (in different languages) can also be enabled. The playback of all the selected contents will be accurately synchronized, regardless on the number and types of consumption devices being used.

DEMO PROPOSAL

The demo proposal consists of showcasing the interaction, personalization, accessibility and inter-device synchronization features of the consumption part of the end-to-end system. For that purpose, the contents from the opera recording will be used. The demo setup will be composed of a main TV on which the director-controlled view and sound

will be presented, and companion devices (tablets, smartphones and HMDs) that will enable a dynamic selection of, and navigation between, the different camera views, as desired. In addition, headphones will be provided to be able to experience with the position- and view-dependent audio presentation. A demo video can be watched at: <https://goo.gl/869LF7>

ACKNOWLEDGMENTS

This work has been funded by European Union’s Horizon 2020 program, under agreement n° 688619 (ImmersiaTV project) and n° 761974 (ImAc project).

REFERENCES

1. V. Vinayagamoorthy, R. Ramdhany, M. Hammond, “Enabling Frame-Accurate Synchronised Companion Screen Experiences”, ACM TVX '16, Chicago, Illinois (USA), June 2016
2. F. Boronat, D. Marfil, M. Montagud, J. Pastor, “HbbTV-Compliant Platform for Hybrid Media Delivery and Synchronization on Single- and Multi-Device Scenarios”, IEEE Transactions on Broadcasting, vol. PP, no. 99, pp. 1-26, 2017
3. J. A. Núñez, S. Malewski, S. Fernández, J. Llobera, “Production and Delivery of Video for Multi-device Synchronized Playout”, ACM TVX '17, Hilversum (The Netherlands), June 2017.
4. M. Montagud, F. Boronat, J. González, J. Pastor, “Web-based Platform for Subtitles Customization and Synchronization in Multi-Screen Scenarios”, ACM TVX '17, Hilversum (The Netherlands), June 2017.
5. A. Brown, J. Turner, J. Patterson, A. Schmitz, M. Armstrong, M. Glancy, “Subtitles in 360-degree Video”, ACM TVX '17, Hilversum (The Netherlands), June 2017.



Figure 1. Multi-Screen Scenario with Omnidirectional Media.



Figure 2. Presentation of Subtitles in 360° Environments.