Disruptive Approaches for Subtitling in Immersive Environments

Chris. J. Hughes

School of Computer Science, University of Salford, Manchester, UK, c.j.hughes@salford.ac.uk

Mario Montagud

i2CAT Foundation, Barcelona, Spain; University of Valencia, Spain, mario.montagud@i2cat.net

Peter tho Pesch

Institut für Rundfunktechnik GmbH, Munich, Germany, thopesch@irt.de

ABSTRACT

The Immersive Accessibility Project (ImAc) explores how accessibility services can be integrated with 360° video as well as new methods for enabling universal access to immersive content. ImAc is focused on inclusivity and addresses the needs of all users, including those with sensory or learning disabilities, of all ages and considers language and user preferences. The project focuses on moving away from the constraints of existing technologies and explores new methods for creating a personal experience for each consumer. It is not good enough to simply retrofit subtitles into immersive content: this paper attempts to disrupt the industry with new and often controversial methods.

This paper provides an overview of the ImAc project and proposes guiding methods for subtitling in immersive environments. We discuss the current state-of-the-art for subtitling in immersive environments and the rendering of subtitles in the user interface within the ImAc project. We then discuss new experimental rendering modes that have been implemented including a responsive subtitle approach, which dynamically re-blocks subtitles to fit the available space and explore alternative rendering techniques where the subtitles are attached to the scene.

CCS CONCEPTS

• *Human-centered computing~Accessibility technologies* • *Human-centered computing~Accessibility design and evaluation methods*

KEYWORDS

Accessibility, Subtitling, 360° video, Immersive video

1 Introduction

ImAc seeks to create new open source tools for the content industry i.e., broadcasters, to create immersive and personalised sign language, enhanced audio description and subtitling services for 360° content. The project consortium brings together nine organisations that offer a wealth of interdisciplinary skills for the project and understanding of the subtitling workflow from conceptualisation through to publication. The consortium has gathered the trend-setters from the accessible media community (broadcasters, research institutions, universities and end user associations) with an intersectional approach.

Accessibility is often only considered after a technology has matured to meet the demand of the mass market, resulting in a significant barrier to around 80 million disabled citizens across Europe. The number of people affected is increasing, as the ageing process is a demographic trend: Europe's population is getting older. The total population in the EU is projected to increase from 511 million in 2016 to 520 million in 2070, an increase by 1.8%. However, the old-age dependency ratio (people aged 65 and above relative to those aged 15 to 64) in the EU is projected to increase by 21.6 percentage points, from 29.6% in 2016 to 51.2% in 2070 [1].

There is increasing pressure on broadcasters to provide access services particularly as discrimination on the ground of disability is prohibited against by the UN Convention on the Rights of Persons with Disabilities [2]. Many citizens with hearing impairments rely on subtitles to understand the content, and at the same time there is also a large proportion and increasing number of users who choose to consume videos without the sound turned on - for example 85% of Facebook videos are watched on mute [3]. Subtitles have been shown to improve comprehension [4], making content available from foreign or non-native languages and studies have also shown that consumers are more likely to watch a video all the way through if they are using subtitles [5]. Subtitling online content also provides a source of metadata, which can be used in indexing for searching and classification.

Modern Teletext subtitles first appeared on BBC Television in 1979 and live subtitles were first produced in 1984. Despite the move towards digital television and online distribution, the Teletext format persists in most broadcast systems. As a result, its technical limitations, such as a 38-character line length and restricted positioning [6] still constrain the way subtitles are made for television. Generally, subtitles are formed as blocks of text split into 1, 2 and sometimes 3 lines, and colour is used to signify different speakers. The subtitle can be justified to the left, centre or the right of the screen and the Teletext line number sets vertical position. Each block is transmitted in the television signal along with information as to when it is displayed and removed. Television subtitles are transmitted either as Teletext in the form of text, or as DVB Subtitles where the block is sent as a rendered image [7].

2 ImAc Project

This section discusses the ImAc project in order to understand the backgrounds of this paper and the potential of its contributions. We discuss previous projects within the scope of ImAc, then describe our methodology and end-to-end platform being developed under the umbrella of ImAc. The contributions of this paper are being implemented in real production and broadcast services, beyond basic prototypes or proofs of concepts.

2.1 Relationship with other projects

ImAc gains from the expertise, insights and contributions from other related projects. On the one hand, HbbTV4all project [8] has addressed accessibility in the emerging connected hybrid broadcast broadband media ecosystem, in the ecosystem of the Hybrid Broadcast Broadband TV (HbbTV) standard [9]. ImAc is seeking the same success story within immersive environments. On the other hand, ImmersiaTV has developed an end-to-end platform to enable customizable and immersive multi-screen TV experiences [10]. These contributions will be augmented in ImAc to also efficiently integrate access services.

2.2 Methodology

ImAc is built on three 3 pillars: 1) requirements gathering, 2) development and integration; and 3) validation and dissemination. A simplified diagram of the chosen methodology is illustrated in Figure 1. The activities are driven by a user-centric methodology, in which end-users, professional users and stakeholders are involved at every stage of the project through the organization of workshops, focus groups, evaluations, and the attendance to events. This allows determining with high precision the accessibility requirements, desired features and the scenarios being demanded. The insights gathered from the user-centric activities will determine the platform specification, the required new technologies and/or extensions to the existing ones that are necessary to meet these requirements. Likewise, an essential premise of ImAc is to build the developments by taking into account as much as possible the state of the art broadcast workflows, technology and standards. This will maximize re-usability, inter-operability and the changes of successful deployment and exploitation. In this context, it is noteworthy to remark that the ImAc consortium is proposing novel contributions to various standardization bodies, such as W3C, MPEG and ISO.

The pilot and dissemination actions conducted in the project allow validating its contributions, but also refining them based on the obtained results and gathered feedback.

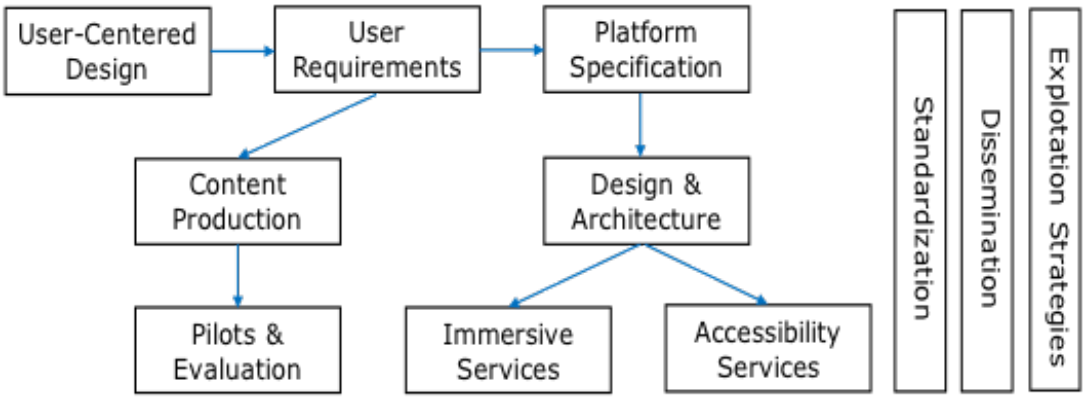

Figure 1: The user-Centric Methodology followed in ImAc.

2.3 End-to-End ImAc System

The ImAc platform is divided into four main parts, including the necessary steps from media production to media consumption. The key modules / functional blocks in these parts are indicated in Figure 2, where green colour indicates these components are under the umbrella of ImAc, orange indicates they have been developed in other projects (e.g. ImmersiaTV [10]), and white that are out-of-scope of the project, but form part of typical end-to-end workflows.

The main parts of the platform are:

Content Production: (web-based) tools for the production, authoring and editing of accessibility contents, and for their integration with classical and immersive media services. These tools provide the required subtitle and signalization information to be provided to the service provider.

Service Provider: different components where the management of programs is handled, the additional (immersive and accessibility) contents are linked to the main TV programs and play-out is scheduled. A key contribution of ImAc in this part is the Accessibility Content Manager (ACM), the component through which the immersive contents are uploaded, the creation of accessibility content is managed, and the preparation of contents (see below) is triggered.

Content Preparation & Distribution: the preparation of contents (e.g. multi-quality encoding, packaging/segmentation, signalling…) for the appropriate delivery via various technologies, such as DVB and IP-based CDNs (e.g. using DASH [11]).

Content Consumption: a web-based player for the presentation of the immersive (360º video and spatial audio) and accessibility contents (subtitles, audio description and sign language interpreting) in a personalized manner, based on the particular needs and/or preferences of the users. The player can be run on traditional consumer devices (e.g. Connected TVs, PCs, laptops, tablets and smartphones) and on Virtual Reality (VR) devices (e.g. Head Mounted Displays or HMDs). This part also includes the proper technology to enable multi-screen scenarios in a synchronized and interactive manner, using both fully web-based and HbbTV-compliant technology.

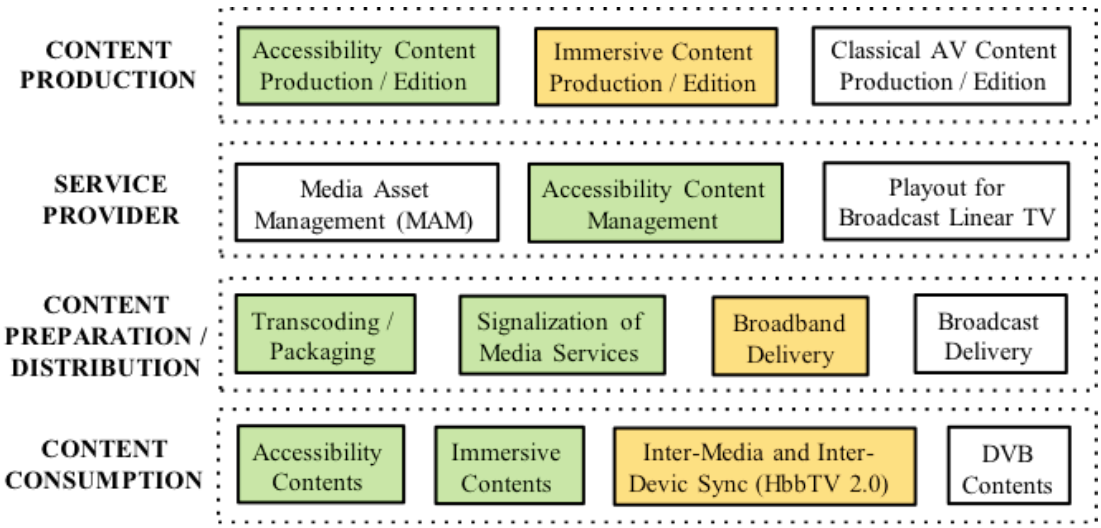

Figure 2: The main components of the ImAc platform.

3 Subtitle Presentation[1]

When consuming immersive media (i.e. 360º video) well-presented subtitles can contribute to a higher e-inclusion, but is more complex than for traditional media (i.e. 2D video). On the one hand, there is more information to process and users can get overwhelmed. On the other hand, the presentation of contents is no longer purely time-based, but involves a spatial dimension, determined by both the free user's exploration and the dynamic positions where the main actions are taking place (e.g. location of the target speaker, who can even move over time) around the 360º area.

For the proposed subtitle rendering, we assume a media playback system with at least three degrees of freedom, such that the user chooses the area of media he is watching. The displayed media is a part of the available media (e.g. a 360º video). Detaching the displayed image from the media source affects subtitle presentation or, in other words, raises questions regarding the desired response of subtitle behaviour to these environment changes.

Speaker identification, eligibility and immersion (often called presence) have been considered in [12], [13] and [14]. Subtitle presentation will always come with a trade off between factors such as speaker identification, eligibility, freedom to look around and perceived immersion. However graphical elements in view may be used to draw the attention away from the media towards helper objects. It is assumed that is necessary to identify each speaker, by colour and to aid navigation, but too many visual elements can be confusing. There must also always be a substitute for audio.

Therefore, ImAc is developing and testing appropriate solutions for both rendering modes and guiding methods for subtitles in 360º videos, while not having a negative impact on immersion and assisting users in a better comprehension of the story. This will contribute to a better accessibility and Quality of Experience (QoE).

3.1 Rendering Modes

At a first stage, the more common approach – subtitles are rendered at a fixed position in the displayed picture – was enriched with cues to support users regarding speaker identification and scene navigation. These cues act as guiding mechanisms and are described in the next chapter.

[1] A demo video showing some personalization features and the discussed guiding methods for subtitles in 360º videos can be watched at: https://goo.gl/Kn9QMx

Additionally, ImAc suggests an alternative rendering mode which main characteristic is that the subtitle is not related to the displayed picture (i.e. the screen), but to the media source or the scene. Different terms are used to describe this behavior, like fixed-positioned subtitles (MPEG OMAF [15]), world-referenced (conference paper [13]) or dynamic subtitles. We will stick with the term world-referenced subtitles in this paper.

Rothe et al. [13] conducted tests with what they call world-referenced subtitles and compared this presentation mode to screen-referenced subtitles. Their result doesn't find that one option is significantly preferred over the other. However, in terms of comfort, world-referenced subtitles led to a better result.

World-referenced subtitles in general conflict with a basic user requirement that a user must always be able to read the subtitle text. This is not hiven for these subtitles. Another obvious side effect is the limitation in the liberty of exploring twordhe scene, since looking away from the speaker will move the subtitle outside the field of view. This is described by [13].

As a consequence to this drawback, Schmierl et al. [14] suggested to build on world-reference subtitles and add guidance mechanism and second subtitle presentation when looking away. That means an additional subtitle is rendered in the presentation mode "screen-referenced" when the speaker and corresponding (world-referenced) subtitle moves out of view. The screen-referenced subtitle (when present) is complemented by a guiding mechanism. Two different guiding mechanisms were tested and compared to the presentation without guiding mechanism. Schmierl found that speakers can be found better, when a guiding mechanism is active. The two guiding mechanisms were rated si milarly.

The ImAc project exceeds the direction of study [14] and explores further assistance modes to compensate the disadvantages of world-referenced subtitles. It might be assumed, that most content has a main focus where the action is happening and the viewer is expected to attend most of the time. An indication that supports this theory is that attention guiding in general is a research topic for 360° content. As a result, we suggest optimizing subtitle presentation for this (main) direction and as a consequence accept less comfort for times where the user looks around and explores the scene. It is further assumed that discomfort may be caused by switching between subtitle behaviours (world-referenced / screen-referenced).

Figure 3: Main subtitle, attached to speaker.

In the suggested mode, subtitles will be presented as world-referenced in align with the speaker if they are visible at the time the subtitle presentation starts, as shown in figure 3. Otherwise, an additional subtitle will be rendered similar to the mode "appear in front, then fixed" described in [13] as shown in figure 4. The suggested presentation mode will be compared to the scene-referenced mode with activated guiding. User preference for one presentation mode may be personal preference and/or depend on the content type.

Figure 4: Assistance subtitle with preliminary guiding mechanism that leads to main subtitle and speaker. The subtitle will appear in the current viewing direction at a predefined height and includes a guiding mechanism "arrow" as described below. It will remain at this world-position.

3.2 Guiding Methods

Regardless of the rendering mode in use, presentation or guiding modes need be provided in order to assist the users in finding the action(s) / speaker(s) associated in the 360º area with the current subtitle frames being displayed. This is especially important if subtitles are aimed at viewers with hearing impairments, or when the audio cannot be listened to (e.g. noisy or public environments). When the audio cue is missing, support on how to locate the speakers and the main actions in the 360º scene becomes necessary.

To address these issues, four different guiding methods are being considered in ImAc:

- None. It means that no guiding method is used, and just the subtitle frames are displayed.
- Arrows. It consists of adding arrows to the left / right of the subtitle frames, indicating the direction towards the associated audio-visual elements in the 360º area see Figure 5.
- Radar. It consists of adding a dynamic radar to indicate where the associated speaker or main action is in the 360º see Figure 6.
- Auto-positioning. It consists of automatically adjusting the user's Field of View based on the position of the associated action(s) / speaker(s). This method can be applied to every subtitle frame by adding spatial information (metadata specified in the subtitling editor developed in ImAc).

Figure 5: Subtitles with arrows (with intermittence effect) as a guiding method. When this position is inside the user's Field of View, the arrows are hidden. Likewise, an intermittence effect can be added to the arrows to catch the attention of the user.

Figure 6: Subtitles with radar indicating the current consumers field of view. This allows the user to understand their position within in the 360º environment.

Limited research has been done, however some studies have tested different transitions, their impact on immersion and motion sickness in 360º environments [16] and methods for guiding the user's focus [17]. In general, no clear conclusions about preferred guiding methods have been obtained in these works. This is also the hypothesis in the tests to be conducted in ImAc: the most proper presentation mode may depend on the specific users' profiles, their sensorial capacities and preferences (e.g. young users may prefer the radar, auto-positioning may be preferred in tablet modes, arrows are the most simple method…). Therefore, adaptability and personalization become essential in this context, and are supported in the ImAc player.

3.3 Responsive Subtitles

The traditional presentation methods are limited by preserving the structure of the subtitles file. Using a responsive subtitle approach [18] allows further customisation by providing rules which allow the

subtitles to be dynamically re-blocked. This approach is particularly effective when adapting content from traditional television displays into an immersive environment, such as rendering the subtitle as speech bubbles attached to a character, or for instance if you if you wish to reduce the width of the subtitles in order to make room for graphics as shown in figure 7.

Figure 7: Responsive subtitles allow the rendering area to be dynamically changed to make space for graphics, or other accessibility services such as a signer.

As part of the ImAc project we have followed practices used in responsive web design to prototype a JavaScript library for generating responsive subtitles. This adopts the principles of text-flow and line length informed by semantic mark-up along with styles to control the final rendering.

ImAc uses Timed-Text Markup Language (TTML), one of W3C's standards regulating timed text on the Internet to store subtitle data with IMSC (TTML Profiles for Internet Media Subtitles and Captions), a file format specifically for representing subtitles and captions. Our library provides an extension to IMSC.js (a JavaScript library for rendering IMSC1 Text and Image Profile documents to HTML5) where IMSC documents can be loaded into a Timed-Text (TT) object. We then restructure the TT-objects based on line character width and line count as shown in figure 8.

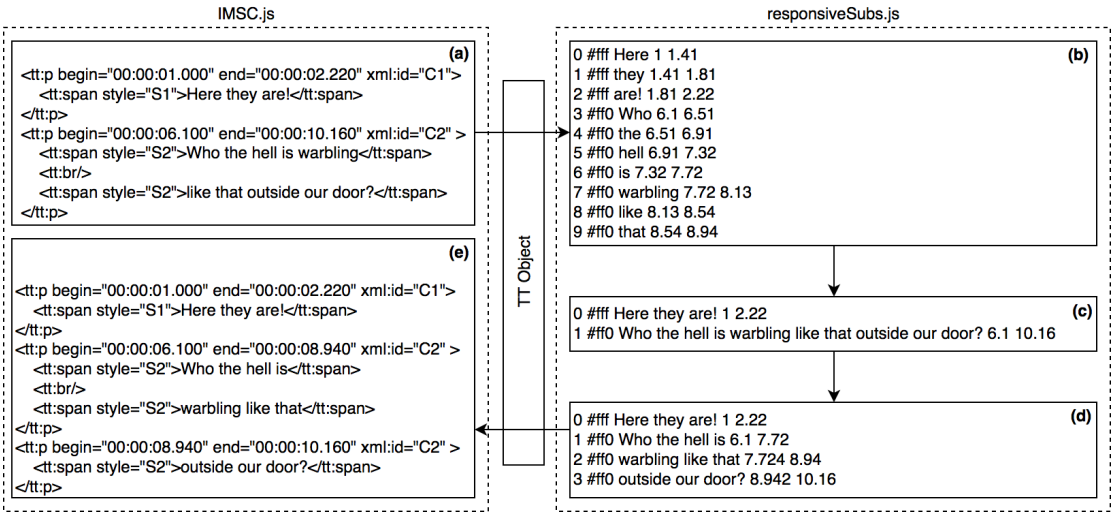

Figure 8: (a) IMSC.js converts the TTML document into a TT Object. (b) The responsive library atomizes the TT object into words, preserving an interpolated time and style for each word. (c) The words are reconstructed into phrases split by a pause in the dialogue or a change of speaker. (d) The phrases are subdivided using a best-fit algorithm to meet the line length requirements. (e) a new TT objet is generated with IMSC.js.

By working directly with TT-objects allows this library to be simply connected to any application which already uses the standard IMSC.js implementation allowing customization controls to be retrofitted, such as font size as shown in figure 9. The library can also be used in non-linear VR applications, as shown in figure 10.

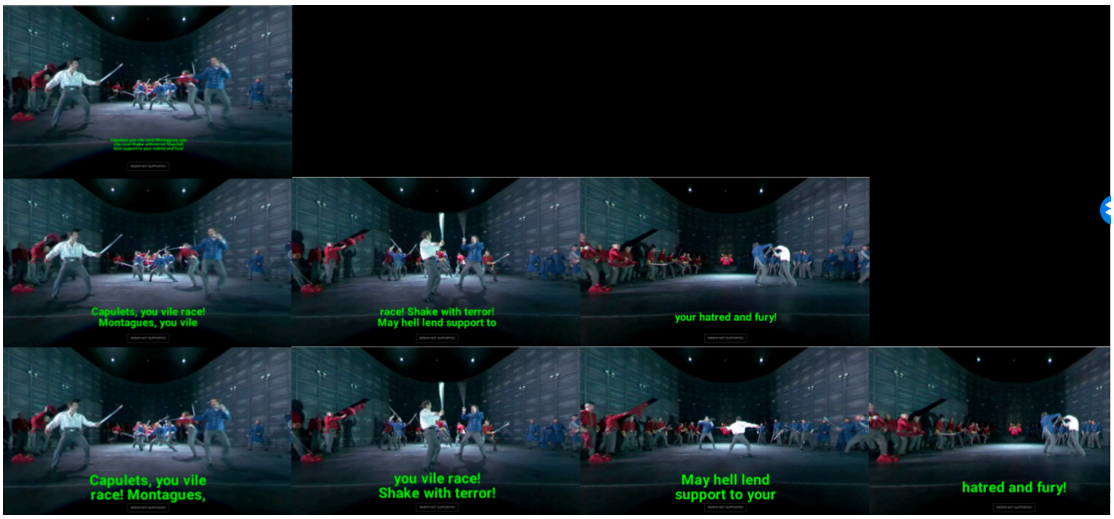

Figure 9: Responsive subtitles re-blocked based on consumers font size settings.

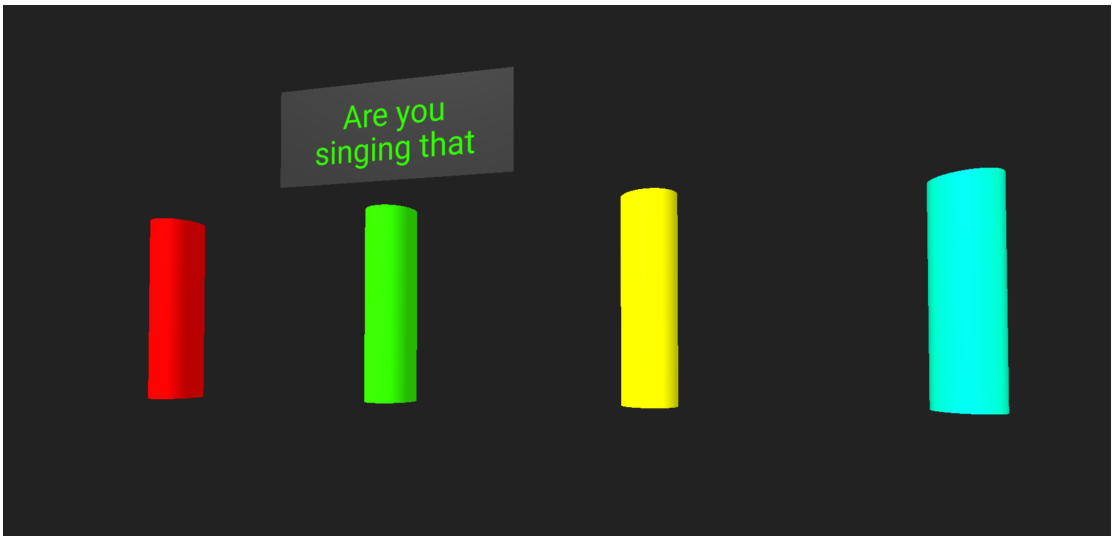

Figure 10: Responsive subtitles can also be used in non-linear VR applications, such as this high contrast reproduction of a 360^{0} video. The responsive subtitle library has automatically identified each unique character in the scene and represented each character as a clear cylinder in the VR scene.

4 Conclusions

The work-in-progress presented in this paper has been positively received during pre-pilot testing with target group representatives within the ImAc project and a full cross-national pilot study is now planned. In this paper we have identified many of the extra challenges faced in subtitling immersive content and presented the range of novel subtitle rendering and guiding methods, which have been implemented as part of the ImAc player in order to address them. We have also discussed how responsive subtitling can be a solution for dynamically adapting traditional subtitle content to new display technologies.

The research generated by the ImAc project is continuing to provide contributions to standardization (MPEG, W3C) and the development of a reference player. It is also planned for the ImAc player to be integrated into a full production environment.

ACKNOWLEDGMENTS

This work has been conducted as part of the ImAc project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement 761974.

REFERENCES

[1] European Commission, *The 2018 Ageing Report: Economic and Budgetary Projections for the EU Member States (2016-2070)*, Institutional Paper 079. May 2018. Brussels. ISBN 978-92-79-77460-7

[2] UN General Assembly, *Convention on the Rights of Persons with Disabilities : resolution / adopted by the General Assembly*, 24 January 2007, A/RES/61/106, available at: https://www.refworld.org/docid/45f973632.html [accessed 20 March 2019]

[3] Sahil Patel, *85 percent of Facebook video is watched without sound*, Dididay UK Article, May 2016, https://digiday.com/media/silent-world-facebook-video/ [accessed 20 March 2019]

[4] Latifi, Mehdi, Mobalegh, Ali, Mohammadi, Elham, *Movie subtitles and the improvement of listening comprehension ability: Does it help?,* The Journal of Language Learning and Teaching 1 / 2 (July 2016): 18-29.

[5] Hayati, Abdolmajid & Mohmedi, Firooz. (2009). The effect of films with and without subtitles on listening comprehension of EFL learners. British Journal of Educational Technology. 42. 181 - 192. 10.1111/j.1467-8535.2009.01004.x.

[6] M. Armstrong, A. Brown, M. Crabb, C. J. Hughes, R. Jones and J. Sandford, "Understanding the Diverse Needs of Subtitle Users in a Rapidly Evolving Media Landscape," in *SMPTE Motion Imaging Journal*, vol. 125, no. 9, pp. 33-41, Nov.-Dec. 2016. doi: 10.5594/JMI.2016.2614919

[7] Crabb, M, Jones, R, Armstrong, M and Hughes, CJ 2015, Online news videos : the UX of subtitle position , in: 17th International ACM SIGACCESS Conference on Computers & Accessibility, 26-28 October 2015, Lisbon, Portugal.

[8] P. Orero, C. A. Martín, M. Zorrilla, "HBB4ALL: Deployment of HbbTV services for all", IEEE BMSB'15, Ghent (Belgium), June 2015.

[9] Hybrid Broadcast Broadband TV (HbbTV) 2.0.2 Specification, HbbTV Association Resource Library, https://www.hbbtv.org/resource-library, February 2018.

[10] J. A. Núñez, M. Montagud, I. Fraile, D. Gómez, S. Fernández, "ImmersiaTV: an end-to-end toolset to enable customizable and immersive multi-screen TV experiences", Workshop on Virtual Reality, co-located with ACM TVX 2018, Seoul (South Korea), June 2018.

[11] ISO/IEC 23009-1: 2012. Information Technology. Dynamic Adaptive Streaming over HTTP (DASH). Part 1: Media Presentation Description and Segment Formats. April 2012.

[12] Andy Brown, Jayson Turner, Jake Patterson, Anastasia Schmitz, Mike Armstrong, and Maxine Glancy. 2017. Subtitles in 360-degree Video. In *Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video* (TVX '17 Adjunct). ACM, New York, NY, USA, 3-8. DOI: https://doi.org/10.1145/3084289.3089915

[13] S. Rothe, K.Tran, H. Hussmann, “Positioning of Subtitles in Cinematic Virtual Reality”, ICAT-EGVE 2018 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments, November 2018
[14] M. Schmierl, S.Rothe, “Methods of Speaker Identification for Subtitles in Cinematic Virtual Reality“ (Bachelor Thesis), November 2018
[15] WD on ISO/IEC 23000-20 Omnidirectional Media Application Format
[16] L. Men, N. Bryan-Kinns, A.S. Hassard, Z. Ma, “The impact of transitions on user experience in virtual reality”, IEEE Virtual Reality (VR) 2017, Los Angeles (USA), March 2017.
[17] Y. Lin, Y. Chang, H. Hu, H. Cheng, C. Huang, M. Sun, “Tell Me Where to Look: Investigating Ways for Assisting Focus in 360° Video”, ACM CHI '17, Denver (USA), March 2017.
[18] Hughes, CJ , Armstrong, M, Jones, R and Crabb, M 2015, Responsive design for personalised subtitles, in: The 12th Web for All Conference, 18-20 May 2015