UAB
Universitat Autònoma de Barcelona

Accessible VR 360º subtitles
The impact of subtitles on the reception of immersive content using eye tracking and questionnaire data

Marta Brescia-Zapata

PhD in Translation and Intercultural Studies

Accessible VR 360º subtitles

The impact of subtitles on the reception of immersive content using eye tracking and questionnaire data

PhD Thesis presented by:

Marta Brescia-Zapata

PhD supervisors:

Dr. Pilar Orero

Dr. Chris Hughes

2023

Departament de Traducció i d'Interpretació i d'Estudis de l'Àsia Oriental

UAB Universitat Autònoma de Barcelona

Universitat Autònoma de Barcelona
Departament de Traducció i d'Interpretació
i d'estudis de l'Asia Oriental
Doctorat en Traducció i Estudis Interculturals

Department of Translation and Interpreting
and East Asian Studies
PhD in Translation and Intercultural Studies

Accessible VR 360º subtitles

The impact of subtitles on the reception of immersive content using eye tracking and questionnaire data

PhD Thesis presented by:

Marta Brescia Zapata

Supervised by:

Dr Pilar Orero

Dr Chris Hughes

December, 2023

«Estudiar para ignorar menos»
Sor Juana Inés de la Cruz

"Are you done? Maybe it's funnier with subtitles"
Roy, K. (2023, 04, 23). Succession, Season 4, Episode 5

Acknowledgments

Lo difícil de hacer una lista de agradecimientos es terminarla, por extensión y por olvido, así que aquí va mi lista incompleta. Desen todos los olvidados por agradecidos.

A quien primero tengo que dar las gracias es a Anna Matamala, por animarme a pedir la beca que me llevaría a estos casi cuatro años de doctorado; y a Pilar Orero, por haberme enseñado (casi) todo lo que hace falta para ser investigadora (y por introducirme en el culto del *less is more*).

A mis compañeras del MRA les debo una ronda de abrazos y de cafés (de la cafetería de los ferros, claro), por hacer el que viaje de casi una hora desde mi casa hasta Bellaterra merezca la pena. Gracias, Mel, por apuntarte siempre a todas las locuras académicas y ser la mejor compañera a la hora de la siesta. Gracias, Sarah, por las recomendaciones musicales y los croissants de mantequilla. Gracias, Chiara, por los debates pasados y futuros. Gracias, Mariona, por tener siempre la sonrisa a punto y el *sticker* perfecto para cada situación.

Este tiempo de tesis me ha dado la oportunidad de coger varios (muchos) aviones, la mayoría de ellos para encontrarme en alguna parte del mundo con el *dream team,* las únicas personas por las que me sentaría en un Starbucks y sin las cuales esta tesis nunca habría existido. *Thanks*, Chris por cederme tu despacho para hacer pruebas y por todos los litros de café que compartimos en Salford. *Thanks*, Andrew por enseñarme a usar un repositorio svn y por las cervezas de vainilla en LA. *Dzięki*, Krzys, por acercarme a los misterios de la estadística y por los brindis de licor café.

Y hablando de viajes, en el viaje de la edad me he cruzado con compañeras que de alguna manera han estado siempre ahí en este recorrido hacia la tesis. Gracias, Iman, por ser mi casa siempre que vuelvo a Granada. Gracias, Ana, por diseñar conmigo la portada de esta tesis.

Por otro lado, tengo otro tipo de amigos a los que debo agradecer profundamente por salvaguardar mi salud mental durante estos últimos años. Gracias Jordi, Rafa y Fran Soria por abrir bares que podrían ser mi casa también. Gracias a todas las profesoras y amigas de Pole Dance Factory por sostenerme (literalmente) cada semana. Gracias a las *chalistas*, por crear un espacio mágico donde he podido reencontrarme con la poesía.

Por último, me gustaría dedicarle esta tesis a mi amiga Elena, la persona que más veces ha leído estas páginas aparte de mí misma.

A mi madre y a mi padre, por apoyarme siempre en todas mis decisiones, aunque muchas veces no tuviesen muy claro qué estaba haciendo exactamente o si estaba tomando las decisiones correctas. A mi hermana Ana, por poner siempre sus manos a mi disposición para hacer realidad cualquier cosa que se me pasa por la cabeza. A mi abuelo Rafael, para que pueda leer su nombre en este libro. A mi abuela Ana, que ya no podrá hacerlo.

A Uxío, por ser el faro que me ha guiado e iluminado para que esta tesis llegara a buen puerto. *No es manía ni locura, esto que tengo contigo, no es manía ni locura…*

Barcelona, 10 de octubre de 2023

Table of Contents

Índex of figures

Índex of tables

Abbreviation and Acronym Glossary

AD	Audio Description
ANOVA	Analysis of Variance
AOI	Areas-of-Interest
AVMSD	Audiovisual Media Service Directive
AVT	Audiovisual Translation
CRPD	Convention for the Rights of Persons with Disabilities
EAA	European Accessibility Act
EBU	European Broadcasting Union
H1	Hypothesis 1
H2	Hypothesis 2
H3	Hypothesis 3
HMD	Head Mounted Display
IV	Independent Variable
MA	Media Accessibility
MLM	Multilevel Linear Models
Ofcom	Office of Communications (UK)
SDH	Subtitling for the Deaf and Hard of Hearing
TS	Translation Studies
UK	United Kingdom
UN	United Nations
UX	User Experience
VE	Virtual Environment
VR	Virtual Reality
WAD	Web Accessibility Directive
XR	eXtended Reality

CHAPTER 1

1. Introduction

Translation Studies (TS) has been on a quest to define its "name and nature" (Holmes, 1972, p. 68) since its inception as an academic discipline in the 1970s. It has evolved from early descriptivism (Nida, 1975; Reiß & Vermeer, 1996; Nord, 1997) to more experimental approaches (Di Giovanni & Gambier, 2018; Orero et al., 2018). Audiovisual Translation (AVT), which was born in the second half of the 20th century, is one of the most vibrant disciplines within the field of TS. As defined by Gambier (2013, p. 45), AVT is a branch of translation studies concerned with "the transfer of multimodal and multimedia speech into another language and/or culture." It implies the intermodal translation of audiovisual texts, that is, translation between different semiotic modes (images, sounds, written words, spoken words, etc.) through specific audiovisual translation modalities (subtitling, dubbing, voice-over...) Research on AVT has increased in recent years thanks to the proliferation of norms and guidelines (Pedersen, 2020), and the implication of both professionals and end-users in the field (Romero-Fresco, 2013; Perego et al., 2015). Furthermore, AVT has been influenced by the emergence of novel technologies, the diversification of audiences, and the promotion of equality and media accessibility (MA) policies (Chaume, 2013). These developments have led to the proliferation of AVT modes that are also known as accessibility services, e.g., audio description (AD) (Maszerowska, Matamala, & Orero, 2014; Fryer, 2016) and subtitling for the deaf and hard-of-hearing (SDH) (Neves, 2005; Matamala & Orero, 2010; Romero-Fresco, 2015).

When talking about accessibility it is fundamental to mention "the right to equal access to information and culture and the need for special conditions for those who do not have sufficient access to messages that are conveyed via audiovisual media" (Neves, 2005, p. 15-16). The concept of participation in social life on a level playing field for all was introduced in 2006 with the Convention on the Rights of Persons with Disabilities (CRPD) issued by the United Nations (UN). The CRPD came into force in 2008, and by June 2023 it had 164 signatories and involves 187 parties (UNCRPD, n.d.) leading to various stakeholders showing interests in the concept —from lawmakers, through broadcasters, to end users. In the European context, the CRPD led to the development of three directives, namely, the Audiovisual Media Service Directive (AVMSD), the Web Accessibility Directive (WAD), and the European Accessibility Act (EAA). These policies, combined with European Standard EN 17161 (2019) and EN 301549 (2019), have advanced the implementation of accessibility services in all forms of multimedia content. Although the CRPD states that these services are meant for 'persons with disabilities', a report by the UK media regulator Office of Communications (Ofcom, 2017) defend that the audience who benefits from these services is heterogenous. Accessibility services can benefit to a wider audience, including the elderly, children, foreigners, or tourists (see Krejtz et al., 2012; Walczak, 2016; Szkriba, 2022). Thus, accessibility should no longer be perceived in a clinical taxonomy concerning only a specific group but rather to cater to the diverse needs of the audience (Greco & Romero-Fresco, 2022).

This PhD is framed within the universal model of accessibility. Its origins are linked to the launch in 2020 of the EU H2020 project TRACTION, under Grant Agreement Nº 870610. TRACTION was a collaborative research project between academia and the industry, which aimed at contributing to the renewal of opera as a territory of cultural and social inclusion. The partners of the consortium range from opera companies (Liceu, Irish National Opera), an art school (SAMP), two universities (Universitat Autònoma de Barcelona, Dublin City University), and two research institutes (VICOM and CWI) along with an immersive 360º content producer (VRI). Within the project, three exploratory operas were co-created involving diverse communities in three different cities: *La gata perduda*/The lost cat, a community opera co-created with 300 local residents from Raval neighbourhood (Barcelona) led by Liceu Opera House; *O tempo (Somós Nós)*/Time (As we are), a community opera co-created with a youth prison community in Leiria led by SAMP; and *As a nGnách*/Out of the Ordinary, a immersive opera co-created with Irish rural communities led by Irish National Opera. Given to the immersive, bilingual nature of this last opera, the challenge of providing access to all audiences became the focal point for the present dissertation.

The main area of study in this research is subtitling, adopted as a universal service to provide access to audiovisual media (see §1.2.2). It deals with subtitling from an aesthetical and technical point of view, taking as reference works by Arnáiz-Urquiza (2012) and Fox (2016, 2018). It focuses on the reception of subtitles displayed in immersive environments —a topic that has gained interest in recent years from both industry stakeholders (EBU, 2017) and academic researchers (Agulló et al., 2018; Bouwels, 2022; Brown et al., 2017; Montagud et al., 2019; Rothe et al., 2018). By investigating solutions to visualise subtitles in 360º environments, this PhD contributes to promoting inclusive media experiences and enhances accessibility in immersive content.

The main novelty of this dissertation lies in the methodological framework. While previous studies focused on assessing presence and user preferences using only subjective measures, this PhD follows a mixed-method approach that combines self-reported measures with eye-tracking technology (see §1.3). The effective combination of both objective and subjective methods allowed the design of a new methodological framework capable of testing different subtitle solutions in 360º environments. To ensure the scientific validity and applicability of the findings, tests with users were conducted across different countries, including the United Kingdom, Spain, Poland, and Slovenia. Following Oncins and Orero (2020), the main criteria for selecting participants was their reading skills instead of their hearing disabilities. Regarding the background, this PhD takes into consideration all lessons learned during the ImAc project[1], where most of the challenges posed by immersive formats when creating accessible experiences were identified and defined (Montagud et al., 2021). The present dissertation, which is presented as a compendium of publications (see §1.4), comprises both theoretical (Chapter 2 and 3) and empirical studies (Chapter 4, 5, and Annex 1). The last chapters detail the development and results of the present work.

This dissertation is registered in the PhD programme in Translation and Intercultural Studies (Doctorat en Traducció i Estudis Interculturals) at the Departament of Translation, Interpreting and East Asian Studies (Departament de Traducció i d'Interpretació I d'Estudis de l'Asia Oriental) of the Universitat Autònoma de Barcelona. It was written within the framework of the H2020 TRACTION project [Grant Agreement Nº 870610]. It was partially supported by funding from the Catalan Government [2021SGR00077].

1.1. Objectives and hypotheses

The main goals of this dissertation are:

1. **To design and validate a mixed methodological framework for testing subtitles in VR 360º media content using eye-tracking technology in real time.**

The following hypothesis (H) has been established:

H1: A pre-established experimental set up yield to objective quantitative data when exploring subtitle solutions in VR 360º media content.

[1] https://www.imacproject.eu/

2. **To evaluate two different subtitling solutions in VR 360º media content:**

– Position of subtitles: head-locked or fixed.
– Colour of subtitles: monochrome or coloured.

The rationale behind choosing position as the first independent variable (IV) is related to previous studies which identified it as one of the main challenges when implementing subtitles in 360º videos (Brown et al., 2018, Rothe et al., 2018). Unlike traditional 2D content, 360º videos do not offer a static image, but rather allow viewers to decide where to look within the 360º sphere. Different solutions where designed (Brown, 2017) and tested (e.g., Brown et al., 2018; Rothe et al., 2018; Agulló & Matamala, 2020) identifying head-locked subtitles as the preferred option. To verify these results, the present dissertation uses eye-tracking to analyse viewers behaviour in immersive media when watching subtitles place in different positions (for details, see Chapter 5; Brescia-Zapata et al., forthcoming b, and Annex 1).

The motivation behind selecting colour as the second IV is related to previous studies that highlighted the need to guide users to the sound source of the subtitles in VR 360º media content (e.g., Agulló & Matamala, 2019). To address this issue, the ImAc project designed and developed several guiding mechanisms, and tests results indicated that an arrow was the preferred method (Agulló et al., 2019). Nonetheless, the difficulty of identifying the speaker when reading subtitles was already a problem in 2D content. Mälzer-Semlinger (2015) proposed to attribute a consistent colour to each character throughout the entire film, a solution yet to be tested in VR 360º subtitles. Testing the validity of this colour-based solution in immersive environments is one of the main objectives of this dissertation (for details, see Chapter 5; Brescia-Zapata et al., forthcoming b, and Annex 1).

Based on the above, the following hypotheses (H) have been established:

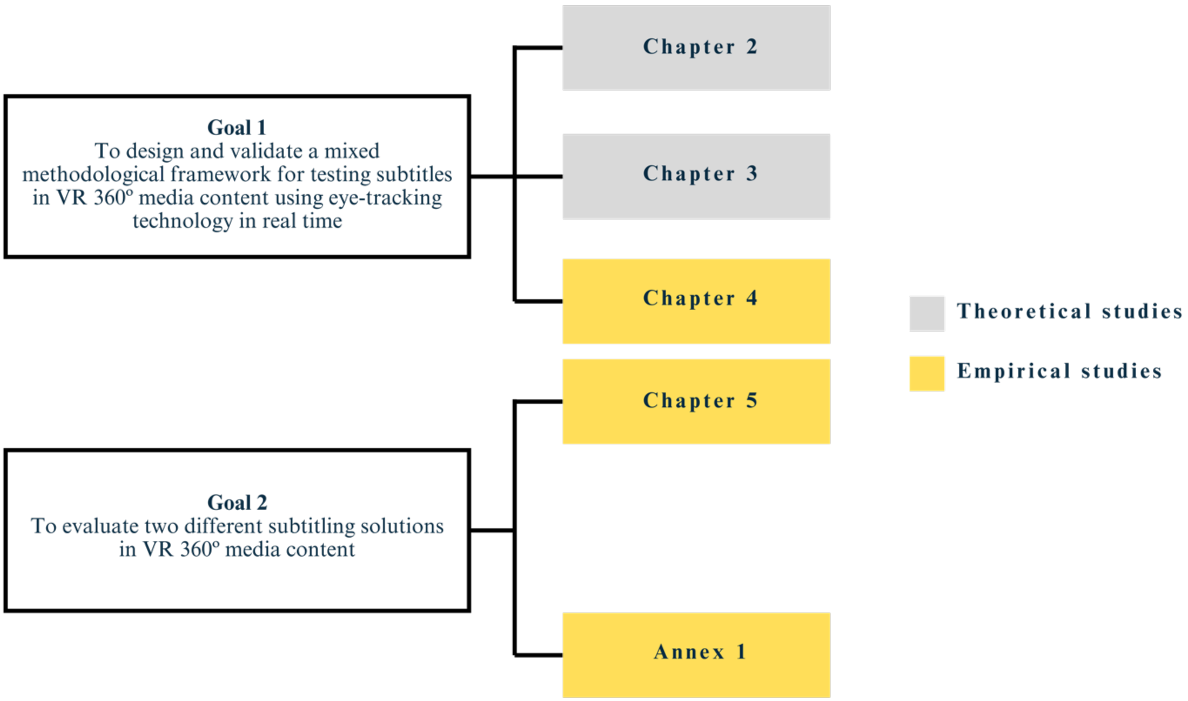
H2: Head-locked subtitles allow for better scene comprehension compared to fixed subtitles. Head-locked subtitles are the preferred option.

H3: Coloured subtitles allow for better character recognition compared to monochrome subtitles. Coloured subtitles are the preferred option.

The main goals are addressed in different articles (see Figure 1).

Figure 1

Summary of the present PhD objectives and corresponding chapters



The first main goal is addressed in Articles 1, 2, and 3. Article 1 establishes a theoretical foundation while Article 2 conducts a critical analysis of existing technologies and prior research in the field. Article 1 is titled *Accessibilities in immersive environments* and is schedule to be published by Comares. Article 2 is titled *The present and the future of accessibility services in VR360 players* and has already been published in *inTRAlinea*. These articles are presented in Chapter 2 and Chapter 3, respectively. Article 3 describes and validates the methodological framework, titled *VR 360º subtitles: Designing a test suite with eye-tracking technology*. It has been published in *The Journal of Audiovisual Translation* and is presented in Chapter 4. The second main goal is addressed in Article 4, and further extended in Annex 1. Article 4, titled *Subtitles in 360º video. Results from an eye-tracking experiment,* is scheduled to be published in *Perspectives* and is presented in Chapter 5. The structure of this dissertation is summarised in §1.5.

1.2. Theoretical framework

The present research is mainly grounded in concepts from AVT Studies, more specifically from Media Accessibility (MA), subtitling, and SDH. Due to the nature of the study, this PhD dissertation adopts a multidisciplinary approach, as it incorporates the concept of immersive storytelling from the field of communication studies. Moreover, this research uses eye-tracking technology, a type of methodology used in psychology and behavioural sciences for measuring eye movements in order to study attentional processes.

In the following subsections, the concepts of subtitling, media accessibility, SDH, and immersive storytelling will be treated separately, specifying their connection within the scope of this dissertation. The first three concepts will be covered within the theoretical framework of TS; the concept of eye-tracking will be covered in §1.4, as part of the methodology subsection.

1.2.1. AVT Studies: Media Accessibility

Accessibility in AVT studies was addressed for the first time by Gambier (2003, p. 179) in relation to screen translation, who declared that "the key word in screen translation is now accessibility." Gambier (ibid) classifies AVT services into two groups: dominant types (interlingual subtitling, dubbing, consecutive interpreting, simultaneous interpreting, voice-over, free commentary, simultaneous translation, and multilingual production), and challenging types (translating scenario/script, intralingual subtitling, live subtitling, surtitling, and audio description). Regarding research in AVT, Gambier (2003, p. 187) underlines the need for reception studies, arguing that "experimental methods could provide insights into the effects of particular subtitle features."

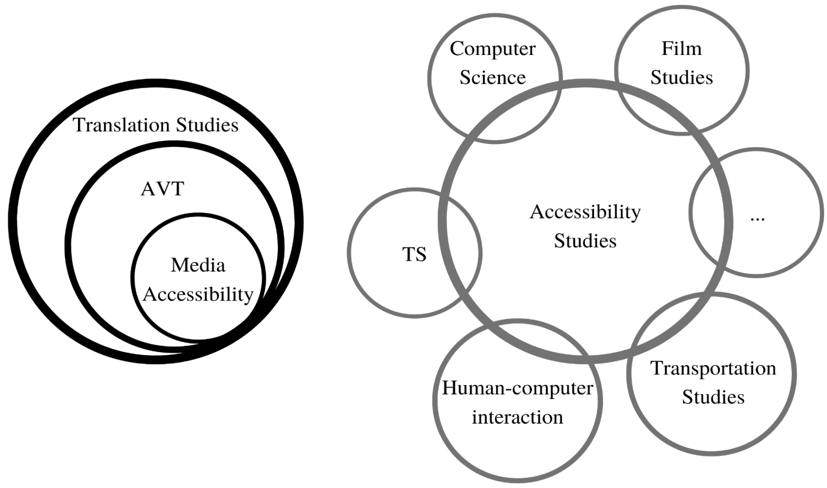
Díaz-Cintas (2005) relies on Jakobson's taxonomy (1959) to justify why modalities such as SDH and audio description (AD) can be studied within the scope of AVT. According to Díaz-Cintas (ibid, p. 4), "accessibility in our field means making an audiovisual programme available to other people that otherwise could not have access to it, always bearing in mind a degree of impairment on the part of the receiver". In the same article, Díaz-Cintas (ibid, p. 4) goes a step further highlighting the need for both linguistic and sensory support "to facilitate the access to an otherwise hermetic source of information and entertainment".

Orero and Matamala (2007, p. 269) identify media accessibility as "a growing subfield of Audiovisual Translation Studies" and confirm SHD and AD as the prominent modalities in this field. Likewise, Greco (2016, p. 23) pointed out that subtitling and AD pertain to MA as "AVT is the field where MA has been developing as a discipline for the last decade". In the same work, Greco (ibid, p. 23) defines MA as "a set of theories, practices, services, technologies and instruments providing access to audiovisual media content for people that cannot, or cannot properly, access that content in its original form."

The relationship between AVT Studies and MA has been explained by Greco (2019) in three accounts. The first frames MA as a sub-area of AVT concerned with AVT for persons with sensory disabilities and includes a restricted set of modalities (Orero 2004; Díaz-Cintas & Remael 2007). The second makes MA overlaps with AVT itself (Orero & Matamala, 2007). The third account has been called the Universalist Account and allows for the inclusion of other groups (e.g., persons with cognitive disabilities, persons with learning disabilities, persons with psychosocial disabilities) and other access services (e.g., Clean Audio, Easy Language, or web accessibility). Greco (ibid) concludes that when AVT meets accessibility, it goes beyond TS, becoming a part of a wider, transversal field that includes other areas of research and practices focused on accessibility (see Figure 2).

Figure 2

Relation between accessibility and AVT



Note. Author's own elaboration based on "Accessibility studies: abuses, misuses and the method of poietic design," by G. M. Greco, 2019, *HCI International 2019 – Late Breaking Papers. HCII 2019. Lecture Notes in Computer Science*, *11786*, 15-27 (https://doi.org/10.1007/978-3-030-30033-3_2).

Based on this reflection, Greco (2019, p. 18) offers a second, extended definition of MA that reads as follows: "access to media and non-media objects, services and environments through media solutions, for any person who cannot or would not be able to, either partially or completely, access them in their original form." This approach is based on human diversity and goes in line with the belief that different people have different access needs, and that one size does not fit all anymore.

The current study aligns with the second definition by Greco (2019), which emphasises the dynamic and evolving nature of MA. It also goes in line with Gambier (2003), as it uses experimental methods; and with Díaz-Cintas (2005), as it understands that accessibility as a mean to provide access to any type of audiovisual content. From a content perspective, this thesis focuses on immersive media. As immersive technologies are still evolving, this research highlights the need for the MA community to seize the opportunity to integrate accessibility from the onset (Romero-Fresco, 2015). From a user perspective, this study posits that accessible services should cater to a diverse audience, moving beyond earlier definitions that considered them only for "people with disabilities" (UNCRPD, n.d.) As stated by the International Telecommunication Union,

> rather than treating accessibility as important solely for addressing the needs of people with disabilities, elderly people or those with temporary impairments, accessibility should be regarded as a universal requirement that aims to ensure that Information and Communication Technologies (ICTs) of all kinds can be used with ease by people with the widest range of capabilities. In this way, all the world's people, whatever their individual abilities or disabilities, stand to benefit from an 'accessible ICT world' (Orero, 2016, p. 250.)

This approach was also supported by Romero-Fresco (2022, p. 18) who declared that "the universal has been instrumental in increasing and improving quantity and quality in MA." Beyond its scholarly implications, this approach fosters social integration and increases the access to culture, two cornerstones at the core of this dissertation.

1.2.2. AVT Studies: subtitles and SDH

Over the years, subtitling has been defined in many ways and classified from different points of view. From a classical, linguistic perspective, subtitles can be intralingual (e.g., an English programme with English subtitles) or interlingual (e.g., an English programme with Spanish subtitles). According to Linde and Kay (1999), the difference lies in subtitle users. In the first case, subtitles aim at people with hearing disabilities, and in the second, at non-native language users. From a technical point of view, subtitles can be either open or closed. Open subtitles always are burnt into the image and visible to all viewers, whereas closed subtitles can be turned on and off (Szarkowska, 2020). Closed subtitles used to be considered intralingual, whereas open subtitles were associated with interlingual translation (Robson, 2004; Gambier, 2003).

Following the former classification, Gambier (2003) defines SDH as intralingual subtitles which are used by deaf and hard of hearing people or migrants who need to improve their skills in a foreign language. Díaz Cintas (2005, p. 4) also identifies SDH as a type of intralingual subtitling:

> intralingual subtitling or captioning provides a written rendering on the screen of the dialogues uttered by the characters as well as complementary information to help deaf viewers identify speakers and gain access to paralinguistic information and sound effects that they cannot hear from the soundtrack.

According to this definition, SDH target audiences are deaf and/or hard of hearing people. This statement relates to Linde and Kay's idea that there are two types of subtitling based on the languages involved in the translation process. As stated by Díaz-Cintas (2003, p. 199), "intralingual subtitles, which simply transfer from a source language (SL) to a target language (TL) and intralingual subtitles (also known as captioning) where there is no change of language."

The language issue was also addressed by Pereira (2005) from a different perspective, considering SDH as a TAV modality that might also occur between languages. SDH interlingual subtitles were further studied by Szarkowska (2013, p. 68), who stressed out that "the common denominator of both intra- and interlingual SDH is that in addition to dialogue, they also include extra information about speakers, music and sounds important to the understanding of the film."

The arrival of the technological age motivated a more recent definition put forth by Neves (2018), who introduces two novel concepts: "enriched content" and "responsive design." The term 'enriched' refers to supplementary elements incorporated into subtitles to cater for the specific users' needs and preferences. The term 'responsive' stands up for the standardised attributes that facilitate the transfer of subtitles across various platforms and media. The term "enriched responsive subtitles" acknowledges the evolving nature of subtitles across diverse media environments and fits in the context of the present study as it allows for new subtitle features that are present in VR. As explained by Neves (2018, p. 91),

> the effectiveness of SDH revolves around three main criteria: readability, understanding and enjoyment. Achieving these will guarantee a fulfilling 'user experience'—

a concept used in reference to human-computer interaction but that can also be applied to the active consumption of subtitled audiovisual content.

The arrival of the technological age motivated a more recent definition by Neves (2018), who introduced two novel concepts: "enriched content" and "responsive design." The term 'enriched' refers to supplementary elements incorporated into subtitles to cater for the specific users' needs and preferences. The term 'responsive' stands up for the standardised attributes that facilitate the transfer of subtitles across various platforms and media. The term "enriched responsive subtitles" acknowledges the evolving nature of subtitles across diverse media environments and fits in the context of the present study as it allows for new subtitle features that are present in VR. As explained by Neves (2018, p. 91),

> the effectiveness of SDH revolves around three main criteria: readability, understanding and enjoyment. Achieving these will guarantee a fulfilling 'user experience'—a concept used in reference to human-computer interaction but that can also be applied to the active consumption of subtitled audiovisual content.

More recently, Agulló (2020) proposed a three-category subtitling classification based on where the emphasis lies: product-oriented (e.g., Díaz-Cintas & Remael, 2007; Gottlieb, 1994), process-oriented (e.g., Gambier, 2003), and user-oriented (e.g., Karamitroglou, 1998). In the same work, Agulló (2020, p. 25) offers an updated definition that reads as follows:

> Subtitling is an instrument that serves the purpose of supporting users in understanding any type of audiovisual content by displaying fragments of written text (integrated in the images and conveying the relevant linguistic or extralinguistic aspects of the content for its understanding) that are legible, readable, comprehensible and accessible for the intended users.

According to Agulló (ibid), this definition intentionally omits references to interlingual/intralingual subtitles as well as to users' abilities/disabilities. It focuses on the implementation of audiovisual experiences in a manner that maximise inclusivity and cater to the needs of a broader audience (Zárate, 2021). The result is a more universal approach, inspired by the UCT model (Suojanen et al., 2015). This dissertation aligns with Agullo's definition, as it is founded on the premise that subtitling serves as a universal service for accessing AV content. It holds the view that the traditional distinction between subtitling (for viewers who can hear) and SDH (for viewers with hearing impairment) is no longer relevant.

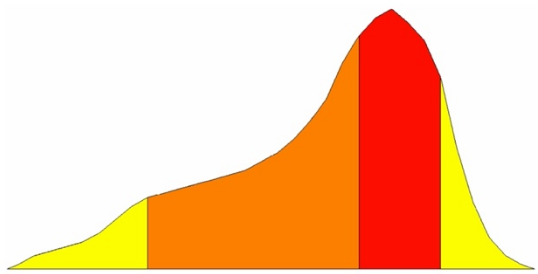
It also understands subtitling as a customisable service, that can adapt its attributes (language, character identification, sounds, etc.) depending on the audience needs and preference.

1.2.3. Immersive storytelling

The roots of traditional narratives trace back to Aristotle (330 BC), who provided the earliest analysis of what (throughout later centuries) became known as traditional drama. Following his ideas and theories, classic, 2D narratives are divided in four phases: exposition, ascension, climax, and conclusion (Fuhrmann, 1992), as illustrated in Figure 3. The person experiencing the story follows one story line path, and the only possible interactions are to move forward, to go back, or exit the narrative.

Figure 3

Classical story line with the four phases: exposition, ascension, climax, and conclusion

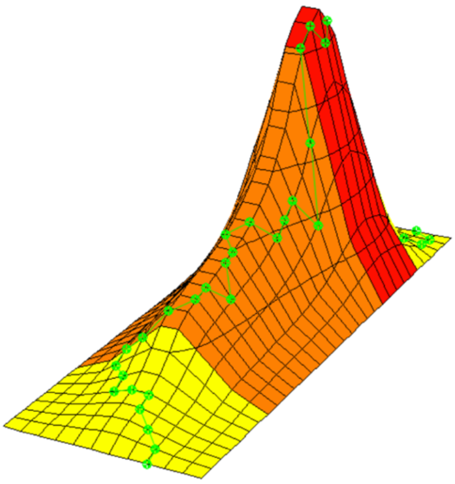


Note. From "Expanding the storyline," by P. Hoffmann & M. Herczeg, 2004, *Museums and the Web 2004, Proceedings*.

Immersive storytelling follows the same strategies, but it offers a dimension which is unavailable in traditional storytelling. It involves a multi-sensory experience that places the audience at the centre of the narrative. In contrast to traditional 2D media, the user ceases to be a passive observer and becomes an active participant with the ability to make choices and affect the outcome of the story (Skult & Smed, 2020). As explained by Hoffmann and Herczeg (2006, p. 38), "the two dimensional story line is expanded with at least one more dimension so it can no longer be called a (story) line" (see Figure 4). This non-linear narrative where space becomes essential is inspired by immersive theatre (Eglinton, 2010), which allows their audience to freely watch and move as they choose. That said, immersive storytelling—also known as digital or interactive storytelling—can be defined as an experience that engages the audience in a three-dimensional narrative space, enabling them to explore, interact, and influence the progression of the story.

Figure 4

Immersive storytelling: example for a story path



Note. Note. From “Expanding the storyline,” by P. Hoffmann & M. Herczeg, 2004, *Museums and the Web 2004, Proceedings*.

Immersive storytelling is linked to XR technologies, which have opened new avenues for creators and audiences to engage with narratives in innovative ways. On the one hand, XR technologies are considered expressive media with unique narrative representation qualities (Bruni et al., 2022). Also, narratives delivered through XR can “create [...] a greater emotional nexus” (Cantero de Julián et al., 2020, p. 418), and may encourage greater empathy and engagement with the issues presented. On the other hand, this immersive, interactive new media has introduced five challenges to storytelling that were not considered in traditional, 2D narratives: presence, interactivity, narrative structure, user comfort, and technical limitations.

The idea of presence can be defined as “the (psychological) sense of being in the virtual environment (VE)” (Slater & Wilbur, 1997, p. 605). When defining presence, it is important to make a distinction with the concept of immersion. According to Slater (2003), presence refers to perceptual and psychological responses to the system while immersion is directly related to some qualities of the technology. To achieve presence there are three factors that content creators must consider (Shubert et al., 2001): (i) the feeling of being physically transported to the VE, instead of watching it from outside, (ii) the focus on what is happening in the VE, instead of paying attention to the real environment, and (iii) the perception of the virtual world as the real world. If any of these elements are off, it can break the feeling of presence and ruin the experience.

Interactivity plays a fundamental role in immersive media, however, harmonising the interactivity of the medium with the interactivity of the narrative is considered one of the main challenges for narratives in XR (Bruni et al., 2022). For example, giving the user a large amount of freedom to explore and interact with the virtual world might result in missing important elements of the narrative (Danieau et al., 2017). Although some techniques such as the control of the depth of field (Hillaire et al., 2008) have been explored, how to direct the attention and guide the user in immersive content is still a work in progress and a pending lesson for content creators and filmmakers.

Digital technologies together with the need to adapt storytelling to the user's actions have taken the possibilities of interactive narratives to a new level (Ryan, 2000). In this way, narrative structures such as traditional linear storytelling in which the character is not driven by the user may not work in immersive media (Bruni et al., 2022). This opens a window for the exploration of interactive storytelling where user agency plays a crucial role in shaping the narrative experience.

Immersive media can be physically and emotionally intense which can lead to user discomfort or disorientation. Some viewers might experience VR sickness or dizziness when watching VR content (LaViola, 2000), which affects the feeling of presence. Moreover, exposing users to lengthy, overly complex narratives with many characters talking at the same time can further degrade the quality of the VR experience or even make it impossible to follow the plot, especially for novice users.

Technical factors may also disturb the illusion of being part of the story (Elmezeny et al., 2018). Immersive environments require high-quality graphics, sound, and other multimedia elements. Additionally, different hardware platforms may have different technical limitations, requiring storytellers to create multiple versions of their experience to ensure compatibility across devices.

Regarding applications, immersive narrative experiences can be found in multiple fields, including entertainment, healthcare, education, and journalism. In the entertainment sector, the filmmaking industry has been prolific in developing immersive experiences, evident through the inclusion of dedicated categories in film festivals such as the Venice International Film

Festival[2] or Cannes[3]. The video game industry has explored the possibilities of VR narratives, exemplified by titles such as "Lone Echo"[4] (2017) or "Tokyo Chronos"[5] (2019). In healthcare, immersive narratives have been investigated as valuable tools for training health professionals (Moreau et al., 2018), and have demonstrated potential in paediatric healthcare and rehabilitation (Arane et al., 2017). Immersive storytelling digital tools have also been integrated into mainstream teaching practices, e.g., to raise environmental awareness and foster empathy with people with disabilities (McDonagh & Brescia-Zapata, 2023). Furthermore, the capacity of immersive experiences to introduce new perspectives allowing users to become active participants in the story can also be exploited in the journalism industry as a way to make people care about what happens in the world (Domínguez, 2017). In the European context, some H2020 projects such as MediaVerse[6] and GreenSCENT[7] are exploring the potential of immersive journalism combined with the integration of accessibility services.

The present dissertation explores immersive content, a new area in narrative design that lacks well-defined working methods. According to Rouse and Barba (2017), prototyping and user testing are considered crucial elements in the design process of immersive narratives. To address this need, this dissertation adopts a novel web-based prototyping framework (Hughes et al., 2020), which enables real-time subtitle editing and visualization in 360º videos (see Chapter 4). To ensure the validity of the findings, the testing materials—stimuli—used during the experiments adhere to a pre-established set of requirements. The next section provides detailed information about the methodology followed during these studies.

1.3. Methodology

The methodology followed in this dissertation is explained in detail in each corresponding publication. However, for a better understanding and to give cohesion to the present work a summary is provided in the introductory chapter. To achieve the goals and objectives in §1.1 the following studies were carried out:

[2] https://web9.labiennale.org/sel/1vr.aspx
[3] https://www.marchedufilm.com/programs/cannes-xr/
[4] https://readyatdawn.com/game-list/lone-echo.html
[5] https://tokyochronos.com/en/
[6] https://mediaverse-project.eu/
[7] https://www.green-scent.eu/

Goal 1: To design and validate a mixed methodological framework for testing subtitles in VR 360º media content using eye-tracking technology in real time.

First, two descriptive studies were carried out to find the gaps. The first study takes a broader approach and explores the requirements needed to achieve accessible experiences in immersive environments (see Chapter 2). The second study focuses on commercial media players that support immersive content and analyses how and to what extent they integrate accessibility services (see Chapter 3).

Second, a methodological framework for rendering and testing subtitles in 360º was designed. Then, a pilot study was conducted to test the accuracy and reliability of the new methodology. The study had four stages: a questionnaire on demographic data, an eye-tracking test using 360º stimuli, a post-stimuli questionnaire, and a focus group. The data analysis was mainly qualitative accompanied by descriptive statistics of the answers obtained though the questionnaires and eye movements captured during the study (see Chapter 4).

Goal 2: To evaluate two different subtitling solutions in VR 360º media content.

Based on the information and feedback obtained from goal 1, a larger experiment was conducted to test two subtitling solutions in 360º videos: position and colour (see Chapter 5 and Annex 1). Regarding position, the two conditions were head-locked (subtitles that are always visible and are displayed in front of the viewer, following head movements) and fixed (subtitles appear near to the speaking characters and remain fixed to the scene). For colour, we tested monochrome (all subtitles in the same colour white on grey background) and colour (four colour were used to identify the different characters, following ISO 20071-23:2018 standard). The combination of these two subtitle features was treated as IV, constituting four experimental conditions: fixed colour, fixed monochrome, head-locked colour, head-locked monochrome. The main purpose of this experiment was to gather feedback from users evaluating general comprehension of immersive content, subjective effort and frustration, and attention allocation over subtitles and scene Areas-of-Interest (AOI).

1.3.1. Research tools

Reception studies in the context of TS has witnessed significant growth in recent years, fuelled by the increasing popularity of eye-tracking as a research tool and the rising interest within the scholarly community regarding accessibility (Orero et al., 2018). However, as highlighted by Meister (2018), researchers often apply different types of data, methods, and theory, yet they frequently fail to harness the potential of a comprehensive and integrated methodological framework. To address this limitation, the present dissertation adopts a mixed method approach combining questionnaires and eye-tracking instruments. The analysis of viewers preferences though questionnaires combined with technology-based empirical research allows researchers to measure the impact of different formats on the intake of audiovisual information (see Romero-Fresco, 2015). An example is a recent study by Black (2020) that measures children's reception of standards vs. integrated titles by applying Gambier's model of the three Rs (2013): *response* is measured via eye-tracking data; *reaction* is measured via scene recognition and content comprehension tests; and *repercussion* is assessed via post-experiment questionnaires and interviews. In line with this, the current study combines physiological process metrics (eye movements), performance metrics (scene comprehension), and subjective self-reports (task-load and preferences), in order to provide a comprehensive understanding of the audience experience.

1.3.1.1. Subjective methods: Questionnaires and focus groups

In AVT studies, eye-tracking studies have often been combined with post-stimuli questionnaires (Mangiron, 2016; Bosch, Soler-Vilageliu & Orero, 2020). For this study, a pot-stimuli customised questionnaire was designed combining content comprehension questions with two previously validated questionnaires: NASA Task Load Index (NASA-TLX) and Igroup Presence Questionnaires (IPQ). Following Black (2020), the content comprehension questions were designed to assess participants' comprehension of the visual, auditory verbal and non-verbal information in the clips. The questionnaire contained six multiple choice questions. Each question had five possible responses: one correct answer, two plausible distractors, and an 'I don't remember' option.

To assess the level of presence experienced by participants, the questionnaire incorporated two specific questions derived from the IPQ developed by Schubert (2003). The IPQ is grounded in a framework that explains the cognitive processes behind the sense of presence in virtual environments, particularly the construction of the own body in these environments and the attenuation of immediate sensory input (Schubert et al., 2001). The use of the IPQ in this study aligns with previous studies exploring the integration of subtitles in 360º environments (Agulló et al., 2018; Agulló & Matamala, 2020).

NASA-TLX is a multidimensional scale developed by Hart and Staveland (1988) to measure subjective workload. Its six subscales, namely mental demand, temporal demand, effort, performance, frustration level, and physical demand, comprehensively capture various aspects of workload perception. For the specific purpose of this study, physical demand was deemed irrelevant and therefore not included. While the scale has demonstrated its reliability in evaluating subjective translation difficulty in previous research (Sun and Shreve, 2014; Liu et al., 2019), its implementation in other fields within TAV studies remains limited. Given the objective of this dissertation to capture participants' subjective perception of workload when reading subtitles in 360º environments, the NASA-TLX scale has been considered appropriate and thus incorporated into the questionnaires.

In addition to this questionnaire, qualitative methods were also implemented in the studies that conform this dissertation. For instance, a pre-stimuli questionnaire was employed to collect demographic data (e.g., usage and attitudes towards digital media, VR, and subtitles), along with an ad-hoc questionnaire to gather feedback regarding users preferences. To enhance the comprehensiveness of the research design, a focus group was convened during the pilot test phase to collect insights and perspectives prior to the main experiment. These qualitative approaches not only enhance the quantitative data but also complement the information gathered through eye-tracking technology, as further elaborated in the subsequent section.

1.3.1.2. Objective methods: Eye tracking

In the latest years, eye-tracking has emerged as a prominent and widely used psychological tool in the field of MA, as evidenced by its extensive adoption (Hermosa-Ramírez, 2021). This preference is well-founded, considering the valuable insights it offers into attention, attention distribution, and scene processing (Orero et al., 2018), which are also pertinent to the present dissertation.

According to Fox (2018), eye-tracking studies on subtitles aim to gain insights into two aspects. Firstly, they focus on understanding what elements in the visual scene are perceived by analysing fixation duration. Secondly, they examine the processing of audiovisual text, including the reading speed, the extent of thoroughness in subtitle reading, the duration of fixations on different parts of the subtitles, the distribution of visual attention between the image and the subtitles, and ultimately, the overall impact of subtitles on the audience's comprehension of the film. Regarding measurements, recent eye-tracking studies on subtitle reception (d'Ydewalle et al., 1991; d'Ydewalle & De Bruycker, 2007; Moran, 2012; Caffrey, 2012; Ghia, 2012; Kruger & Steyn, 2014) have provided valuable insights into mean fixation duration, time to first fixation, fixations per subtitle, dwell time, and gaze control. Additionally, Romero-Fresco (2015) introduced the concept of viewing speed, which refers to the pace at which viewers engage with audiovisual content, encompassing the processing of subtitles, accompanying images, and sound. His study revealed a close relationship between presentation speed and the division of attention between image and subtitle reading.

Although there is a considerable number of research on eye-tracking studies focusing on subtitling in general (see Perego, 2012; Kruger et al., 2015), there remains a scarcity of studies examining subtitles presented in immersive environments. However, with the advent of VR equipment incorporating built-in eye trackers, there is an opportunity to capture objective performance measures that shed light on how subtitles are viewed in such environments. For real-time use of gaze with subtitles, Sidenmark et al. (2019) developed three novel techniques addressing subtitle depth conflicts in VR, i.e., backdrop blur, depth shift, and a hybrid of the two. Their study concluded that head-locked subtitles attached to the bottom of the field of view (similar to traditional 2D subtitles), proved to be effective. However, the study did not provide a detailed description of how the subtitles were rendered in the VR environment, leaving some gaps in the understanding of the rendering process. In a more recent study, Gadin (2021) used eye-tracking equipment similar to that used in this dissertation (HTC Vive Pro Eye) to assess text legibility in VR. However, subtitled 360º videos used in the study were pre-rendered, limiting the dynamic presentation of the text. Furthermore, it remains unclear whether the gaze data collected in the study were specifically analysed for fixations within subtitles. While the study conducted basic gaze analysis, including saccade amplitude, fixation, and fixation duration, the lack of dedicated gaze-on-subtitle analysis makes it challenging to discern specific instances of subtitle viewing.

This dissertation explores the use of gaze in VR in the context of subtitle evaluation. It delves into the evaluation of subtitles in the context of virtual reality (VR) by leveraging gaze analysis. The gaze data is captured directly from the SRAnipal SDK within Unity, enabling the detection of fixations on subtitles as AOIs. While the current processing of gaze analysis is performed offline, the methodology allows for potential real-time implementation, as discussed further in Chapter 4.

1.4. Structure of the dissertation

This dissertation is presented as a compilation of four research publications. One is a chapter of an academic book, and the other three are scientific articles that have been published or accepted for publication in peer-reviewed academic journals. All publications are available in their original version in Annex 2. The only changes made were to the format, such as the numbering of figures and tables and font style, while the content of each publication remains unchanged.

This thesis is organised into eight chapters. Its structure is described below in detail.

Chapter 1 presents the context for the research. It opens with an introduction discussing the motivation for the work (Section 1). Then it introduces the main objectives and hypotheses (Section 1.1), provides insights into the theoretical framework of the research (Section 1.2), and outlines the methodological aspects of the research (Section 1.3).

Chapter 2 and 3 constitute the theoretical part while Chapter 4 and 5 constitute the empirical research.

Chapter 2 includes Article 1 (Brescia-Zapata, forthcoming a). It discusses the main challenges that arise when integrating accessibility services into immersive environments. Some of the primary hurdles include the compatibility of assistive technologies with immersive platforms and the lack of guidelines to ensure a uniform and accessible user experience across various immersive environments. It also provides an overview of the fundamental criteria needed to generate accessible, immersive content, in addition to presenting cutting-edge case studies regarding the seamless integration of these services into 360º videos. The primary objective of this research is to advance the standardisation of accessibility service integration within the immersive media domain.

Chapter 3 includes Article 2 (Brescia-Zapata, 2021). It investigates to what extent the main accessibility services are incorporated into the most prevalent commercial VR360 players. Additionally, it examines the development of a fully customisable and accessible player, created based on the EU standard EN17161 and a user-centric approach. The paper starts exploring the relation between technology and AVT studies, followed by a brief survey of existing XR content. Then it provides a list of VR360 players that are commercially available and evaluates their incorporation of the primary accessibility services, including SDH, AD, and SL. The last section presents the player developed by the ImAc project, centred on accessibility.

Chapter 4 includes Article 3 (Brescia-Zapata et al., 2022). It describes the methodological framework designed for the first experimental setup aimed at incorporating creative subtitles in 360 environments. It also describes the choices made regarding the stimuli (e.g., sound, duration, and storyboard). The analysis reveals that evaluating subtitles in immersive media contexts is a multidimensional task that entails both linguistic and artistic aspects. An adaptable framework that enables comparison and contrast of different functionalities is essential to this process. The paper also outlines the difficulties encountered when generating stimuli to explore meaningful viewing behaviours of users using eye-tracking technology. The findings of this exploratory study offer insights into future research endeavours using eye-tracking technology to study subtitles in immersive media.

Chapter 5 includes Article 4 (Brescia-Zapata et al., forthcoming b). It highlights the potential of immersive environments, such as 360º videos, to enhance media accessibility for individuals who are deaf or hard-of-hearing and to promote cultural inclusivity through language translation. The objective of this paper is to compare two conditions of subtitles in 360º videos: position (head-locked vs fixed) and colour (monochrome vs colour). The analysis relies on a novel triangulation of data from three complementary methods: psycho-physiological attentional process measures (eye movements), performance measures (media content comprehension), and subjective task-load and preference measures (self-report). The findings suggest that head-locked, coloured subtitles can serve as good practice guidelines for subtitling in 360º videos.

Chapter 6 includes a summary of the thesis as requested in a compendium dissertation.

Chapter 7 discusses the results as well as the practical and theoretical implications of this work. It also offers ideas for future studies in the field.

Chapter 8 presents the bibliography. This is added as a separate chapter to provide a unified presentation of various citation styles and to offer updated versions of the references quoted in the publications.

The last chapter is dedicated to annexes. Annex 1 serves as an extension of Chapter 5, providing a compilation of data obtained from deaf and hard of hearing participants.

Annex 2 contains the original version of the articles (as published or accepted for publication). Annex 3 encompasses the ethical committee documentation. Annex 4 includes the questionnaires used in the experiments, while Annex 5 contains the digital and multimedia resources. All these elements have been included in accordance with the regulations for PhD theses that comprise a compendium of academic publications.

CHAPTER 2

Article 1: Accessibilities in immersive media environments

Accessibilities in Immersive Media Environments

Marta Brescia Zapata

Department of Translation and Interpretation, Universitat Autònoma de Barcelona, Barcelona, Spain

Abstract

Research on immersive media content has been intense in the last few years. In this context, 360º videos have become very popular as they are a cheap and effective way to provide immersive experiences. While it is possible to find plenty of accessibility solutions and guidelines for traditional media services, there is a lack of standards regarding accessibility in immersive media. During recent years, both audiovisual translation (AVT) and media accessibility (MA) studies have spotlighted these technologies, discussing how to integrate accessibility services into immersive environments. Moreover, these technologies are gaining popularity due to the COVID-19 crisis; as they enable interactive, hyper-personalised and engaging experiences anytime and anywhere. Thinking about how to implement accessibility services is the next logical step to guarantee universal access to the new virtual world.

This chapter will discuss the main challenges that arise when integrating accessibility services into immersive environments. Minimum requirements for creating accessible, immersive content will also be presented, as well as examples of the latest experiments on how these services could be integrated within 360º videos. The main goal of this work is to contribute to standardise the integration of accessibility services in immersive media.

Keywords: media accessibility, immersive media, virtual reality, 360º videos

Resumen

Durante los últimos años, los entornos inmersivos han sido un campo de investigación muy prolífico. En este contexto, los vídeos 360º son una manera barata y efectiva de crear experiencias inmersivas. Si bien es posible encontrar muchas guías y estándares de servicios de accesibilidad en medios tradicionales, en los entornos inmersivos no sucede lo mismo. El mundo de la traducción audiovisual (TAV) y de la accesibilidad a los medios (MA) también se

ha centrado en estas tecnologías con el objetivo de investigar cómo incorporar los servicios de accesibilidad en los entornos inmersivos. Una de las ventajas de las tecnologías inmersivas, que se ha magnificado durante la crisis derivada de la COVID-19, es que permiten disfrutar de experiencias interactivas y personalizables en cualquier momento y en cualquier lugar. Implementar los servicios de accesibilidad en este tipo de entornos es el paso lógico para garantizar el acceso universal a este nuevo mundo virtual.

En este capítulo se discutirán los retos y desafíos que se plantean a la hora de integrar los servicios de accesibilidad en entornos inmersivos. También se presentarán los requisitos mínimos para crear contenidos inmersivos accesibles para todos, así como ejemplos de los últimos experimentos sobre la integración de estos servicios en videos 360º. Este trabajo pretende contribuir a estandarizar la integración de los servicios de accesibilidad en los entornos inmersivos.

Palabras clave: accesibilidad, medios inmersivos, realidad virtual, videos 360º

2.1. Introduction

Immersive media have the potential to drastically transform our lives: from the way we work and communicate to the way we understand and experience the world (Creed *et al.,* 2023). Immersive technology is already a reality, with only a few pieces left to fall into place before it grows to become something that encompasses lives and cultures all over the world. We live surrounded by digital media, and more importantly, we must interact with digital media for many basic daily activities such as work, training, shopping, and communication. The COVID-19 pandemic further exposed and exacerbated digital inequalities (Beaunoyer *et al*., 2020), also known as the digital divide. The gap produced by the digital divide —evidenced by related concepts like digital inclusion, digital participation, digital skills, media literacy and media accessibility— needs to be reduced because it is one of the most challenging problems facing the information society (Aissaoui, 2021).

To get the most out of the information and communication technology and now that immersive environments are becoming mainstream, the next logical step is thinking about accessibility. The availability of affordable technology to both produce and consume media content was the challenge when 3D was considered. Early concerns about the individual nature of this format were quickly dispelled by the possibility of projecting immersive media, as seen in the Klimt art exhibitions (2021)[8] or Die Soldaten opera (2018)[9]. In the past, the only way to consume eXtended reality (XR) was with some special glasses (head-mounted display) which proved challenging in terms of the amount of time a person could stand to wear them, due to the possible side effects on their health such as nausea and dizziness. More recently, the possibility of a shared end-to-end system for the production and delivery of photorealistic, social, immersive experiences has become a reality (Gunkel *et al.,* 2018).

In the last decade, the UN decided to challenge the effects of the digital divide with the most vulnerable groups and people with disabilities. The Convention on the Rights of Persons with Disabilities (CRPD) was signed by almost all countries worldwide and it had an impact in Europe with the development of three EU Directives: the Audiovisual Media Service Directive (AVMSD), the Web Accessibility Directive (WAD), and the European Accessibility Act (EAA). A commonality across the different directives is the recommendation to thinking of accessibility in the design phase of any product or system. This means that Europe believes in the successful development of accessibility services in the development phase of a new technology. This has been true for many EU funded projects such as DTV4ALL (https://dea.brunel.ac.uk/dtv4all/), when Europe went from analogue to digital; and in Hbb4ALL (http://www.hbb4all.eu/), when Internet technology converged with broadcasting. Europe has continued funding research in that direction with the project called ImAc (https://www.imacproject.eu), which explored how accessibility services could be efficiently integrated with immersive media. In particular, ImAc aimed to ensure that immersive media experiences are inclusive across different languages, addressing the needs not only of those with hearing and low vision problems, but also of people with cognitive or learning difficulties, newcomers, people with low literacy, and the aged, in an interactive and personalised manner. The next challenge is to extrapolate all the ImAc learnings to digital immersive storytelling platforms while developing accessibility further.

This paper will start with a brief historical overview of the transition from traditional to immersive media experiences. It will also introduce the main concepts of XR, as well as a discussion on the main differences between these two environments. Section 3 will focus on the integration of accessibility with immersive media, describing four key areas where XR

8 For further information about the Klimt exposition, please consult https://www.expo-klimt.be/en/ [retrieved December 20, 2021]
9 For further information about Die Soldaten opera, please consult https://lafura.com/en/works/die-soldaten/ [retrieved December 20, 2021]

accessibility should be considered: accessibility of the XR contents, media players, user interfaces, and devices. The last section is a brief reflection on what has been analysed in the chapter, and an open door to a near future in which all traditional and immersive media are born accessible.

2.2. Traditional media vs new media environments

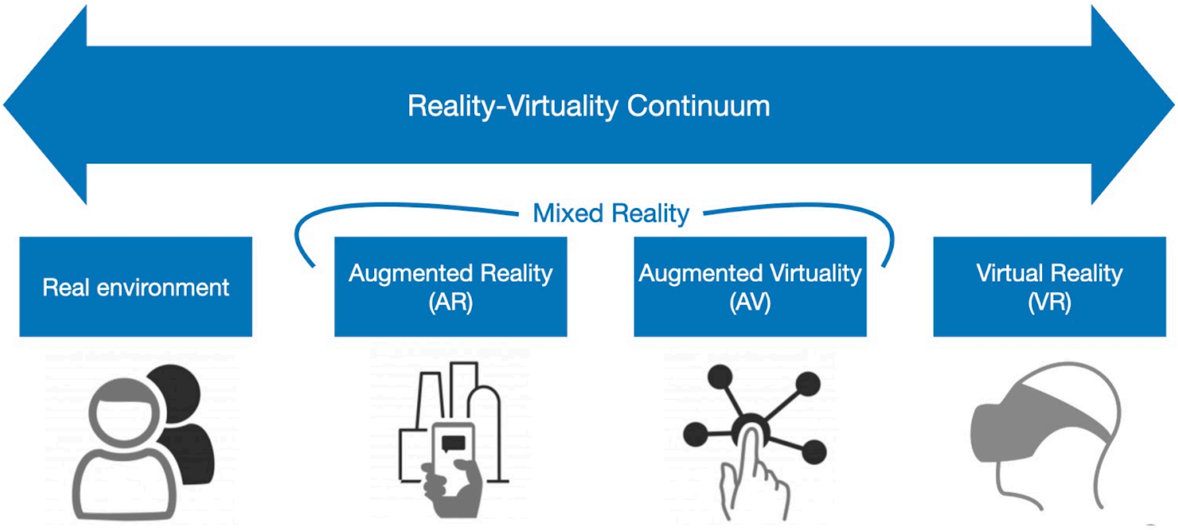
The idea of Virtual Reality (VR) began to take shape long before the appearance in 1956 of the first 3D device, the Sensorama (Puerta Domínguez, 2016). Until then, the concept was limited to science fiction, and all the media content was generated as 2D or flat images —also known as traditional media. Cinema was born at the end of the 19th century, and movie productions were black and white and mostly silent (Crafton, 1999). At the time, technology was not developed enough to allow for music or dialogue in these movies, however, they were accompanied by other elements, such as external music (piano or orchestra) and intertitles (slides that appeared between scenes or in the middle of a scene to provide contextual information for the spectators). With time, technology became more advanced, and sound began to be included in movies: from simple soundtracks or sound effects to dialogues. The expansion into cinema in colour took a little longer, but thanks to Technicolor, its popularity increased, and in the 1950s, black and white movies almost completely disappeared (Molina-Siles *et al.,* 2013).

Technology has continued to develop since then, and still seeks the same goal: to reproduce what we see and hear in our environment with the same quality as a human is capable of. The quest to develop a real digital environment that simulates reality led to the failed attempt to adopt 3D imaging as the mainstream display for audiovisual products in the 1960s. According to Agulló and Matamala (2019, pp. 219), this failure "may have opened a door for VR and 360º content as a new attempt to create engaging immersive experiences."

XR is a term referring to all the technologies that combine real and virtual environments and human-machine interactions generated by computer technology. XR includes the entire Reality-Virtuality Continuum (see Figure 5), also known as "Milgram's Continuum"*,* a concept that was first introduced in 1994 by Paul Milgram to describe all possible combinations of reality and computer-generated images. On the left side of the continuum there is the Real environment, unmodified images just as our eyes see them. On the opposite side, there is the Virtual Reality or VR, where the user is immersed in a computer-generated experience and cannot see the real world. In this case, the content can be accessed using VR devices (Head-Mounted Displays like Oculus Rift, HTC Vive, Samsung Gear, etc.) and can also be consumed as a CAVE (Cave Automatic Virtual Environment), which uses high-resolution projection screens to deliver 360º visual experiences. The area between the real environment and the virtual environment represents Mixed Reality (MR), with various degrees of overlap between reality and virtuality.

Figure 5

Reality-Virtuality Continuum



Note. Author's own elaboration based on "A taxonomy of mixed reality visual displays," by P. Milgram & F. Kishino, *IEICE TRANSACTIONS on Information and Systems*, *77*(12), 1321–1329.

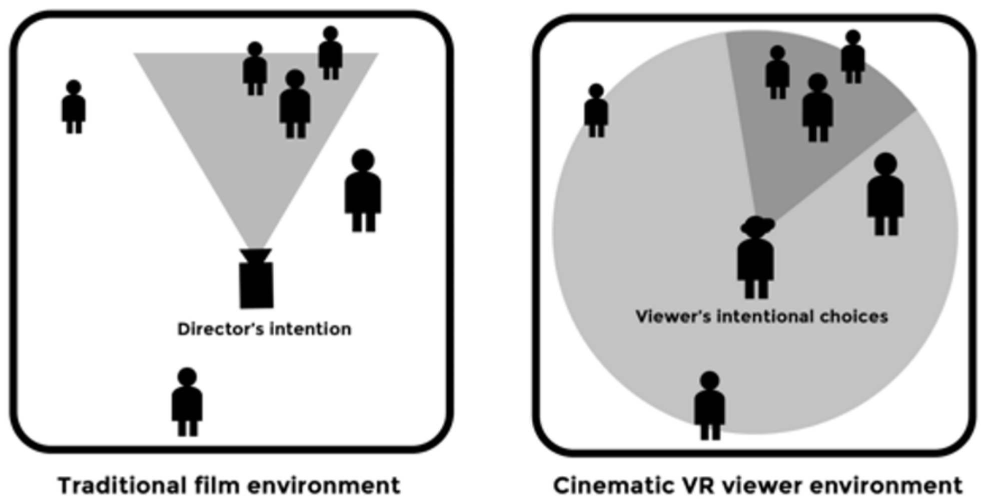
Immersive technologies are mainly designed to elicit presence, the experience of 'being' or 'acting', when physically situated in another place (Slater & Wilbur, 1997). This feeling refers to tricking at least three of the five senses of your brain into believing that you have been taken to another place. Still, immersive videos are limited in nature as they can be unrealistic or make people motion or simulator sick leading to a suboptimal experience that turns people off the immersive experience. According to Slater and Wilbur (1997), the process of achieving spatial presence is divided into three phases: place illusion ("I am here"), plausibility ("this is happening"), and body ownership ("it is my body").

Traditional media (static 2D imagery and 2D video) is the best media format for storytelling, while immersive media (XR experiences) invite users to interact with the content, and thus to modify it. In XR the narrative can be developed using space and movement over the temporal period. This is called “spatial storytelling” and consists of engaging the viewers inside a spatial environment that contains non-linear narratives that can be discovered by exploring the space provided.

Along with XR's new narratives, the main challenge for immersive experiences is that the field of view (FoV) is limited, and the interactor cannot pay attention to the entire visual scene in the given time (Aylett & Louchart, 2003). Traditionally, filmmakers have relied on four core tools to tell stories: cinematography, sound, *mise-en-scène* (the arrangement of the scenery), and editing (Skult & Smed, 2020). These techniques that largely define the current movie-making industry are not suitable for XR experiences. The interactor is free to explore the scene by looking wherever they want, which is why all these core approaches need to be reconsidered. As shown in Figure 6, in XR the viewer does not have a window to a world, he or she is directly present in the world instead. Orientation plays a crucial role especially in cuts since every cut requires the viewer to re-orientate themselves in the new environment. The pacing and playing should be more like it is in traditional theatre with a focus on cinematography.

Figure 6

Traditional film environment in comparison to an XR environment



Note. From “Making new narrative structures with actor’s eye-contact in cinematic virtual reality (CVR),” by K. Dong-uk, R. Hokyoung, & K. Jieun, *Interactive Storytelling: 11th International Conference on Interactive Digital Storytelling, ICIDS,* 11318, 2018, 343–347.

Another key feature is that immersive media (as any novel technology) can engage the audience at a higher level than traditional media, but this effect may subside once the novelty wears off. This is called "the novelty effect", and a study carried out by Vertebrae (2019) —a 3D and Augmented Reality commerce and media platform— showed that participants with less frequent exposure to XR had a stronger emotional response to the medium, validating the novelty expectation. This effect must be considered as a relevant variable when testing immersive technology and/or media with final users.

To sum up, immersive technologies are at a crossroad in their development, and despite the hurdles, it seems likely that the XR industry will steadily continue to improve these technologies, weaving them into more aspects of our personal and professional lives. Understanding the differences between traditional and new immersive media is essential for AVT and MA professionals and researchers in order to take the wheel and integrate these innovative technologies in their work and experiments.

2.3. Implementing accessibility in new media environments

The most popular media accessibility service is translation, which is in fact language accessibility (Orero, 2022). This service shows the universal nature of media accessibility beyond people with disabilities. Immersive media need to be accessible beyond different languages, addressing not only consumers with hearing and visual impairments, but also those with cognitive difficulties, the elderly, tourists, or for example in Europe migrants and refugees who are literate in their language, but can't read the Latin alphabet (also known as "new Europeans"). The way to cater for the needs of the users who cannot access the multimedia content is by integrating different types of accessibilities within these new formats.

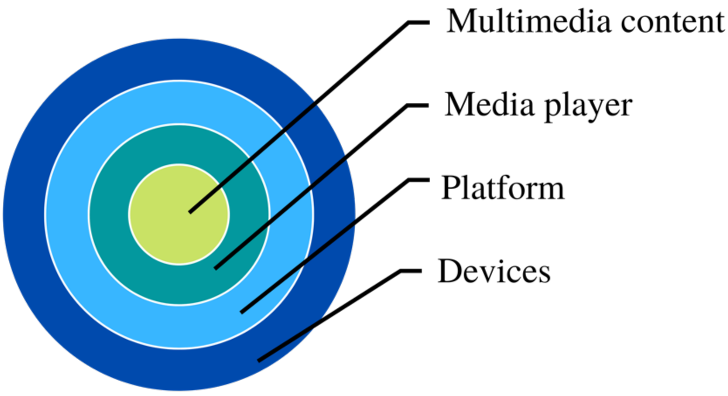
XR is often touted as a democratising way to access new worlds and experiences, but in terms of accessibility, there is still a long way to go. An online survey carried out in 2017 by Alice Wong (Brennan, 2017), a disability activist and founder of the Disability Visibility Project in partnership with Lucasfilm, showed that people with a variety of disabilities —from blindness to cerebral palsy or autism— enjoy using VR and believe it be beneficial, but also experience major accessibility issues. The two biggest issues raised by survey responders were the inability to customise the experience and the necessity to move certain parts of the body.

To achieve fully immersive media experiences, media creators have to consider four components: the content, the media player, the platform, and the device(s).

As depicted in Figure 7 below, the four components can be organised in a concentric circle: the outermost layer correspond to the device(s), the second layer to the “platform”, the third layer to the “media player” and the innermost layer to the “multimedia content”. This circle interrelates the 4 accessibilities, and it can be interpreted as that the only way to access the centre —the accessible content— is by ensuring that the components of the outer layers also take accessibility into account.

Figure 7

The four accessibilities



2.3.1. Content accessibility

During the last few decades, MA and AVT studies have been a prolific area of research. Best practices and standards for 2D traditional media have been established by both public and private entities (Matamala & Orero, 2018). However, studies on access services integrated within immersive content are still in their infancy; and currently, VR lacks agreed-upon guidelines and/or methods for making content accessible. Moreover, studies involving content accessibility and XR need to adopt a multidisciplinary approach as they borrow concepts such as usability and presence, which belong to other disciplines, mainly Psychology or Computer Science. This fact, linked to the lack of standardized solutions and guidelines, has led to the development of non-unified solutions, meeting only specific requirements (Hughes & Montagud, 2021).

2.3.1.1. Subtitling for the deaf and hard of hearing

Still, subtitling is the most popular (and also the cheapest) media access service, since there is no need to record voices as in audio description or dubbing. There are also standardised practices regarding many aspects: number of characters, number of lines, speed, position, etc. Traditionally, subtitles for the deaf and hard of hearing (SDH) are referred to as intralingual subtitles, which provide additional details about ambient noises and speaker identification in addition to a written rendition of the script (ISO20071:23:2019). Research in subtitles for XR is a work in progress, and some results have been published already (Agulló *et al*., 2018; Brown *et al.*, 2018; Montagud *et al.*, 2019; Rothe *et al.*, 2018). Also, different key aspects have been identified so far (Montagud *et al.*, 2021): comfortable viewing fields, presentation modes, guiding methods, and (re-)presentation of non-speech information.

The main challenge when creating SDH in immersive environments is the need to orient the viewer towards the sound's source (i.e., sound effects, or character speaking or not speaking). To facilitate this requirement, location within the 3D space information may be added to each subtitle (Agulló & Matamala, 2019). However, the drawback is that the location is only specified once per caption, and if a person is moving dynamically during this period, the guide could be wrong (Hughes *et al*., 2019). The ImAc project designed and developed several guiding mechanisms (Agulló *et al*., 2019), and test results showed two preferred methods: an arrow positioned to the left or right that directs the user to the target, and a radar circle shown in the user view that identifies both the position of the caption and the relative viewing angle of the user.

All the studies carried out so far regarding SDH within immersive media followed a user-centric methodology and chose people with hearing loss for testing. The British Broadcasting Corporation (BBC) was one of the first research organisations to design subtitles in XR (Brown & Patterson, 2017). They first identified the main challenges when developing subtitles for immersive content (Brown *et al*., 2017) and then designed four solutions for subtitle rendering (Brown, 2017). These rendering modes were tested with several clips (Brown, *et al.,* 2018), and users reported that they preferred subtitles which are always presented in front of the viewer, and that as they turn their head, they follow the movement, staying always in the same location in the headset display (head-locked). This trend is also observed in later studies, such as the one carried out by Rothe *et al.* (2018) and the most recent performed under the umbrella of the ImAc project (Hughes *et al*., 2019).

In the area of standardisation, a W3C community group[10] is focusing on determining and publishing best practices for access, activation, and display settings for captions within different types of immersive media, including AR, VR, and games. They have recently conducted a community survey to gather opinions, but no tests were performed. A small group of users with different hearing levels (Deaf, Hard of Hearing, and Hearing) were asked to evaluate different approaches for captions within immersive environments. Head-locked was identified as the preferred choice, however, it was noted that the most likely reason for this was that it replicated the experience that users were familiar with. It was also acknowledged that it was difficult for users to properly and theoretically evaluate new methods without the opportunity and content to enable them to be experienced.

2.3.1.2. Audio description

As defined in ADLAB guidelines (Remael, Reviers & Vercauteren, 2015), AD "is a service for the blind and visually impaired that renders Visual Arts and Media accessible to this target group offering a verbal description of the relevant (visual) components of a work of art or media product." In traditional audiovisual content such as films or series, AD is delivered between the dialogues, and it is expected not to interfere with music and other important sound effects (Jankowska, 2015). As well as SDH, many studies on this accessibility service have been carried out to date, however, empirical studies in more immersive environments are almost non-existent (Fidyka & Matamala, 2018).

Given this context, a series of focus groups and experiments on how AD could be integrated into XR was conducted under the umbrella of the ImAc project. The experiments focused on two main topics: how to integrate AD into a 360 environment, and key aspects of the technical realisation and production. Regarding the first topic, two aspects were explored in particular: using different types of script and different sound designs or AD presentation modes (Orero *et al.*, 2019).

To assess the quality of experience regarding understanding, enjoyment, and immersion, three approaches were taken to script the AD, including first-person narration, second-person narration, and the traditional standard AD. In a script in first person, the main character becomes the describer and narrates the story from their own perspective. The style and writing are also similar to how the character sounds and speaks. In a second person script, the describer

[10] https://www.w3.org/community/immersive-captions/ [retrieved February 5, 2022]

is sitting next to the viewer describing the scenes, and the style of writing and delivery is casual, informal, and friendly. Lastly, a standard AD is a narration in the third person, like any traditional AD, and the describer objectively sets out what is happening in the scene, what the characters look like, etc.

In terms of audio presentation modes, the ImAc project designed three options that where later introduced in the player developed within the project: a) classic, no positioning; b) static: from a fixed point in the scene, and c) dynamic: coming from the direction of the action. This player also offers extended AD tracks for specific scenes, actions, or objects as well as independent volume settings for AD and main audio tracks.

2.3.1.3. Sign language interpreting

The few efforts devoted to integrating sign language interpretating into immersive content have focused exclusively on representations of 2D avatars. The best example so far is the ImAc player, which allows minimal presentation personalisation in terms of size, position, and language; and also supports non-continuous view streams (showing/hiding the signer window based on the interpreter's activity).

The possibility of introducing a fully customisable 3D avatar with which the user could interact has not been investigated or tested yet. However, according to a statement from the World Federation of the Deaf (WFD) and the World Association of Sign Language Interpreters (WASLI) (2018), animated or digital signing avatars should be avoided as users find them less expressive than a recorded video of humans who can convey the natural quality and skill provided by trained and qualified interpreters. Both organisations insist on using signed avatars only on pre-recordings of real people who are trained and qualified by interpreters and translators.

Integrating sign language interpretating into immersive media is an emerging field, and exploration is encouraged to ensure the future development of quality signing avatars (Saunders *et al.,* 2021). For example, this could be via building a signing avatar that provides a face with fully functioning muscular variables and can successfully parse the nuances of vocal expression and meaning.

2.3.1.4. Emerging services: easy to read, audio subtitling (AST), and web accessibility

Under-resourced accessibility services are new emerging accessibility services that have never been considered within XR. For example, an innovative test on easy-to-read subtitles carried out by the ImAc project revealed that many users, especially the elderly, enjoyed a simplified version of the subtitles when compared to standard ones. This access service can also be implemented within web accessibility, which allows everyone to perceive, understand, navigate, and interact with the Internet and its contents. It also applies to mobile apps and must consider the developments in technology and trends in recent years.

Audio subtitling (AST), also known as spoken subtitles/captions, is defined by ISO/IEC 20071-25 as "captions/subtitles that are read aloud over the audio in a video"; and according to the Access Service Survey published by the European Broadcasting Union (EBU, 2016), it is the least developed access service in Europe. The first attempt to introduce this access service in an immersive environment was for the ImAc project. The ImAc player allows for two audio placement modes: a classic one (no positioning), and a dynamic one (coming from the direction of the speaker). It also has independent volume settings for AST and the main audio tracks, so that the user can customise the volume of each track.

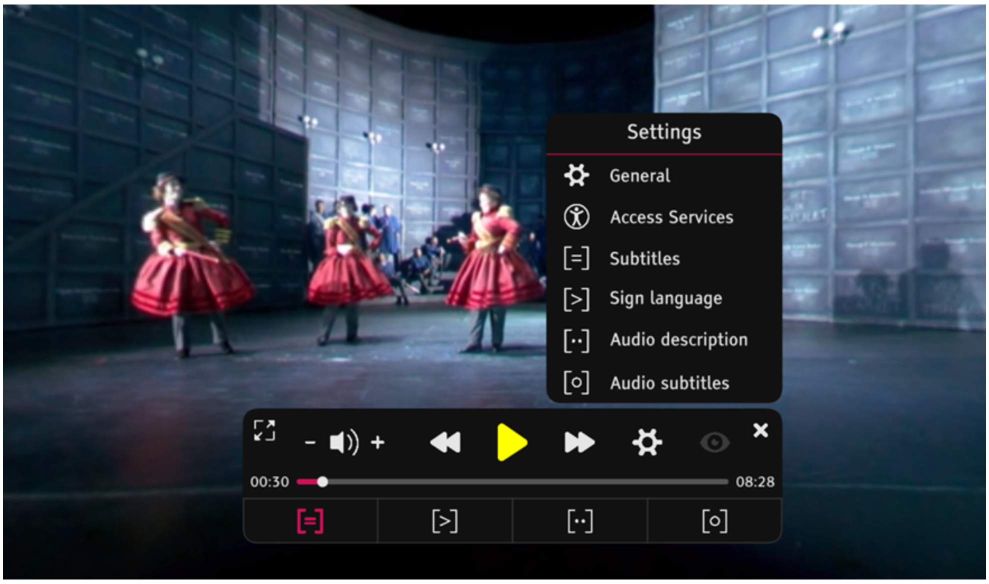
Regarding web accessibility, W3C's Web Content Accessibility Guidelines (WCAG) are the most internationally adopted voluntary web accessibility standard. Even though XR environments are identified as a challenge for accessibility implementation, only a first draft developed as part of the initial findings into potential accessibility-related user needs and requirements for XR is available (see Section 3.3).

2.3.2. Accessible player

As with traditional 2D media, to create and consume subtitles in XR, a subtitle editor and a subtitle player are needed. Currently, there are commercially available immersive video players offering the ability to play 360º videos, but not many of them support accessible services (Brescia-Zapata, 2022). Moreover, the player has to support different media formats, such as traditional 2D video and 360º video, traditional 2D audio and 3D spatial audio, IMSC subtitle files or support for DASH streaming. This problem has already been addressed by the ImAc project, as one of the main outcomes was the development of an accessible player along with a 360º subtitle editor. The player supports audio description, audio subtitles, and sign language, along with other features (Montagud *et al.,* 2019), as can be seen in Figure 8 below.

Figure 8

ImAc player settings



Note. Imac Player screenshot

One of the challenges for testing immersive subtitles is the difficulty users have in properly evaluating new modalities. The reasons are the cost and time needed to create new prototype subtitle presentations to enable users to experience them. To allow for visualising creative subtitles, an XR subtitle web simulator was developed by Hughes *et al.* (2020). This open-source, web-based simulator was designed for the rapid prototyping of subtitles within a 360º space and can be accessed —along with an editor— from the main project area, where all the imported 360º videos are located. These tools take inspiration from the player and the editor developed by the ImAc project.

On the one hand, the ImAc player integrates various accessibility services, although it does not allow much customisation; on the other hand, the player developed by Hughes only allows subtitles, but with a very high degree of customisation. Therefore, the main difference between them is that the ImAc tools are intended to be used by generic audiences (final users), while the tools developed by Hughes are more focused on research and testing.

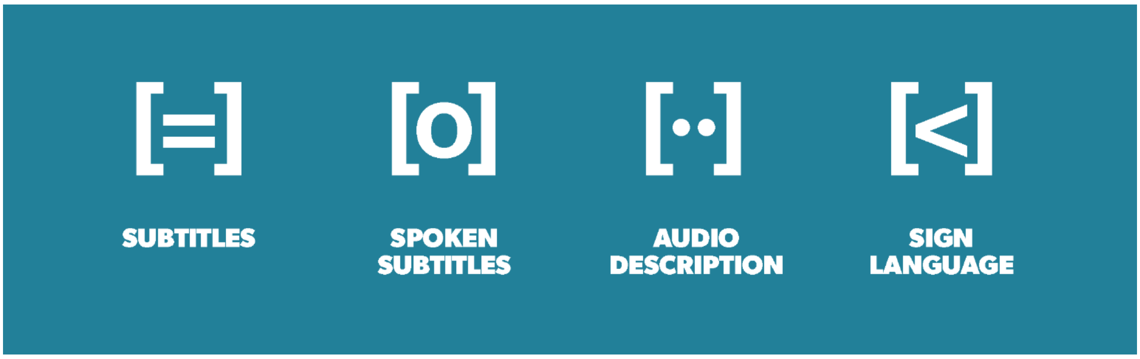
2.3.3. User interface accessibility

Not only the player itself needs to be accessible, but also the platform or portal where the 360º videos are stored, as well as the user interface. It comes as no surprise that XR interfaces are often less accessible than traditional 2D user interfaces on desktop and mobiles, as the additional spatial component can hamper easy use for users with no accessibility needs. Recently, a W3C working group published a first draft on various accessibility-related user needs for XR. It is mainly focused on gaming, but most insights can be extrapolated to any immersive experience which aims at being accessible. This document outlines personalisation, which involves tailoring aspects of the user experience to meet the needs and preferences of the individual user. In this regard, W3C is exploring areas such as expanding the accessibility information that may be supplied by the author, facilitating preference driven individual personalisation and enabling the author to specify key semantics needed to support users with cognitive impairments. This customisation affects very different aspects, from language selection to colour changes or context magnification. Voice command was also identified as a key feature, as it allows users to navigate, interact and communicate with the environment and also with other users.

Furthermore, platforms and portals must always display their catalogue of videos including the available access services (and languages) for each video. This aspect is especially useful, as it allows users to rapidly identify which content they can or cannot consume. For a better comprehension of the controls, icons or symbols used to identify each access service should be consistent and ideally follow established conventions or standards (showing their more commonly known acronyms, like ST, SL, AST, and AD, in English). Otherwise, the symbol or icons should clearly illustrate the functionality of the represented service. A good example is the icons proposed by the Danish Broadcaster (Eiersø, 2018), also supported by the EBU (see Figure 9 below).

Figure 9

Accessibility services icons for a common European standardisation



Note. From © DR Design

2.3.4. Device accessibility

VR hardware typically makes many assumptions about users' abilities that can lead to accessibility problems. For example, most head-mounted displays (HMDs) are quite heavy and require significant strength to wear them, as well as the ability to execute a large range of movements. They may also be problematic for people who wear glasses or hearing aids. Moreover, the surrounding tracking systems (such as eye-tracking) create an infrastructure that needs to be set up and calibrated before using the HMD. Although the most recent generation of VR headsets aims to reduce the number of hardware components, still cables and chargers are remaining included (as is the case with the HTC VIVE PRO headset) also making the equipment hard to transport.

Flexible and customizable design is becoming more common in VR hardware, which could be a trend for accessibility; for instance, many HMDs allow the user to specify the distance between the eyes or between the eyes and the display in order to optimise the experience (Mott *et al.,* 2019). This flexibility is a great framework that may particularly benefit accessibility, i.e., by supporting novel or customized controllers used by all people regardless of their capabilities. This is the case of AR equipment, which is much lighter than VR headsets, and in the future, they may consider incorporating prescription lenses to benefit end-users with low vision. Other accessibility systems could also be integrated into future devices, such as voice control or easy-to-read instructions to facilitate the set-up of the HMDs.

2.4. Conclusions

More than ever, technology and new media formats are enabling the empowerment of end-users, both consumers and prosumers. The interaction of users in the current Information Society plays a fundamental role in the social integration and democratic participation of all citizens. However, technology is still designed with accessibility as an afterthought, far from user-centric design. Allowing easy access to content and guiding the user in control of media services are two of the most recent UN and European media regulations. The only way to comply with European regulations, as well as with user expectations, is to find the best way to integrate accessibility within the new immersive media.

The main idea of this chapter is that for XR to be accessible, both the content and the container must also be accessible. Creating immersive, accessible content is an investment that requires a series of technical knowledge, but that is useless if, in the end, users cannot access this content. Both the XR and MA communities should seize this moment as devices and standards continue to evolve to include accessibility in XR from the ground up. That is the only way to create a more usable and inclusive technological future for users of all abilities. In addition, promoting the concepts of accessibility is also a way to guarantee all people equal access to culture.

Acknowledgments

The author is member of TransMedia Catalonia, a research group funded by Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya, under the SGR funding scheme (ref. code 2021SGR00077). This article reflects only the author's view, and the funding institutions hold no responsibility for any use that may be made of the information it contains. This article is part of Marta Brescia's PhD in Translation and Intercultural Studies at the Department of Translation, Interpreting and East Asian Studies (Departament de Traducció i d'Interpretació i d'Estudis de l'Àsia Oriental) of Universitat Autònoma de Barcelona.

This research has been partially funded by the H2020 project TRACTION (under Grant Agreement 870610).

References

Agulló, Belén, & Matamala, Anna, «Subtitling for the deaf and hard-of-hearing in immersive environments: results from a focus group», *Jostrans. The Journal of Specialised Translation*, vol. 33, 2019, pp. 217-235.

Agulló, Belén, & Orero, Pilar, «3D Movie Subtitling: Searching for the best viewing experience», *Studi di Comunicazione e Mediazione linguistica e culturale COME,* vol. *2*, 2017, nº 1, pp. 91-102.

Agulló, Belén, Matamalana, Anna, & Orero, Pilar, «From disabilities to capabilities: testing subtitles in immersive environments with end-users», *HIKMA*, vol. 17, 2018, pp. 195-220.

Aissaoui, Najeh, «The digital divide: a literature review and some directions for future research in light of COVID-19», *Global Knowledge, Memory and Communication,* 2021. DOI: https://doi.org/10.1108/GKMC-06-2020-0075

Aylett, Ruth, & Louchart, Sandy, «Towards narrative theory of virtual reality», *Virtual Reality*, vol. 7, 2–9, 2003. DOI: https://doi.org/10.1007/s10055-003-0114-9

Beaunoyer, Elisabeth, Dupéré, Sophie, & Guitton, Matthieu J., «COVID-19 and digital inequalities: reciprocal impacts and migration strategies», *Computer in Human Behavior,* vol. 111, 2020, nº 6.

Bernabé Caro, Rocío, & Orero, Pilar, «Easy to read as multimode accessibility services». *Hermeneus. Revista de Traducción e Interpretación*, vol. 21, 2019, pp. 53-74.

Brennan, Dominic, (2017, December 6). *New report from ILMxLab & Disability Visibility Project shares insights on VR accessibility design.* Retrieved from: https://www.roadtovr.com/disability-visibility-project-and-ilmxlab-share-lessons-learned-about-vr-accessibility-design/

Brescia-Zapata, Marta, «The present and future of accessibility services in VR360 players», *inTRAlinea*, vol. 24, 2022.

Brown, Andy (2017, October 26). User testing subtitles for 360º content. Retrieved from: https://www.bbc.co.uk/rd/blog/2017-10-subtitles-360-video-virtual-reality-vr

Brown, Andy, & Patterson, Jake, (2017, October 26). *Designing subtitles for 360º content.* Retrieved from: https://www.bbc.co.uk/rd/blog/2017-03-subtitles-360-video-virtual-reality

Brown, Andy, Turner, Jayson, Patterson, Jake, Schmitz, Anastasia, Armstrong, Michael & Glancy, Maxine (2018, April 30). *Exploring subtitle behavior for 360º video.* Retrieved from: https://www.bbc.co.uk/rd/publications/whitepaper330

Brown, Andy, Turner, Jayson, Patterson, Jake, Schmitz, Anastasia, Armstrong, Mike, & Glancy, Maxine, «Subtitles in 360-degree video». *Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video*, 2017, pp. 3–8. DOI: https://doi.org/10.1145/3084289.3089915

Crafton, Donald, *The talkies: American cinema's transition to sound, 1926-1931*, vol. 4, University of California Press, 1999.

Creed, Chris, Al-Kalbani, Maadh, Theil, Arthur, Sarcan, Sayan, & Williams, Ian, «Inclusive AR/VR: accessibility barriers for immersive technologies». *Universal Access in the Information Society*, 2023. DOI: https://doi.org/10.1007/s10209-023-00969-0

Eiersø, Caroline, (2018, Mai 15). *Smart icons design — common European standardization.* Retrieved from https://www.dr.dk/om-dr/about-dr/smart-icons-design-common-european-standardization

Fidyka, Anita, & Matamala, Anna, «Audio description in 360º videos: results from focus groups in Barcelona and Kraków». *Translation Spaces*, vol. 7, 2018, nº 2, pp. 285-303. DOI: https://doi.org/10.1075/ts.18018.fid

Gunkel, Simon, Stokking, Hans M., Prins, Martin J., van der Stap, Nanda, ter Haar, Frank B. & Niamut, Omar A., (2018, June 15). *Virtual reality conferencing: Multi-user immersive VR experiences on the web.* DOI: http://doi.org/10.5281/zenodo.3530681

Hughes, Chris J., & Montagud-Climent, Mario, «Accessibility in 360° video players». *Multimedia Tools and Applications*, vol. 80, 2021, pp. 30993-31020. DOI: https://doi.org/10.1007/s11042-020-10088-0

Hughes, Chris J., Brescia-Zapata, Marta, Johnston, Matthew, & Orero, Pilar, «Immersive captioning: Developing a framework for evaluating user needs». *IEEE AIVR 2020: 3rd International Conference on Artificial Intelligence & Virtual Reality 2020*, 2020, Retrieved from: http://usir.salford.ac.uk/id/eprint/58518/

Hughes, Chris J., Montagud-Climent, Mario, & tho Pesch, Peter, «Disruptive approaches for subtitling in immersive environments», *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video*, 2019, pp. 216–229. DOI: https://doi.org/10.1145/3317697.3325123

International Organisation for Standardisation. (2018). Information technology — User interface component accessibility — Part 23: Visual presentation of audio information (including captions and subtitles). https://www.iso.org/standard/70722.html

Jankowska, Anna, *Translating Audio Description Scripts*, Berlin, Peter Lang Verlag, 2015. DOI: https://doi.org/10.3726/978-3-653-04534-5

Ko Dong-uk, Ryu Hokyoung, & Kim Jieun, «Making new narrative structures with actor's eye-contact in cinematic virtual reality (CVR)», *Interactive Storytelling: 11th International Conference on Interactive Digital Storytelling, ICIDS,* vol. 11318, 2018, pp. 343–347.

Matamala, Anna, & Orero, Pilar, «Standardising accessibility: Transferring knowledge to society», *Journal of Audiovisual Translation,* vol. 1, 2018, pp. 139–154. DOI: https://doi.org/10.47476/jat.v1i1.49

Matamala, Anna, & Orero, Pilar, *Listening to Subtitles. Subtitling for the Deaf and Hard-of-Hearing,* Berna, Peter Lang, 2010.

Milgram, Paul, & Kishino, Fumio, «A taxonomy of mixed reality visual displays», *IEICE TRANSACTIONS on Information and Systems*, vol. *77*, nº12, 1994, pp. 1321-1329.

Molina-Siles, Pedro Javier, Piquer Cases, Juan Carlos, Cortina Maruenda, Francisco Javier, «El color en los comienzos del cine. De la aplicación manual al Technicolor», *Libro de actas del X Congreso Nacional del Color*, 2013.

Montagud-Climent, Mario, Fraile Vila, Isaac, Meyerson, Einar, Genís, María, & Fernández Langa, Sergi (2019, September 19). Imac player: Enabling a personalized consumption of accessible immersive content. DOI: https://doi.org/10.6084/m9.figshare.9879254.v1

Montagud-Climent, Mario, Soler-Vilageliu Olga, Fraile Vila, Isaac, & Fernández Langa, Sergi, «VR360 subtitling: Requirements, technology, and user experience», *IEEE Access*, vol. 9, 2021, pp. 819-2838. DOI: https://doi.org/10.1109/ACCESS.2020.3047377.

Mott, Martez, Cutrell, Edward, Gonzalez Franco, Mar, Holz, Christian, Ofek, Eyal, Stoakley, Richard, & Ringel Morris, Meredith, «Accessible by Design: An Opportunity for Virtual Reality», *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, 2019, pp. 451-454. DOI: https://doi.org/10.1109/ISMAR-Adjunct.2019.00122

Orero, Pilar, «Translation, accessibility, and minorities», *The Routledge Handbook of Translation and Media* (pp. 384-399). London, Taylor Francis, 2022.

Orero, Pilar, Rai, Sonali, & Hughes, Chis, J., «The challenges of audio description in new formats», *Innovation in Audio Description Research*. London, Routledge, 2019.

Puerta Domínguez, Simón, «Del color al sensorama. Cine e industria cultural», *Razón crítica*, 1, 170–205, 2016. DOI: http://dx.doi.org/10.21789/25007807.1141

Remael, Aline, Reviers, Nina, & Vercauteren, Gert (eds.) *Pictures Painted in Words: ADLAB Audio Description Guidelines*, Trieste, EUT. 2015.

Rothe, Sylvia, Tran, Kim, & Hussmann, Heinrich, «Positioning of subtitles in cinematic virtual reality», *International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*, 2018. DOI: https://doi.org/10.2312/egve.20181307

Saunders, Benjamin, Camgoz, Necati Cihan, & Bowden, Richard, (2021, July 23). Mixed SIGNals: Sign language production via mixture of motion primitives. *Proceedings of the International Conference on Computer Vision (ICCV).* Retrieved from https://openresearch.surrey.ac.uk/esploro/outputs/conferencePaper/Mixed-SIGNals-Sign-Language-Production-via/99609266002346

Skult Natasha, & Smed Jouni, «Interactive storytelling in eXtended Reality: Concepts for the design», *Game user experience and player-centered design. International series on computer entertainment and media technology*, Springer, Cham, 2020. DOI: https://doi.org/10.1007/978-3-030-37643-7_21

Slater, Mel, & Wilbur, Sylvia, «A framework for immersive virtual environments five: Speculations on the role of presence in virtual environments», *Presence: Teleoper. Virtual environ.* Vol. *6,* 1997, pp. 603–616. DOI: https://doi.org/10.1162/pres.1997.6.6.603

Vertebrae. (2109). Immersive vs traditional media. How interactivity and realism influence persuasions. Retrieved from: https://cdn.nrf.com/sites/default/files/2020-02/Vertebrae_3d-AR-eCommerce_Whitepaper-2019%20%281%29_0.pdf

World Federation of the Deaf and World association of sign language interpreters. (2018). WFD and WASLI statement on use of signing avatars. Retrieved from: https://docs.google.com/viewerng/viewer?url=https://wfdeaf.org/wp-content/uploads/2017/02/WFD-and-WASLI-Statement-on-Avatar-FINAL-14032018-Updated-14042018-1.pdf&hl=en_US

XR Accessibility User Requirements. Retrieved from: https://www.w3.org/TR/xaur/

CHAPTER 3

Article 2: The present and the future of accessibility services in VR360 players

The Present and Future of Accessibility Services in VR360 Players

Marta Brescia Zapata

Department of Translation and Interpretation, Universitat Autònoma de Barcelona, Barcelona, Spain

Abstract

Technology moves fast and making it accessible turns it into an endless game of cat and mouse. Immersive content has become more popular over the years and VR360 videos open new research avenues for immersive audiovisual experiences. Much work has been carried out in the last three years to make 360º videos accessible (Agulló and Orero 2017; Fidyka and Matamala 2018; Agulló and Matamala 2019) and the chase continues. Unfortunately, VR360 videos cannot yet be classified as accessible because subtitles and/or audio description are not always available. This paper focuses on analysing to what extent the main accessibility services are integrated into the most popular and available commercial VR360 players. It also analyses the development of a fully customisable and accessible player which has been created following the EU standard EN17161 and a user centric approach. The first part deals with the relationship between technology and audiovisual translation (AVT) studies, followed by a brief overview of existing eXtended Reality (XR) content. Section 3 provides a list of the VR360 players available commercially and compares them in relation to the main accessibility services: subtitling for the deaf and the hard-of-hearing (SDH), audio description (AD) and sign language (SL). The last section presents a new player developed by the ImAc project, which has accessibility at its heart.

Keywords

Accessibility, virtual environment, 360º video, subtitling, audio description, sign language interpreting, personalisation

3.1. Introduction

Immersive media such as virtual reality (VR) and augmented reality (AG) are technologies which have the potential to transform the way we work, communicate, and experience the world. VR is capable of transforming and innovating traditional sectors such as manufacturing industries, construction, and healthcare. It can also revolutionise education, culture, travelling and entertainment. Immersive environments have enjoyed much popularity in video games and other commercial applications such as simulators for kitchen design or surgical training (Fida et al. 2018; Parham et al. 2019). While tools to generate these environments have been available for some time—the most popular being Unity— photographs and videos are growing in number. This is associated with the availability of cameras to record in 360º. In all cases, the aim is to provide an immersive and engaging audiovisual experience for viewers. The two major types of VR content are 360º videos (web-based) and 3D animations (synthetic or Unity-based). This article deals only with the former.

In this context, 360º video (aka VR360 video) has become a simple, cheap, and effective way to provide VR experiences (Montagud et al. 2020b). The potential of VR360 videos has led to the development of a wide variety of players for all platforms and devices (Papachristos et al. 2017); like computers, smartphones and Head Mounted Displays (HMDs). Traditionally, accessibility has been considered by the media as an afterthought, despite many voices asking for it to be included at the design stage of any process. Moreover, the lack of standardisation and guidelines in this novel medium has resulted in non-unified solutions, focusing only on specific requirements.

This situation has served as the motivation for exploring to what extent have accessibility services been integrated in the available most popular commercial VR360 players. This study is a necessary first step towards identifying the advantages and challenges in this novel field. The second contribution of this paper is the presentation of a fully accessible player, developed under the umbrella of the EU H2020 funded project ImAc[11].

The structure of this paper will now be presented. Section 1 deals with the relationship between new technologies and audiovisual translation (AVT) studies by giving a general overview, with a focus on the media accessibility (MA) field. To do so, this section is divided into two parts: the first outlines the state-of-the-art by summarising the main accessibility guidelines; the second focuses on new technologies as a catalyst, making the interaction between the final users and the main accessibility services (namely AD, SDH and SL) possible. Section 2 offers a general overview of XR content, focusing on VR, AR and 360º video. Section 3 analyses the degree of accessibility in the main commercial VR360 players that are available. After presenting the solutions offered by existing VR360 players, a fully accessible player (the ImAc

[11] See https://www.imac-project.org.uk/ [retrieved 11/02/2021]

player) is presented in Section 4. Finally, Section 5 offers a discussion on existing limitations, improvements and solutions provided by the ImAc player, as well as some ideas and avenues for future work.

3.2. Immersive media and AVT: an overview

The unceasing advances in Information and Communication Technologies (ICT) open the door to new fascinating opportunities within MA studies, despite also posing a series of challenges that this discipline has never before faced. This section deals with how the development of new technologies has affected the field of AVT. Firstly, by revisiting the existing accessibility guidelines and recommendations; and secondly, by analysing the importance of researching the tools and technologies needed to generate accessible content.

3.2.1. Accessible guidelines and recommendations

It is not yet possible to talk about immersive content "for all" because audio description (AD), subtitling for the deaf and hard-of-hearing (SDH) and sign language (SL) interpreting —among others— are not always available for 2D content, let alone 3D or XR.

Regarding legislation, in 2008 the United Nations (UN) issued the Convention on the Rights of Persons with Disabilities (CRPD), currently ratified by 181 countries. In Article 30 (section 1 b), the convention states that "State Parties […] shall take all appropriate measures to ensure that persons with disabilities: […] (b) Enjoy access to television programmes, films, theatre and other cultural activities, in accessible formats." This convention has helped to promote the proliferation of accessibility services in media content. AVT, and more specifically MA (Remael et al. (eds) 2014; Greco 2016), is the field in which research on access to audiovisual content has been carried out in the last few years, generally focusing on access services such as audio description (AD), subtitling for the deaf and hard-of-hearing (SDH) or sign language (SL) interpreting, among others (Matamala and Orero 2010; Arnáiz-Uzquiza 2012; Romero-Fresco 2015; Fidyka and Matamala 2018).

The CRPD (UN 2006), together with the Convention on the Protection and Promotion of the Diversity of Cultural Expressions (UNESCO 2005) created legal repercussions within the scope of the accessibility of cultural goods in the European Union (EU). These two conventions resulted in two Directives and one Act that demand accessibility services not only on websites, services and spaces; but also for content and information offered at all cultural venues and events (Montagud et al. 2020a). The three pieces of legislation are:

1. The EU Directive on the Accessibility of Websites and Mobile Applications transposed into the law of each EU member state by September 2018. It is based on Web Content Accessibility Guidelines (WCAG) 2.0 guidelines, and references EN301549 as the standard which will enable websites and apps to comply with the law.

2. The Audiovisual Media Services Directive (AVMSD) approved in 2018 gave member states 21 months to transpose it into national legislation. It addresses key issues, for example: rules to shape technological developments that preserve cultural diversity and protect children and consumers whilst safeguarding media pluralism.

3. The European Accessibility Act which takes the form of a legally binding Directive for all member states. It is a law that aims at making many EU products and services (smartphones, computers, TV programs, e-books, websites, mobile apps etc.) more accessible for persons with disabilities.

These three pieces of EU legislation demand accessibility services for information and content and at all cultural venues and events; as opposed to only websites, services and spaces. As is reviewed in the following sections, all audiovisual products need to be accessible, including those providing XR content for their audience.

3.2.2. The impact of new technologies in AVT and MA studies

Technical advances have not only modified media applications, but accessibility services and profiles also. Even end users have been affected by the advent of new ways to access media content. Traditionally, AVT and MA research has focused on the analysis of translated audiovisual texts and their many transformations or *transadaptations* (Gambier 2003). Some years ago, the translated audiovisual text was static in the sense that one format was distributed and consumed by all. Audiences are now allowed to make decisions about the media they consume, beyond the two traditional values: level of sound and image contrast. These days, technology allows for media personalisation (Orero forthcoming; Orero et al. forthcoming; Oncins and Orero 2020). When watching media content on TV or YouTube, subtitles can be read in different sizes, positions, and colours (Mas Manchón and Orero 2018). The speed of the content reproduction can also be altered, and the audiovisual text is increasingly changing according to individual needs. Media accessibility has shifted from a one-size-fits-all approach to customisation. Technology has allowed for this change. The Internet of Things lets customers decide the way in which each media object is set-up. Artificial Intelligence is now integrated and understands individual choices, needs and preferences.

On the one hand, technology is the basis of the tools used for creating or adapting content; on the other, it is the basis for consuming content (Matamala 2017). Regarding content creation and adaptation, there is a wide range of professional and amateur translation software for both subtitling (e.g., Aegisub, Subtitle Workshop, VisualSubSync, WinCaps, SOftwel, Swift Create, Spot, EZTitles, etc.) and AD (e.g., Softel Swift ADePT, Fingertext, MAGpie2, Livedescribe or YouDescribe). It is not that easy to find studies regarding preferences or comparative analysis of the previously mentioned tools. Concerning subtitling, Aulavuori (2008) analyses the effects of subtitling software on the process. Regarding AD, Vela Valido (2007) compares the existing software in Spain and the USA. Oncins et al. (2012) present an overview of existing subtitling software used in theatres and opera houses and propose a universal solution for live media access which would include subtitling, AD and audio subtitling, among other features.

The tool used by the users to consume audiovisual media is the media player. The choice of device (mobile phone, TV, PC, smart watch, tablet), type of content and the available accessibility services determine the level of personalisation (Gerber-Morón et al. 2020). Subtitles, for example, are displayed differently according to screen size, type of screen, and media format. The choice of the subtitle is made through the media player settings, making understanding its capabilities a basic departure point when analysing translated immersive media content.

The existence of such a wide range of technical solutions opens the door to many research questions. It remains to be seen how new developments will affect the integration of accessible services into existing tools to adapt to the needs of end users: everyone has the right to access the information provided by media services (including immersive experiences). This paper offers a valuable resource to content providers seeking to improve their products, users with accessibility needs in choosing the player that best suits their requirements, and to researchers setting out to identify the challenges and possibilities that this new field poses to AVT and MA.

3.3. The rise of VR and AR: VR360 video becomes mature

In this section, a general overview of VR, AR and VR360 videos will be given, highlighting the impact that these new technologies may have on our society at different levels.

Although immersive content production is still at an early stage and is generally used in professional environments such as hospitals and universities; in the future it might be used in the daily lives of ordinary users. As with all new technologies, VR will make our lives easier: from going to the supermarket to online shopping (Lee and Chung 2008). Immersive environments are the new entertainment experiences of the 21st century, from museums (Carrozzino and Bergamasco 2010) and theatres to music events such as opera (Gómez Suárez and Charron 2017). They allow users to feel as if they are being physically transported to a different location. Though the most popular applications are cultural representations, it is a great tool for larger audiences and functionalities such as leisure, sport (Mikami et al. 2018), tourism (Guttentag 2010) and health (Rizzo et al. 2008).

There are various solutions that can provide such an experience, such as stereoscopic 3D technology which has re-emerged in films during the last ten years (Mendiburu 2009). Nevertheless, this format is nothing new. It has been available since the 1950s, but the technology has not been ready to deliver quality 3D (González-Zúñiga et al. 2013). Both the quality of the immersive experience and the sense of depth depend on the display designs, which for 3D content are diverse and lacking in standards (Holliman et al. 2011). However, stereoscopy did not become the main display for AV products; perhaps due to the lack of standardisation, the intrusive nature of 3D, and uncomfortable side effects such as headaches or eyestrain (Belton 2012). According to Belén Agulló and Anna Matamala (2019), “the failure to adopt 3D imaging as mainstream display for AV products may have opened a door for VR and 360º content, as a new attempt to create engaging immersive experiences". VR stands for ‘virtual reality’ and it takes on several different forms, 360º video being one of them. However, VR and 360º videos are two different mediums (see Table 1). In 360º video, multi-camera rigs (often static) are used to record live action in 360º, giving the consumer a contained perspective of a location and its subjects. VR renders a world in which, essentially, the consumer operates as a natural extension of the creator’s environment, moving beyond 360º video by enabling the viewer to explore and/or manipulate a malleable space. In 360º video, the consumer is a passenger in the storyteller’s world; in VR, the consumer takes the wheel. The storyteller directs the viewer’s gaze through this situational content by using elemental cues such as light, sound and stage movement. The traditional notion of the fourth wall has been eliminated.

Table 1

Differences between VR video and 360º video

	Virtual reality	360º video
Photography	Digital environment	Live action
Mobility	Immersive world that you can walk around in.	360º view from camera's perspective. Limited to filmmaker's camera movements.
Video timeline	Video can progress through a series of events. Experiences can be held in an existing world to be explored by the user (6 degrees of freedom).	Video progresses on a timeline created by the filmmaker's camera movements (3 degrees of freedom).
Platforms	A full experience requires an HMD.	Available on 360º compatible players (desktop and mobile).
Story	The filmmaker does not control the physical location of the viewer in the built environment and must capture attention and motivate the user to travel in the direction of the events of the story.	The filmmaker controls the physical location of the camera but must capture the attention of viewers to direct the story.

Note. Author's own elaboration based on Ullman, S.

Other less commercial immersive technologies are mixed and augmented reality. Paul Milgram and Fumio Kishino (1994: 1321) define those terms as:

> Mixed Reality (MR) visual displays […] involve the merging of real and virtual worlds somewhere along the 'virtuality continuum' which connects completely real environments to completely virtual ones. Probably the best known of these is Augmented Reality (AR), which refers to all cases in which the display of an otherwise real environment is augmented by means of virtual (computer graphic) objects.

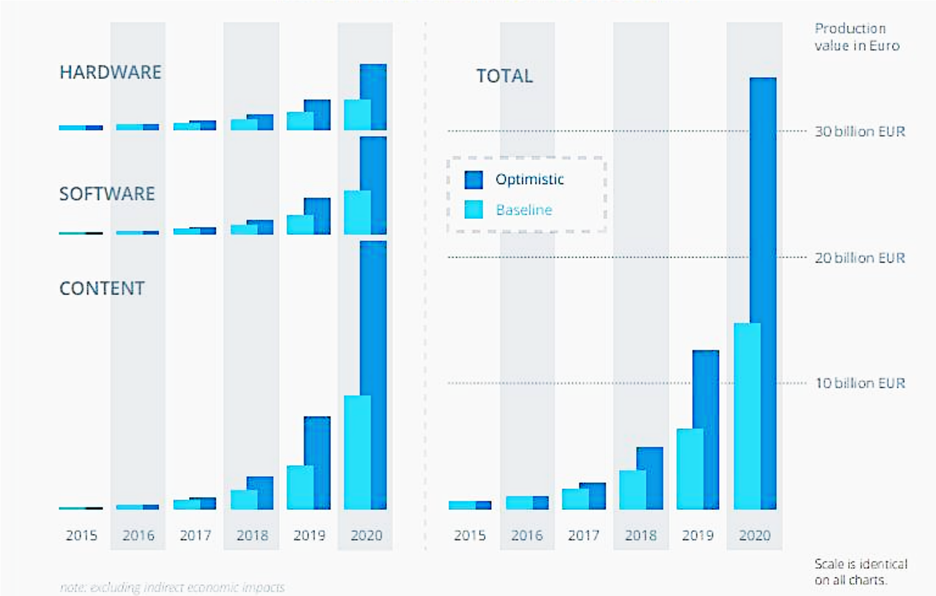
Julie Carmigniani and Borko Furht (2001: 3) define AR as "a real-time direct or indirect view of a physical real-world environment that has been enhanced by adding virtual computer-generated information to it." The properties of AR are that it "combines real and virtual objects in a real environment; runs interactively, and in real time; and registers (aligns) real and virtual objects with each other" (Azuma et al. 2001: 34).

It is not easy to outline a state-of-the-art when talking about immersive environments as the development in VR technology is happening at an unprecedented speed. Moreover, Covid-19 has helped to accelerate virtual experiences. While it is early to evaluate, with the writing of this paper taking place during lockdown, what is now considered "the new normal" will have strong non-presential and virtualization elements. Systems and applications in this domain are presented on a daily or weekly basis. AR/VR technology makes use of sensory devices to either virtually modify a user's environment or completely immerse them in a simulated environment. Specially designed headsets and glasses can be used for visual immersion, while handhelds and wearables offer tactile immersion. Optical devices such as Facebook's Oculus Rift and Sony's PlayStation Virtual have shipped millions of units as consumers look to explore the possibilities offered by virtual environments. Nevertheless, the adoption rate for AR/VR devices is relatively low when compared to other consumer electronics, though many of the world's biggest technological companies see the promise of AR/VR technology and have begun to allocate significant budgets to develop it.

Stable growth of the VR and AR markets is expected both in Europe and around the world, as can be seen in Figure 10. According to a report by Ecorys, the total production value of the European VR & AR industry was expected to increase to between €15 billion and €34 billion by 2020 and to directly or indirectly account for 225,000 to 480,000 jobs. Also, wider supply chain impacts are expected to indirectly increase the production value to between €5.5 billion and €12.5 billion and generate an additional 85,000-180,000 jobs. Due to the strong growth of content-related VR activities, the share of Europe in the global market is expected to increase.

Figure 10

Europe: Future growth driven by content



Note. From Gartner, 2018

The 360º video has been around for several years with differing levels of sophistication and polish. However, with the advancement of camera technology combined with the development of software for handling the images, 360º video is being used in an increasing number of ways.

2018 was considered the Year of 360º video. Indeed, the trend of watching 360º videos on web browsers, tablets or mobiles has increased. Even if PCs, tablets or mobiles are currently still the main devices for watching 360º video, the use of VR Headsets is starting to grow as well. There are three main factors that could explain why the 360º video market is considered to be already developed: 360º video capture devices are more sophisticated and affordable, the increasing number of web and mobile players (YouTube, Facebook, Twitter and the use of mobile phones as HMDs), and the slightly decreasing price of VR headsets (HDM).

VR360 video is now being used by all kinds of people and organisations for sharing immersive stories and extreme experiences in stunning locations. Estate agents, airlines and the hospitality industry are embracing VR360 video to show off their goods; while broadcasters, educators and social media platforms are also experimenting with the format. Many videographers are now learning about VR360 video creation for various purposes and platforms, for example:

- Immersive journalism. The New York Times has started publishing The Daily 360º, a short 360 news report – often in 4K resolution – with multiple cuts between static shots, and a reporter acting as the narrator. Done well, 360º adds a unique eyewitness feel to storytelling and reporting.

- 360º time-lapses. Many 360º cameras allow the capture of time-lapses, which are sequences of 360º images recorded at set intervals to record changes that take place slowly over time. Speed up the frames and the effect can be exceptional for something as delicate as the Milky Way emerging at night.
- Live 360º video cameras can now capture and live-stream spherical imagery and most popular online and social media platforms have recently been updated to support 360º content. Viewers can now tune-in to live 360º broadcasts and download the video afterwards.

3.4. VR360 players: web and mobile apps

In this section, a list of the main web VR360 players will be presented. Each player will be analysed based on how, and to what extent, they integrate the main accessibility services.

There is an increasing number of VR360 videos available over the internet and the majority of browsers support them, meaning we can enjoy watching VR content online with no need for a VR device. However, it is still necessary to download a 360º video player which can support them. Regular video players such as Windows Media Player do not support them, although they probably will in the future. Nevertheless, there are some players aside from regular video formats that also support VR videos. Most of them are available on the web, but the majority fail in terms of accessibility. The table below shows a comparative analysis of accessibility services offered by the most popular executable players.

The selection has been made for two different approaches and from a descriptive perspective. Firstly, following a top-down approach, specialized media such as magazines and papers were taken as a reference to elaborate on the first draft of the list. Secondly, users' opinions and comments in different online forums were taken into account to complete the final list (see Appendix 1 – Table 2).

GOM is a South Korean product from the Gretech Corporation, one of the best-known video players, mainly used for playing 'regular' videos but also supports 360º video. It is a multilingual player, as you can select the language of the player. It can play 360º videos downloaded to your computer and also directly from YouTube. It is possible to search and upload subtitles and to adjust several features, such as style and position. This player offers the ability to load two separate subtitle files: one displayed at the top of the screen and the other at the bottom.

Codeplex Vr Player is an experimental open-source VR Media Player for Head-Mounted Display devices like Oculus Rift. Not only does it provide the function of playing VR videos, but it also lets users watch 2D and even 3D videos. Its user interface is designed to be intuitive which makes it very easy to use. It provides a free version and a professional paid version.

Total Cinema 360 Oculus Player is a VR App for Oculus Rift developed by Total Cinema 360. It is specifically designed to capture fully interactive, live action spaces in high quality 360º video. It comes equipped with four demonstration videos that allow you to pause, zoom and adjust eye distance. You can also upload your own 360º video content for use with Total Cinema 360.

RiftMax was developed in Ireland by the VR software developer Mike Armstrong. It is not only a VR video player, as it allows you to interact with other people in scenarios such as parties or film screenings. It also enhances video with good effects that come out of the screen. It does not support subtitles and is only available in English.

SKYBOX was developed in the UK by Source Technology Inc and supports all stereo modes (2D and 3D, or 180º and 360º). You can choose multiple VR theatres when you are watching 2D or regular 3D videos; including Movie Theater, Space Station and Void. It supports all VR platforms: Oculus, Vive, Gear VR and Daydream. SKYBOX supports external text-based subtitles (.srt/.ssa/.ass/.smi/.txt) when you are watching a non-VR video. If a video has multiple audio tracks, you can click on the "Track" tab and select the corresponding audio track.

VR Player is a Canadian product developed by Vimersiv Inc. It is specially designed for playing virtual reality videos and is a popular program among Oculus Rift users. It plays not only VR, but also 2D and 3D videos. It opens media from multiple resources, such as YouTube URLs or cloud-based services like Dropbox. It allows "floating subtitles" for watching foreign immersive videos. So far, this player is only available in the English version.

Magix VR-X is a German product from MAGIX. The player supports Android, iOS and Windows with Oculus Rift, HTC Vive and Microsoft Mixed Reality. There are six languages available: English, Spanish, French, German, Italian and Dutch. Captions are only available in the Premium version (Photostory Premium VR).

Simple VR was developed in Los Angeles. It provides users with the simplest functions and can serve as a typical media player for users. You can play, stop and pause VR video through simple controls. In addition, it has a super enhancement mode which can improve the fidelity, contrast and detail of VR videos. It also allows something called "splitters" which take multi-track video files (such as .mkv) and feed the video decoder with specific video/audio tracks and subtitles that you can configure.

After a quick analysis of the players which allow subtitle files to be loaded, it can be observed that these files are rendered as a 2D overlay onto the video window. As there is still no specific subtitle file for VR360 videos, there is no information about where in the 360 scene the subtitle relates to. Regarding AD, some of the players provide support for selecting alternative audio tracks which can be used for playing the AD track. However, there is no mechanism for mixing the AD over the existing audio track in the player. Concerning SL, none of the players provide any mechanism for adding this access service. They also do not offer the possibility to overlay an additional video stream, which could be used for the SL service.

3.5. ImAc project and player

This section will present and analyse the VR360 video player created under the umbrella of the H2020 funded ImAc project following the Universal Design approach and the "Born Accessible" concept. The design departed from user specifications and took accessibility requirements into account in the development.

ImAc[12] was a European project funded by the European Commission that aimed to research how access services (subtitling, AD, audio subtitles, SL) could be integrated in immersive media. The project aimed to move away from the constraints of existing technologies into an environment where consumers could fully customise their experience (Agulló, 2020). The key action in ImAc is to ensure that immersive experiences address the needs of different kinds of users. One of the main features of the ImAc project was the user-centred methodological approach (Matamala et al. 2018), meaning that the design and development of the system and tools were driven by real user needs, continuously involving users in every step. The player was developed after gathering user requirements from people with disabilities in three EU

[12] See https://www.imac-project.eu/ [retrieved 15/08/2020]

countries: Germany, Spain and UK. User input was gathered in two iterations through focus groups and pre-pilot actions.

The first step in the user centric methodology was to define the profile of the end users. Two different profiles were created: professional user and advanced home user. Professional users were considered to be those who would use the tools at work: IT engineers, graphic designers, subtitlers, audio describers and sign language interpreters (signers). On the other hand, the advanced home users were people with disabilities who consumed the media content: the deaf, hard-of-hearing, blind, low vision users, and the elderly. To successfully profile home users, a number of considerations were taken into account beyond disability, such as level of technological knowledge and VR environments. This was decided in order to engage home users in an open conversation regarding their expectations and match them accordingly with the innovation. Only users with knowledge or experience in either functional diversity or technology were consulted. Other profiling features of the home users were oral/written languages (Catalan, German, Spanish and English) and three visual-gestural languages (Catalan Sign Language, German Sign Language and Spanish Sign Language). Other significant profiling factors were level of expertise in the service that the participant was testing (audio description, audio subtitling, sign language, subtitling), sensorial functionality (deaf, hard-of-hearing, blind, low vision) and age. As some degree of hearing or vision loss can often be linked to age, the elderly were included in the home users' category.

Once the two groups of end users (advanced home and professionals) had been defined, they formulated two user requirements: home and professional requirements. The former described the functions exposed by the ImAc services towards consuming media, and the latter described the functions from a working perspective. Three versions or iterations of the requirements were carried out. The first version was based on focus groups in which the user scenarios created by the ImAc partners were evaluated by both professional and home users. User scenario refers to what the already identified user would be experiencing and how (e.g., how the interface deals with AD depending on angle of visualisation). The second was conducted after the pre-pilot tests, where prototypes of accessible immersive media content were presented to the target group of home users (i.e., for the visual access services, the tests focused on the preferred size of the area to display the services and the preferred ways of guiding the users to the speaker). In this way, more extensive feedback on specific issues could be gathered and the home user requirements were subsequently fine-tuned. The final iteration took place after the demonstration pilots which involved both professional and home users. The resulting list of final requirements provides the basis for the further development and quality assurance of the ImAc platform.

After compiling all the information regarding the end user profile and requirements, the user interface (UI) was designed. The aim was to have a concept that was flexible enough to extend the settings later, based on the results of the user testing. The main challenge was to integrate four services with a large number of settings, while avoiding a long and complex menu. The UI design to access accessibility services in the ImAc player was based on existing players (legacy players from catch up TV services, web players for video-on-demand and streaming services and VR players); taking these as a starting point, a design for a "traditional UI" was developed. An "enhanced accessibility UI" was developed in parallel, the two were combined, and the resulting ImAc player UI offers aspects of both. The implementation of the UI allows access to the accessibility services in the ImAc portal and the player reflects their status after feedback from the user tests.

The success of the ImAc player and the reason why it has been presented as the main example of an accessible player is because it follows both the "Universal design" and "Born accessible" concepts. The first term comes from the European Standard EN 17161 (2019) 'Design for All - Accessibility following a Design for All approach in products, goods and services - Extending the range of users.' It specifies the requirements in design, development and provision of products, goods and services that can be accessed, understood, and used by the widest range of users, including persons with disabilities. Along with the European Standard EN17161, there is an EU standard for accessible technologies: the EN301549 (version 3.1.1.: 2019). These two standards secure the concepts of Universal Design and Born Accessible. According to Pilar Orero (2020: 4): "the concept of Born Accessible closely follows EU legislation and has been proven to be successfully integrated in R&D activities and developments."

3.6. Conclusions

More than ever, technology is enabling the empowerment of end users both as consumers and prosumers. However, technology is still designed with accessibility as an afterthought: away from user centric design. User interaction in today's Information Society plays a key role in the full social integration and democratic participation of all its citizens. Enabling easy access to content and guiding the user in controlling media services are two of the most recent UN and

European media regulations. At the same time, UIs should also be accessible, as the demand for guidance is especially high for accessibility services. Some groups of users need to activate an accessibility service such as subtitles before they can consume the media content. In general, the default setting for a media service is to have all accessibility services switched off. Therefore, it is very important that activating and controlling the accessibility services is made as easy as possible. This article has focused on the accessibility of the media players which are currently available commercially to show the way towards full accessibility in VR, at a time when VR content production is beginning. The article would like to raise awareness of the real possibility of generating accessible VR content from the point of production.

As we have been able to verify by analyzing the most widespread players available commercially, none of them provide access to the full set of accessibility services. VR players do not focus on accessibility services at all and they have not been created departing from user needs. Following the Born Accessible principle to avoid this basic problem, accessibility (and multilingualism) must be considered at the design stage of any process. Most of the players that have been analysed are only available in English, leaving out the users with other linguistic realities. The documented media players that support access to accessibility services do not use aligned conventions for their icons/representation. Using a universal set to represent accessibility services is desirable, and for that reason ImAc uses a set of icons proposed by the Danish Radio for all users in all countries.

To finish, we are living a global change of the consumer landscape due to Covid-19. Confinement measures have changed user demands and moved them further towards the use of online services. The so-called “new normal” will bring a significant increase in remote online activities and virtual experiences will be essential in the near future. Marketing, simulation, leisure, training, and communication will need to adapt to new needs, as well as to associate with them. The promotion of the Universal Design and Born Accessible concepts can have a significant role in achieving these goals and supporting the full democratic participation of all people, while protecting their social rights.

Appendix 1

Table 2

VR360 players facing accessibility

	SUBTITLE (Deaf and hard-of-hearing)	**AUDIO DESCRIPTION** (Blind and visually-impaired)	**SIGN LANGUAGE AVATAR** (Deaf)	**MULTILINGUAL** (Linguistically-impaired)
GOM Player[13]	Yes	No* (select audio track)	No	Yes
Codeplex VR Player[14]	No	No	No	No
Total Cinema 360 Oculus Player[15]	No	No	No	No
RiftMax VR Player[16]	No	No	No	No
SKYBox VR Video Player[17]	Yes (only in non-VR video)	No (select audio track)	No	No
VR Player[18]	Yes	No	No	No
Magix[19]	No (only in the premium version)	No	No	Yes
Simple VR[20]	Yes	No	No	No

Acknowledgements

The author is member of TransMedia Catalonia, a research group funded by Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya, under the SGR funding scheme (ref. code 2017SGR113). This article reflects only the authors' views and the funding institutions hold no responsibility for any use that may be made of the information it contains). This article is part of Marta Brescia's PhD in Translation and Intercultural Studies at the Department of Translation, Interpreting and East Asian Studies

13 See https://www.gomlab.com/ [retrieved 05/08/2020]
14 See https://archive.codeplex.com/?p=vrplayer [retrieved 05/08/2020]
15 See http://www.totalcinema360.com/ [retrieved 05/08/2020]
16 See https://www.wearvr.com/apps/riftmax-theater [retrieved 05/08/2020]
17 See https://skybox.xyz/en/ [retrieved 05/08/2020]
18 See http://www.vrplayer.com/ [retrieved 05/08/2020]
19 See https://www.magix.com/us/apps/vrx-player/ [retrieved 05/08/2020]
20 See http://simplevr.pro/ [retrieved 05/08/2020]

(Departament de Traducció i d'Interpretació i d'Estudis de l'Àsia Oriental) of Universitat Autònoma de Barcelona.

References

Agulló, Belén (2020) "Technology for subtitling: a 360-degree turn", *Hermeneus*, no. 22.

Agulló, Belén, and Anna Matamala (2019) "The challenge of subtitling for the deaf and hard-of-hearing in immersive environments: results from a focus group", *The Journal of Specialised Translation*, no. 32, 217–235.

Agulló, Belén, and Pilar Orero (2017) "3D Movie Subtitling: Searching for the best viewing experience", *CoMe - Studi di Comunicazione e Mediazione linguistica e culturale*, Vol. 2, 91–101, URL: https://ddd.uab.cat/record/183656 (accessed 8 August 2020).

Arnáiz Uzquiza, Verónica (2012) "Los parámetros que identifican el subtitulado para sordos: análisis y clasificación" *MonTi: Monografías de Traducción e Interpretación*, no. 4, 103–32. URL: https://www.raco.cat/index.php/MonTI/article/view/301235 (accessed 1 August 2020).

Aulavuori, Katja (2008) *Effects of the digital subtitling software on the subtitling process: a survey among Finnish professional television subtitlers*, MA thesis, University of Helsinki.

Azuma, Ronald, Yohan Baillot, Reinhold Behringer, Steven K. Feiner, Simon Julier, and Blair MacIntyre (2001) "Recent advances in augmented reality", *IEEE Computer Graphics and Applications, 21*(6), 34–47.

Belton, John (2012) "Digital 3D cinema: Digital cinema's missing novelty phase", *Film History: An International Journal*, 24(2), 187–195.

Carmigniani, Julie, and Borko Furht (2011) "Augmented reality: An overview" in *Handbook of Augmented Reality*, Borko Furht (ed), Florida, Springer: 3–46.

Carrozzino, Marcello, and Massimo Bergamasco (2010) "Beyond virtual museums: Experiencing immersive virtual reality in real museums" *Journal of cultural heritage,* no. 11, 4:452–458.

EN 17161 (2019) "Design for All - Accessibility following a Design for All approach in products, goods and services - Extending the range of users", URL: https://standards.cen.eu/dyn/www/f?p=204:110:0::::FSP_PROJECT,FSP_ORG_ID:62323,2301962&cs=1D28CFDC66E7CEF3CE441294CAA9FEABE (accessed 15 August 2020).

Fida, Benish, Fabrizio Cutolo, Gregorio di Franco, Mauro Ferrari, and Vicenzo Ferrari (2018) "Augmented reality in open surgery", *Updates in Surgery, no.* 70, 389–400.

Fidyka, Anita, and Anna Matamala (2018) "Audio description in 360º videos. Results from focus groups in Barcelona and Kraków", *Translation Spaces*, Vol. 7, no. 2, 285–303.

Gambier, Yves (2003) "Introduction. Screen transadaptation: perception and reception", *The Translator*, no. 9(2), 171–189

Gerber-Morón, Olivia, Olga Soler-Vilageliu, and Judith Castellà (2020) "The effects of screen size on subtitle layout preferences and comprehension across devices", *Hermēneus. Revista de Traducción e Interpretación*, Vol. 22, 157–182.

Gómez Suárez, Mónica, and Jean-Philippe Charron (2017) "Exploring concert-goers' intention to attend virtual music performances", *AIMAC 14 International Conference on Arts and Cultural Management*, Beiging.

González-Zúñiga, Diekus, Jordi Carrabina, and Pilar Orero (2013) "Evaluation of Depth Cues in 3D Subtitling", *Online Journal of Art and Design*, no.1, 3:16–29.

Greco, Gian Maria (2016) "On Accessibility as a Human Right, with an Application to Media Accessibility" in *Researching Audio Description. New Approaches*, Anna Matamala and Pilar Orero (eds), UK, Palgrave Macmillan: 11–33.

Guttentag, Daniel (2010) "Virtual reality: Applications and implications for tourism" *Tourism Management*, Vol. 31 (5), 637–651.

Holliman, Nick, Neil A. Dodgson, Gregg E. Favalora, and Lachlan Pockett (2011) "Three-dimensional displays: A review and applications analysis", *IEEE Trans. Broadcast*, *57*(2), 362–371.

International Organization for Standardization (2015) *ISO/IEC TS 20071-20 Information Technology — User interface component accessibility — Part 20: Software for playing media content,* ISO.

Lee, Kun, and Namho Chung (2008) "Empirical analysis of consumer reaction to the virtual reality shopping mall", *Computers in Human Behavior*, no. 24, 88–104.

Mas Manchón, Lluis and Pilar Orero (2018) "New Subtitling Possibilities: Testing Subtitle Usability in HbbTV", *Translation Spaces*, 7(2), 263–284.

Matamala, Anna, and Pilar Orero (eds) (2010) *Listening to Subtitles. Subtitles for the Deaf and Hard of Hearing,* Peter Lang, Berna.

Matamala, Anna (2017) *"Mapping audiovisual translation investigations: research approaches and the role of technology"* in Audiovisual translation-Research and use, Mikolaj Deckert (ed), Frankfurt, Peter Lang: 11-28.

Matamala, Anna, Pilar Orero, Sara Rovira-Esteva, Helena Casas-Tost, Fernando Morales Morante, Olga Soler Vilageliu, Belén Agulló, Anita Fidyka, Daniel Segura Giménez, and Irene Tor-Carroggio (2018). "User-centric approaches in Access Services Evaluation: Profiling the End User", *Proceedings of the Eleventh International Conference on Language Resources Evaluation,* 1–7.

Mendiburu, Bernard (2009) *3D movie making: Stereoscopic digital cinema from script to screen*, Taylor and Francis.

Mikami, Dan, Kosuke Takahashi, Naoki Saijo, Mariko Isogawa, Toshitaka Kimura, and Hideaki Kimata (2018) "Virtual Reality-based Sports Training System and Its Application to Baseball", *NTT Technical Review*, no. 16(3).

Milgram, Paul, and Fumio Kishino (1994) "A taxonomy of mixed reality visual displays", *IEICE Transactions on Information* Systems, no. 77, 1321–1329.

Montagud, Mario, Pilar Orero, and Anna Matamala (2020a) "Culture 4 all: accessibility-enabled cultural experiences through immersive VR360 content", *Personal and Ubiquitous Computing*, URL: https://link.springer.com/article/10.1007/s00779-019-01357-3#citeas (accessed 5 August 2020).

Montagud, Mario, Pilar Orero, and Sergi Fernández (2020b) "Immersive media and accessibility: Hand in hand to the future", *ITU Journal*, URL: https://www.itu.int/pub/S-JOURNAL-ICTS.V3I1-2020-2 (accessed 11 February 2021).

Oncins, Estel·la, Oscar Lopes, Pilar Orero, and Javier Serrano (2013) "All together now: a multi-language and multisystem mobile application to make live performing arts accessible", *Journal of Specialised Translation*, no.20, 147-164.

Oncins, Este·la, and Pilar Orero (2020) "No audience left behind, one App fits all: an integrated approach to accessibility services", *JosTrans*, no. 34.

Orero, Pilar (2020) *Born Accessible: Beyond raising awareness*, URL: https://imac-project.eu/wp-content/uploads/2020/05/Policy-Paper_final.pdf (accessed 5 August 2020).

Orero, Pilar (forthcoming) *"Audio Description Personalisation"* in The Routledge Handbook of Audio Description, Christopher Taylor and Elisa Perergo (eds), London, Taylor Francis.

Orero, Pilar, Rai Sonali, and Chris Hughes (forthcoming) *"The challenges of audio description in new formats"* in Innovation in Audio Description Research, Sabine Braun (ed.), London, Routledge.

Papachristos, Nikiforos M., Ioannis Vrellis, and Tassos A. Mikropoulos (2017) "A Comparison between Oculus Rift and a Low-Cost Smartphone VR Headset: Immersive User Experience and Learning", *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*, 477-481, doi: 10.1109/ICALT.2017.145.

Parham Groesbeck, Eric G. Bing, Anthony Cuevas, Boris Fisher, Jonathan Skinner, Richard Sullivan, and Mulindi H. Mwanahamuntu (2019) "Creating a low-cost virtual reality surgical simulation to increase surgical oncology capacity and capability", *Ecancermedicalscience,* no. 13.

Remael, Aline, Pilar Orero, and Mary Carroll (eds) (2014) *Audiovisual Translation and Media Accessibility at the Crossroads: Media for all 3*, Amsterdam, Rodopi.

Rizzo, Albert, Maria Schultheis, Kimberly A. Kerns, and Catherine Mateer (2004) "Analysis of assets for virtual reality applications in neuropsychology", *Neuropsychological Rehabilitation*, no. 14:1-2, 207–239.

Romero-Fresco, Pablo (2015) *The reception of Subtitles for the Deaf and Hard of Hearing in Europe*, Berlin, Peter Lang.

UN (2006) *Convention on the Rights of People with Disabilities*, URL: https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with- disabilities.html (accessed 15 August 2020).

UNE-EN 301549 (2019) *Accessibility requirements for ICT products and services.*

UNESCO (2005) *Convention on the Protection and Promotion of the Diversity of Cultural Expressions*, URL: https://en.unesco.org/creativity/convention/texts (accessed 15 August 2020).

Vela Valido, Jennifer (2007) *Software de audiodescripción: análisis comparativo de los programas existentes en España y EE. UU*, MA thesis, Universitat Autònoma de Barcelona.

CHAPTER 4

Article 3: VR 360º subtitles: Designing a test suite with eye-tracking technology

VR 360° Subtitles. Designing a Test Suite with Eye-Tracking Technology

Marta Brescia Zapata

Universitat Autònoma de Barcelona, Barcelona, Spain

Krzysztof Krejtz

SWPS University of Social Sciences and Humanities, Warsaw, Poland

Pilar Orero

Universitat Autònoma de Barcelona, Barcelona, Spain

Andrew T Duchowski

Clemson University, USA

Chris J Hughes

University of Salford, Salford, UK

Abstract

Subtitle production is an increasingly creative accessibility service. New technologies mean subtitles can be placed at any location on the screen in a variety of formats, shapes, typography, font size, and colour. The screen now allows for accessible creativity, with subtitles able to provide novel experiences beyond those offered by traditional language translation. Immersive environments multiply 2D subtitle features to produce new creative viewing modalities. Testing subtitles in eXtended Reality (XR) has expanded existing methods to address user needs and enjoyment of audiovisual content in 360º viewing displays. After an overview of existing subtitle features in XR, the article describes the challenges of generating subtitle stimuli to test meaningful user viewing behaviours, based on eye-tracking technology. The approach for the first experimental setup for implementing creative subtitles in XR using eye-tracking is outlined in line with novel research questions. The choices made regarding sound, duration and storyboard are described. Conclusions show that testing subtitles in immersive media environments is both a linguistic and an artistic endeavour, which requires an agile framework fostering contrast and comparison of different functionalities. Results of the

present, preliminary study shed light on the possibilities for future experimental setups with eye-tracking.

Keywords

Subtitles, immersive environments, 360º videos, testing, eye-tracking

4.1. Introduction

Immersive media technologies such as Virtual Reality (VR) and 360° videos are increasingly prevalent in society. Their potential has placed them in the spotlight of the scientific community for research and education. Industry has also adopted them not only in the entertainment sector, but also for communication, arts, and culture, which has attracted more and mixed audiences (Montagud et al., 2020). At present, these technologies are gaining popularity very fast due to the COVID-19 crisis as they enable interactive, hyper-personalised, and engaging experiences anytime and anywhere. Moreover, 360° videos, also known as immersive or VR360 videos, are a cheap and effective way to provide VR experiences. Specialised multi-camera equipment that can capture a 360° or 180° Field of View (FoV) instead of the limited viewpoint of a standard video recording, is used to produce content. VR360 videos can be enjoyed both via traditional devices (PC, laptops, smartphones) or VR devices (Head-Mounted Displays). They can also be consumed as a CAVE (Cave Automatic Virtual Environment), which uses high-resolution projection screens to deliver 360◦ visual experiences.

Immersive environments (in eXtended Reality, or XR) are generally used as an umbrella term referring to hardware, software, methods, and experience in Augmented Reality (AR) or VR or in general Mixed Reality (MR). The main goal of any immersive content is to make people believe that they are physically present (Slater & Wilbur, 1997). According to Rupp et al. (2016, p. 2108), VR360 videos can allow for "highly immersive experiences that activate a sense of presence that engages the user and allows them to focus on the video's content by making the user feel as if he or she is physically a part of the environment". Immersive videos, however, can also produce negative effects such as motion or simulator sickness, possibly turning people away from VR as a medium (Smith, 2015).

Like any other type of media content, 360º media experiences should be accessible. In almost all media assets, accessibility is added as an afterthought during the postproduction phase, despite many voices asking for accessibility in the creation process (Mével, 2020; Romero-

Fresco, 2013). For this research we focus on subtitling, where standardised practices have emerged (Matamala & Orero, 2018), rather than covering different accessibility services. In 2D subtitles, the main aspects to consider are position, character identification, speed, number of lines, and number of characters (Bartoll, 2004; Díaz-Cintas & Remael, 2007; Gottlieb, 1995). Nevertheless, some Audiovisual Translation (AVT) studies have challenged traditional subtitling practices, encouraging more creative and integrated subtitles (Foerster, 2010; Fox, 2018; McClarty, 2012, 2014).The production of creative subtitles requires technology since such subtitles may change any of their paratextual features like the font or size or colour, but also where they are positioned, and more so in immersive environments where 2D features do not apply (Hughes et al., 2015; Lee et al., 2007). The integration of subtitles in XR is yet to be defined, and multiple challenges have emerged. Subtitles should be generated "in an immersive, engaging, emotive and aesthetically pleasing way" (Brown et al., 2017, p.1), always considering accessibility and usability.

Beyond the challenge of subtitle text creation, XR requires direction to the sound source, as it may be outside the current audience viewpoint. Guiding and readability require the subtitler to preview and tweak formal aspects (Hughes & Montagud, 2020; Orero et al., 2020). This has led to the design of a new, web-based, prototyped framework that generates subtitles in 360º videos. The present article aims to identify how to display such subtitles for an optimal viewing experience. The framework allows for methods used in existing solutions (Brown & Patterson, 2017; Montagud et al., 2019; Rothe et al., 2018) to be easily contrasted and compared, as well as for the quick implementation of new ideas for user testing. After an overview on subtitle features in XR, the article describes the challenges of generating subtitle stimuli to test meaningful user viewing behaviours, based on eye-tracking technology. The approach for the first experimental set up for implementing creative subtitles in XR using eye-tracking is presented, in line with the stated research questions.

4.2. An overview of subtitles in immersive environments

Even though XR media was first introduced in the world of videogames, thanks to the development of 360º recording equipment these technologies are now expanding to videos (Hughes et al., 2020a). There are a few significant differences between content created within 2D and 3D environments. 2D means that the content is rendered in two dimensions (flat), while 3D content has depth and volume which allows a rich visual experience. According to Skult and Smed (2020, p.451), "the key challenge for XR is that the FoV is limited and the interactor cannot pay attention to the entire virtual scenery at once.” The immersive experience, as in real life, moves from passive to active with the user becoming the centre of the story “creating a greater emotional nexus” (Cantero de Julián et al. 2020, p.418). In a play or opera, the action takes place on the proscenium. However, another activity somewhere in the theatre may distract from that narrative, such as the noise of a lady unwrapping sweets two rows away. The audience in VR has freedom of movement. They can also determine the time spent in any area of interest or field of vision and decide on where to focus their attention. This freedom affects subtitle reading since the development of the narrative may be random, decided by the viewer. Similarly, in VR, the aim is for immersiveness and the concept of presence and engagement are central, with the ultimate goal of being a witness to the narrative from a first-person viewpoint. This breaks with the concept of a passive audience that reads subtitles, following what Jenkins et al. (2015) define as *spreadable* reading, meaning that the audience spreads its attention across the image as the linear narrative is displayed. In VR images and sound surround, there is no linear narrative and passive viewing moves towards interaction or *drillable* viewing as in video games or transmedia products. In this context, Mittell (2009) explains that “spreadable media encourages horizontal ripples, accumulating eyeballs without necessarily encouraging more long-term engagement. Drillable media typically engage far fewer people but occupy more of their time and energies in a vertical descent into a text’s complexities.” These features are theoretical principles that have yet to be tested.

The value of VR lies in its potential to tamper with both time and space; hence the experience relies on the viewer. This has a direct effect on the way subtitles are consumed. A person may be watching one part of the scene while there is a person speaking away from the viewing field. A hearing person may be able to locate the sound source but someone with hearing loss will need to be guided.

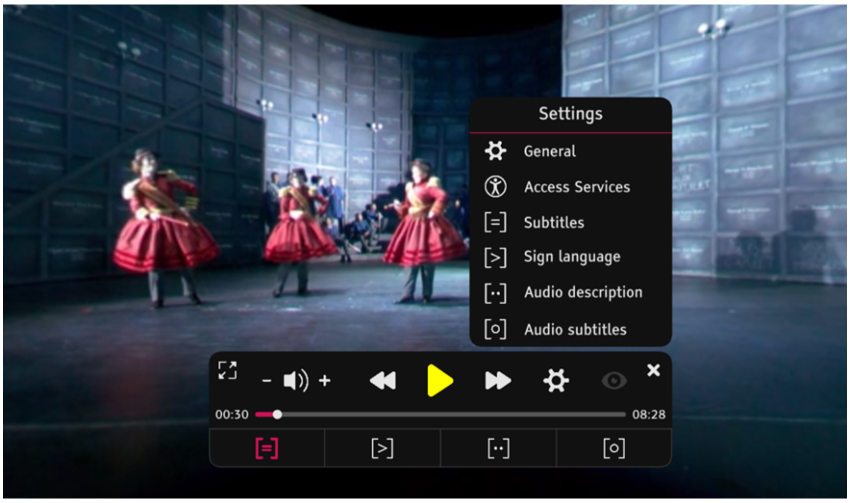
Another feature that is different from 2D resides in the way media is accessed. There are two options: using CAVE (the naked eye) or a device such as a Head-Mounted Display (HMD). This is a type of display device or monitor that is worn over the head and allows the user to be immersed in whatever experience the display is meant for. The 360º environment accessed when wearing the HMD may be an animation (such as a video game or an animated movie) or live action (such as a movie or a documentary). Depending on the type of media, the content will be installed on a PC or on the HMD itself or stored on the cloud. As Internet speed improves, media content streamed from the web is becoming increasingly popular.

As in traditional 2D media, to create and consume subtitles in XR, a subtitle editor and a subtitle player are needed. Although immersive video players offering the ability to play VR360 video are commercially available, not many of them support accessible services (Brescia-Zapata, 2022). The player needs to be accessible, and the user has to activate the display accessibility. The interface or menu also needs to display the choice of accessibility services available, and finally, the interaction with the terminal or device also needs to be accessible. All these features show the complex ecosystem required for a true XR accessible experience. This, linked to the lack of standardised solutions and guidelines, has led to the development of non-unified solutions, meeting only specific requirements (Hughes & Montagud, 2020). The majority of players seem to have inherited features from the traditional 2D world, instead of addressing the specific features of 360º environments.

This situation served as an inspiration for initiatives like the European H2020 funded Immersive Accessibility (ImAc) project[21] that explored how accessibility services and assistive technologies can be efficiently integrated with immersive media, focusing on VR360 video and spatial audio. Under the umbrella of this project, both an accessible player and a subtitle editor were developed. The accessibility-enabled 360º ImAc player supports audio description, audio subtitles and sign language along with other features (Montagud et al., 2019) as can be seen in Figure 11.

Figure 11

ImAc player settings



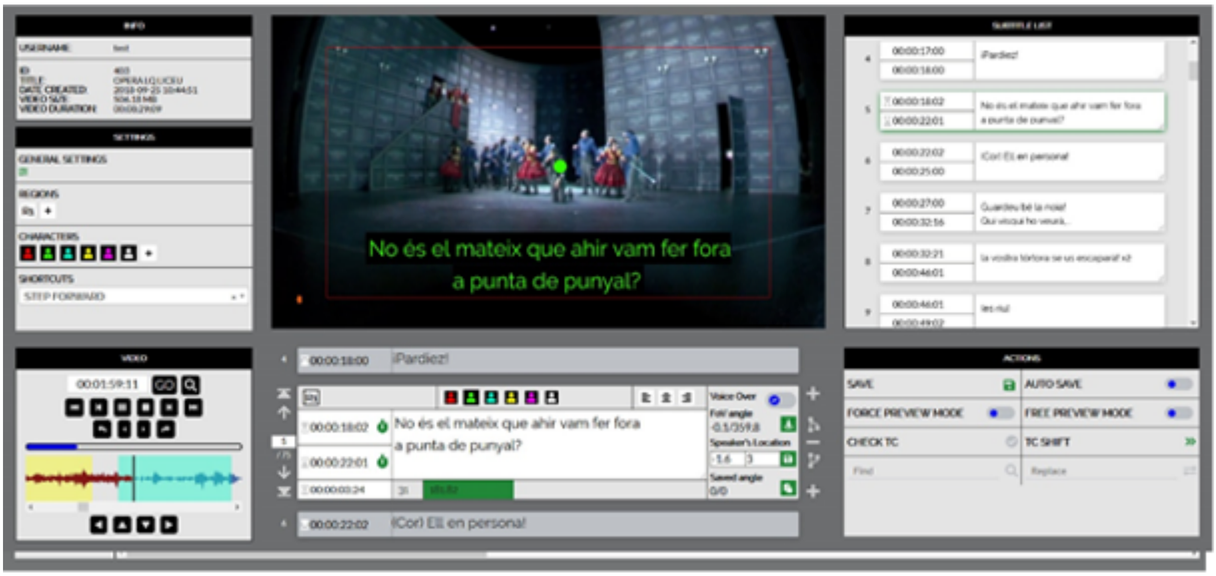
Note. Imac Player screenshot

[21] http://www.imac-project.eu

In contrast, the ImAc subtitle editor is a commercial web-based editor, and its interface is similar to that of any traditional subtitle editor, as can be seen in Figure 12. The main innovations are related to the FoV in VR360 video, i.e. the extent of observable environment the user is able to see at any given moment. It includes navigation buttons for FoV in spherical space to move up, down, left and right. There is also a button which moves the FoV to the angle where the speaker of the current subtitle is located. The editor also allows the FoV angle to be changed using the naviga tion buttons in the video control area or moving the mouse with the left button over the video. By default, the video initially has the current angle as longitude: 0.00º and latitude: 0.00º. In addition, the voiceover option can be marked when there is no speaker in the 360º scene.

Figure 12

Immersive subtitle editor developed in ImAc



The basic tools to create and consume accessible VR content are now commercially available, e.g., a VR subtitle editor and a VR subtitle player. What is evident is that unless different display modes can be produced, they cannot be tested, and this is one of the shortcomings of the ImAc project which was concluded recently and focused on traditional subtitles projected on immersive environments (Hughes et al., 2020b).

4.2.1. Related work

Excluding works that have added (sub)titles at post-editing stages, only three recent studies have focused on investigating subtitles in immersive environments. All the studies followed a user-centric methodology and chose people with hearing loss for testing. Reading skill was not considered within the demographic data.

The British Broadcasting Corporation (BBC) was one of the first research organisations to design subtitles in XR (Brown & Patterson, 2017). The BBC research team first identified the main challenges when developing subtitles for immersive content and based on these, the following four solutions for subtitle rendering were developed by Brown et al. (2017):

- Evenly spaced: subtitles are placed into the scene in three fixed positions, equally spaced by 120° around the video and slightly below the eye line;
- Follow head immediately: the subtitles are presented as a 'head-up display' always in front of you, and slightly below straight ahead. As you turn your head, the subtitle moves with you, always at the same location in the headset display;
- Follow head with lag: the subtitles follow head direction, but only for larger head movements: if you look slightly left or right the subtitle stays in place, but a head movement of a greater amplitude will cause the subtitle to catch up with your head orientation;
- Appear in front, then fixed: each subtitle is placed in the scene in the direction you are looking at the time when it appears and remains fixed in that location in the scene until it disappears.

These four rendering modes were tested with several clips (Brown, 2017), and users reported that while it was easy to locate the evenly spaced captions, they preferred the head-locked options (see Table 3). Head-locked subtitles resemble most traditional ecstatic 2D subtitles, always visible at the bottom of the screen. These results come as no surprise since for years now subtitle testing in Europe has shown that people like what they are used to, even if the data demonstrate that the solution is not ideal, as proved with eye-tracking tests (Mas Manchón & Orero, 2018; Romero-Fresco, 2015).

Table 3

Numbers of people (and percentages) who selected each behaviour as their favourite or least favourite behaviour.

Behaviour	Favourite	Least favourite
Evenly spaced	1 (4%)	5 (38%)
Follow head immediately	10.5 (44%)	3 (23%)
Follow with lag	7 (29%)	2 (15%)
Appear in front, then fixed	5.5 (23%)	3 (23%)

Note. Least favourite was not specifically requested, so was not available for all participants.

The second study (Rothe et al., 2018) compared the two presentation modes: fixed and head-locked subtitles. Although no conclusive results were found, in terms of comfort (i.e., presence, VR sickness and task load) fixed subtitles led to slightly better results even though fixed captions in general mean that users may not always be able to see the caption as it may be outside their FoV.

The third study, performed under the umbrella of the H2020-funded ImAc project (Hughes et al., 2019), revealed the need to guide users to the sound source of the subtitle (i.e., a sound effect or a character speaking or not speaking). To facilitate this requirement, location within the 3D space information was added to each subtitle (Agulló & Matamala, 2019). This allowed for different modes to be developed which could guide the user to where the speaker was located (Agulló et al., 2019). However, this had the drawback that the location was only specified once per caption, and if a person was moving dynamically, this could affect the exactness of the guiding feature (Hughes et al., 2019). Nevertheless, the ImAc project designed and developed several guiding mechanisms, and test results showed two preferred methods:

- ImAc Arrow: an arrow positioned left or right directs the user to the target;
- ImAc Radar: a radar circle is shown in the user's view. This identifies both the position of the caption and the relative viewing angle of the user.

In the area of standardisation, a W3C Community Group[22] is focusing on developing new standards for immersive subtitles. They have recently conducted a community survey to gather opinions, but no tests were performed. A small group of users with different hearing levels (Deaf, Hard of Hearing, and Hearing) were asked to evaluate each of the identified approaches for subtitles within immersive environments. Head-locked was clearly identified as the preferred choice. However, it was noted that this was a likely outcome since it replicated the experience that users were familiar with, as indicated above. It was also acknowledged that it was difficult for users to evaluate new methods theoretically without the opportunity to experience them while accessing content. Although all agreed that the head-locked option should be set as default, the respondents maintained that other choices should be made available. Other suggestions included changing the font size and colour and the number of lines (two lines being the default). Consequently, the need to develop a framework enabling delivery of the full experience of each captioning mode, in an environment where an extensive user study could be conducted, would be a priority prior to testing.

4.3. Methodology for a pilot study

Conducting a pilot study before launching a full spectrum study is always desirable. The goal of such a pilot study is not only to try to ensure that the survey questions operate well, but also that the research procedures and measures are adequate and reliable (Bryman, 2004). Especially when research aims to substantiate the validity of a new framework and/or involve the use of novel technology (such as eye-tracking in VR), a pilot study is crucial to ensure that both methodology and the study design are accurate and reliable. The preparation stage for this pilot study involved four main steps: user profile definition, selection of the testing material, implementing the material within the new framework and design of the test procedure itself.

The procedure followed by the current study consisted of four stages: an introduction, a questionnaire on demographic information, an eye-tracking test using 360º immersive videos, and a focus group. The main aims were (1) to test a new framework for subtitle presentation in 360º videos, (2) to obtain feedback regarding expectations, recommendations, and preferences from users when consuming subtitles and (3) to explore the visual attention distributions between subtitles and movie scenes while watching videos in VR. To do so, three different

[22] https://www.w3.org/community/immersive-captions/

subtitle modes were implemented: mode 1 (following ImAc results), mode 2 (following Fox, 2018 studies) and mode 3 (fully customised).

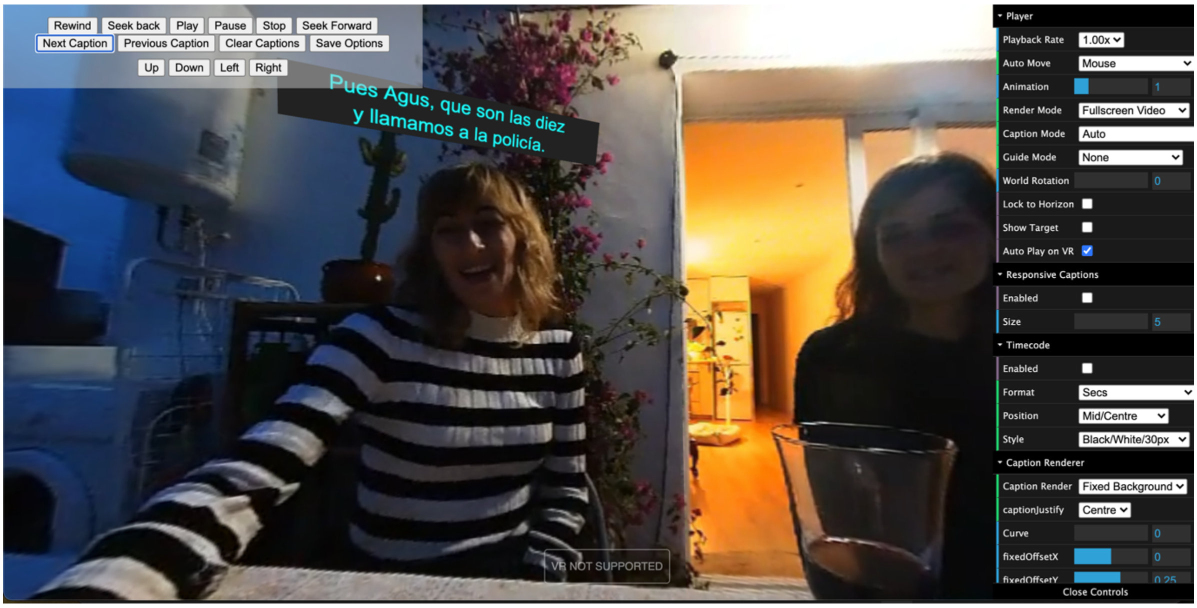
Before starting the pilot study and taking the previous work in the field as a reference, the following hypothesis was formulated: Fixed, near to the mouth subtitles will allow viewers to spend more time exploring the image instead of reading the subtitles than head-locked subtitles.

4.3.1. The live web testing framework

One of the challenges for testing immersive subtitles is the difficulty of having users evaluate new modalities properly because of the cost and time needed to create new prototype subtitle presentations so that users can experience them. To this end, an XR subtitle web simulator was developed by Hughes et al. (2020b). This web-based simulator was designed for rapid prototyping of subtitles within a 360º space, as can be seen in Figure 12 below.

Figure 12

Open-source demo player developed as part of this study

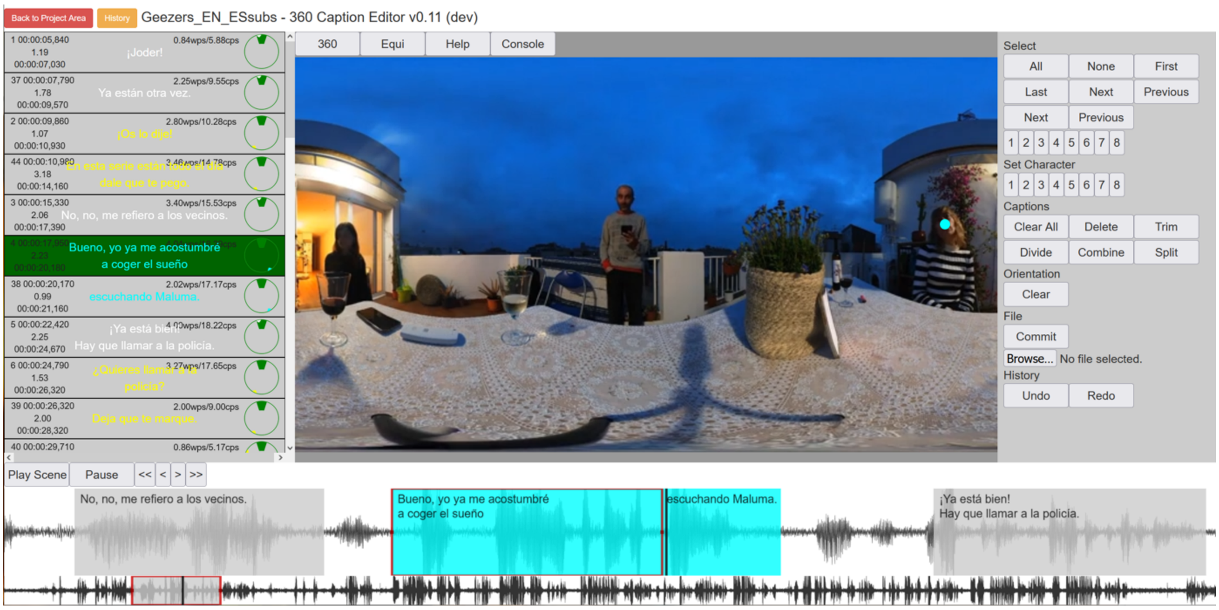


This new framework allows for instant immersive subtitle production in up to nine different modes: four of them are fixed, where the subtitle is rendered relative to a fixed location in the world, generally at the position of the character speaking, and five are head-locked, where the subtitle is rendered relative to the user's viewpoint. The main idea behind this demo player is to allow as much personalisation as possible (i.e. subtitle display, placement, timing, render mode, guiding mechanism, etc.); this way, any feature may be activated to define and test subtitles within 360º videos.

Along with this XR subtitle simulator, a web-based editor was also developed (see Figure 13), which allows previously created subtitles to be imported in .srt format or subtitles to be created from scratch. On the one hand, each subtitle can be associated with a character ("Set Character" button), and, on the other hand, each subtitle must have an associated position (FoV), i.e. the place in the 360º scene where it should appear.

Figure 13

Open-source editor developed as part of this study



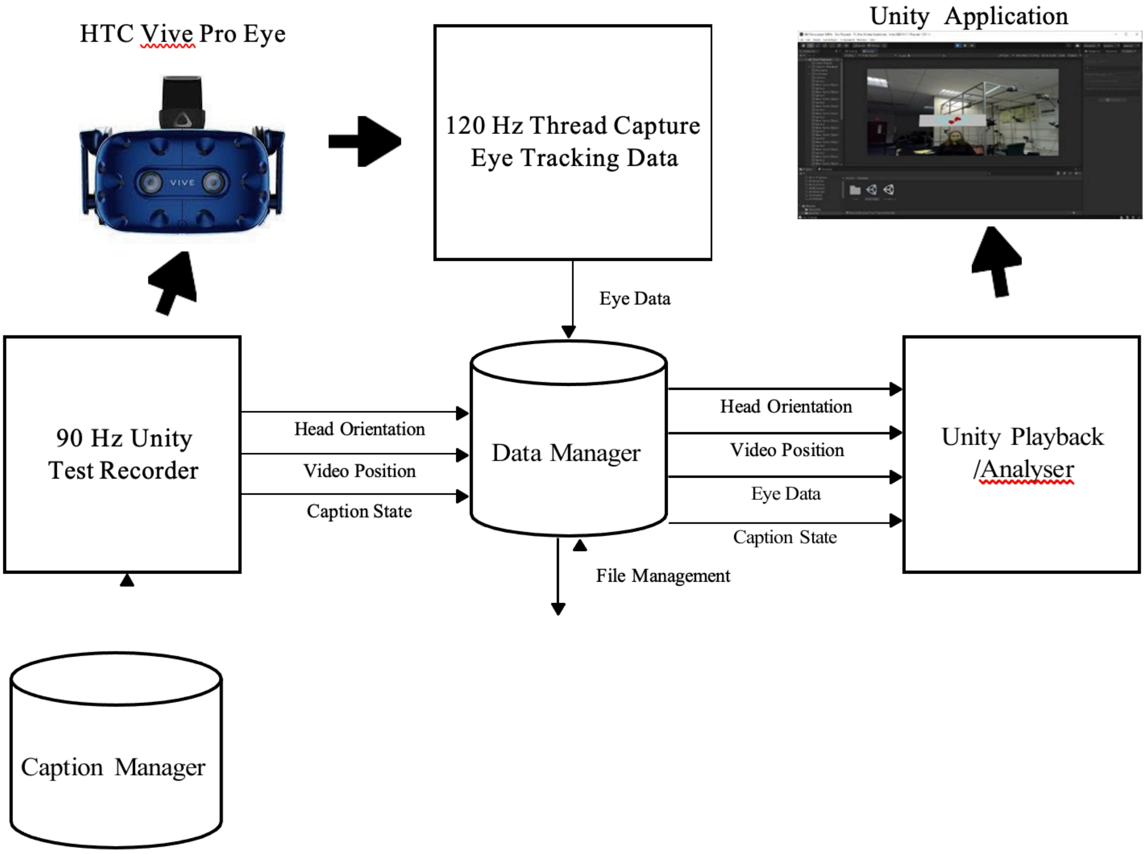
Both the demo player and the editor are open-source and can be accessed from a main project area where all the imported 360º videos are located. These tools take their inspiration from the player and the editor developed in the ImAc project. The main difference between them is that ImAc tools are intended to be used by generic audiences (final users), while the tools used in this study are more focused on research and testing.

4.3.2. System Architecture

To enable recording of gaze within 360º video, the live web testing framework developed by Hughes et al. (2020a, 2020b) was ported to Unity 3D to allow the display of 360º video content and to capture data from the eye tracker built into the VR device. A new system architecture emerged, as depicted by the schematic in Figure 14.

Figure 14

Eye-tracking VR system architecture



The system architecture was developed to utilise the HTC Vive Pro Eye, which contains an eye tracker from Tobii built into the display. The application uses two Unity assets, one specifically optimised for recording and the other for playback. At the centre of the architecture is a Data Manager, designed to store all test data. It also handles file management and can generate the output data in a variety of formats as required.

The recording application allows for a specified 360º video to be played with the captions fixed in the scene. During the test, each event and data is logged into the data manager as it becomes available and timestamped. In order to be able to replay a user viewing session, the system needs to record head orientation, video (frame) position, gaze data (raw and analysed, see below) as well as the subtitle caption state, i.e., which caption from the accompanying subrip format (.srt) file was being displayed and where.

The playback application allows for the data to be retrieved from the Data Manager and the entire test to be replayed. This offers the opportunity to change the analysis process or to include additional Areas of Interest (AOIs) and allows for the analysis to be repeated. It also allows for visual analysis by overlaying the eye data onto the video following capture.

One technical difficulty that had to be overcome was synchronisation of gaze data with video and subtitle data. Gaze data is sampled at 120 Hz while the Unity display refresh rate is 90 Hz. Thus, on average 1.3 gaze samples are expected on any given frame. To enable synchronisation from data streams of different rates, a separate eye-tracking data thread was created to collect gaze data captured at 120 Hz, ensuring no loss of eye movement samples. System playback can be set to either the speed of video or eye tracker, with gaze data drawn atop the projected video, as shown in Figure 15.

Figure 15

Gaze recording in VR showing varying elements of gaze to subtitle: (a) saccade in mid-flight, (b) saccade landing site with slight undershoot, (c) saccade to midpoint of subtitle, (d) fixation within a subtitle.

(a) *saccade flight* (b) *saccade landing below text* (c) *saccade landing on text* (d) *fixation on word*

4.3.3. Participants

The size of the group was determined in accordance with the pilot nature of the study (Bryman, 2004, p. 507). In the beginning, 7 participants were expected, but due to complications related to the Covid-19 pandemic, only five appeared (2 male and 3 female). All participants were professionals from the Arts, Sciences or Humanities fields, staying for a few weeks at the residence Faberllull in Olot. The average age was 40 (SD = 8.37) and all of them had completed a postgraduate university degree. All were active professionally (1 social worker, 1 cultural manager, 1 music therapist, 1 pre-doctoral researcher, 1 project manager). All participants spoke Spanish and at least one other language.

All participants were familiar with using computers and mobile devices. Two participants reported having previous experience with VR. Most of the participants declared watching different TV content with subtitles at least occasionally (only one of them claimed that she/he never turned subtitles on).

4.3.4. Study materials

One of the main concerns of the study was to find appropriate material for testing. Due to the difficulty of finding royalty-free material that met the needs of the study, a homemade 360º video was recorded using an Insta360 One X2 camera. The duration was 3 minutes and 45 seconds. The camera was settled in the centre of the action and three characters were positioned around the camera so that the action took place throughout the 360º space. The characters followed a script to avoid overlaps because if two characters located at different points in the 360º scene spoke at once, it would be almost impossible for the user to read the subtitles.

There were three types of subtitles yielding three experimental conditions:

- Mode 1: following ImAc results. Same font and colour (b&w) for all the characters, with a grey background and head-locked.
- Mode 2: following Fox 2018 studies. Same font and colour (b&w) for all the characters, without background and near the mouth.
- Mode 3: fully customised. Different font and colour for each character, with a grey background and near the mouth.

Other conditions, for example subtitles for non-human sound sources will be presented in future tests.

4.3.5. Procedure

The study included the following stages. First, participants were welcomed by the facilitator, who briefly explained the aim of the project. The session took place in a meeting room divided into two separate spaces. On one side there was a large TV screen, a computer connected to the screen and chairs for the participants. On the other side, an improvised eye-tracker lab was installed with a computer and a pair of HTC Vive Pro HMD. One researcher took notes and summarised the conclusions in real-time. Second, the aim of the focus group was explained to the participants, and they were asked to sign informed consent forms. The third step consisted of filling in a short questionnaire on demographic information. Finally, the session began. To trigger the discussion, the facilitator gave a short introduction to VR and 360º content and explained how subtitles are integrated within 360º content, showing VR glasses to the participants.

The eye-tracking technology was introduced, as it is integrated within the VR glasses and was one of the data collecting methods in the study. The facilitator explained that 360º content can also be accessed on a flat TV screen using a mouse to move around the 360º scene. Different types of subtitles were presented to give users some idea about how creative subtitling can be implemented in immersive content and to stimulate their imagination.

Then, each participant used the HTC Vive Pro HMD to watch a short video with audio in English and subtitled into Spanish. In total there were three rounds, always using the same video but a different subtitle mode each time. The order of the participants was determined randomly. Immediately after each visualisation, participants filled out a short questionnaire with questions on content understanding, subtitling preferences, and the task load index (NASA-TLX).

After the last round the focus group took place. Together with the stimuli, the facilitator used a list of guiding questions grouped under major topics to generate the participants' reactions. A balance between an open-ended and a structured approach was sought, and the result was a lively discussion in which interesting ideas came up.

4.4. Pilot study results

The data analysis of the study was mainly qualitative accompanied by descriptive statistics of the post-study questionnaire and eye movements captured during the study (see Figure 6).

4.4.1. Movie content understanding

To check the understanding of the stimuli movies we averaged the accuracy of responses to questions about the content separately for each condition. The highest average accuracy was obtained for the movie with fully customised subtitles ($M = 0.64$, $SD = 0.26$). Average accuracy for movies with subtitles in mode 1 ($M = 0.52$, $SD = 0.18$) and mode 2 ($M = 0.52$, $SD = 0.36$) were the same.

Additionally, when asked about the description of the scenes presented in the movie, participants used, on average, slightly more words after watching the movie in mode 1 ($M = 22.2$, $SD = 12.99$) than mode 3 ($M = 18$, $SD = 8.34$). The smallest number of words used in the description after watching the movie was in mode 2 ($M = 16.20$, $SD = 9.01$).

Qualitative analysis of responses during the focus group interviews showed that some of the participants could not understand the plot until the third visualisation of the clip. This could be related to a learning effect, but also because 3 of the participants had no previous experience with subtitled immersive content. Furthermore, another participant commented that sometimes it was difficult to follow the story because she was distracted exploring the 360º scene. The participant who was the least familiar with new technologies (and the least interested in the immersive format) noted that paying attention to the story stressed her and that she tried to distract herself during the visualisations.

4.4.2. Subtitle readability

The participants were asked whether they had been able to read the subtitles after watching each movie. Two responded 'yes', two 'no' and one was 'not sure' for mode 1. In mode 2, two responded 'yes' and three 'no'. The least readable subtitles seemed to be in mode 3. Three participants noted they were not able to read them; only one responded 'yes' and one participant was 'not sure'. When asked to estimate the percentage of subtitles that they were able to read, the differences were very small: 70% in mode 1, 68% in mode 2, and 67% in mode 3. Both results seem to suggest a slight preference for the subtitles in mode 1 as the most readable ones.

These results comply with the qualitative data extracted from the focus group, since most participants agreed that mode 3 was difficult to read. Only one of the participants noted that she liked the coloured text, and there was a brief discussion about the possibility of customising the subtitles further. Regarding the grey background, there was no consensus: some of the participants found subtitles with no background hard to read, others found them less intrusive. One participant highlighted the reading pace in general, arguing that some captions disappeared 'too soon' forcing the user to read faster.

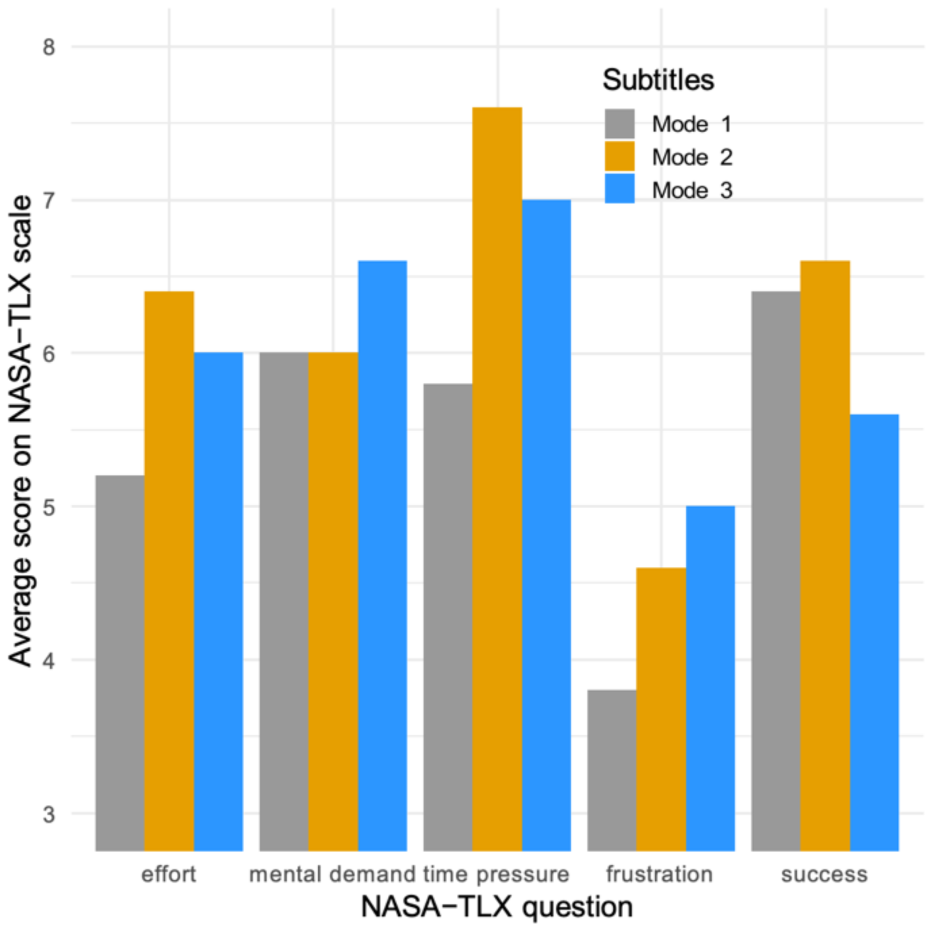
4.4.3. Self-reported task load

To collect self-reports on the effort elicited by the task of watching stimuli movies in different subtitle modes, the NASA-TLX scale with five questions was analysed (see Figure 16). Subjective evaluation of effort while watching videos in different subtitle modes also suggests a preference for mode 1 (M = 5.20, SD = 2.39). In the participants' opinion, more effort was required to read subtitles in modes 2 (M = 5.80, SD = 2.17) and 3 (M = 6.00, SD = 1.87). However, evaluation of mental load shows a different pattern, namely that modes 1 (M = 6.00, SD = 2.35) and 2 (M = 6.00, SD = 2.35) were equally less demanding than mode 3 (M = 6.60, SD = 2.30). Participants also evaluated reading modes 1 (M = 6.40, SD = 1.52) and 2 (M =

6.60, SD = 2.30) with greater perceived success than mode 3 (M = 5.6, SD = 2.19). These results are not surprising considering the average responses regarding how time-pressured participants felt. Subtitles in mode 2 caused the highest experience of time pressure (*M* = 7.60, *SD* = 0.89); this was lower in mode 3 (*M* = 7.00, *SD* = 0.71) and lowest in mode 1 (*M* = 5.80, *SD* = 2.17). The perceived level of frustration/stress was lowest when watching the video in mode 1 (*M* = 3.80, *SD* = 2.17), greater in mode 2 (*M* = 4.60, *SD* = 3.21), and greatest in mode 3 (*M* = 5.00, *SD* = 2.35).

Figure 16

Task load self-reports with NASA-TLX scale while watching videos in different subtitle modes



4.4.4. Attention distribution and cognitive effort while reading captions and scene viewing

Gaze was captured as it traversed subtitles when reading the text displayed within the quadrilaterals that contained them. The analysis of the eye movement signal relies on fixation detection, which in turn depends on saccade detection. Fixations are detected within the raw eye movement signal following Nyström and Holmqvist (2010) and by using the Savitzky-Golay filter for velocity-based (I-VT (Salvucci & Goldberg, 2000)) event detection Savitzky and Golay (1964).

The current system architecture allows for detection of fixations falling within arbitrarily defined Areas of Interest (AOIs), including polygons defined over actors and more importantly over quadrilaterals (quads) used to display subtitles as well as quads defined over individual words, see Figure 17 below.

Figure 17

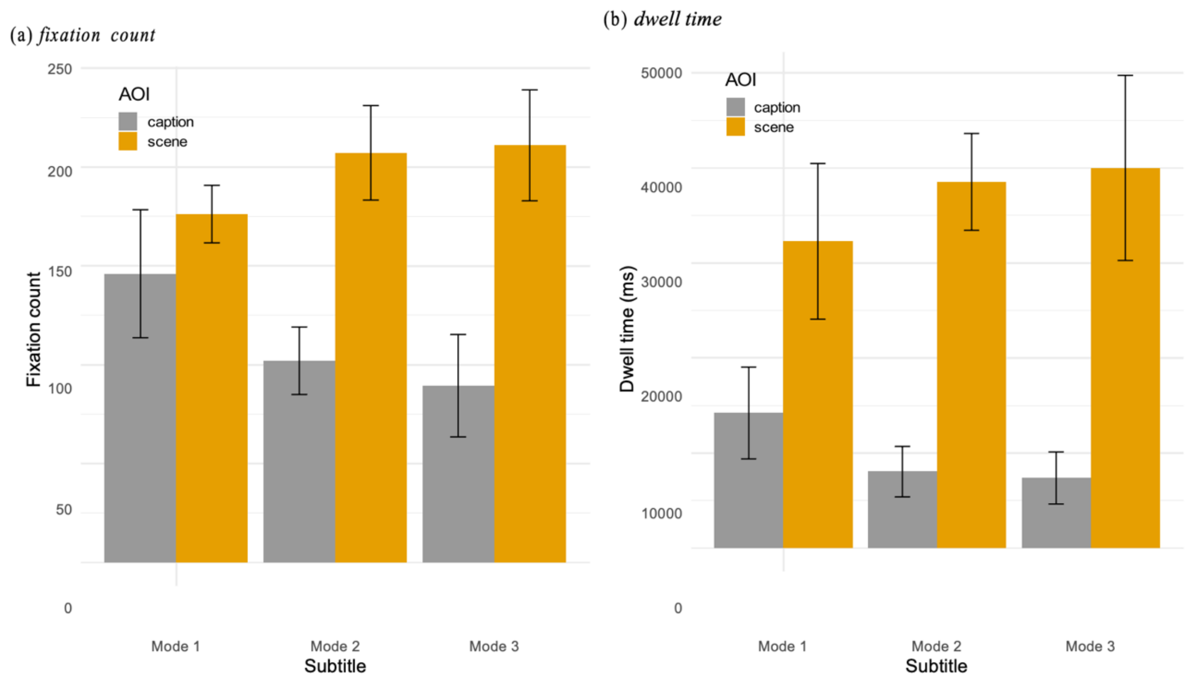
Gaze recording showing fixations over Areas of Interest: (a) actor body, (b) subtitle quad, and (c) individual word

The eye movement analysis aimed first at capturing differences in attention to captions and visual scenes in terms of fixation count and dwell time as dependent variables. Descriptive statistics show that in all conditions most fixations were on the visual scene rather than on subtitles. However, the difference is less important for the video in mode 1.

Participants exhibited more fixations on captions (M = 145.8, SD = 32.37) than on the visual scene (M = 176.0, SD = 14.54) in mode 1 compared to modes 2 (for caption M = 101.80, SD = 17.08; for scene M = 207.00, SD = 23.92) and 3 (for caption M = 89.20, SD = 25.90; for scene M = 210.80, SD = 28.05) see Figure 18(a). A similar pattern is observed when analysing dwell time. On average, participants dwelled more on captions than on the visual scene in mode 1 than in modes 2 or 3, see Figure 18(b). Participants appeared to allocate more attention to captions when viewing subtitles in mode 1 than in modes 2 or 3.

Figure 18

Visual attention distribution over captions and visual scenes while watching video with different types of subtitles



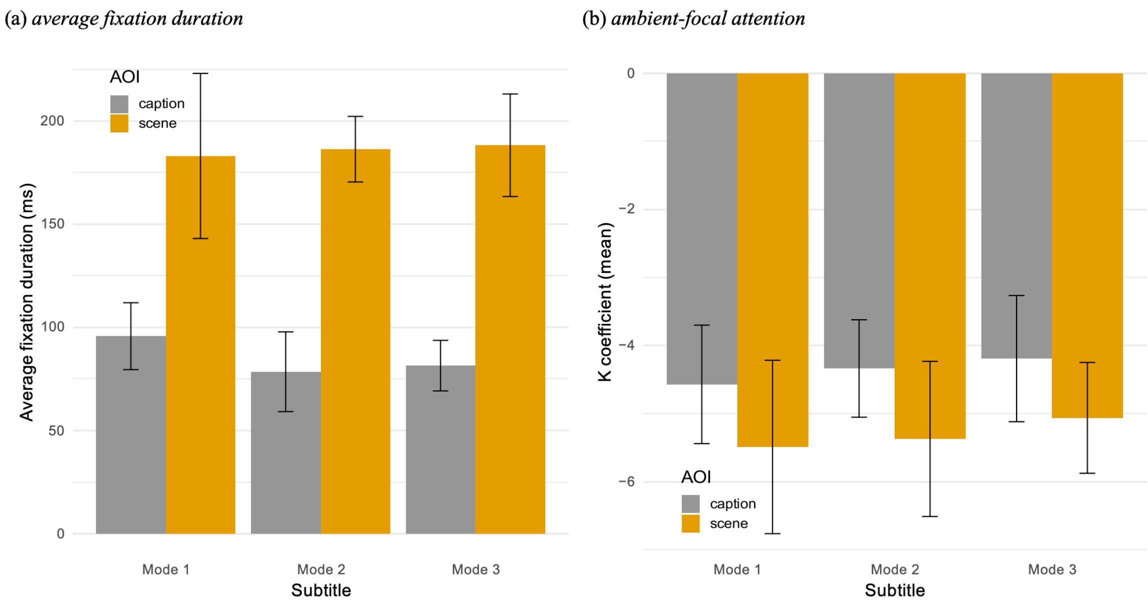
Note. Attention distribution is depicted by two metrics: (a) shows fixation counts over captions and visual scene, (b) shows dwell time of captions and visual scene fixating. Bars height represents mean values and whiskers represent ±1*SD*.

We also examined cognitive effort while processing information from captions or visual scene based on average fixation duration (following the *eye-mind assumption* (Just & Carpenter, 1976)) and focus of attention with coefficient *K* (Krejtz et al., 2016), which captures the temporal relation between fixation duration and subsequent saccade amplitude. $K > 0$ indicates focal viewing while $K < 0$ suggests ambient viewing. Focal attention is usually related to higher cognitive effort when processing complex visual or text stimuli (Duchowski et al., 2020; Krejtz et al., 2017; Krejtz et al., 2018). Analysis of descriptive statistics on average fixation duration showed that the visual scene triggered longer average fixation durations than captions in all modes. However, the difference in average fixation durations between visual scene and caption is smallest in mode 1.

Moreover, fixation duration on subtitles in mode 1 (M = 95.73, SD = 16.22) is much longer than on subtitles in either mode 2 (M = 78.47, SD = 19.35) or mode 3, (M = 81.43, SD = 12.24) see Figure 19(a). Coefficient K showed that viewers were not as focused when reading captions in mode 1 ($M = -4.57$, $SD = 0.87$) compared to mode 2 ($M = -4.34$, $SD = 0.72$) or 3 ($M = 4.19$, $SD = 0.92$), see Figure 19(b). Both fixation duration and coefficient K suggest highest cognitive effort along with decreased focal processing when processing subtitles in mode 1.

Figure 19

Cognitive processing of textual and visual information from captions and visual scenes while watching video with different types of subtitles: (a) shows average fixation duration as a metric for cognitive effort, (b) shows K coefficient as a metric ambient-focal attention.



Note. Bars height represents mean values and whiskers represent ±1*SD*.

4.4.5. Focus group insights

A qualitative analysis was carried out on the notes taken during the focus group (the last part of the session, after the participants had watched the video with the different subtitle options). The notes were thoroughly revised and tagged using Atlas.ti. This procedure allowed us to identify three areas that can be associated with the quantitative analysis (subtitle readability, task load, and movie content understanding). The analysis also allowed us to define user preferences and identify aspects on which there was consensus among users and issues on which opinions diverged.

In terms of preference, most of the participants agreed that mode 2 was the easiest to read with one participant suggesting adding a colour code to mode 2 (like mode 3). The second preferred option was mode 1 (selected by 2 participants). The main problem in mode 1 seems to have been the difficulty in identifying the character speaking at each given moment.

Regarding creative subtitles, all the participants agreed it is a great idea to dramatise what is said and that this can add much visual beauty to the content. One of the participants noted that, in some cases, so much creativity distracted her from the content itself.

As in previous studies reviewed earlier, participants highlighted the lack of direction to guide people to the source of the sound (guiding mechanisms). Some of them mentioned that they missed human interaction when watching the immersive content, and that they felt isolated when wearing the HMD for the first time.

4.5. Discussion

The first hypothesis we wanted to validate was whether fixed, near to the mouth subtitles allow viewers to spend more time exploring the image instead of reading the subtitles than head-locked subtitles do. Although the present pilot study cannot yield conclusive evidence, eye movement data together with focus group insights seem to support this hypothesis. Interestingly, eye movement data appear to be consistent with qualitative insights from the focus group, suggesting that participants tend to prefer fixed subtitles near to the mouth of the speaking character (mode 2). These results differ from those obtained in previous studies, in which participants opted for head-locked subtitles.

The results of self-reported cognitive load during movie watching with different subtitle modes suggest a slight preference (less perceived mental effort and higher perceived success in reading captions) for mode 1 (b&w font for all characters, grey background and head-locked) over modes 2 and 3. However, results carry a large statistical variance and cannot be interpreted decisively. The results may also be biased by a lack of randomisation in order of presentation and learning effect of the questions during the experimental procedure. Future studies must employ tighter experimental control over stimulus presentation order (e.g. via randomisation or counterbalancing).

Eye movement analysis sheds light on attention allocation (captions vs. scene) and perception. Identification of fixations showed that participants allocated more attention to captions and less to the visual scene when viewing subtitles in mode 1 than in modes 2 or 3. Process measures (average fixation and ambient-focal coefficient) suggest higher cognitive effort paired with the less focal processing of subtitles in mode 1.

Subtitles in modes 2 and 3 appear to outperform mode 1 as they may be less distracting from scenes in the movie, but they also seem to require less cognitive effort when focused on reading. We do not know, however, whether mode 2 or 3 is easier to read and less distracting when movie watching. This issue needs to be addressed in a study with more experimental control and a larger sample.

The visualisation of the eye movements analysed, specifically saccades, drawn in red in Figure 20, expose the inadequacy of the velocity-based filtering approach. The I-VT method, while computationally efficient and generally applicable to traditional desktop displays, tends to ignore head-induced gaze movement when captured in the VR HMD. It is likely that a better model of eye and head coupling is required (Guitton et al., 1990), e.g., a fixation detection algorithm suitable for immersive environments (Llanes-Jurado et al., 2020).

Figure 20

The view and scanpath for each participant (rows: participant 1-5, columns: mode 1-3, 00:22)

Note: General observations can be drawn, such as that participant 1, although finding the captions in modes 1 and 3, was lost in mode 2 and can be observed saccading between the mouths of the wrong characters, trying to identify the character speaking. The second participant can be seen fixating on the speaking character's mouth rather than reading the caption in modes 1 and 2. Also Participants 3, 4 and 5 can be observed reading the captions in modes 1 and 2, but not mode 3.

4.6. Conclusions

Immersive environments simulate reality to heighten the immersive experience. Attention should be paid when drawing on results from previous studies on subtitle reading performance in 2D to VR environments. Prior research on evaluation of subtitles has mainly focused on subtitle style, speed of display, and positioning and has largely been qualitative. In VR we still need to define the subtitle research challenges. Investigating how subtitles are read is one logical quest but finding subtitles when two people are interacting from different fields of vision is also a candidate for testing. The tests carried out in this study have shown that we need some basic understanding of media presentation in VR and user behaviour and habits when consuming media in VR, where the narrative is no longer lineal. VR media environments present us with new variables to consider when testing for optimal subtitle presentation. The objective is to find the least disruptive subtitle reading experience for protecting immersivity and the simulation of reality. To understand the visual presentation of subtitles in VR, a framework was developed along three basic presentation modes for 360º videos, which were all piloted. Our contribution is thus two-fold: on the one hand, there is our presentation of the VR subtitle framework and, on the other hand, a new method for triangulation of psycho-physiological (eye movements) self-reports and qualitative (focus group discussions) analyses. To the best of our knowledge, this is the first attempt to advance these two directions when discussing subtitle presentation in VR 360º videos.

Immersive environments need new subtitle presentation modes and reading patterns. This article described the first pilot study using a comprehensive methodological environment to test subtitles in immersive environments. The novel testbed includes a subtitle editor and a VR system designed specifically to collect eye movement data as visual attention is distributed over 360º videos containing subtitles. Both the framework and the methodology tested in this first pilot study can be used to collect quantitative and qualitative behavioural data when viewing subtitled 360º media. Future studies involving more participants are expected to yield new insights and lead to subtitle standardisation in immersive media environments.

References

Agulló, B., & Matamala, A. (2019). Subtitling for the deaf and hard-of-hearing in immersive environments: Results from a focus group. *Journal of Specialised Translation, 32.*

Agulló, B., Montagud, M., & Fraile, I. (2019). Making interaction with virtual reality accessible: Rendering and guiding methods for subtitles. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 33*(4), 416–428. https://doi.org/10.1017/ S0890060419000362

Bartoll, E. (2004). Parameters for the classification of subtitles. In P. Orero (Ed.). John Benjamins. https://doi. org/10.1075/btl.56.08bar

Brescia-Zapata, M. (2022). The present and future of accessibility services in vr360 players. *inTRAlinea, 24.*

Brown, A. (2017). User testing subtitles for 360° content. https://www.bbc.co.uk/rd/blog/2017-10-subtitles- 360-video-virtual-reality-vr

Brown, A., & Patterson, J. (2017). Designing subtitles for 360º content. https://www.bbc.co.uk/rd/blog/2017- 03-subtitles-360-video-virtual-reality

Brown, A., Turner, J., Patterson, J., Schmitz, A., Armstrong, M., & Glancy, M. (2017). Subtitles in 360-degree video. *Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video*, 3–8. https://doi.org/10.1145/3084289.3089915

Bryman, A. (2004). *Social research methods*. Oxford University Press.

Cantero de Julián, J. I., Calvo Rubio, L. M., & Benedicto Solsona, M. Á. (2020). La tenue apuesta por los vídeos en 360º en las estrategias transmedia de las televisiones autonómicas españolas. *Revista Latina de Comunicación Social*, (75), 415–433. https://doi.org/https://doi.org/10.4185/RLCS-2020-1433

Díaz Cintas, J., & Remael, A. (2007). *Audiovisual translation: Subtitling*. St. Jerome.

Duchowski, A. T., Krejtz, K., Zurawska, J., & House, D. H. (2020). Using microsaccades to estimate task difficulty during visual search of layered surfaces. *IEEE Transactions on Visualization and Computer Graphics, 26*(9), 2904–2918. https://doi .org/10.1109/TVCG.2019.2901881

Foerster, A. (2010). Towards a creative approach in subtitling: A case study. In J. Díaz-Cintas, A. Matamala, & J. Neves (Eds.), *New insights into audiovisual translation and media accessibility: Media for all 2* (pp. 81–98). Rodopi.

Fox, W. (2018). Can integrated titles improve the viewing experience? Investigating the impact of subtitling on the reception and enjoyment of film using eye-tracking and questionnaire data. Language Science Press. https://doi.org/10.5281/zenodo.1180721

Gottlieb, H. (1995). Establishing a framework for a typology of subtitle reading strategies - viewer reactions to deviations from subtitling standards. *Communication Audiovisuelle et Transferts Linguistiques - FIT Newsletter,* (14/3-4), 388–409.

Guitton, D., Munoz, D. P., & Galiana, H. L. (1990). Gaze Control in the Cat: Studies and Modeling of the Coupling Between Orienting Eye and Head Movements in Different Behavioral Tasks. *Journal of Neurophysiology, 64*(2), 509–531. https://doi.org/10.1152/jn.1990.64.2.509

Hughes, C., Armstrong, M., Jones, R., & Crabb, M. (2015). Responsive design for personalised subtitles. *Proceedings of the 12th International Web for All Conference*, 1–4. https://doi.org/10.1145/2745555.2746650

Hughes, C., Brescia-Zapata, M., Johnston, M., & Orero, P. (2020a). Immersive captioning: Developing a framework for evaluating user needs. *IEEE AIVR 2020: 3rd International Conference on Artificial Intelligence & Virtual Reality 2020*. http://usir.salford.ac.uk/id/eprint/58518/

Hughes, C., & Montagud, M. (2020b). Accessibility in 360° video players. *Multimedia Tools and Applications*, 1–28. https://doi.org/10.1007/s11042-020-10088-0

Hughes, C., Montagud Climent, M., & tho Pesch, P. (2019). Disruptive approaches for subtitling in immersive environments. *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video*, 216–229. https://doi.org/10.1145/3317697.3325123

Jenkins, H., Ford, S., & Green, J. (2015). Cultura transmedia. la creación de contenido y valor en una cultura en red. Gedisa.

Just, M. A., & Carpenter, P. A. (1976). The role of eye- fixation research in cognitive psychology. *Behavior Research Methods & Instrumentation, 8*(2), 139–143. https://doi.org/10.3758/BF03201761

Krejtz, K., Çöltekin, A., Duchowski, A., & Niedzielska, A. (2017). Using Coefficient K to Distinguish Ambient/Focal Visual Attention During Map Viewing. *Journal of Eye Movement Research, 10*(2). https://doi.org/10.16910/jemr.10.2.3

Krejtz, K., Duchowski, A., Krejtz, I., Szarkowska, A., & Kopacz, A. (2016). Discerning Ambient/Focal At- tention with Coefficient K. *ACM Transactions on Applied Perception, 13*(3), 11:1–11:20. https://doi. org/10.1145/2896452

Krejtz, K., Wisiecka, K., Krejtz, I., Holas, P., Olszanowski, M., & Duchowski, A. T. (2018). Dynamics of Emotional Facial Expression Recognition in Individuals with Social Anxiety. *Proceedings of the 2018 ACM Symposium on Eye-tracking Research & Applications*, 43:1–43:9. https://doi.org/10.1145/3204493.3204533

Lee, D. G., Fels, D. I., & Udo, J. P. (2007). *Emotive captioning. Comput. Entertain*, *5*(2). https://doi.org/10.1145/1279540.1279551

Llanes-Jurado, J., Marín-Morales, J., Guixeres, J., & Alcañiz, M. (2020). Development and Calibration of an Eye-Tracking Fixation Identification Algorithm for Immersive Virtual Reality. *Sensors, 20*(17). https://doi.org/10.3390/s20174956

Mas Manchón, L., & Orero, P. (2018). Usability tests for personalised subtitles. *Translation spaces, 7*(2), 263–284. https://doi.org/10.1075/ts.18016.man

Matamala, A., & Orero, P. (2018). Standardising accessibility: Transferring knowledge to society. *Journal of Audiovisual Translation, 1,* 139–154. https://doi.org/10.47476/jat.v1i1.49

McClarty, R. (2012). Towards a multidisciplinary approach in creative subtitling. *MonTI: Monografías de Traducción e Interpretación*, 133–153. https://doi.org/10.6035/MonTI.2012.4.6

McClarty, R. (2014). In support of creative subtitling: Contemporary context and theoretical framework. *Perspectives, 22*(4), 592–606. https://doi.org/10.1080/0907676X.2013.842258

Mével, P. A. (2020). Accessible paratext: Actively engaging (with) d/deaf audiences. *Punctum. International Journal of Semiotics, 6*(01). http://doi.org/10.18680/hss.2020.0010

Mittell, J. (2009). *Forensic fandom and the drillable text.* Spreadable media. https://spreadablemedia.org/essays/mittell/index.html#.YgzhV-5_o4g

Montagud, M., Fraile, I., Meyerson, E., Genís, M., & Fernández, S. (2019). Imac player: Enabling a personalized consumption of accessible immersive contents. https://doi.org/10.6084/m9.figshare.9879254.v1

Montagud, M., Orero, P., & Matamala, A. (2020). Culture 4 all: Accessibility-enabled cultural experiences through immersive vr360 content. *Personal and Ubiquitous Computing, 24*(6), 887–905. https://doi.org/10.1007/s00779-019-01357-3

Nyström, M., & Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eye-tracking data. *Behaviour Research Methods, 42*(1), 188–204. https://doi.org/10.3758/BRM.42.1.188

Orero, P., Hughes, C. J., & Brescia-Zapata, M. (2020). Evaluating subtitle readability in media immersive environments. DSAI 2020: 9th International Conference on Software development and Technologies for Enhancing Accessibility and Fighting Info-exclusion, 51–54. https://doi.org/10.1145/3439231.3440602

Romero-Fresco, P. (2013). Accessible filmmaking: Joining the dots between audiovisual translation, accessibility and filmmaking. *The Journal of Specialised Translation, 20*, 201–223.

Romero-Fresco, P. (2015). *The Reception of Subtitles for the Deaf and Hard of Hearing in Europe.* Peter Lang. https://www.peterlang.com/view/title/36324

Rothe, S., Tran, K., & Hussmann, H. (2018). Positioning of subtitles in cinematic virtual reality. https://doi.org/ 10.2312/egve.20181307

Rupp, M., Kozachuk, J., Michaelis, J., Odette, K., Smither, J., & McConnell, D. (2016). The effects of immersiveness and future vr expectations on subjective- experiences during an educational 360 video. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 60,* 2108–2112. https://doi.org/10.1177/1541931213601477

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the 2000 Symposium on Eye-Tracking Research & Applications*, 71–78. https://doi.org/10. 1145/355017.355028

Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry, 36*(8), 1627–1639. http://pubs.acs.org/doi/abs/10.1021/ac60214a047

Skult, N., & Smed, J. (2020). Interactive storytelling in extended reality: Concepts for the design. In B. Bostan (Ed.), *Game user experience and player-centered design* (pp. 449–467). Springer International Publishing. https://doi.org/10.1007/978-3-030-37643-7_21

Slater, M., & Wilbur, S. (1997). A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators & Virtual Environments, 6,* 603–616. https://doi.org/10.1162/pres.1997.6.6.603

Smith, W. (2015). Stop calling google cardboard's 360-degree video 'vr'. *Wired.* https://www.wired.com/ 2015/11/360-video-isnt-virtual-reality/

CHAPTER 5

Article 4: Subtitles in 360º video. Results from an eye-tracking experiment

Subtitles in VR 360º Video. Results from an Eye-Tracking Experiment

Marta Brescia Zapata

Universitat Autònoma de Barcelona, Barcelona, Spain

Krzysztof Krejtz

SWPS University of Social Sciences and Humanities, Warsaw, Poland

Pilar Orero

Universitat Autònoma de Barcelona, Barcelona, Spain

Andrew T Duchowski

Clemson University, USA

Chris J Hughes

University of Salford, Salford, UK

Abstract

Virtual and Augmented Reality, collectively known as eXtended Reality, are key technologies for the next generation of human–computer–human interaction. In this context, 360º videos are becoming ubiquitous and especially suitable for providing immersive experiences thanks to the proliferation of affordable devices. This new medium has an untapped potential for the inclusion of modern subtitles to foster media content accessibility (Gejrot et al., 2021), e.g., for the deaf or hard-of-hearing people, and to also promote cultural inclusivity via language translation (Orero, 2022). Prior research on the presentation of subtitles in 360º videos relied on subjective methods and involved a small number of participants (Brown et al., 2018; Agulló, 2019; Oncins, 2020), leading to inconclusive results.

The aim of this paper is to compare two conditions of subtitles in 360º videos: position (head-locked vs fixed) and colour (monochrome vs colour). Empirical analysis relies on novel triangulation of data from three complementary methods: psycho-physiological attentional process measures (eye movements), performance measures (media content comprehension), and subjective task-load and preferences (self-report measures). Results show that head-locked coloured subtitles are the preferred option.

Keywords

Immersive environments, 360º video, media accessibility, subtitles, captions, eye-tracking

5.1. Introduction

Immersive content is meant to give the user the illusion of being 'physically present' (Slater & Wilbur, 1997), and can provide benefits in a variety of sectors, such as entertainment, communication, learning, arts, and culture (Liberatore & Wagner, 2021; Montagud-Climent et al., 2020). 360º videos—also known as immersive or VR360 videos—offer great potential in providing engaging media experiences. Already in 2017, 49 % of the public broadcasters who responded to the European Broadcasting Union (EBU, 2017) report on the use of VR declared offering 360º content. The most popular devices to access this immersive content are head-mounted displays (HMD). According to the newsletter XR today (Greener, 2022), 'in 2020, roughly 57.4 people owned a VR headset in the US, although in 2022, this figure increased by 37.7 million.' The demand for VR headsets is expected to increase due to the adoption of VR technology in enterprise, industry, and education sectors.

These new immersive media environments must be accessible for all to fulfil existing accessibility legislation in most world regions. Standards such as the EN301459 recommend Universal Design when developing any system or product. This requirement is becoming mainstream with advice from the United Nations Convention on the Rights of Persons with Disabilities (CRPD, 2006), and now it is also an issue of political will and moral obligation, thanks to the Audiovisual Media Services Directive (AVMSD) and the European Accessibility Act (EAA). These pieces of legislation adopted a user-centric approach, a method that is at the heart of Human Rights towards full democratic participation in society by all, which, in the twenty-first century, depends on access to media. Following the aforementioned legislation and academic research (Romero-Fresco, 2013; Udo & Fels, 2010) any system or process should be designed with accessibility in mind from the outset, leading to a born accessible system that avoids expensive and complex afterthought solutions.

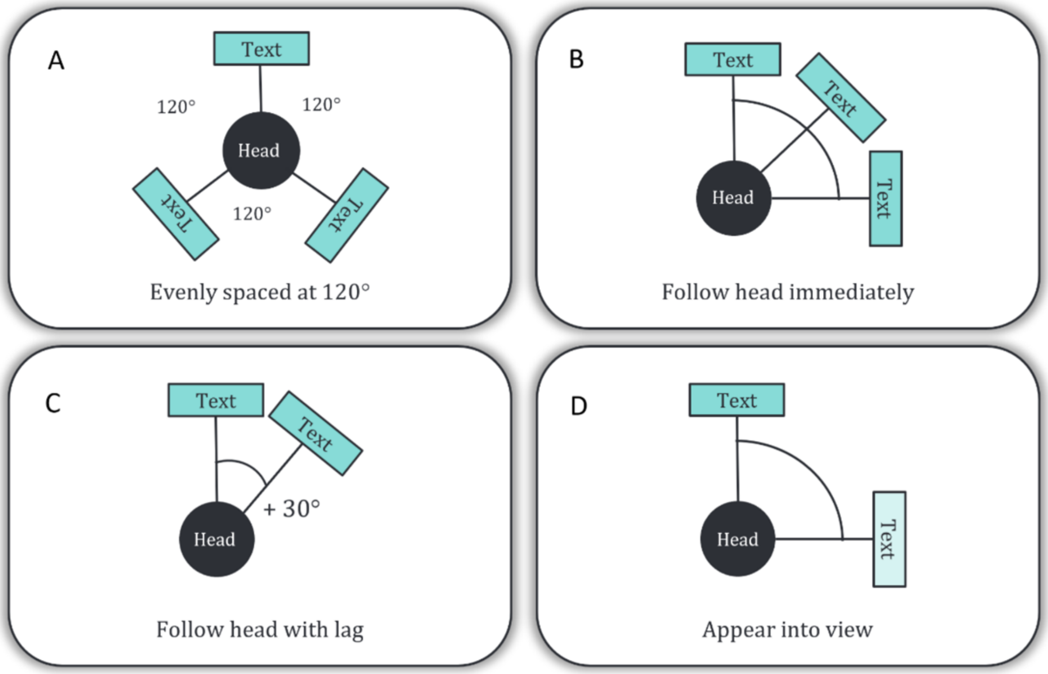
This paper focuses on subtitling, where standardised practices have been adopted in the context of 2D non-immersive media (ISO 20071:23; Matamala & Orero, 2018). Although the trend towards mixed methods in translation accessibility studies is becoming more popular, publications still fail to discuss the mixed-method nature of the study in depth (Hermosa-Ramírez, 2022). Prior reception studies on the evaluation of subtitles in 360º have largely been based on subjective measures, using questionnaires, interviews, and focus groups (Agulló & Orero, 2017; Brown et al., 2018; Rothe et al., 2018; Fidyka & Matamala, 2018; Agulló & Matamala, 2020). Objective, psychophysiological measures (such as eye movements, heart rate, and electrodermal activity) have been largely adopted in the context of tests for 2D non-immersive subtitles (Krejtz et al., 2016; Kruger, 2016; Liao et al., 2021; Szarkowska & Gerber-Morón, 2019), with only a recent contribution by Ibourk and Al-Aldwan (2019), who included for the first time eye-tracking technology in a small study on immersive subtitling. The present study adopts a mixed-method design, with psycho-physiological process metrics (eye movements), performance metrics (scene comprehension), and subjective self-reports (task-load and preferences), followed by result triangulation. In this article, a critical review of previous studies related to subtitling in immersive media is presented, followed by the characteristics of the study and the methodology. To finish, the results, a discussion, and conclusions are presented.

5.2. Background

Studies on subtitles in VR are relatively recent. Understanding the displaying mode for subtitles started at the British Broadcasting Corporation (BBC), which first identified the main challenges when developing subtitles for immersive content (Brown et al., 2017). As a result, four solutions for subtitle rendering were designed (Brown, 2017) and tested with 24 hearing participants (Brown et al., 2018). These four conditions were (a) fixed positioned, evenly spaced, (b) follow head immediately (also called Always Visible), (c) follow with lag, and (d) appear in front, then fixed position (see Figure 21). Users reported a preference for subtitles which are always presented in front of the viewer, following head movements: Always Visible subtitles.

Figure 21

Four static subtitle solutions designed and tested by Brown et al. (2018)



Note. Author own elaboration based on Brown et al. (2018)

Rothe et al. (2018) compared Brown et al. (2018) Always Visible subtitles with a new type called dynamic subtitles: placed near the speaker. They go a step further by measuring simulator sickness and task workload. They tested with 34 hearing participants and received feedback from one deaf participant (although this data was not included in their analysis). This study obtained similar results to those of its predecessor and pointed out that "additional usage of an eye tracker could lead to more detailed results in the analysis of the viewing direction" (Rothe et al., 2018, p. 214).

The H2020 European-funded project ImAc23, based on the results of Brown et al. (2018) and Rothe et al. (2018), added location information within the 3D space to each subtitle (Agulló & Matamala, 2019). Results from Agulló and Matamala (2019) helped to draft user requirements, with a priority for designing a guiding mechanism to help deaf and hard-of-hearing users to locate the sound source related to the subtitle. The project designed and tested several guiding mechanisms with six hearing and two deaf participants (Agulló, Montagud, & Fraile, 2019). Results showed two preferred guiding methods: an arrow positioned to the left or right of the

23 https://www.imacproject.eu/

subtitle directing the user to the sound source, and a radar circle shown in the user field of view that identifies both the position of the sound source and the relative viewing angle of the user. Guiding mechanisms can also be found in videogames, such as The Last of Us: Part II (Myers, 2020), which "includes a guide arrow to direct the user to the location of the character speaking" (Hughes et al., 2020). It is worth noting that the replication of the Agulló at al. (2019) study with a larger number of participants (N = 40: 27 hearing, 20 with hearing loss, six hearing impaired and seven deaf) showed that always-visible subtitles with arrows were the preferred option (Agulló & Matamala, 2020).

On the one hand, all studies to this point conclude that the preferred visualisation mode is head-locked, centred, bottom subtitle: a trend that replicates conventions established for 2D, non-immersive content as defined in ISO/IEC 20071-23:2018 (ISO, 2018) or UNE-153010:2012 (AENOR, 2003), among others. On the other hand, all studies focus on two subtitle features: sound source and subtitle position. All previous studies have three limitations: (a) demography samples are small in number, (b) the methodology is based solely on subjective opinions and answers (in-depth interviews, focus groups, and/or questionnaires), and (c) unstandardised stimuli videos in terms of length of video, language of dialogue in video, language of users, etc.

The most recent work in this field has adopted a new web-based prototyping framework (Hughes et al., 2020). The framework allows for real time subtitle editing and visualisation in 360º videos. It takes into consideration all the previously reported work and is based on two mechanisms for subtitle rendering: (a) head-locked, where the subtitle is rendered relative to the user's viewpoint, and (b) fixed, where the subtitle is rendered relative to a fixed location in the world, usually close to the speaking character. This framework offers the opportunity of instant evaluation which was tested by Brescia-Zapata et al. (2022) in a pilot study leading to a full experiment, the results of which are presented in this paper. The experiment received ethics clearance from the authors' home institutions, according to the ethics and privacy regulations of the H2020 EU funded TRACTION project24.

The present study was conceived to further clarify which is the best visualisation mode for subtitles in immersive environments for all kinds of users. The aim is two-fold: (a) to detail the design of controlled experiments testing the influence of subtitles on users' attention allocation,

[24] https://www.traction-project.eu/

and immersive content comprehension, and (b) to gather feedback regarding preferences of two characteristics of subtitles in immersive content: position and colour. Special attention has been put on the following methodological aspects: (a) diverse sample in terms of demography and subtitle usage habits (b) a triangulation of psychophysiological, qualitative and questionnaire methods, and (c) controlled stimuli. It focuses on two hypotheses: the first is related to subtitle positioning and the second is related to subtitle colour.

H1: Head-locked subtitles are easier to follow but more intrusive than fixed subtitles.

H2: Coloured subtitles are helpful to identify a speaking character in a 360º narrative.

5.3. Method

5.3.1. Experimental design and independent variables

To test the research hypotheses, the experimental study was conducted with 2 x 2 mixed design with two independent variables (IV): subtitle position (head-locked vs fixed) and subtitle colour (monochrome vs colour). The first one was treated as a between-subject IV and the later one as a within-subjects IV. The order of the experimental conditions was counterbalanced.

The two conditions for subtitle position were the following.

(a) Head-locked subtitles which are always visible and are displayed in front of the viewer (see Figure 22), following head movements. Head-locked subtitles are equivalent to static-follow in Brown et al. (2017), to static subtitles in Rothe et al. (2018), and to always-visible in Agulló and Matamala (2020).

Figure 22

Head-locked subtitles attached to the FoV in Vacations video stimuli

(b) Fixed subtitles which appear near to the speaking characters and remain fixed to the scene (Figure 23). Fixed subtitles are equivalent to 120 degrees in Brown et al. (2017), and to dynamic subtitles in Rothe et al. (2018).

Figure 23

Fixed subtitles attached to one position in the sphere in Vacations video stimuli

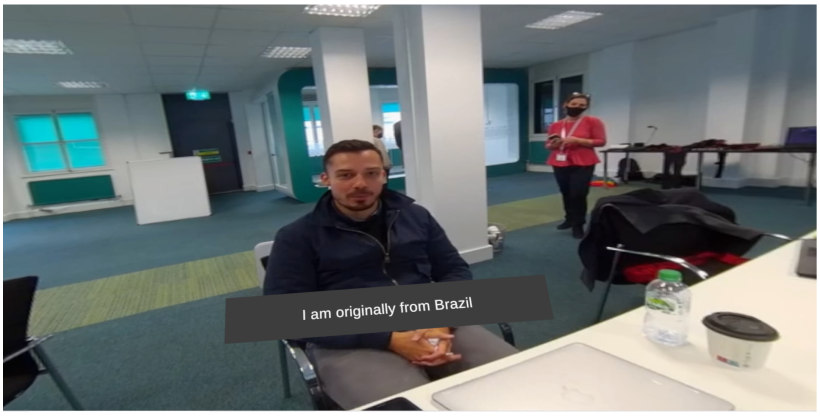
The main reason for testing these two conditions is that results from previous studies (Brown et al., 2018; Rothe et al., 2018; Agulló et al., 2018) only included qualitative data obtained through questionnaires and focus groups.

The two conditions for subtitle colour were the following.

(a) Monochrome (see Figure 24). All the subtitles for each character were in the same colour white on grey background.

Figure 24

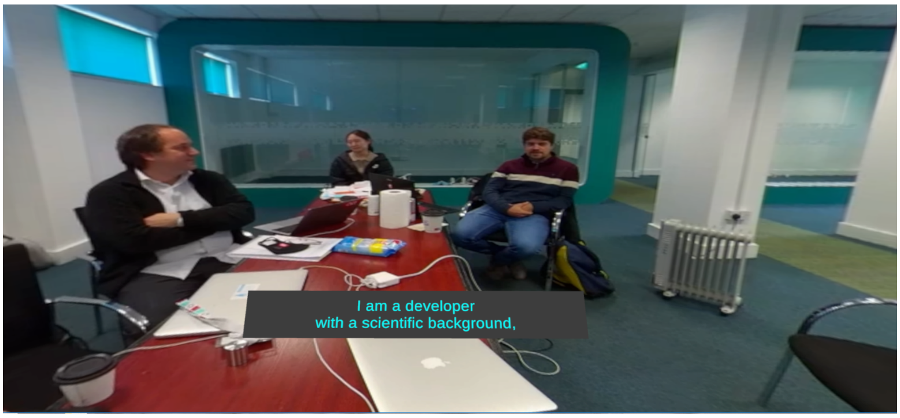
Monochrome subtitles in TRACTION video stimuli



(b) Colour (see Figure 25). Four colours have been used to identify the different characters: white, yellow, cyan, and green. The colour coding follows the ISO 20071-23:2018 standard.

Figure 25

Coloured subtitles in TRACTION video stimuli



The objective of examining these two conditions is to evaluate the effectiveness of colour coding, which is suggested by certain guidelines, such as BBC (2022) as a means of distinguishing between speakers and has demonstrated an enhancement in the immersive experience of multilingual 2D content (Szarkowska & Boczkowska, 2022). However, despite its potential benefits, this feature has not been treated as a variable in previous studies on immersive content (Brown et al., 2018; Rothe et al., 2018; Agulló et al., 2018).

5.3.2. Participants

In total, 73 volunteers took part in the experiment at three locations: $N = 24$ in Universitat Autònoma de Barcelona (Spain), $N = 24$ in Salford University (UK), and $N = 25$ in SWPS University of Social Sciences and Humanities (Poland). Participants received no incentives for the study. There were no significant differences between the countries in age of the participants $F(2,69) = 2.26$, $p > 0.1$ (see Table 4).

Table 4

Participants' demographic characteristics, attitude, and usage of VR technology in each location

	Spain	UK	Poland
Number (*N*)	24	24	25
Age	(*M*=33.96, *SD*=11.18)	(*M*=29.41, *SD*=10.34)	(*M*=27.40, *SD*=6.68)
Gender	17 female	4 female	19 female
Age difference between genders	*t*(21)=0.52, *p*=0.61	*t*(22)=0.70, p=0.49	*t*(23)=0.94, *p*=0.36
Occupation	Students (N=11); Researchers/lecturers (N=11); Practitioners (N=2)	Students (N=18); Researchers/lecturers (N=3); Data Analysts/engineers (N=4)	Students (N=16); Researchers/lecturer s (N=4); Journalist (N=2); Analyst (N=1)
Vision	Corrected (*N*=12), Uncorrected (N=12)	Corrected (*N*=12), Uncorrected (N=15)	Corrected (N=7), Uncorrected (N=18)
Handedness	Right (*N*=24)	Right (*N*=21)	Right (N=25)
VR interest	(*M*=4.13, *SD*=0.76)	(*M* = 3.96, *SD* = 0.65)	(*M* = 3.00, *SD* = 1.22)
VR experience	(*M*=0.87, *SD*=1.18)	(*M*=0.81, *SD*=1.24)	(*M*=1.24, *SD*=2.01)
Digital device daily usage	(*M*=3, *SD*=1.38)	(*M*=3.59, *SD*=1.28)	(*M*=2.91, *SD*=0.91)
Attitude to subtitles (answers to "I always turn subtitles on")	*(M*=3.97, *SD*=1.16)	*(M*=3.88, *SD*=1.32)	*(M*=4.01, *SD*=0.97)

The three locations of the study were chosen in order to ensure the robustness of the obtained results, as each country has a different language translation tradition. Spain has traditionally been considered a dubbing country (Ballester Casado, 1998; Chaume, 2012; Gil Ariza, 2004), although the presence of subtitles has increased in recent years (Matamala et al., 2017). Poland is generally considered a stronghold of voice over (Gottlieb, 1998; Valdeón, 2022), coexisting with other audiovisual translation modes, such as dubbing, subtitling, audio description and subtitling for the deaf and the hard-of-hearing (Szarkowska, 2009). The UK belongs to a large anglophone audiovisual market and is neither a classical "subtitling" nor "dubbing" country (Luyken, 1991), as most audiovisual content is produced in English.

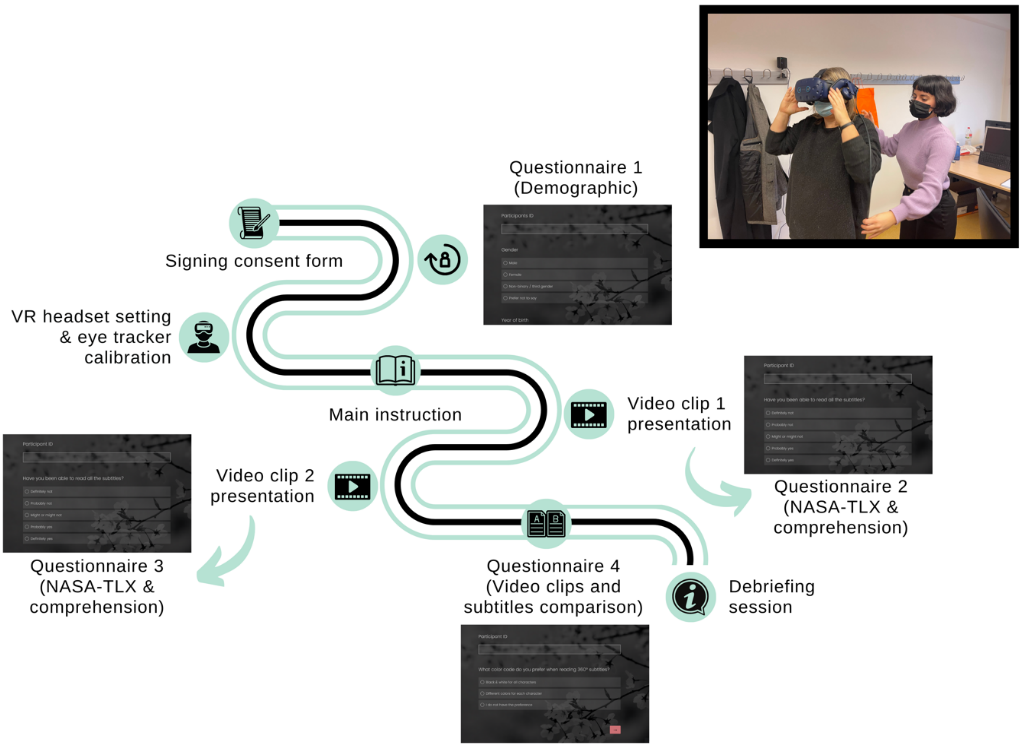
All participants in the sample had above average reading skills. We can also consider them as digital media savvy based on their declared daily usage of digital media and devices although most participants had never experienced VR content before (see Table 4).

5.3.3. Experimental procedure

The experimental procedure was the same in the three locations of the present study (see Figure 26).

Figure 26

Experimental procedure schematic timeline and settings (top-right)



The experimental procedure was executed individually in the laboratory facilities. After signing the consent form, participants started with the demographic and usage and attitudes questionnaire. Next, they were familiarised with the VR headset. The HMD built-in eye tracker was calibrated. The following main part of the experiment consisted of three steps:

1. watching first video stimuli followed by a comprehension and NASA-TLX questionnaire;

2. watching second video stimuli followed by comprehension and NASA-TLX questionnaire; and

3. a questionnaire comparing subtitles in both viewed videos.

After the experiment participants were thanked and debriefed. A facilitator was present during the experiment providing information as required, helping the participants to correctly wear the HMD, and assisting with the completion of the questionnaires (see Figure 6, top-right). All the questionnaires were formulated using Qualtrics system and completed using a desktop computer from the laboratory. The order of video stimuli presentation was counterbalanced; thus participants were presented with the two videos in random order, one with monochrome and the other with coloured subtitles, but the same subtitle position in both cases (either head-locked or fixed). During video stimuli presentation the participants' eye movements were recorded.

5.3.4. Experimental materials

5.3.4.1. Stimuli video

For the present study two custom 360º stimuli videos were recorded using an Insta360 One X2 camera. The first is called *Vacations* and was of a family, speaking Arabic, discussing their vacation plans. In the second stimuli video, called *TRACTION*, a group of researchers introduce themselves, each speaking in their own language (Spanish, Korean, Catalan, Portuguese, and English). *Vacations* is one min and 18 s long, and *TRACTION* is 2 min and 32 s long. Both videos followed the same features.

1. Speaking characters were placed all around the scene, exploiting 360º movement and immersion,
2. there were no overlaps between speaking characters. If two characters located at different points in the 360º scene spoke at once, it would be almost impossible for the user to read the subtitles,
3. language spoken was different to that of participants. This feature was added after the pilot study (Brescia-Zapata et al., 2022), as during the data collection, several participants claimed that they did not need to read the subtitles because they understood the original audio.

Both videos were subtitled in English, Spanish, and Polish to meet the experimental design (two colour schemes, two positioning schemes). The number of subtitles is the same in all languages: in *Vacations* there are 27 subtitles and in *TRACTION* 39. Participants at UAB were presented with video stimuli subtitles in Spanish, participants at Salford University were presented with video stimuli subtitled in English, and participants at Warsaw University were presented with video stimuli in Polish. Subtitles followed the ISO 20071-23:2018 standard and the font type selected was Liberation Sans. Another feature that differentiates the pilot study from this experiment is the grey background box used in both videos. Although some previous studies proved that it obstructs important parts of the image (Brown et al., 2018), results from the pilot test revealed that "some of the participants found subtitles with no background hard to read" (Brescia-Zapata et al., 2022).

5.3.4.2. Questionnaires

Four questionnaires were designed to gather information about demographics, task-load, comprehension, and preference. The first questionnaire had 20 questions to gather demographic data (sex, age, handedness, occupation) and data related to usage and attitudes towards digital media (e.g., "Which of these devices do you use on a daily basis?"), VR (e.g., "How often do you watch virtual reality (VR) content for instance, 360º videos?"), and subtitles (e.g., "When subtitles are available, do you turn them on?", "How many hours a day do you watch subtitled content?")

After each stimuli video, participants filled out a questionnaire with six multiple choice questions on content comprehension. These questions were customised for each stimuli video. After comprehension questions, participants completed a task-load index (NASA-TLX) (Hart, 2006) consisting of six questions on personal evaluation of task load (e.g., mental load, physical demands, time pressure, etc).

After viewing two stimuli videos and completing the subsequent questionnaires, participants answered three questions on subtitles preferences ("What colour code do you prefer when reading 360º subtitles?", "Explain in your own words why you prefer the option selected in the previous question", and "Do you think that the subtitles obstruct important parts of the image?")

5.3.5. Experimental equipment

The experimental procedure was prepared with a framework developed by Hughes et al. (2020) and ported to Unity 3D. We used the HTC Vive Pro Eye headset, which contains a Tobii eye tracker built-in with 120 *Hz* sampling rate. Vive's eye tracking accuracy estimation is 0.5º–1.1º

(Sipatchin et al., 2020). The HMD has two AMOLED screens, with a resolution of 2,880 x 1,600 pixels in total, a display refresh rate of 90 Hz, and a FoV of 110º.

5.4. Results

Hypotheses testing was based on a series of Multilevel Linear Models (MLMs) with random effects for subjects and for stimuli videos. Each test we started with the Null model with random effects in it only, then consecutive models added fixed effects for country, subtitle position and colour. When also needed, the interaction of position and colour were added as fixed factor. Each consecutive model was tested against the previous one with the chi-square test to check if a new fixed term improves the fit of the model. All the statistical analyses were performed in R language for statistical computing (R Core Team, 2020).

H1 was "head-locked subtitles are easier to follow but more intrusive than fixed subtitles". To test this hypothesis, we evaluated general comprehension of the content, subjective effort and frustration, and attention allocation over subtitles and scene Areas-of-Interest (AOI). In detail, we expected that head-locked positioning would cause better comprehension of the content, lower frustration, and effort. We also expected that head-locked positioning would cause relatively more fixations on subtitles AOI than fixed positioning.

H2 was "coloured subtitles are helpful to identify a speaking character in a 360º narrative". To test this hypothesis, we evaluated comprehension of the content. We expected that coloured subtitles allowed participants to identify speaking characters more easily than monochrome subtitles.

5.4.1. Influence of subtitle positioning and colour code on content comprehension and self-reported effort

To test the prediction that head-locked positioning would lead to better content comprehension, we ran MLM models on the general comprehension score from the comprehension questionnaires (see Figure 27a and Appendix Table 6 for model detailed tables and Appendix Table 7 for estimated means for the interaction between subtitle position and colour).

Figure 27

Content comprehension (Fig. 7a) and self-evaluation of perceived effort (Fig. 7b) depending on subtitle position (fixed or head-locked) and colour.

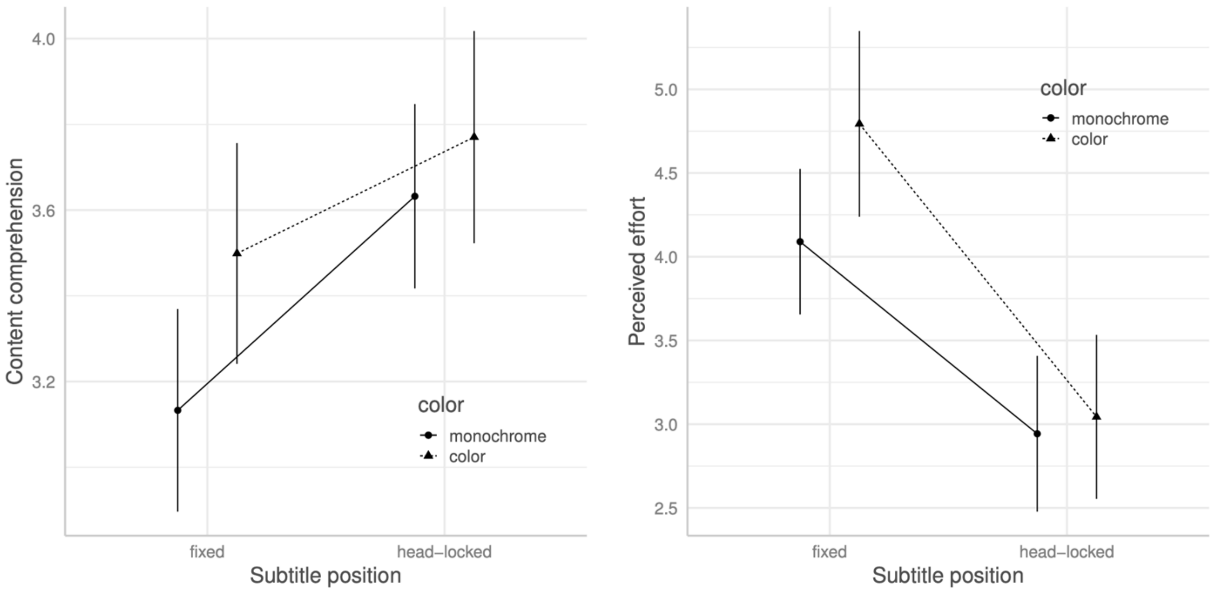


Fig. 27a. Content comprehension depending on subtitle position (fixed or head-locked) and colour.

Fig. 27b. Self-evaluation of perceived effort depending on subtitle position (fixed or head-locked) and colour.

Note. Bar height represents the estimated means and whiskers ±1*SE*.

The null model with random effects for subject and video showed *SD* = 0.10 for subject intercepts and *SD* = 0.32 for video intercept, and *SD* = 0.99 for residuals. The intercept of the model was significantly higher than zero (b_0 = 3.52, $t(3.12)$ = 13.70, $p < 0.001$).

The country fixed effect did not influence the model fit (AIC = 1775.1, $\chi^2(2) = 0.74$, p = 0.69). Contrary to expectations, subtitle position did not improve significantly the model fit (AIC = 1778.5, $\chi^2(1) = 2.38$, p = 0.12). However, in line with expectations, subtitle colour significantly improved the model fit ($\chi^2(1) = 8.64$, p = 0.003). The full model includes also the interaction of subtitle colour and position as a fixed effect, but it did not influence the fit ($\chi^2(1) = 1.91$, p = 0.17).

The full model fixed effects parameters showed that subtitle colour significantly improved comprehension of the movie ($b = 0.37$, $t(497.99) = 3.05$, $p < .001$).

The estimated mean for colour subtitles was ($M = 3.63$, $SE = 0.31$) and for monochrome subtitles ($M = 3.38$, $SE = 0.31$).

Head-locked subtitle position improved comprehension but was only marginally significant, (b=0.50, t(88.23) = 1.93, $p = 0.06$), while the interaction of subtitle colour and position did not alter significantly the movie comprehension (b= −0.23, t(498.68) = −1.38, $p = 0.17$).

To test the prediction that head-locked positioning would cause lower effort a similar analysis on comprehension MLM approach was used (see Figure 27b and Appendix Table 8 for model detailed tables and Appendix Table 9 for estimated means for the interaction between subtitle position and colour). The null model showed SD = 2.20 for subject intercepts and SD = 1.01 for video intercept, and SD = 1.64 for residuals. The intercept of the model was statistically significant (b_0 = 3.68, t(2.55) = 4.80, p = 0.02).

Comparison of models with fixed factors added showed that country did not improve the model (AIC = 2247.5, $\chi^2(2) = 2.84$, $p = 0.24$) but both the subtitle position (AIC = 2241.9, $\chi^2(1) = 7.58$, $p = 0.005$) and subtitle colour did (AIC = 2236.9, $\chi^2(1) = 7.05$, $p = 0.008$) as well as an interaction term between the colour and position of the subtitles (AIC = 2234.4, $\chi^2(1) = 4.43$, $p = 0.04$).

The analysis of the full model fixed parameters showed that colour subtitles in general add significant effort to the task ($\beta = 0.70$, t(471.01) = 3.37, $p < 0.001$). Head-locked subtitles lower significantly the perceived effort ($\beta = -1.15$, t(79.09) = −2.21, $p = 0.03$). Significant interaction term shows that head-locked subtitles lower the effort mainly when they are presented in colour ($\beta = -0.60$, t(469.41) = −2.11, $p = 0.04$).

5.4.2. Influence of subtitle positioning on visual attention directed toward subtitles

We hypothesised that fixed positioning causes more fixations and longer fixation time over subtitles. To test these predictions, we applied MLM analyses.

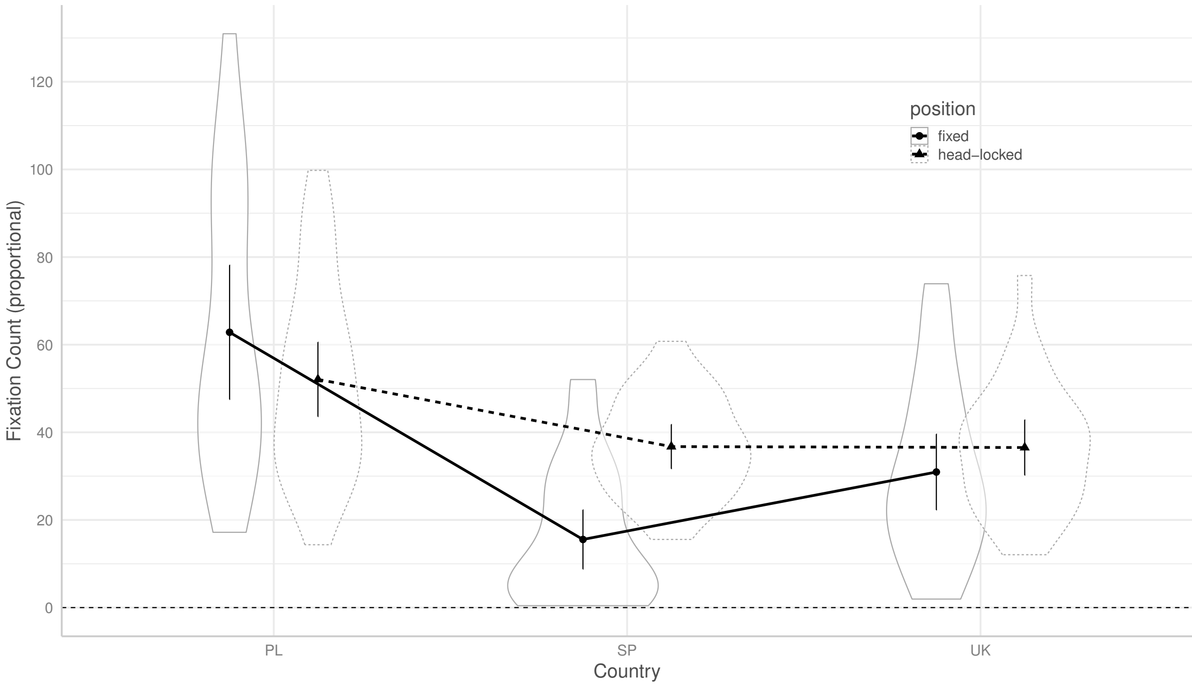
5.4.2.1. Fixation count

The first analysis used proportional fixation count (the number of fixations divided by the number of words in each subtitle) as a dependent variable. The MLM started with the null model with random effects of participants and stimuli video in it.

In the next steps, the model was updated with the fixed effects of country, subtitle position, subtitle colour, and interaction fixed effects of position and colour, position, and country. The models were tested against each other with chi-squared statistics to see if the fixed effect significantly improved the model fit (see Figure 28 and also Appendix Table 10 for model detailed tables and Appendix Table 11 for estimated means for the interaction between subtitle position and country).

Figure 28

Proportional fixation count on subtitles AOI depending on subtitle position (fixed or head-locked) and country (Poland, Spain or UK)



Note. Bar height represents the estimated means and whiskers ±1*SE*. Violin shapes represent the distribution of the data points in each experimental condition in different countries.

The null model showed *SD* = 16.20 for subject intercepts and *SD* = 19.30 for video intercept, and *SD* = 13.32 for residuals. The intercept of the null model was not statistically significant ($\beta_0 = 39.06$, $t(1.04) = 2.83$, $p = 0.21$).

Models' comparison shows the significant effect on the model fit of a country as a fixed effect: $\chi2(2) = 47.03$, p < 0.001, AIC = 1217.10, BIC = 1234.90. Also, the interaction effect of country and position significantly improved the model fit: $\chi2(2) = 19.53$, p < 0.001, AIC = 1165.28, BIC = 1197.87. The final models' total psycho-R2 = 0.78.

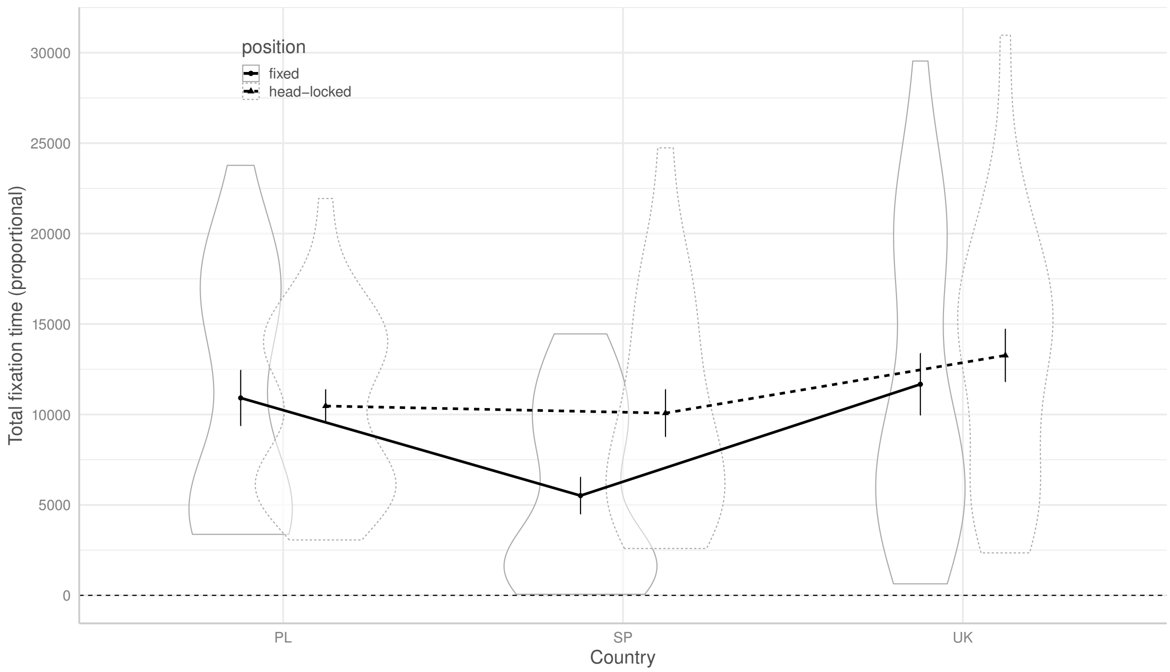
The full model fixed parameters showed that fixed subtitles significantly lower the proportional fixation count compared to head-locked (β = -–2.84, t(91.72) = -2.39, p = 0.02). This effect was quantified by country. The pairwise comparisons of means showed that the differences between fixed and head-locked subtitles were significant in Spain (t(66.6) = –4.22, p < 0.001) and in Poland (t(65.0) = 2.20, p = 0.03) but not in the UK (t(65.3) = –1.15, p = 0.26). Note that the difference in fixation count between head-locked and fixed subtitles in Polish sample was in favour of head-locked, unlike in the Spanish sample.

5.4.2.2. Total fixation time

The second MLM analysis used proportional total fixation time (total fixation time divided by visibility time of subtitle) as a dependent variable. The procedure of testing and modelling random and fixed effect structure was identical to the analysis of fixation count described in subsection 5.4.2.1. (see Figure 29 and also Appendix Table 12 for model detailed tables and Appendix Table 13 for estimated means for the interaction between subtitle position and country).

Figure 29

Proportional total fixation time on subtitles AOI depending on subtitle position (fixed or head-locked) and country (Poland, Spain, or UK)



Note: Bar height represents the estimated means and whiskers ±1*SE*. Violin shapes represent the distribution of the data points in each experimental condition in different countries. The interaction effect is not statistically significant in the full model.

The null model (with random effects only) for total fixation time showed *SD* = 3537.14 for subject intercepts, *SD* = 6696.95 for video intercept, and *SD* = 3238.18 for residuals. The intercept of the null model was not statistically significant (β_0 = 10322.92, $t(1.02) = 2.17$, $p = 0.27$).

The fixed effect of the country significantly improved the model fit ($\chi^2(2) = 16.96$, $p < 0.001$, $AIC = 2808.60$, $BIC = 2826.40$) as well as the fixed effect of subtitle position ($\chi^2(1) = 4.50$, $p = 0.03$, $AIC = 2806.10$, $BIC = 2826.90$). Additionally, the interaction term between country and subtitle position was marginally significant ($\chi^2(2) = 5.83$, $p = 0.054$, $AIC = 2808.10$, $BIC = 2840.70$). Because there was only a marginal model fit improvement, the interaction between country and subtitle positioning is hard to interpret.

However, the parameters of the model with the country and position fixed terms (with pseudo-R^2 (total) = 0.85) clearly show that the fixed subtitles gained more total fixation time than head-locked subtitles ($\beta = 1861.30$, $t(58.07) = 2.11$, $p = 0.04$). Significant effect of country shows also that participants in Spain spent significantly less time reading subtitles than participants in Poland ($\beta = 12718.09$, $t(68.28) = -2.51$, $p = 0.01$). See Figure 9 for the estimated means of proportional total fixation time for both effects.

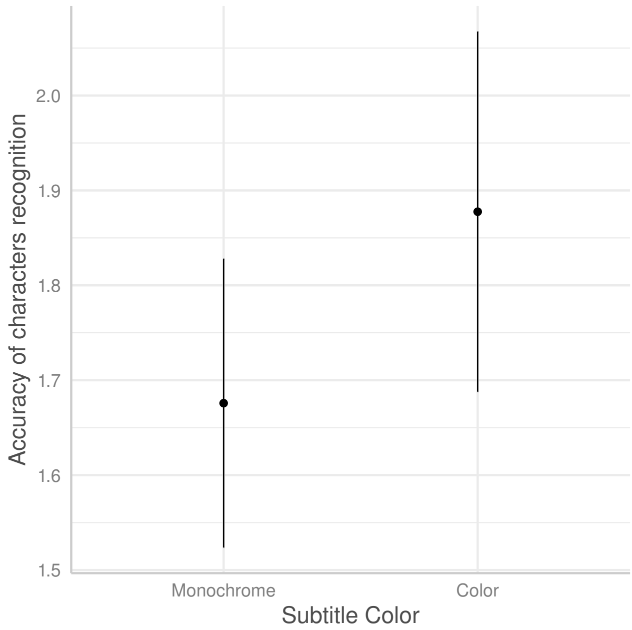
5.4.3 Influence of subtitle colour on character recognition

Null multilevel model on character recognition, with random effects only for participants and stimuli video, showed *SD* = 0.67 for subject random effect and *SD* = 1.60 for video, and *SD* = 0.69 for residuals. The intercept of the null model was not statistically significant (β_0 = 1.78, $t(1.01) = 1.57$, $p = 0.36$).

The models fitting with the fixed effects for country, subtitle position, subtitle colour as well as interactions between colour and position, and position and country showed that subtitle colour code significantly improved the model fit ($\chi^2(1) = 12.26$, $p < 0.001$, $AIC = 1378.32$, $BIC = 1413.14$, pseudo-R^2(total) = 0.86). The model parameters showed that colour subtitles significantly improves the recognition of characters ($\beta = 0.20$, $t(3.51) = 3.51$, $p < 0.001$). The estimated average recognition of characters with monochrome subtitles was significantly lower ($M = 1.68$, $SE = 1.12$) than with colour subtitles ($M = 1.88$, $SE = 1.12$), see Figure 30 and Appendix Table 14 for detailed parameters of the model.

Figure 30

Estimated means of fixed effect of subtitle colour on accuracy of characters' recognition



Answers to the question "what colour code do you prefer when reading 360º subtitles?" showed no significant preference between colour and monochrome subtitles ($p > 0.5$). Regarding coloured subtitles, one participant stated that "colours distracted me from what they [the characters] were actually saying", but, on the contrary, another participant stated that "it's easier to know which one is talking, especially because you aren't looking to all of them at the same time".

These general results are in line with those obtained in Poland, but not with those obtained in Spain and the UK (see Table 5).

Table 5

Subtitles colour preferences between countries

	Condition		
	Monochrome	Coloured	No preference
Poland	42.9 %	40.8 %	16.3 %
Spain	39.1 %	56.5 %	4.3 %
UK	58.3 %	25 %	16.7 %
Total	46.8 %	40.5 %	12.6 %

The differences in subtitle colour preference between countries were statistically significant ($p = 0.02$). Spain participants stated a preference for coloured subtitles, while UK participants stated a preference for monochrome subtitles.

5.5. Discussion

The present study examined 360º video although immersive subtitles can extend to any media asset. It focused on two hypotheses: the first related to subtitle positioning and the second related to subtitle colour. Regarding subtitle positioning, head-locked subtitles were subjectively evaluated as causing less cognitive effort in comparison to fixed subtitles. However, eye tracking data showed that head-locked subtitles engaged more attention from the viewer than fixed subtitles. Head-locked subtitles gained significantly higher fixation counts and total fixation time, both of which are proportional to the number of words in the subtitle and visibility time of the subtitle, respectively. These results suggest that head-locked subtitles are more cognitively engaging, despite their subjective evaluation. The latter might be interpreted via the familiarity heuristic (Ashcraft and Radvansky, 2014) which states that people usually prefer things or situations that they are familiar with. However, it is important to note that this conclusion is specific to the Spanish sample. In the Polish sample, the difference in fixation count between head-locked and fixed subtitles favoured head-locked subtitles, which contrasts with the findings in the Spanish sample.

Subtitle positioning does not significantly affect video content comprehension. There was only a marginally significant effect of head-locked subtitles improving comprehension. The factor which influenced content comprehension significantly was the subtitles' colour scheme. Use of the different subtitle colours for each character video character improved character recognition and general comprehension; however, colour subtitles were evaluated by participants as causing more effort. Nevertheless, this is the first experiment testing subtitle colour in 360º videos, and it would be interesting to carry out further research on this topic to confirm our results.

The present article points out interesting evidence of discrepancies between self-report measures and more objective, psychophysiological, and comprehension metrics. Thus, we postulate to incorporate psychophysiological and performance methods and indicators in future studies on immersive media accessibility. It also points out the need for test environments or tools similar to what was proposed by Hughes et al. (2020) and which were used in this empirical user study.

An additional contribution of this study is the comparison of the different audiences from the sample. Replicating the experiment in three countries with different language translation traditions in audiovisual media allowed us to work with a heterogeneous sample and report differences between the participants. We found some significant differences both in perceived effort/frustration and AOI fixations depending on the country. Although it is not possible to define tendencies with the number of participants tested, it might be interesting to consider the preferences of people from different countries and cultures when testing subtitled immersive media.

Even though it was not the aim of the study, some participants mentioned the problem of speaker or audio source localisation when answering the questionnaires. This is in line with earlier work (Hughes et al. 2019, Rother et al., 2018; Agulló et al. 2019) calling for user guiding mechanisms such as arrows, radar, or other means. Although arrows have been shown to be the option preferred by users (Agulló, 2020), eye-tracked evaluation of such means holds potential for future work in the field.

5.6. Conclusions

The present paper presented a review of the main studies regarding subtitles in immersive content and the main challenges posed by this new medium. The first aim of the study was to detail the design of controlled experiments involving immersive environments and subtitles. The methodology followed in the three locations where the experiment took place has been described in detail, including the design and IV, the participants, the procedure, the materials, and the equipment. This allows other researchers to replicate the experiment in other locations and/or using different video stimuli.

The second aim was to gather feedback regarding two characteristics of subtitles in immersive content: position and colour. To confirm the hypotheses, feedback on preferences and task-load was gathered, together with eye movement data. The results have shown that head-locked subtitles, which are always visible and positioned in front of the viewer to align with their head movements, attracted more attention from viewers. However, it should be noted that this finding was observed specifically within the Spanish sample, and the observed trend in the Polish sample pointed in the opposite direction. In all samples, it was observed that participants consistently showed a preference for coloured subtitles as opposed to monochrome subtitles. This preference can be attributed to the enhanced ability to identify the speaking character facilitated by the use of colour. Although colour subtitles were perceived as requiring more effort, particularly when presented in a fixed position, they yielded improved comprehension of the content, especially in the fixed position, and better recall of the characters. Importantly, the impact of colour on comprehension and recall was consistent across the different country samples, indicating that it is not influenced by cultural factors.

The present study had some limitations regarding the type of content. To obtain meaningful results, two videos were recorded in a non-professional environment with non-professional actors. This might have had an impact on the results and a replication of this study using professional content (e.g., a short movie, a documentary, a news program, or interviews) is encouraged.

Head-locked, coloured subtitles could be considered as good practice guidelines when subtitling in 360º videos. On the one hand, results about the position confirm previous studies, in which results were obtained through Igroup Presence Questionnaires (Agulló and Matamala, 2020). On the other hand, results about subtitle colour are not conclusive and further research in this regard is encouraged.

The study also highlighted the importance of the materials used for testing to ensure the ecological validity of the experiment. Special attention must be paid to the three languages involved in each experiment: the language of the film, the language of the subtitles, and the language of the participants. This problem has already been identified in other studies involving testing of different subtitle implementations, and various solutions have been adopted: the audio was muted (Kurzhals et al., 2017; Agulló and Matamala, 2020) or manipulated (Rothe et al., 2018). When testing with deaf and/or hard-of-hearing participants one more language needs to be added depending on whether they are oralists or if they use sign language (signers).

To conclude, this is a first-of-its-kind analysis of eye-tracked visual attention on subtitle processing in 360º media content, with a focus on subtitle position and subtitle colour. Detailed examination of eye movements over subtitles combined with self-reports and performance measures has allowed an analysis of visual attention distribution and scene comprehension. We believe that the present work contributes to the novel frontiers on the implementation of accessibility services in immersive media studies.

Acknowledgments

This study is part of the article-based PhD thesis of Marta Brescia-Zapata in the Department of Translation and Interpreting at Universitat Autònoma de Barcelona (UAB) within the PhD program in Translation and Intercultural Studies. This research has been partially funded by the H2020 projects TRACTION (under Grant Agreement 870610) and MEDIAVERSE (under Gran Agreement 957252); and it is based upon work from COST Action LEAD ME, CA19142, supported by COST (European Cooperation in Science and Technology). The Commission's support for this publication does not constitute an endorsement of the contents, which reflects the views of the authors only, and the Commission cannot be held responsible for any use which may be made of the information contained herein. Marta Brescia-Zapata and Pilar Orero are members of TransMedia Catalonia, an SGR research group funded by "Secretaria d'Universitats I Recerca del Departament d'Empresa.Coneixement de la Generalitat de Catalunya" (2021 SGR 00077).

References

Agulló, B. & Matamala, A. (2019). Subtitling for the deaf and hard-of-hearing in immersive environments: Results from a focus group. *The Journal of Specialised Translation*, *32*, 217-235. Retrieved from http://www.jostrans.org/issue32/art_agullo.pdf

Agulló, B. & Matamala, A. (2020). Subtitles in virtual reality: Guidelines for the integration of subtitles in 360º content. *Íkala, Revista de Lenguaje y Cultura*, *25*(3), 643-661. https://doi.org/10.17533/udea.ikala.v25n03a03

Agulló, B., Matamala, A., & Orero, P. (2018). From disabilities to capabilities: Testing subtitles in immersive environments with end users. *HIKMA 17*: 195-220.

Agulló, B., Montagud, M., & Fraile, I. (2019). Making interaction with virtual reality accessible: Rendering and guiding methods for subtitles. *Artificial Intelligence for*

Engineering Design, Analysis and Manufacturing, *33*(4), 416-428. https://doi.org/10.1017/S0890060419000362

Agulló, B. & Orero, P. (2017). 3D Movie subtitling: Searching for the best viewing experience. *CoMe - Studi di Comunicazione e Mediazione linguistica e culturale, 2,* 91-101. Retrieved from https://ddd.uab.cat/record/183656

Ashcraft, M. H. & Radvansky, G. A. (2014). *Cognition*. Pearson Education, Boston, sixth edition.

Asociación Española de Normalización y Certificación (AENOR). (2003). *Norma UNE 153010: Subtitulado para personas sordas y personas con discapacidad auditiva. Subtitulado a través del teletexto.* https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma/?Tipo=N&c=Q21 N0029761

Ballester Casado, A. (1998). *Traducción y nacionalismo. La recepción del cine americano en España a través del doblaje, 1920-1948.* [Translation and Nationalism. The Reception of American Cinemain SPain through Dubbing, 1920-1948]. Comares.

BBC. (2022). *Subtitle guidelines*. Version 1.2.1. https://www.bbc.co.uk/accessibility/forproducts/guides/subtitles/. Accessed 25 April 2023.

Brescia-Zapata, M., Krejtz, K., Orero, P., Duchowski, A. T., & Hughes, C. J. (2022). VR 360◦ subtitles: Designing a test suite with eye-tracking technology. *Journal of Audiovisual Translation, 5*(2). (In press.)

Brown, A. (2017, October 26). *User testing for 360º content.* BBC. https://www.bbc.co.uk/rd/blog/2017-10-subtitles-360-video-virtual-reality-vr

Brown, A., Turner, J., Patterson, J., Schmitz, A., Armstrong, M., & Glancy, M. (2017). Subtitles in 360-degree video. *Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video.* The Netherlands, 3–8. https://doi.org/10.1145/3084289.3089915

Brown, A., Turner, J., Patterson, J., Schmitz, A., Armstrong, M., & Glancy, M. (2018). *Exploring subtitle behaviour for 360° video* (White Paper WHP 330) https://www.bbc.co.uk/rd/publications/whitepaper330

Chaume, F. (2012). *Audiovisual translation: Dubbing.* Manchester: St Jerome Publishing.

European Broadcasting Union (EBU). (2017). Virtual reality: How are public broadcasters using it? https://www.ebu.ch/publications/virtual-reality-how-are-public-broadcasters-using-it

Fidyka, A., & Matamala, A. (2018). Audio description in 360º videos: Results from Focus Groups in Barcelona and Kraków. *Translation Spaces, 7*(2), pp. 285–303. http://doi.org/10.1075/ts.18018.fid

Gejrot, E., Kjellstrand, S., & Laurin, S. (2021). New realities unlocking the potential of XR for persons with disabilities. Tech. rep., Funka Nu AB, Stockholm, Sweden.

Gil Ariza, M. C. (2004). A case study: Spain as a dubbing country. *Translation journal*, 29. http://translationjournal.net/journal/29movies.html

Gottlieb, H. (1998). Subtitling. In M. Baker (Ed.) *Routledge encyclopaedia of translation studies.* London and New York: Routledge, 244-8.

Greener, R. (2022, December 19). *Virtual reality statistics to know in 2023.* XR today. https://www.xrtoday.com/virtual-reality/virtual-reality-statistics-to-know-in-2023/#:~:text=In%202020%2C%20roughly%2057.4%20people,%2C%20industry%2C%20and%20education%20adoption.

Hart, S. G. (2006). Nasa-task load index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50*(9), 904-908. https://doi.org/10.1177/154193120605000909

Hermosa-Ramírez, I. (2022). Physiological instruments meet mixed methods in media accessibility. *Translation Spaces, 11*(1), 38-59. https://doi.org/10.1075/ts.21020.her

Hughes, C., Brescia-Zapata, M., Johnston, M., & Orero, P. (2020). *Immersive captioning: Developing a framework for evaluating user needs.* IEEE AIVR: 3rd International Conference on Artificial Intelligence & Virtual Reality, IEEE.

Hughes, C., Montagud, M., & tho Pesch, P. (2019). Disruptive approaches for subtitling in immersive environments. In *Proceedings of the 2019ACM International Conference on Interactive Experiences for TV and Online Video – TVX '19*. http://doi.org/10.1145/3317697.3325123

Ibourk, E. M., & Al-Adwan, A. (2019). Subtitling virtual reality into arabic: Eye tracking 360-degree video for exploring viewing experience. *Lebende Sprachen, 64*(2), 286-308.

International Organization for Standardisation. (2018). *Information technology — User interface component accessibility* (ISO/IEC 20071) https://www.iso.org/standard/70722.html

Krejtz, I., Szarkowska, A., & Łogińska, M. (2016). Reading function and content words in subtitled videos. *The Journal of Deaf Studies and Deaf Education, 21*(2), 222–232. https://doi.org/10.1093/deafed/env061

Kruger, J. (2016). Psycholinguistics and audiovisual translation. *Target. International Journal of Translation Studies, 28*(2), 276–287. https://doi.org/10.1075/target.28.2.08kru

Kurzhals, K. Cetinkaya, E., Hu, Y., Wang, W., & Weiskopf, D. (2017). Close to the Action: Eye-Tracking Evaluation of Speaker-Following Subtitles, *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*, 6559–6568. https://doi.org/10.1145/3025453.3025772

Liao, S., Yu, L., Reichle, E. D., & Kruger, J. L. (2021). Using eye movements to study the reading of subtitles in video. *Scientific Studies of Reading, 25*(5), 417–435.

Liberatore, M. J., & Wagner, W. P. (2021). Virtual, mixed, and augmented reality: A systematic review for immersive systems research. *Virtual Reality, 25*(3), 773–799. https://doi.org/10.1007/s10055-020-00492-0

Luyken, G. M. (1991). *Overcoming language barriers in television: Dubbing and subtitling for the European audience.* European Institute for the Media.

Matamala, A., & Orero, P. (2018). Standardising accessibility: Transferring knowledge to society. *Journal of Audiovisual Translation, 1*(1), 139–154. https://doi.org/10.47476/jat.v1i1.49

Matamala, A., Perego, E., & Bottiroli, S. (2017). Dubbing versus subtitling yet again? An empirical study on user comprehension and preferences in Spain. *Babel. Revue internationale de la tradution/International Journal of Translation, 63*(3), 423–41. https://doi.org/10.1075/babel.63.3.07mat

Montagud-Climent, M., Orero, P., & Matamala, A. (2020). Culture 4 all: Accessibility-enabled cultural experiences through immersive VR360 content. *Personal and Ubiquitous Computing, 24*(6), 887-905.

Myers, Maddy. (June 12, 2020). "The Last of Us Part 2 review: We're better than this". Polygon. Vox Media. Archived from the original on June 12, 2020. Retrieved July 6, 2022.

Oncins, E., Bernabé, R., Montagud, M., & Arnáiz Uzquiza, V. (2020). Accessible scenic arts and virtual reality: A pilot study with aged people about user preferences when reading subtitles in immersive environments, *MonTI. Monografías de Traducción e Iterpretación 12*(12), 214–41. https://doi.org/10.6035/MonTI.2020.12.07

Orero, P. (2022). Translation, accessibility and minorities. In E. Bielsa (Ed.), *The Routledge handbook of translation and media* (pp. 384-399). Taylor Francis.

Orero, P., Doherty, S., Kruger, J., Matamala, A., Pedersen, J., Perego, E., Romero-Fresco, P., Rovira-Esteva, S., Soler-Vilageliu, O., & Szarkowska, A. (2018). Conducting experimental research in audiovisual translation (AVT): A position paper. *JoSTrans, The Journal of Specialised Translation,* (30), 105–126. https://www.jostrans.org/issue30/art_orero_et_al.pdf

Orero, P., Sonali R., & Hughes, C. (2019). The challenges of audio description in new formats. In S. Braun (Ed.), *Innovation in audio description research*. Routledge.

R Core Team. (2020). R: A language and environment for statistical computing. R foundation for statistical computing. Vienna, Austria. http://www.R-project.org

Romero-Fresco, P. (2013). Accessible filmmaking: Joining the dots between audiovisual translation, accessibility and filmmaking. *The Journal of Specialised Translation, 20*, 201–223.

Rothe, S., Tran, K., & Hussmann, H. (2018). Positioning of subtitles in cinematic virtual reality. In G. Bruder, S. Yoshimoto, & S. Cobb (Eds.). *International Conference on Artificial Reality and Telexistence and Eurographics Sysmposium on Virtual Environments*, *ICAT-EGVE 2018.* The Eurographics Association.

Sipatchin, A., Wahl, S., & Rifai, K. (2020). Eye-tracking for low vision with virtual reality (VR): testing status quo usability of the HTC Vive Pro Eye. *bioRxiv.*

Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and Virtual Environments, 6*(6), 603–616. https://doi.org/10.1162/pres.1997.6.6.603

Szarkowska, A. (2009). The audiovisual landscape in Poland at the dawn of the 21st century. In: A. Goldstein, & B. Golubović (Eds.), *Foreign language movies – dubbing vs. subtitling* (pp. 185–201). Verlag Dr. Kovač.

Szarkowska, A., & Boczkowska, J. (2022). Colour coding subtitles in multigoal films – a reception study. *Perspectives: Studies in Translatology, 30*(3), 520–536. https://doi.org/10.1080/0907676X.2020.1853186

Szarkowska, A., & Gerber-Morón, O. (2019). Two or three lines: A mixed-methods study on subtitle processing and preferences. *Perspectives: Studies in Translatology, 27*(1), 144–164. https://doi.org/10.1080/0907676X.2018.1520267

Udo, J.-P., & Fels, D. I. (2010). Universal design on stage: Live audio description for theatrical performances. *Perspectives: Studies in Translatology, 18*(3), 189–203. http://doi.org/10.1080/0907676X.2010.485683

Valdeón, R. A. (2022). Latest trends in audiovisual translation. *Perspectives: Studies in Translatology*, 30(3), 369–381. https://doi.org/10.1080/0907676X.2022.2069226

Appendices

Tables with detailed statistics for the model on content general comprehension.

Table 6

Model parameters for content general comprehension

	Estimate (β)	*SE*	*t*	*df*	*p*
Intercept	3.25	0.33	9.70	11.10	0.00
Country Spain	–0.11	0.30	–0.37	72.02	0.71
Country UK	–0.22	0.30	–0.76	71.94	0.45
Position head-locked	0.50	0.26	1.93	88.23	0.06
Colour colourful	0.37	0.12	3.05	497.99	0.00
Position head-locked x colour colourful	–0.23	0.16	–1.38	498.68	0.17

Table 7

Estimated means from the model for content general comprehension for interaction effect of subtitle position and colour

Position	Colour	Est. mean	*SE*
fixed	monochrome	3.13	0.34
head-locked	monochrome	3.63	0.33
fixed	colourful	3.50	0.34
head-locked	colourful	3.77	0.33

Table 8

Model parameters for self-perceived effort

	Estimate (β)	*SE*	*t*	*df*	*p*
Intercept	4.57	0.90	5.07	4.52	0.01
Country Spain	–0.31	0.62	–0.50	68.96	0.62
Country UK	–1.12	0.61	–1.83	68.45	0.07

Position head-locked	–1.15	0.52	–2.21	79.09	0.03
Colour colourful	0.70	0.21	3.37	471.01	0.00
Position head-locked X Colour colourful	–0.60	0.29	–2.11	469.41	0.04

Table 9

Estimated means from the model for self-perceived effort for interaction effect of subtitle position and colour

Position	Colour	Est. mean	*SE*
Fixed	monochrome	4.09	1.08
Head-locked	monochrome	2.94	1.07
Fixed	colourful	4.79	1.08
Head-locked	colourful	3.04	1.07

Table 10

Model parameters for proportional fixation number

	Estimate *(β)*	*SE*	*t*	*df*	*p*
Intercept	64.93	14.23	4.56	1.17	0.11
Country Spain	–47.29	5.15	–9.18	66.55	0.00
Country UK	–31.90	5.11	–6.24	65.29	0.00
Position head-locked	–12.84	5.36	–2.39	91.72	0.02
Colour monochrome	–4.20	3.29	–1.28	69.19	0.21
Position head-locked X Colour monochrome	4.18	4.49	0.93	69.46	0.35
country Spain: Position head-locked	31.95	7.00	4.56	65.80	0.00
country UK: Position head-locked	16.34	6.90	2.37	65.11	0.02

Table 11

Estimated means from the model for proportional fixation number for interaction effect of subtitle position and country

Position	Country	Est. mean	*SE*
Fixed	Poland	62.8	14.1
Head-locked	Poland	52.1	14.0
Fixed	Spain	15.5	14.1
Head-locked	Spain	36.7	14.1
Fixed	UK	30.9	14.0
Head-locked	UK	36.5	14.0

Table 12

Model parameters for proportional total fixation time

	Estimate *(β)*	*SE*	*t*	*df*	*p*
Intercept	10756.03	4916.41	2.19	1.13	0.25
Country Spain	–5401.21	1556.41	–3.47	66.92	0.00
Country UK	756.23	1547.46	0.49	65.83	0.63
Position head-locked	–317.28	1575.22	–0.20	83.65	0.84
Colour monochrome	316.28	806.71	0.39	69.40	0.70
Position head-locked: Colour monochrome	–257.78	1100.44	–0.23	69.58	0.82
Country Spain: Position head-locked	5010.05	2118.56	2.36	66.45	0.02
Country UK: Position head-locked	2041.71	2088.48	0.98	65.86	0.33

Table 13

Estimated means from the model for proportional total fixation time for interaction effect of subtitle position and country

Position	Country	Est. mean	*SE*
Fixed	Poland	10914	4900
Head-locked	Poland	10468	4856
Fixed	Spain	5513	4880
Head-locked	Spain	10077	4888
Fixed	UK	11670	4878
Head-locked	UK	13266	4878

Table 14

Model parameters for accuracy of character' recognition

	Estimate *(β)*	*SE*	*t*	*df*	*p*
Intercept	1.53	1.13	1.35	1.05	0.40
Country Spain	0.14	0.21	0.68	68.19	0.50
Country UK	0.12	0.21	0.60	68.10	0.55
Position head-locked	0.12	0.17	0.70	68.11	0.48
Colour colourful	0.20	0.06	3.51	500.42	0.00

CHAPTER 6

6. Summary

Subtitles play a critical role in ensuring access to multimedia content, as well as enabling cultural inclusivity via language translation. Standards and best practices for implementing subtitles have been established for 2D non-immersive media. Nevertheless, these standards need to be re-evaluated to address the new challenges posed by immersive VR/AR 360º videos. Previous research on this topic focused on assessing presence and user preferences using only subjective measures. Furthermore, the consideration of user gaze to evaluate immersive subtitles was almost non-existent.

This dissertation aimed to provide insights into the optimal design choices for immersive subtitles in VR experiences. It used eye-tracking to evaluate subtitling solutions in the context of VR 360º videos. Three main goals were established: 1) To explore the possibilities of integrating accessibility services in new immersive media; 2) To validate a methodological framework for testing subtitles in VR 360º media content using eye-tracking technology; and 3) To evaluate different subtitling solutions (related to position and colour) in VR 360º media content. To fulfil these goals, four studies were carried out. Two were descriptive and two experimental.

The first descriptive study was twofold. On the one hand, it focused on the challenges of integrating accessibility services into immersive environments. On the other hand, it examined the minimum requirements for creating accessible immersive content. The study described the differences between 2D and immersive environments and established the state-of-the-art concerning accessibility integration. The main contribution of this study was the description of the four components needed to achieve fully immersive media experiences, namely the device(s), the platform, the media player, and the multimedia content.

The second descriptive study consisted in analysing the integration of accessibility services in existing VR and 360º video players. It analysed the most popular VR 360º players to assess how and to what extent they incorporate the main accessibility services. The findings revealed that most players are not designed with accessibility by default, as the ISO Standard recommends. However, the study identified one fully accessible but non-commercial player: the ImAc player. Developed as part of the H2020-funded project ImAc, this VR 360º video player follows the EU Standard EN301549 and the ISO/IEC 24751-3:2008, securing the concepts of Universal Design and Born Accessible. In light of this, the ImAc project was

considered the primary reference to elaborate the reception studies conducted for this dissertation.

Third, a reception study was carried out to validate the methodological framework used to analyse the presentation of subtitles in immersive 360º video experiences. The framework had three functionalities: a web-based subtitle player, a web-based subtitle editor, and a VR system designed to collect eye movement data as visual attention is distributed over 360º video containing subtitles. Both the player and the editor drew upon the tools created in the ImAc project. Conducted as a pilot test, the study involved a group of 5 participants without hearing disabilities. The methodological framework implemented in this initial pilot study demonstrated its efficacy in gathering both quantitative and qualitative behavioural data when viewing subtitles in 360º media. These findings proved the framework potential for further application in subsequent studies.

The fourth study evaluated different subtitle solutions in 360º videos concerning position (head-locked vs. fixed) and colour (monochrome vs. coloured). It aimed at providing insights into the optimal design choices for immersive subtitles in VR experiences and used the methodological framework validated in the previous study. In the initial stage, the study involved 73 hearing participants from Warsaw, Salford, and Barcelona. Then, it was extended to include 39 Deaf and HoH participants from Barcelona and Ljubljana. The results contributed to a broader understanding of how immersive subtitles can be useful for any user who needs support to understand the AV content. In terms of position, head-locked subtitles, which are always visible and align with the viewers' head movements, attracted more attention from viewers. In terms of colour, using distinct subtitle colours for each character improved character recognition and overall content comprehension.

CHAPTER 7

7. Discussion and conclusions

The emergence of immersive technologies has led in a transformative era for the audiovisual landscape, redefining how we perceive and engage with multimedia content. On one hand, there is a rise of on-demand content platforms like Netflix, Twitch, or Youtube; an expansion of domains such us videogames, websites, and emerging social media platforms like TikTok, and the development of the Metaverse. On the other hand, the global immersive technology market is growing, with its value predicted to soar from USD 27.41 billion in 2023 to USD 167.75 billion by 2032[25]. It is evident that we are entering a new era in entertainment and communication. This situation has also affected the field of AVT and MA, revealing their essential role in making media content accessible to diverse audiences.

In this context, subtitles are a critical tool for accessing audiovisual content. Whether they serve audiences with hearing loss, people who do not understand the language of the content, or those who just prefer subtitles, the choice to use them in immersive content must be available, just as in 2D non immersive media. Accessibility in immersive media started receiving attention as a research topic in 2018. Although the challenges have already been identified (Montagud-Climent et al., 2021; Creed et al., 2023), there is still a lack of standards or unified solutions. XR experiences present new opportunities to support social interaction in virtual environments, which holds a significant potential to address the digital divide. However, there is a significant gap in current and emergent technology design which illustrates the implicit lack of design consideration for the real needs of users. This thesis is a step forward to bridge this gap between immersive technologies and accessibility.

Presented as a compendium of articles, this dissertation explores subtitling in immersive environments. While each publication stands as an independent piece of research, they explore different elements of the same research question, and therefore can be presented as a unified whole. The first two article settle the theoretical background, while the last two present experimental studies based on a novel mixed-method approach. To provide a good way of verifying the reliability of the findings (Saldanha & O'Brien, 2013), the last two articles implement methodological triangulation of data. Such approach allows to study the richness and complexity of human behaviour from more than one standpoint.

There are three factors that played a significant role throughout the thesis:

[25] https://www.precedenceresearch.com/immersive-technology-market

- New ICT developments. As stated by Remael et al. (2019, p. 131), "AVT has been technology and industry-driven from the start." The evolution of our audiovisual and digital society has led a transformative "technological turn" in AVT, including subtitling (Díaz-Cintas, 2013). Within this evolving landscape, emerging formats like VR or AR have introduce new opportunities and challenges for AVT. This dissertation describes the current situation of accessibility regarding these new formats (see Chapter 2 and 3).

- Research contribution. MA and AVT research are increasingly acknowledging the mixed methods approach applied to studies using physiological instruments (Hermosa-Ramírez, 2021). When studying 2D subtitles reception, eye-tracking has emerged as the predominant choice (e.g., Perego et al., 2010; Szarkowska & Gerber-Morón, 2019; Black, 2020). However, research employing mixed methods to investigate immersive subtitles remains scarce. One of the motivations behind this thesis was to bridge this methodological gap. This is reflected in Article 3 (see Chapter 4).

- TRACTION project. This PhD was written as part of the H2020 TRACTION project, which main aims include fostering innovation in opera production, enhancing audience engagement, and exploring novel ways to use new technologies for the revitalisation of this art form. In the view of forthcoming need to implement accessibility services within artistic productions in immersive environments, this dissertation carried out a study to explore the visualisation of subtitles in 360º content. This is reflected in Article 4 (see Chapter 5) and Annex 1.

The overall evaluation of the results of the research studies carried out within the framework of this dissertation is presented in the next sections. An analysis is carried out to evaluate whether the hypothesis put forward in this research were validated or refuted, and whether the objectives were achieved. To conclude, §7.5 describes the contributions and limitations of the studies and points some directions for future work.

7.1. Theoretical framework: multilayered accessibility in immersive media

This section is related to the following objective:

To design and validate a mixed methodological framework for testing subtitles in VR 360º media content using eye-tracking technology in real time.

In recent years, accessibility within immersive environments has gained significant attention, from both scholars and the industry. An example is the set of user needs and requirements for people with disabilities when using XR technologies published by a W3C working group during the course of this PhD. This fact, together with the proliferation of academic publications on the subject, highlights the growing recognition of the importance of accessibility in immersive experiences.

The rapid evolution of immersive technologies together with the dynamic nature of the media landscape has brought out the need for a review of existing literature and an analysis of the state of the art. This task was carried out in Chapters 2 and 3 to ensure that research remains pertinent to the evolving accessibility requirements of immersive media. Chapter 2 offers a critical examination of the integration of accessibility services within immersive media. The main conclusion from this exploration is that to ensure fully accessible immersive media experiences, it is imperative to address accessibility across four distinct yet interdependent layers: the media content, the media player, the platform, and the device.

As stated by Montagud-Climent et al. (2019, p.7), "providing cutting-edge technology to all citizens is essential to achieve global e-inclusion." This assertion reinforces the idea that accessibility should not be a peripheral concern but rather an integral component embedded at every layer of the immersive media landscape. For example, all efforts dedicate to creating accessible immersive content become futile if the platforms facilitating its consumption do not support access services. Chapter 3 underscores this challenge by revealing that none of the commercially available players provide access to the full set of accessibility services. The solution lies in the adoption of the Born Accessible principle, where accessibility is considered at the design stage of any process. This approach ensures that both content creators and developers work together to embed accessibility as an inherent feature of immersive media, aligning with the broader goal of inclusive digital experiences.

7.2. Methodology: a novel approach for subtitle evaluation in 360º media

This section is related to the following objective:

To design and validate a mixed methodological framework for testing subtitles in VR 360º media content using eye-tracking technology in real time.

Although there are numerous ways of rendering subtitles in 360º video, there is still no straightforward way of evaluating which is the most suitable for universal adoption. While technology has afforded increased personalisation possibilities (Biswas et al., 2021), a critical requirement remains: the need for usability testing (UX) to understand user behaviour patterns and to assess the utility and subjective preferences among various subtitle rendering strategies.

First, it is essential that user testing is conducted with real users and working prototypes. In the context of an innovative and sometimes unfamiliar technology, approaches such paper prototypes and focus group tend to yield limited insights. Users need to familiarise themselves and immerse within the test environment. Previous studies have found that this often results in preference for the status quo or what is familiar from traditional 2D media (Brown et al., 2018; Rothe et al., 2018; Agulló et al., 2018; Montagud-Climent et al., 2021).

Regarding UX evaluation, two forms of response can be solicited from users: subjective impressions and objective measures. Subjective impressions are used to gauge the user's sense of presence, affective response, and potential issues such as motion sickness (Bohdanowicz et al., 2020). These impressions are often obtained through post-experience questionnaires administered after an immersive experience. However, such tools have limitations, as self-reporting can be incomplete and may not capture the real user experience. In contrast, objective measures can provide a more quantitative understanding of user task performance within the immersive environment. Some modern VR HMD now incorporate built-in eye-tracking capabilities, enabling the recording of objective performance metrics on how users interact with subtitles. These metrics may encompass elements such as the number and duration of gaze fixations on subtitles, timing metrics (e.g., time taken to fixate on a subtitle), and the number of fixated words within a subtitle, among others.

In light of the above, this dissertation is grounded in a novel framework (Hughes et al., 2020) that enables contrast and comparison in the evaluation of subtitles. It also adopts a mixed-methods approach, drawing inspiration from previous studies in 2D content that utilised eye-tracking technology (Fox, 2018). This methodology aims to provide a holistic understanding of user interactions with subtitles in 360º environments, bridging the gap between traditional subjective assessments and more precise objective measures.

7.3. Subtitling for immersive media: the relevance of the stimuli

> This section is related to the following objective:
>
> **To evaluate two different subtitling solutions in VR 360º media content:**
>
> Position of subtitles: head-locked or fixed
>
> Colour of subtitles: monochrome or coloured

In the context of 360º video, storytelling is evolving in parallel with traditional mediums. In a study conducted by Elmezeny et al. (2018), it becomes evident that the immersive experience is woven with both technical and narrative elements. From a technical standpoint, immersion is achieved through strategic cues that guide the viewer's attention within the 360º environment. However, the immersive narrative experience extends beyond the application of technical cues. It is influenced by the setting and the dynamic interplay between the story, its characters, and the integration of the viewer. These elements mutually reinforce one another, enriching the viewer's sense of immersion. Nevertheless, the language of 360º storytelling is still being invented (Warren, 2017), and there is the potential for future developments to combine advanced technical aspects with pre-existing narrative approaches.

Within the scope of this study, a contribution lies in the formulation of a set of prerequisites regarding the stimuli used when testing subtitles within 360º videos. These prerequisites encompass both technical and narrative aspects and are designed to emulate the intricate interplay between technical and narrative dimensions found in immersive storytelling. It is worth noting that previous studies exploring subtitling in 360º videos have omitted descriptions of the stimuli used, leaving an assumed absence of specific prerequisites.

The established prerequisites include:

- Strategic character placement: Speaking characters were positioned throughout the 360º scene to better exploit the 360º space and allowing users to engage with the narrative from multiple angles.
- Avoidance of character overlaps: An effort was made to ensure that speaking characters within the scene did not overlap in their dialogue. The reason is that when two characters located at different points within the 360º environment spoke

simultaneously, it could render subtitles unreadable, undermining the immersive experience.

– Language disparity: The language spoken within the immersive content differed from the native language of the study participants, forcing them to read the subtitles to follow the plot.

In essence, these prerequisites provide a structured framework for testing subtitles in 360º environments. By considering both the technical and the narrative dimension the credibility of research outcomes is reinforced. Moreover, such experiments contribute to the ongoing development of immersive storytelling.

7.4. Subtitling strategies in immersive media

This section is related to the following objective:

To evaluate two different subtitling solutions in VR 360º media content:

Position of subtitles: head-locked or fixed

Colour of subtitles: monochrome or coloured

One of the main goals of this dissertation was to evaluate different subtitling strategies in immersive media (360º content). After a review of the existing literature (Chapter 2 and 3) and the development and testing of the methodological framework (Chapter 4), the following strategies were compared:

1. Position: head-locked (subtitles are always visible, displayed in front of the viewer versus fixed) versus fixed (subtitles appear near the speaking character and remain fixed to the scene)

2. Colour: monochrome (all subtitles are displayed in white on a grey background) versus coloured (different colours to identify the different characters)

These strategies were tested in a study with 112 participants: 73 hearing and 39 with hearing loss. The purpose of the test was to verify and complement previous findings by quantifying the impact of these strategies on user behaviour by measuring eye movements. In this study, colours were tested for the first time following the ISO 20071-23:2018 standard. The potential benefits of using colour to distinguish speakers have been demonstrated in previous research involving multilingual 2D content (Szarkowska & Boczkowska, 2022). However, this feature had not been treated as a variable in previous studies on immersive content (Brown et al., 2018; Rothe et al., 2018; Agulló et al., 2018). To conduct this investigation, a mixed-methodology approach was used, combining questionnaires and eye movement recordings.

Two hypotheses were stablished in regard to this goal, and both were confirmed by the results:

H2: Head-locked subtitles allow for better scene comprehension compared to fixed subtitles. Head-locked subtitles are the preferred option.

This hypothesis was formulated based on two premises. On one hand, prior studies have favoured head-locked subtitles as the preferred option when watching immersive content. On the other hand, head-locked subtitles behave similar to traditional 2D subtitles which are placed always in front of the viewer, and at the bottom of the screen. This alignment with familiar habits and practices (Bartoll & Martínez-Tejerina, 2010) led to expect that head-locked subtitles would be easier to follow.

Analysis of eye movement data confirm findings from previous studies, which relied on the Igroup Presence Questionnaires (Agulló and Matamala, 2020). The observed difference in fixation count generally shows a greater proportion of gaze on head-locked than fixed subtitles. However, this can be explained as head-locked subtitles are more readily fixated because they are "always there". Fixed subtitles receive fewer fixations because they first need to be located by the viewer, making them potentially more difficult to process. Analysis of gaze data also shows that fewer fixations fall on unique words within subtitles when they are fixed, suggesting viewers have less time to process subtitles "in full" when they are not always available, e.g., in the head-locked style of presentation.

It is relevant to note that gaze data appears to be in stark contrast to self-reported impressions of subtitle reading performance, for both hearing and people with hearing loss. Participants reported that they had thought they read a greater percentage of the fixed subtitles compared to head locked. Gaze analysis shows the opposite: a larger proportion of fixations was made to head-locked subtitles by both participants groups. What is particularly striking about this finding is that how viewers performed differs from how they thought they had performed, at least in terms of objective evidence provided by eye-tracked data. This may suggest that self-reported, questionnaires and/or focus groups-based analyses may not reflect objective measures when exploring subtitles in immersive environments. This may be attributed to self-serving bias which can lead to discrepancies between self-reported and objective performance (Miller & Ross, 1975).

H3: Coloured subtitles allow for better character recognition compared to monochrome subtitles. Coloured subtitles are the preferred option.

This hypothesis is based on an established practice in traditional 2D content. As suggested by Mälzer-Semlinger (2015), attributing a consistent colour to each character throughout the entire film helps in identifying the speaker when reading subtitles.

Eye-tracked data showed that fixation counts on colour subtitles were significantly greater than in monochrome subtitles. Furthermore, coloured subtitles drew more fixations on unique words compared to monochrome subtitles. This trend was especially pronounced among participants with hearing loss. This can be explained because they are more used to coloured subtitles than hearing participants (Neves, 2007).

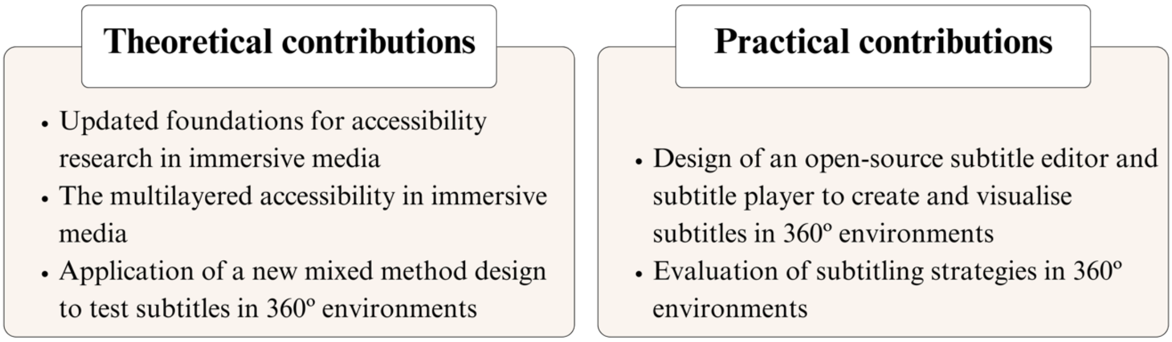
These findings confirm the effectiveness of coloured subtitles in facilitating character recognition and improving the overall reading experience, aligning with the conclusions drawn from traditional 2D content.

7.5. Summary of main contributions, limitations, and future lines of research

The previous subsections summarised all the contributions of this dissertation. On one hand, the theoretical implications that have been discussed include and updated state-of-the-art for accessibility in immersive media with a highlight in the paradigm of multi-layered accessibility within immersive environments. That leads to the design and test of a new mixed method approach used for testing subtitles in 360º environments. On the other hand, the practical implications have focused on subtitles in 360º video. An open-source subtitle editor and player have been developed, and different subtitle solutions in 360º contents have been tested in four different countries with different media consumption habits. This reception study, the largest so far concerning this topic (N = 112), included both hearing participants and people with hearing loss. The stimuli used adhere to a pre-established set of requirements that ensure the validity of the findings. The last contribution has been presenting recommendations for rendering subtitles in 360º videos (i.e., head-locked, coloured). Both theoretical and practical contributions are summarised in figure 31 below.

Figure 31

Summary of theoretical and practical contributions of this dissertation



The study also faced some limitations. First, technical constraints derived from the newness of the technologies that could not be overcome at the time of the study. For example, the video quality of the stimuli, which could have been further improved using 4K resolution. Also, audio quality could have been improved using binaural audio, a technique that is becoming popular in immersive experiences. These limitations were noted by some of the participants and may have influenced user experiences and their interaction with subtitles. Likewise, the novelty of the technology itself may have affected participant reactions, as many lacked previous experience with immersive technologies. The level of user expertise is a relevant variable in assessing novel technologies, although finding a sufficient number of expert users for in-person studies can be challenging. A potential solution for future research could involve designing a prototype for remote testing, allowing for the recruitment of participants who already have access to the required HMDs at home.

Second, the type of content selected for the study. Finding immersive, royalty-free content that also met the prerequisites for subtitle testing in 360º environments posed a notable challenge. To ensure meaningful results, a specific set of content requirements was established. Then, several videos were recorded in a non-professional setting, featuring non-professional actors. These recordings were captured using an Insta360 One X2 camera and lacked the use of dedicated microphones. The choices regarding content and production conditions might have influenced participants reactions and presence levels during the study.

With the research data collected, numerous opportunities for future work become evident. First, an expansion of the scope from 360º environments to encompass the broader spectrum of XR technology presents many opportunities. In particular AR is becoming more integrated into our daily lives through smartphones and AR glasses. Likewise, the rise of the metaverse opens new avenues for the implementation of accessibility services. In fact, the standardisation agency ITU has initiated a working group on accessibility and inclusion within the metaverse. The efforts devoted to exploring subtitle visualisation within 360º environments in this PhD can be adapted and extended to AR and the metaverse.

Moreover, a recent report by Gartner[26] highlights the essential role of integrating eye-tracking in an HMD for enhancing user experiences in the metaverse. This technology has a range of potential applications, including controlling and navigating immersive content, improving the rendering of graphics for more lifelike experiences, and collecting valuable insights from the users (such as behaviour, emotions, or expressions). Future research can delve deeper into understanding how users engage in such an interactive medium. The problem that arises in this case is the same as that of the Inclusive User Model (Biswas et al., 2021), since in any case, security aspects and local regulations and legislation must be considered.

Regarding future developments, a promising avenue involves the creation of plug-ins designed to incorporate accessibility services when creating spaces within the metaverse. While the current dissertation provides insights into rendering subtitles in immersive spaces, the objective would be to craft a versatile plug-in capable of facilitating the integration of a wide spectrum of accessibility services, ranging from AD and sign language to easy-to-read. Such a plug-in would not only empower content creators to embrace accessibility but also foster a more inclusive metaverse for diverse user communities.

In conclusion, this dissertation has explored accessibility in immersive media, paving the way for inclusive experiences within 360º environments. By highlighting the multilayered accessibility in immersive media and applying a mixed methodological framework, this work contributes to the ongoing evolution of accessibility standards and practices. As we look to the horizon of emerging XR technologies and the metaverse, the lessons learned, and the paths forged in this research can guide future endeavours towards creating truly inclusive digital spaces for all.

[26] https://www.gartner.com/en/documents/4744231

CHAPTER 8

8. Updated bibliography

This is a unified bibliography for the dissertation, following referencing rules from APA 7th Edition.

Agulló, B. (2020). *Subtitling in immersive media: A user-centered study* [Doctoral dissertation, Universitat Autònoma de Barcelona]. http://hdl.handle.net/10803/670322

Agulló, B. (2020). Technology for subtitling: A 360-degree turn. *Hermeneus. Revista de Traducción e Interpretación*, *22.*

Agulló, B., & Matamala, A. (2019). Subtitling for the deaf and hard-of-hearing in immersive environments: results from a focus group. *Jostrans. The Journal of Specialised Translation, 32*, 217–235.

Agulló, B., & Matamala, A. (2020). Subtitles in virtual reality: Guidelines for the integration of subtitles in 360º content. *Íkala, Revista De Lenguaje Y Cultura, 25*(3), 643–661. https://doi.org/10.17533/udea.ikala.v25n03a03

Agulló, B., & Orero, P. (2017). 3D movie subtitling: Searching for the best viewing experience. *Studi di Comunicazione e Mediazione linguistica e culturale COME, 2*(1), 91–102.

Agulló, B., Matamala, A., and Orero, P. (2018). From disabilities to capabilities: Testing subtitles in immersive environments with end users. *HIKMA 17*, 195–220. https://doi.org/10.21071/hikma.v17i0.11167

Agulló, B., Montagud, M., & Fraile, I. (2019). Making interaction with virtual reality accessible: Rendering and guiding methods for subtitles. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 33*(4), 416–428. https://doi.org/10.1017/ S0890060419000362

Aissaoui, N. (2021). The digital divide: a literature review and some directions for future research in light of COVID-19. *Global Knowledge, Memory, and Communication.* https://doi.org/10.1108/GKMC-06-2020-0075

Álvarez de Morales Mercado, C. (2015). El subtitulado para personas sordas como discurso narratológico. *Sendebar, 26*, 57–81.

Arane, K., Behboudi, A., & Goldman, R.D. (2017). Virtual reality for pain and anxiety management in children. *Canadian Family Physician, 63*(12), 932–934.

Arnáiz Uzquiza, V. (2012). Los parámetros que identifican el subtitulado para sordos: análisis y clasificación. *MonTi: Monografías de Traducción e Interpretación, 4,* 103–32. https://www.raco.cat/index.php/MonTI/article/view/301235

Arnáiz-Uzquiza, V. (2012). *Subtitling for the deaf and the hard-of-hearing: Some parameters and their evaluation* [Doctoral dissertation, Universitat Autònoma de Barcelona]. https://www.tdx.cat/handle/10803/117528

Ashcraft, M. H., & Radvansky, G. A. (2014). *Cognition*. Pearson Education.

Aulavuori, K. (2008). *Effects of the digital subtitling software on the subtitling process: a survey among Finnish professional television subtitlers*. [Master's thesis, University of Helsinki].

Aylett, R., & Louchart, S. (2003). Towards narrative theory of virtual reality. *Virtual Reality*, *7*, 2–9. https://doi.org/10.1007/s10055-003-0114-9

Azuma, R., Baillot, Y., Behringer, R., Feiner, S.K., Julier, S., & MacIntyre, B. (2001). Recent advances in augmented reality. *IEEE Computer Graphics and Applications, 21*(6), 34–47.

Ballester Casado, A. (1998). *Traducción y nacionalismo. La recepción del cine americano en España a través del doblaje, 1920-1948.* [Translation and Nationalism. The Reception of American Cinema in Spain through Dubbing, 1920-1948]. Comares.

Bartoll, E. (2004). Parameters for the classification of subtitles. In P. Orero (Ed.), *Topics in audiovisual translation* (pp. 53–60). John Benjamins Publishing Company. https://doi.org/10.1075/btl.56.08bar

Bartoll, E., & Martínez-Tejerina, A. (2010). The positioning of subtitles for the deaf and hard of hearing. In A. Matamala, & P. Orero (Eds.), *Listening to subtitles: Subtitles for the deaf and hard of hearing* (pp. 69–86). Peter Lang.

BBC. (2022). *Subtitle guidelines*. Version 1.2.1. https://www.bbc.co.uk/accessibility/forproducts/guides/subtitles/

Beaunoyer, E., Dupéré, S., & Guitton, M. J. (2020). COVID-19 and digital inequalities: Reciprocal impacts and migration strategies. *Computer in Human Behavior, 111*(6).

Belton, J. (2012). Digital 3D cinema: Digital cinema's missing novelty phase. *Film history: An international journal*, *24*(2), 187–195.

Bernabé Caro, R., & Orero, P. (2019). Easy to read as multimode accessibility services. *Hermeneus. Revista de Traducción e Interpretación*, *21,* 53–74.

Biswas, P., Orero, P., Swaminathan, M., Krishnaswamy, K., & Robinson, P. (2021). Adaptive acccessible AR/VR systems. *CHI '21 Extended Abstracts on Human Factors in Computing Systems.*

Black, S. (2020). Could integrated subtitles benefit young views? Children's reception of standard and integrated subtitles: a mixed methods approach using eye tracking. *Perspectives, 30*(3), 503–519. https://doi.org/10.1080/0907676X.2020.1849324

Bolaños-García-Escribano, A., Díaz-Cintas, J., & Massidda, S. (2021). Subtitlers on the cloud: The use of professional web-based systems in subtitling practice and training. *Revista Tradumàtica. Tecnologies de la Traducció, 19*, 1–21.

Bosch-Baliarda, M., Soler-Vilageliu, O., & Orero, P. (2020). Sign Language Interpreting on TV: A Reception Study of Visual Screen Exploration in Deaf Signing Users. *MonTi: Monografías de Traducción e Interpretación, 12*, 108–143. https://doi.org/10.6035/MonTI.2020.12.04

Bouwels, L. (2022). *Accessible subtitles in 6DoF 3D virtual reality: How preference of subtitle presentation modes relates to presence and VR sickness* [Master's thesis, Breda University of Applied Sciences].

Brennan, D. (2017, December 6). *New report from ILMxLab & Disability Visibility Project shares insights on VR accessibility design.* .https://www.roadtovr.com/disability-visibility-project-and-ilmxlab-share-lessons-learned-about-vr-accessibility-design/

Brescia-Zapata, M. (2022). The present and future of accessibility services in VR360 players, *inTRAlinea, 24.*

Brescia-Zapata, M., Krejtz, K., Orero, P., Duchowski, A. T., & Hughes, C. J. (2022). VR 360º subtitles: Designing a test suite with eye-tracking technology. *Journal of Audiovisual Translation, 5*(2), 233–258. https://doi.org/10.47476/jat.v5i2.2022.184

Brescia-Zapata, M. (in press). Accessibilities in immersive media environments. In M. J. Varela Salinas & C. Plaza Lara (Eds.), *Aproximaciones teóricas y prácticas a la accesibilidad desde la traducción y la interpretación.* Comares.

Brescia-Zapata, M., Krejtz, K., Orero, P., Duchowski, A. T., & Hughes, C. J. (in press). Subtitles in VR 360º video. Results from an eye-tracking experiment. *Perspectives: Studies in Translation Theory and Practice.* https://doi.org/10.1080/0907676X.2023.2268122

Brown, A. (2017, October 26). *User testing subtitles for 360º content. BBC.* https://www.bbc.co.uk/rd/blog/2017-10-subtitles-360-video-virtual-reality-vr

Brown, A., & Patterson, J., (2017, October 26). *Designing subtitles for 360º content.* BBC. https://www.bbc.co.uk/rd/blog/2017-03-subtitles-360-video-virtual-reality

Brown, A., Turner, J., Patterson, J., Schmitz, A., Armstrong, M. & Glancy, M. (2018, April 30). *Exploring subtitle behavior for 360º video.* https://www.bbc.co.uk/rd/publications/whitepaper330

Brown, A., Turner, J., Patterson, J., Schmitz, A., Armstrong, M., & Glancy, M. (2017). Subtitles in 360-degree video. *Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video,* 3–8. https://doi.org/10.1145/3084289.3089915

Bruner, J. (2004). Life as narrative. *Social research: An international quarterly 71*(3), 691–710. http://doi.org/10.1353/sor.2004.0045

Bruni, L. E., Kadastik. N., Pedersen, T. A., & Dini, H. (2022). Digital narratives in Extended Realities. In M. Alcañiz, M. Sacco, & J. G. Trump (Eds.), *Roadmapping Extended Reality: Fundamentals and Applications* (pp. 35-62). Scrivener Publishing LLC.

Bryman, A. (2004). *Social research methods*. Oxford University Press.

Caffrey, C. (2012). Using an eye-tracking tool to measure the effects of experimental subtitling procedures on viewer perception of subtitled AV content. In E. Perego (Ed.), *Eye tracking in audiovisual translation* (pp. 223–258). Aracne. http://doi.org/10.4399/97888548491369

Cantero de Julián, J. I., Calvo Rubio, L. M., & Benedicto Solsona, M. Á. (2020). La tenue apuesta por los vídeos en 360º en las estrategias transmedia de las televisiones autonómicas españolas. [The weak bet on videos in 360º in the transmedia strategies of Spanish autonomic televisions]. *Revista Latina de Comunicación Social*, *75*, 415–433. https://doi.org/https://doi.org/10.4185/RLCS-2020-1433

Carmigniani, J., and Furht, B. (2011). Augmented reality: An overview. In B. Furht (Ed.), *Handbook of Augmented Reality* (pp. 3–46). Springer.

Carrozzino, M., & Bergamasco, M. (2010). Beyond virtual museums: Experiencing immersive virtual reality in real museums. *Journal of cultural heritage, 11*(4), 452–458.

Chaume, F. (2012). *Audiovisual Translation: Dubbing.* St Jerome Publishing.

Chaume, F. (2013). The turn of audiovisual translation: New audiences and new technologies. *Translation Spaces 2*(1), 105–123. https://doi.org/10.1075/ts.2.06cha

Civera, C., & Orero, P. (2010). Introducing icons in subtitles for deaf and hard of hearing: Optimising reception? In A. Matamala, & P. Orero (Eds.), *Listening to subtitles: Subtitles for the deaf and hard of hearing* (pp. 149-162). Peter Lang.

Crafton, D. (1999). *The talkies: American cinema's transition to sound, 1926-1931*. University of California Press.

Creed, C., Al-Kalbani, M., Theil, A., Sarcan, S., & Williams, I. (2023). Inclusive AR/VR: Accessibility barriers for immersive technologies. *Universal Access in the Information Society*. https://doi.org/10.1007/s10209-023-00969-0

d'Ydewalle, G., & De Bruycker, W. (2007). Eye movements of children and adults while reading television subtitles. *European Psychologist 12*(3). 196–205.

d'Ydewalle, G., Praet, C., Verfaillie, K. & Van Rensbergen J. (1991). Watching subtitled television: Automatic reading behaviour. *Communication Research 18*(5), 650–660.

Danieau, F., Guillo, A., & Dore, R. (2017). Attention guidance for immersive video content in head-mounted displays. *2017 IEEE Virtual Reality (VR)* (pp. 205-206). http://doi.org/10.1109/VR.2017.7892248.

Di Giovanni, E., & Gambier, Y. (Eds.) (2018). *Reception studies and audiovisual translation*. John Benjamins.

Díaz-Cintas, J. (2003). Audiovisual translation in the third millennium. In G. Anderman, & M. Rogers (Eds.), *Translation today: Trends and perspectives* (pp. 192–204). Multilingual Matters.

Díaz-Cintas, J. (2005). Audiovisual translation today. A question of accessibility for all. *Translating Today*, 3–5.

Díaz-Cintas, J. & Remael, A. (2007). *Audiovisual translation: Subtitling. (Translation practices explained)*. St. Jerome Publishing.

Domínguez, E. (2017). Going beyond the classic news narrative convention: The background to and challenges of immersion in journalism. *Frontiers in Digital Humanities, 4*, 1–11. https://doi.org/10.3389/fdigh.2017.00010

Duchowski, A. T., Krejtz, K., Zurawska, J., & House, D. H. (2020). Using microsaccades to estimate task difficulty during visual search of layered surfaces. *IEEE Transactions on Visualization and Computer Graphics, 26*(9), 2904–2918. https://doi .org/10.1109/TVCG.2019.2901881

Eglinton, A. (2010). Reflections on a Decade of Punchdrunk Theatre. *TheatreForum 37,* 46–55.

Eiersø, C. (2018, Mai 15). *Smart icons design — common European standardization.* https://www.dr.dk/om-dr/about-dr/smart-icons-design-common-european-standardization

Elmezeny, A., Edenhofer, N., & Wimmer, J. (2018). Immersive storytelling in 360-degree videos: An analysis of interplay between narrative and technical immersion. *The Journal of Virtual Worlds Research, 11*(1). https://doi.org/10.4101/jvwr.v11i1.7298

EN 17161. (2019). *Design for All - Accessibility following a Design for All approach in products, goods and services - Extending the range of users.* https://standards.cen.eu/dyn/www/f?p=204:110:0::::FSP_PROJECT,FSP_ORG_ID:62323,2301962&cs=1D28CFDC66E7CEF3CE441294CAA9FEABE

European Broadcasting Union (EBU). (2017). *Virtual reality: How are public broadcasters using it?* https://www.ebu.ch/publications/virtual-reality-how-are-public-broadcasters-using-it

Fida, B., Cutolo, F., di Franco, G., Ferrari, M., & Ferrari, V. (2018). Augmented reality in open surgery. *Updates in Surgery, 70,* 389–400.

Fidyka, A., & Matamala, A. (2018). Audio description in 360º videos: Results from focus groups in Barcelona and Kraków. *Translation Spaces*, *7*(2), 285–303. https://doi.org/10.1075/ts.18018.fid

Foerster, A. (2010). Towards a creative approach in subtitling: A case study. In J. Díaz-Cintas, A. Matamala, & J. Neves (Eds.), *New insights into audiovisual translation and media accessibility: Media for all 2* (pp. 81–98). Rodopi.

Fox, W. (2016). Integrated titles: an improved viewing experience? In S. Hansen-Schirra, & S. Grucza (Eds.), *Eyetracking and Applied Linguistics* (pp. 5–30). Language Science Press.

Fox, W. (2018). *Can integrated titles improve the viewing experience? Investigating the impact of subtitling on the reception and enjoyment of film using eye tracking and questionnaire data.* Language Science Press. https://doi.org/10.5281/zenodo.1180721

Fryer, L. (2016). *An introduction to audio description. A practical guide.* Routledge.

Fuhrmann, M. (1992). *Die Dichtungstheorie der Antike*. Wissenschaftliche Buchgesellschaft, Darmstadt.

Gadin, V. (2021). *Factors for good text legibility: Eye-tracking in virtual reality.* [Master thesis]. Uppsala Universitat.http://www.diva-portal.org/smash/get/diva2:1572637/FULLTEXT01.pdf

Gambier, Y. (2003). Introduction: Screen Transadaptation: Perception and Reception. *The translator, 9*(2) 171-18. https://doi.org/10.1080/13556509.2003.10799152

Gambier, Y. (2013). The position of audiovisual translation studies. In C. Millán & F. Bartrina (Eds.), *The Routledge Handbook of Translation studies* (pp. 45–59). Routledge.

Gejrot, E., Kjellstrand, S., & Laurin, S. (2021). New realities unlocking the potential of XR for persons with disabilities. Tech. rep., Funka Nu AB.

Gerber-Morón, O., Soler-Vilageliu, O., & Castellà, J. (2020). The effects of screen size on subtitle layout preferences and comprehension across devices. *Hermēneus. Revista de Traducción e Interpretación*, *22*, 157–182.

Ghia, E. (2012). The impact of translation strategies on subtitle reading. In E. Perego (Ed.), *Eye-tracking and audiovisual translation* (pp. 155–181). Aracne.

Gil Ariza, M. C. (2004). A case study: Spain as a dubbing country. *Translation journal*, *29.* http://translationjournal.net/journal/29movies.html

Gómez Suárez, M., & Charron, J-P. (2017). Exploring concert-goers' intention to attend virtual music performances. [Presentation]. *AIMAC 14 International Conference on Arts and Cultural Management*, Beijing.

González-Zúñiga, D., Carrabina, J., & Orero, P. (2013). Evaluation of Depth Cues in 3D Subtitling. *Online Journal of Art and Design*, *1*(3), 16–29.

Gottlieb, H. (1992). Subtitling – a new university discipline. In C. Dollerup, C. & A. Loddegaard (Eds.), *Teaching translation and interpreting* (pp. 161-170). John Benjamins Publishing Company.

Gottlieb, H. (1994). Subtitling: Diagonal translation. *Perspectives: Studies in Translation Theory and Practice, 2*(1), 101-121. http://doi.org/10.1080/0907676X.1994.9961227

Gottlieb, H. (1995). Establishing a framework for a typology of subtitle reading strategies - viewer reactions to deviations from subtitling standards. *Communication Audiovisuelle et Transferts Linguistiques - FIT Newsletter, 14*(3-4), 388–409.

Gottlieb, H. (1998). Subtitling. In M. Baker (Ed.), *Routledge Encyclopaedia of Translation Studies* (pp. 244–248). Routledge.

Greco, G. M. (2016). On Accessibility as a Human Right, with an Application to Media Accessibility. In A. Matamala & P. Orero (Eds.), *Researching audio description. New approaches* (pp. 11–33). Palgrave Macmillan. https://doi.org/10.1057/978-1-137-56917-2_2

Greco, G. M. (2019). Accessibility Studies: Abuses, Misuses and the Method of Poietic Design. In C. Stephanidis (Ed.), *HCI International 2019 – Late Breaking Papers. HCII 2019. Lecture Notes in Computer Science, 11786*, 15-27. Springer. https://doi.org/10.1007/978-3-030-30033-3_2

Greener, R. (2022, December 19). *Virtual reality statistics to know in 2023.* XR today. https://www.xrtoday.com/virtual-reality/virtual-reality-statistics-to-know-in-2023/#:~:text=In%202020%2C%20roughly%2057.4%20people,%2C%20industry%2C%20and%20education%20adoption.

Guitton, D., Munoz, D. P., & Galiana, H. L. (1990). Gaze Control in the Cat: Studies and Modeling of the Coupling Between Orienting Eye and Head Movements in Different Behavioral Tasks. *Journal of Neurophysiology, 64*(2), 509–531. https://doi.org/10.1152/jn.1990.64.2.509

Gunkel, S., Stokking, H. M., Prins, M. J., van der Stap, N., ter Haar, F. B. & Niamut, O. A. (2018, June 15). *Virtual reality conferencing: Multi-user immersive VR experiences on the web.* http://doi.org/10.5281/zenodo.3530681

Guttentag, D. (2010). Virtual reality: Applications and implications for tourism, *Tourism Management*, *31*(5), 637–651.

Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50,* 904–908.

Hart, S. G., & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati, *Advances in Psychology, 52,* 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

Hermosa-Ramírez, I. (2022). Physiological instruments meet mixed methods in Media Accessibility. *Translation Spaces, 11*(1), 38–59. https://doi.org/10.1075/ts.21020.her

Hillaire, S., Lecuyer, A., Cozot, R., & Casiez, G. (2008). Depth-of-field blur effects for first-person navigation in virtual environments. *IEEE Computer Graphics and Applications, 28*(1), 47–55. http://doi.org/10.1109/MCG.2008.113

Hoffmann, P., & Herczeg, M. (2006) Hypervideo vs. Storytelling. Integrating narrative intelligence into Hypervideo. In S. Göbel, R. Malkewitz, & I. Iurgel (Eds.), *Technologies for Interactive Digital Storytelling and Entertainment* (pp. 37–48).

Holliman, N., Dodgson, N. A., Favalora, G. E., & Pockett, L. (2011). Three-dimensional displays: A review and applications analysis, *IEEE Trans. Broadcast*, *57*(2), 362–371.

Holmes, J. S. (1972). The Name and Nature of Translation Studies. In J. S. Holmes (Ed.), *Translated! Papers on Literary Translation and Translation Studies* (pp. 66–80). Rodopi. http://doi.org/10.7202/000580ar

Hughes, C. J., & Montagud-Climent, M. (2021). Accessibility in 360° video players. *Multimedia Tools and Applications, 80,* 30993–31020. https://doi.org/10.1007/s11042-020-10088-0

Hughes, C. J., Armstrong, M., Jones, R., & Crabb, M. (2015). Responsive design for personalised subtitles. *Proceedings of the 12th International Web for All Conference*, 1–4. https://doi.org/10.1145/2745555.2746650

Hughes, C. J., Brescia-Zapata, M., Johnston, M., & Orero, P. (2020). Immersive captioning: Developing a framework for evaluating user needs. *IEEE AIVR 2020: 3rd International Conference on Artificial Intelligence & Virtual Reality 2020*. http://doi.org/10.1109/AIVR50618.2020.00063

Hughes, C. J., Montagud-Climent, M., & tho Pesch, P. (2019). Disruptive approaches for subtitling in immersive environments. *Proceedings of the 2019 ACM International*

Conference on Interactive Experiences for TV and Online Video, 216–229. https://doi.org/10.1145/3317697.3325123

Ibourk, E. M., & Al-Adwan, A. (2019). Subtitling Virtual Reality into Arabic: Eye Tracking 360-Degree Video for Exploring Viewing Experience. *Lebende Sprachen, 64*(2), 286-308.

International Organization for Standardisation. (2018). *Information technology — User interface component accessibility — Part 23: Visual presentation of audio information (including captions and subtitles).* (ISO/IEC 20071) https://www.iso.org/standard/70722.html

International Organization for Standardization (2015) *ISO/IEC TS 20071-20 Information Technology — User interface component accessibility — Part 20: Software for playing media content,* ISO.

Jakobson, R. (1959). On linguistic aspects of translation. In Brower, R. A. (Ed.), *On Translation* (pp. 232–239). Harvard University Press.

Jankowska, A. (2015). *Translating Audio Description Scripts*. Peter Lang Verlag. https://doi.org/10.3726/978-3-653-04534-5

Jenkins, H., Ford, S., & Green, J. (2015). *Cultura transmedia. la creación de contenido y valor en una cultura en red.* Gedisa.

Jensema, C., McCann, R., & Ramsey, S. (1996). Closed-captioned television presentation speed and vocabulary. *American Annals of the Deaf, 141*(4), 284–292.

Just, M. A., & Carpenter, P. A. (1976). The role of eye- fixation research in cognitive psychology. *Behavior Research Methods & Instrumentation, 8*(2), 139–143. https://doi.org/10.3758/BF03201761

Karamitroglou, F. (1998). *A proposed set of subtitling standards in Europe*. https://translationjournal.net/journal/04stndrd.htm

Ko Du., Ryu, H., Kim, J. (2018). Making new narrative structures with actor's eye-contact in cinematic virtual reality (CVR). In R. Rouse, H. Koenitz, & M. Haahr (Eds.), *Interactive Storytelling: 11th International Conference on Interactive Digital*

Storytelling, ICIDS 2018, 11318, (pp. 343–347). https://doi.org/10.1007/978-3-030-04028-4_38

Krejtz, I., Szarkowska, A., & Łogińska, M. (2016). Reading Function and Content Words in Subtitled Videos. *The Journal of Deaf Studies and Deaf Education, 21*(2), 222–232. https://doi.org/10.1093/deafed/env061

Krejtz, I., Szarkowska, A., Krejtz, K., Walczak, A., & Duchowski, A. (2012). Audio description as an aural guide of children's visual attention: evidence from an eye-tracking study. *Proceedings of the ACM Symposium on Eye Tracking Research and Applications Conference* (pp. 99–106). ACM

Krejtz, K., Çöltekin, A., Duchowski, A., & Niedzielska, A. (2017). Using Coefficient K to Distinguish Ambient/Focal Visual Attention During Map Viewing. *Journal of Eye Movement Research, 10*(2). https://doi.org/10.16910/jemr.10.2.3

Krejtz, K., Duchowski, A., Krejtz, I., Szarkowska, A., & Kopacz, A. (2016). Discerning Ambient/Focal At- tention with Coefficient K. *ACM Transactions on Applied Perception, 13*(3), 11:1–11:20. https://doi. org/10.1145/2896452

Krejtz, K., Wisiecka, K., Krejtz, I., Holas, P., Olszanowski, M., & Duchowski, A. T. (2018). Dynamics of Emotional Facial Expression Recognition in Individuals with Social Anxiety. *Proceedings of the 2018 ACM Symposium on Eye-tracking Research & Applications*, 43:1–43:9. https://doi.org/10.1145/3204493.3204533

Kruger, J. (2016). Psycholinguistics and Audiovisual Translation. *Target, 28*(2), *276–287.* https://doi.org/10.1075/target.28.2.08kru

Kruger, J. L., & Steyn, F. (2014). Subtitles and eye tracking: Reading and performance. *Reading Research Quarterly 49*(1), 105–120. https://doi.org/10.1002/rrq.59

Kruger, J. L., Szarkowska, A., & Krejtz, I. (2015). Subtitles on the Moving Image: An Overview of Eye Tracking Studies. *Refractory: A Journal of Entertainment Media, 25*, 1–14.

Kurzhals, K. Cetinkaya, E., Hu, Y., Wang, W., & Weiskopf, D. (2017). Close to the Action: Eye-Tracking Evaluation of Speaker-Following Subtitles, *Proceedings of the 2017 CHI*

Conference on Human Factors in Computing Systems (CHI '17), 6559–6568. https://doi.org/10.1145/3025453.3025772

LaViola, J. J. (2000). A discussion of cybersickness in virtual environments. *ACM Sigchi Bulletin, 32*(1), 47–56. http://doi.org/10.1145/333329.333344

Lee, D. G., Fels, D. I., & Udo, J. P. (2007). *Emotive captioning. Comput. Entertain*, *5*(2). https://doi.org/10.1145/1279540.1279551

Lee, K., & Chung, N. (2008). Empirical analysis of consumer reaction to the virtual reality shopping mall. *Computers in Human Behavior*, *24,* 88–104.

Liao, S., Yu, L., Reichle, E. D., and Kruger, J. L. (2021). Using eye movements to study the reading of subtitles in video. *Scientific Studies of Reading, 25*(5), 417–35.

Liberatore, M. J., & Wagner, W. P. (2021). Virtual, mixed, and augmented reality: a systematic review for immersive systems research. *Virtual Reality, 25,* 773–799. https://doi.org/10.1007/s10055-020-00492-0

Linder, K. (2016). *Student uses and perceptions of closed captions and transcripts: Results from a national study.* Oregon State University Ecampus Research Unit.

Liu, Y. M., Zheng, B. H., & Zhou, H. (2019). Measuring the difficulty of text translation: the combination of text-focused and translator-oriented approaches. *Targets, 31*, 125–149. https://doi.org/10.1075/target.18036.zhe

Llanes-Jurado, J., Marín-Morales, J., Guixeres, J., & Alcañiz, M. (2020). Development and Calibration of an Eye-Tracking Fixation Identification Algorithm for Immersive Virtual Reality. *Sensors, 20*(17). https://doi.org/10.3390/s20174956

Luyken, G. M. (1991). *Overcoming Language Barriers in Television: Dubbing and Subtitling for the European Audience.* European Institute for the Media.

Mälzer-Semlinger, N. (2015). It's crime time: about the use of colour in SDH. In A. Jankowska, & A. Szarkowska (Eds.), *New Points of View on Audiovisual Translation and Media Accessibility* (pp. 267–278). Peter Lang Verlag. https://doi.org/10.3726/978-3-0353-0761-0

Mangiron, C. (2013). Subtitling in game localisation: a descriptive study. *Perspectives: Studies in Translation Theory and Practice, 21*(1), 42–56. doi:10.1080/0907676X.2012.722653

Mangiron, C. (2016). Reception of Game Subtitles: An Empirical Study. *The Translator, 22*(1), 72–93. https://doi.org/10.1080/13556509.2015.1110000

Mas Manchón, L., & Orero, P. (2018). Usability tests for personalized subtitles. *Translation Spaces, 7*(2), 263–284. https://doi.org/10.1075/ts.18016.man

Maszerowska, A., Matamala, A., & Orero, P. (2014). *Audio description. New perspectives illustrated.* John Benjamins.

Matamala, A. (2017). Mapping audiovisual translation investigations: research approaches and the role of technology. In M. Deckert (Ed.), *Audiovisual Translation-Research and use* (pp. 11–28). Peter Lang.

Matamala, A., & Orero, P. (2018). Standardising accessibility: Transferring knowledge to society. *Journal of Audiovisual Translation, 1*, 139-154. https://doi.org/10.47476/jat.v1i1.49

Matamala, A., & Orero, P. (eds) (2010). *Listening to subtitles: Subtitles for the deaf and hard of hearing.* Peter Lang.

Matamala, A., Orero, P., Rovira-Esteva, S., Casas-Tost, H., Morales Morante, F., Soler Vilageliu, O., Agulló, B., Fidyka, A., Segura Giménez, D., and Tor-Carroggio, I. (2018). User-centric approaches in Access Services Evaluation: Profiling the End User, *Proceedings of the Eleventh International Conference on Language Resources Evaluation,* 1–7.

Matamala, A., Perego, E., and Bottiroli, S. (2017). Dubbing versus subtitling yet again? An empirical study on user comprehension and preferences in Spain. *Babe, 63*(3), 423–441. https://doi.org/10.1075/babel.63.3.07mat

McClarty, R. (2012). Towards a multidisciplinary approach in creative subtitling. *MonTI: Monografías de Traducción e Interpretación*, *4,* 133–153. https://doi.org/10.6035/MonTI.2012.4.6

McClarty, R. (2014). In support of creative subtitling: Contemporary context and theoretical framework. *Perspectives, 22*(4), 592–606. https://doi.org/10.1080/0907676X.2013.842258

McDonagh, S., & Brescia-Zapata, M. (2023). Combining XR, accessibility, and sustainability in the classroom: results of an exploratory study. In A. Fegely & T. Cherner (Eds.), *Bridging the XT technology-to-practice gap* (pp. 67–79). AACE2023

Meister, L. (2018). On methodology: How mixed methods research can contribute to translation studies. *Translation Studies, 11*(1), 66–83. https://doi.org/10.1080/14781700.2017.1374206

Méndez, R. (2017). Subtítulos y videojuegos. En busca de una norma que mejore la experiencia del usuario. *Quaderns de cine, 12*, 87–103.

Mendiburu, B. (2009). *3D movie making: Stereoscopic digital cinema from script to screen*. Taylor and Francis.

Mével, P. A. (2020). Accessible paratext: Actively engaging (with) d/deaf audiences. *Punctum. International Journal of Semiotics, 6*(1). http://doi.org/10.18680/hss.2020.0010

Mikami, D., Takahashi, K., Saijo, N., Isogawa, M., Kimura, T., & Kimata, H. (2018). Virtual Reality-based Sports Training System and Its Application to Baseball. *NTT Technical Review*, *16*(3).

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, *77*(12), 1321–1329.

Miller, D. T., & Ross, M. (1975). Self-Serving Biases in the Attribution of Causality: Fact or Fiction? *Psychological Bulletin, 82*(2), 213–225.

Mittell, J. (2009). *Forensic fandom and the drillable text.* Spreadable media. https://spreadablemedia.org/essays/mittell/index.html#.YgzhV-5_o4g

Molina-Siles, P. J., Piquer Cases, J. C., & Cortina Maruenda, F. J. (2013). El color en los comienzos del cine. De la aplicación manual al Technicolor. *Libro de actas del X Congreso Nacional del Color*.

Montagud-Climent, M., Fraile Vila, I., Meyerson, E., Genís, M., & Fernández Langa, S. (2019, September 19). *Imac player: Enabling a personalized consumption of accessible immersive content.* https://doi.org/10.6084/m9.figshare.9879254.v1

Montagud-Climent, M., Orero, P., & Fernández, S. (2020b). Immersive media and accessibility: Hand in hand to the future. *ITU Journal*. https://www.itu.int/pub/S-JOURNAL-ICTS.V3I1-2020-2

Montagud-Climent, M., Orero, P., & Matamala, A. (2020a). Culture 4 all: accessibility-enabled cultural experiences through immersive VR360 content. *Personal and Ubiquitous Computing, 24*(6), 887–905. https://doi.org/10.1007/s00779-019-01357-3

Montagud-Climent, M., Orero, P., & Fernández, S. (2020). Immersive media and accessibility: hand in hand to the future. *ITU Journal: ICT Discoveries, 3*(1), 1–8.

Montagud-Climent, M., Soler-Vilageliu, O., Fraile Vila, I., & Fernández Langa, S. (2021). VR360 subtitling: Requirements, technology, and user experience. *IEEE Access, 9*, 819–2838. https://doi.org/10.1109/ACCESS.2020.3047377

Moran, S. (2012). The effect of linguistic variation on subtitle reception. In E. Perego (Ed.), *Eye tracking in audiovisual translation* (pp. 183–222). Aracne.

Moreau, K. A., Eady, K., Sikora, L., & Horsley, T. (2018). Digital storytelling in health professions education: a systematic review, *BMC Medical Education, 18,* 1–9. https://doi.org/10.1186/s12909-018-1320-1

Mott, M., Cutrell, E., Gonzalez Franco, M., Holz, C., Ofek, E., Stoakley, R., & Ringel Morris, M. (2019). Accessible by Design: An Opportunity for Virtual Reality. *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, 451–454. https://doi.org/10.1109/ISMAR-Adjunct.2019.00122

Myers, M. (2020, June 12). *The Last of Us Part 2 review: We're better than this.* Polygon. Vox Media.

Neves, J. (2005). *Audiovisual translation: subtitling for the deaf and hard of hearing.* [Doctoral dissertation]. University of Roehampton. http://hdl.handle.net/10400.8/409

Neves, J. (2008). Ten Fallacies about Subtitling for the d/Deaf and the Hard of Hearing. *Journal of Specialised Translation, 10*, 128–143.

Neves, J. (2018). Subtitling for the deaf and the hard of hearing audiences: moving forward. In L. Pérez González (Ed.), *The routledge handbook of audiovisual translation* (pp. 82–95). Routledge.

Nida, E. A. (1975). *Language Structure and Translation: Essays.* Stanford University Press.

Nord, C. (1997). *Translating as a purposeful activity. Funcionalist approaches explained.* St. Jerome.

Nyström, M., & Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eye-tracking data. *Behaviour Research Methods, 42*(1), 188–204. https://doi.org/10.3758/BRM.42.1.188

Ofcom. (2006). *Television access services review.* Office of Communications.

Ofcom. (2017). *Ofcom's Code on Television Access Services*, 1–25.

Oncins, E., & Orero, P. (2020). No audience left behind, one App fits all: an integrated approach to accessibility services. *Jostrans, 34,* 192–211.

Oncins, E., Bernabé, R., Montagud, M., & Arnáiz Uzquiza, V. (2020). Accessible scenic arts and virtual reality: A pilot study with aged people about user preferences when reading subtitles in immersive environments, *MonTI. Monografías de Traducción e Iterpretación 12*(12), 214–41. https://doi.org/10.6035/MonTI.2020.12.07

Oncins, E., Lopes, O., Orero, P. & Serrano, J. (2013). All Together Now: A multi-language and multi-system mobile application to make live performing arts accessible. *JosTrans, 20,* 147-164.

Orero, P. (2016). From DTV4ALL to HBB4ALL: Accessibility in European Broadcasting. In A. Matamala & P. Orero (Eds.), *Researching Audio Description* (pp. 249–267). Palgrave Macmillan.

Orero, P. (2020). *Born Accessible: Beyond raising awareness*. https://imac-project.eu/wp-content/uploads/2020/05/Policy-Paper_final.pdf

Orero, P. (2022). Audio Description Personalisation. In C. Taylor, & E. Perego (Eds.), *The Routledge Handbook of Audio Description* (pp. 407-419). Routledge. https://doi.org/10.4324/9781003003052-31

Orero, P. (2022). Translation, accessibility, and minorities. In E. Bielsa (Ed.), *The Routledge Handbook of Translation and Media* (pp. 384-399). Taylor Francis.

Orero, P. (ed.) (2004). *Topics in Audiovisual Translation.* John Benjamins.

Orero, P., & Matamala, A. (2007). Accessible opera: overcoming linguistic and sensorial barriers. *Perspectives. Studies in Translation Theory and Practice, 15*(4), 262–277. https://doi.org/10.1080/13670050802326766

Orero, P., and Tor-Carroggio, I. (2018). User requirements when designing learning e-content: interaction for all. In E. Kapros and M. Koutsombogera (Eds.), *Designing for the User Experience in Learning Systems* (pp. 105-121). Springer.

Orero, P., Doherty, S., Kruger, J. L., Matamala, A., Pedersen, J., Perego, E., Romero-Fresco, P., Rovira-Esteva, S., Soler-Vilageliu, O., & Szarkowska, A. (2018). Conducting experimental research in audiovisual translation (AVT): A position paper. *Journal of Specialised Translation, 30*(30), 105–126. https://ddd.uab.cat/record/195062

Orero, P., Hughes, C. J., & Brescia-Zapata, M. (2020). Evaluating subtitle readability in media immersive environments. *DSAI 2020: 9th International Conference on Software development and Technologies for Enhancing Accessibility and Fighting Info-exclusion,* 51–54. https://doi.org/10.1145/3439231.3440602

Orero, P., Rai, S., & Hughes, C., J. (2019). The challenges of audio description in new formats. In S. Braun (Ed.), *Innovation in Audio Description Research*. Routledge.

Papachristos, N. M., Vrellis, I., & Mikropoulos, T. A. (2017). A Comparison between Oculus Rift and a Low-Cost Smartphone VR Headset: Immersive User Experience and Learning. *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*, 477–481. http://doi.org/10.1109/ICALT.2017.145

Parham, G., Bing, E.G., Cuevas, A., Fisher, B., Skinner, J., Mwanahamuntu, M., & Sullivan, R. (2019). Creating a low-cost virtual reality surgical simulation to increase surgical oncology capacity and capability. *Ecancermedicalscience, 13*, 910. https://doi.org/10.3332/ecancer.2019.910

Patel, S. (17 May 2016). *85 percent of Facebook video is watched without sound.* Digiday https://digiday.com/media/silent-world-facebook-video/

Pedersen, J. (2011). *Subtitling norms for television: An extrapolation focussing on extralinguistic cultural references.* John Benjamins.

Pedersen, J. (2018). From old tricks to Netflix: How local are interlingual subtitling norms for streamed television? *Journal of Audiovisual Translation, 1*(1), 81–100.

Pedersen, J. (2020). Audiovisual Translation Norms and Guidelines. In Ł. Bogucki & M. Deckert, (Eds.), *The Palgrave Handbook of Audiovisual Translation and Media Accessibility. Palgrave Studies in Translating and Interpreting.* Palgrave Macmillan. https://doi.org/10.1007/978-3-030-42105-2_21

Perego, E. (Ed.) (2012). *Eye tracking in audiovisual translation.* Aracne.

Perego, E., Del Missier, F., & Bottiroli, S. (2015). Dubbing versus subtitling in young and older adults: cognitive and evaluative aspects. *Perspectives, 23*(1), 1–21.

Perego, E., Del Missier, F., Porta, M., & Mosconi. M. (2010). The cognitive effectiveness of subtitle processing. *Media Psychology, 13*(3), 243-272. http://doi.org/10.1080/15213269.2010.502873

Pereira, A. (2005). El subtitulado para sordos: estado de la cuestión en España. *Quaderns. revista de traducción, 12*, 161–172. Retrieved from https://ddd.uab.cat/pub/quaderns/11385790n12/11385790n12p161.pdf

Pérez-González, L. (2009). Audiovisual Translation. In M. Baker, & M. Saldanha (Eds.), *The Routledge Encyclopaedia of Translation Studies* (pp. 13-20). Routledge.

Puerta Domínguez, S. (2016). Del color al sensorama. Cine e industria cultural, *Razón crítica*, *1,* 170–205. http://dx.doi.org/10.21789/25007807.1141

R Core Team. (2020). R: A language and environment for statistical computing. R foundation for statistical computing. Austria. Retrieved from: http://www.R-project.org

Reiß, K., & Vermeer, H. J. (1996). *Grundlegung einer allgemeinen Translationstheorie.* De Gruyter.

Remael, A., Orero, P., & Carroll, M. (eds.) (2014). *Audiovisual Translation and Media Accessibility at the Crossroads: Media for all 3*. Rodopi.

Remael, A., Orero, P., Black, S., & Jankowska, A. (2019). From translators to accessibility managers: How did we get there and how do we train them? In M. Tolosa Igualada & A. Echeverri (Eds.), Porque algo tiene que cambiar. La formación de traductores e intérpretes: Presente & futuro [Because something should change: Present & Future Training of Translators and Interpreters]. *MonTI, 11,* 131–154. http://dx.doi.org/10.6035/MonTI.2019.11.5

Remael, A., Reviers, N., & Vercauteren, G. (eds.) (2015). *Pictures Painted in Words: ADLAB Audio Description Guidelines*. EUT.

Rizzo, A., Schultheis, M., Kerns, K. A., & Mateer, C. (2004). Analysis of assets for virtual reality applications in neuropsychology. *Neuropsychological Rehabilitation*, *14*(1-2), 207–239.

Robson, G. D. (2004). *The closed captioning handbook.* Elsevier.

Romero-Fresco, P. (2013). Accessible filmmaking: joining the dots between audiovisual translation, accessibility and filmmaking. *The Journal of Specialised Translation, 20,* 201-223.

Romero-Fresco, P. (2018). In support of a wide notion of media accessibility: Access to content and access to creation. *Journal of Audiovisual Translation, 1*(1), 187–204. https://doi.org/10.47476/jat.v1i1.53

Romero-Fresco, P. (ed.) (2015). *The Reception of Subtitles for the Deaf and Hard of Hearing in Europe*. Peter Lang.

Romero-Fresco, P., & Dangerfield, K. (2022). Accessibility as a conversation. *Journal of Audiovisual Translation, 5*(2), 15–34. https://doi.org/10.47476/jat.v5i2.2022.228

Rothe, S., Tran, K., and Hussmann, H. (2018). Dynamic subtitles in cinematic virtual reality. *ACM TVX 2018*, *Seoul*, *Republic of Korea*, 209–214. https://doi.org/10.1145/3210825.3213556

Rothe, S., Tran, K., and Hussmann, H. (2018). Positioning of Subtitles in Cinematic Virtual Reality. In Bruder, G., Yoshimoto, S., & Cobb, S. (Eds.), *International Conference on Artificial Reality and Telexistence and Eurographics Sysmposium on Virtual Environments*, ICAT-EGVE 2018. The Eurographics Association. https://doi.org/10.2312/egve.20181307

Rouse R., & Barba, E. (2017). Design for emerging media: How MR designers think about storytelling, process, and defining the field. In N. Nunes, N, I. Oakley, & V. Nisi (Eds.), *Interactive Storytelling: 10th International Conference on Interactive Digital Storytelling, ICIDS 2017, Funchal, Madeira, Portugal, November 14–17, 2017, Springer-Verlag, Lecture Notes in Computer Science*, *10690,* 245–258. http://doi.org/10.1007/978-3-319-71027-3_20

Rupp, M., Kozachuk, J., Michaelis, J., Odette, K., Smither, J., & McConnell, D. (2016). The effects of immersiveness and future vr expectations on subjective- experiences during an educational 360 video. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 60,* 2108–2112. https://doi.org/10.1177/1541931213601477

Ryan, M. L. (2000). *Narrative as Virtual Reality: Immersion and Interactivity in Literature and Electronic Media*. JHU Press.

Saldanha, G., & O'Brien, S. (2013). *Research methodologies in translation studies.* St. Jerome Publishing

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the 2000 Symposium on Eye-Tracking Research & Applications*, 71–78. https://doi.org/10. 1145/355017.355028

Saunders, B., Camgoz, N. C., & Bowden, R. (2021, July 23). Mixed SIGNals: Sign language production via mixture of motion primitives. *Proceedings of the International Conference on Computer Vision (ICCV).* Retrieved from https://openresearch.surrey.ac.uk/esploro/outputs/conferencePaper/Mixed-SIGNals-Sign-Language-Production-via/99609266002346

Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry, 36*(8), 1627–1639. http://pubs.acs.org/doi/abs/10.1021/ac60214a047

Schubert, T. (2003). The sence of presence in virtual environments: A three component scale measuring spatial presence, involvement and realness. *Zeitschrift für Medienpsychologie, 15*, 69–71. http://doi.org/10.1026//1617-6383.15.2.69

Schubert, T., Friedmann, T., Regenbrecht, H. (2001). The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators and Virtual Environments*, *10*(3), 266–281. https://doi.org/10.1162/105474601300343603

Sidenmark, L., Kiefer, N., & Gellersen, H.-W. (2019). Subtitles in Interactive Virtual Reality: Using Gaze to Address Depth Conflicts. *Workshop on Emerging Novel Input Devices and Interaction Techniques, IEEE VR 2019*.

Sipatchin, A., Wahl, S., & Rifai, K. (2020). Eye-tracking for low vision with virtual reality (VR): testing status quo usability of the HTC Vive Pro Eye. *bioRxiv.*

Skult, N., & Smed, J. (2020). Interactive storytelling in extended reality: Concepts for the design. In B. Bostan (Ed.), *Game user experience and player-centered design* (pp. 449–467). Springer International Publishing. https://doi.org/10.1007/978-3-030-37643-7_21

Slater, M. (2003). A note on presence terminology. *Presence connect, 3*(3), 1–5.

Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and virtual environments, 6*(6), 603– 616. https://doi.org/10.1162/pres.1997.6.6.603

Smith, W. (2015). Stop calling google cardboard's 360-degree video 'vr'. *Wired.* https://www.wired.com/ 2015/11/360-video-isnt-virtual-reality/

Sun, S., & Shreve, G. M. (2014). Measuring translation difficulty: An empirical study. *Target, 26*(1), 98–127. https://doi.org/10.1075/target.26.1.04sun

Suojanen, T., Koskinen, K., & Tuominen, T. (2015). *User-centered translation.* Routledge.

Szarkowska, A. (2009). The audiovisual landscape in Poland at the dawn of the 21st century. In A. Goldstein, & B. Golubović (Eds.), *Foreign language movies – dubbing vs. subtitling* (pp. 185–201). Verlag Dr. Kovač.

Szarkowska, A. (2013). Towards interlingual subtitling for the deaf and the hard of hearing. *Perspectives. Studies in Translatology, 21*(1), 68–81. https://doi.org/10.1080/0907676X.2012.722650

Szarkowska, A. (2020). Subtitling for the Deaf and the Hard of Hearing. In Ł. Bogucki & M. Deckert (Eds.), *The Palgrave Handbook of Audiovisual Translation and Media Accessibility. Palgrave Studies in Translating and Interpreting.* Palgrave Macmillan. https://doi.org/10.1007/978-3-030-42105-2_13

Szarkowska, A., & Boczkowska, J. (2022). Colour coding subtitles in multigoal films – a reception study. *Perspectives: Studies in Translatology, 30*(3), 520–536. https://doi.org/10.1080/0907676X.2020.1853186

Szarkowska, A., & Gerber-Morón, O. (2019). Two or Three Lines: A Mixed-Methods Study on Subtitle Processing and Preferences. *Perspectives: Studies in Translatology, 27*(1), 144–164. https://doi.org/10.1080/0907676X.2018.1520267

Szarkowska, A., Silva, B., & Orrego-Carmona, D. (2021). Effects of subtitle speed on proportional reading time. Re-analysing subtitle reading data with mixed effects models. *Translation, Cognition & Behavior, 4*(2), 305–330. https://doi.org/10.1075/tcb.00057.sza

Szkriba, S. (2022). Accessible cinema for older adults. Can an app for blind viewers benefit the sighted population? *Journal of Audiovisual Translation, 5*(2), 134–151.

Tamayo, A. (2016). Reading speed in subtitling for hearing impaired children. An analysis in Spanish television. *The Journal of Specialised Translation, 26*, 275–294. Retrieved from http://www.jostrans.org/issue26/art_tamayo.pdf

Udo, J.-P., & Fels, D. I. (2010). Universal design on stage: Live audio description for theatrical performances. *Perspectives: Studies in Translatology, 18*(3), 189–203. http://doi.org/10.1080/0907676X.2010.485683

UN. (2006). *Convention on the Rights of People with Disabilities*. https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with- disabilities.html

UNCRPD. (n.d.) United Nations Convention on the Rights of Persons with Disabilities. http://www.un.org/disabilities/documents/convention/convoptprot-e.pdf

UNE-EN 301549. (2019). *Accessibility requirements for ICT products and services.*

UNESCO. (2005). *Convention on the Protection and Promotion of the Diversity of Cultural Expressions*. https://en.unesco.org/creativity/convention/texts

Valdeón, R. A. (2022). Latest trends in audiovisual translation. *Perspectives, 30*(3), 369–381. https://doi.org/10.1080/0907676X.2022.2069226

Vandaele, J. (2018). Narratology and audiovisual translation. In L. Pérez-González (Ed.), *The Routledge Handbook of Audiovisual Translation* (pp. 225–241). Routledge.

Vela Valido, J. (2007). *Software de audiodescripción: análisis comparativo de los programas existentes en España y EE. UU* [MA thesis]. Universitat Autònoma de Barcelona.

Vertebrae. (2109). *Immersive vs traditional media. How interactivity and realism influence persuasions.* https://cdn.nrf.com/sites/default/files/2020-02/Vertebrae_3d-AR-eCommerce_Whitepaper-2019%20%281%29_0.pdf

Walczac, A. (2016). Foreign language class with audio description: a case study. In A. Matamala, & P. Orero (Eds.), *Researching Audio Description. New Approaches* (pp. 209–226). Palgrave Macmillan.

Warren, M. (2017). *Making your first 360 video? Here are 10 important things to keep in mind.* https://www.filmindependent.org/blog/making-first-360-video-10-important-things-keep-mind

World Federation of the Deaf and World association of sign language interpreters. (2018). *WFD and WASLI statement on use of signing avatars.* https://docs.google.com/viewerng/viewer?url=https://wfdeaf.org/wp-content/uploads/2017/02/WFD-and-WASLI-Statement-on-Avatar-FINAL-14032018-Updated-14042018-1.pdf&hl=en_US

XR Accessibility User Requirements. https://www.w3.org/TR/xaur/

Zárate, S. (2021). *Captioning and Subtitling for d/Deaf and Hard of Hearing Audiences.* UCL Press. https://doi.org/10.14324/111.9781787357105

Zdenek, S. (2015). *Reading sounds: closed-captioned media and popular culture.* The University of Chicago Press.

ANNEXES

ANNEX 1

Additional data analysis

1.1. Results from a Deaf and Hard of Hearing study

The results presented in this section stem from an experiment that adheres to the same experimental design (see §5.3.1) and procedure (see § 5.3.3) presented in Chapter 5. It also uses the same experimental materials (see §5.3.4) and equipment (see §5.3.5).

1.1.1. Participants

This section provides a concise overview of the results obtained after analysing data from 39 participants (26 females, aged: M = 58.31, SD = 14.01) in two different locations: Barcelona, Spain (N = 18, 14 females), aged: (M = 55.72, SD = 16.81) and Ljubliana, Slovenia (N = 21, 12 females), aged: (M = 60.52, SD = 11.04). It is essential to note that all participants included in this study presented some degree of hearing loss. They were recruited by local Deaf and Hard-of-Hearing associations (Catalan association for the promotion of Deaf people–ACCAPS and Deaf and Hard-of-Hearing association of Slovenia–ZDGNS). Participants volunteered with no compensation. They were interested in the VR experience when asked on a 5-point Likert type scale (M=3.95, SD=0.72); they were also in favour of subtitles (M=2.33, SD=1.57).

1.1.2. Results

To analyse gaze-based and self-reported data, a series of mixed-design ANOVA with three fixed factors (independent variables) were used: a between-subjects factor of subtitle position (head-locked vs. fixed), a within-subjects factor of subtitle colour (monochrome vs. colour), and a between-subjects factor of country (Spain vs. Slovenia). For every ANOVA all statistically significant effects were followed by pairwise comparisons with HSD Tukey correction for multiple comparisons.

All statistical analyses were conducted with the use of R, the free software environment for statistical computing and graphics (R Core Team, 2020). For gaze-based analysis, fixations on subtitle AOIs were evaluated.

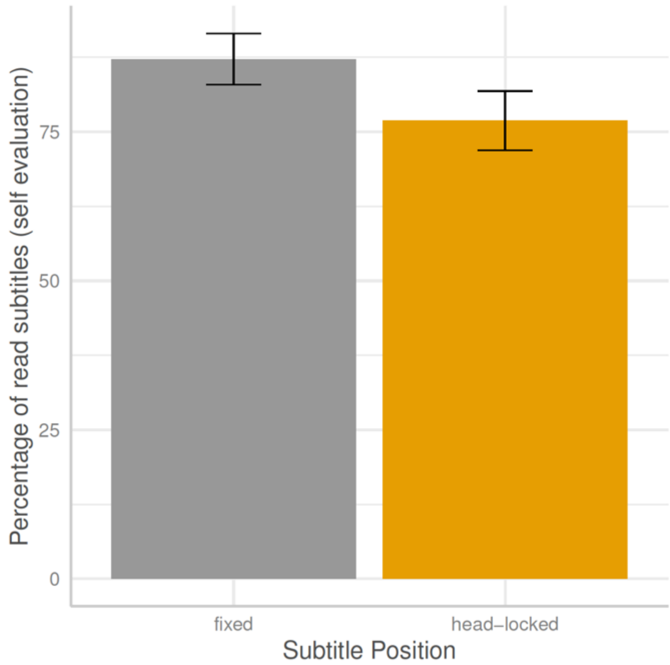
1.1.2.1. Self-evaluation of subtitle readability

A 2x2 mixed-design ANOVA was run on self-evaluation of percentage of read subtitles. Analysis showed that the main effect of subtitle position was statistically non-significant, $F(1,31) = 2.47$, $p = 0.13$. Participants evaluated both types of subtitles as similarly readable, see Figure 32. The main effect of country was marginally significant, $F(1, 31) = 3.35$, $p = 0.077$,

$\eta2$ =0.074, showing slightly higher percentage of read subtitles (via self-evaluation) by the Spanish sample (*M* = 88%, *SE* =4.56) than by the Slovenian sample (*M*=76%, *SE* = 4.74).

Figure 32

Self-evaluation of percentage of read subtitles



Note. Bar height represents estimated means and whiskers represent ±1*SE*

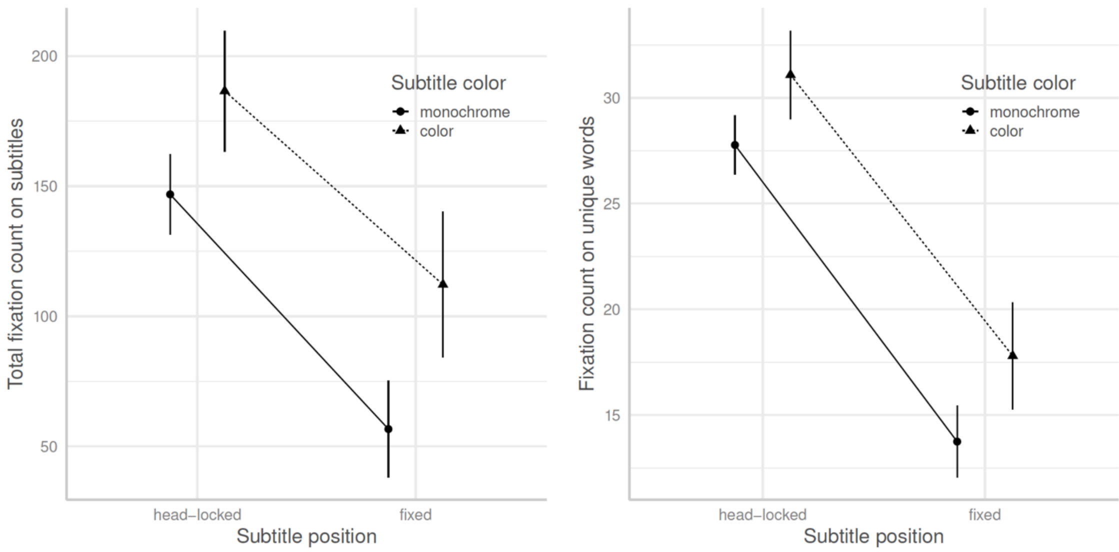
1.1.2.2. Eye-movement-based analysis of subtitle reading

Two Analyses of Variance were conducted of mixed-design (2 × 2 × 2) with subtitle position, colour, and country as fixed factors. First, analysis of total fixation counts on subtitle AOIs showed a statistically significant main effect of subtitle position, F (1, 33) = 12.25, $p < 0.01$, $\eta2 = 0.177$. Head-locked subtitles drew, on average, significantly more fixations (*M*=166.70, *SE*=15.00) than fixed subtitles (*M*=84.40, *SE*=18.10). Effect of subtitle colour was statistically significant, F (1, 33) =5.66, p<0.05, $\eta2$ =0.067. Fixation count on colour subtitles was significantly greater (*M*=149, *SE*=18.20) than on monochrome subtitles (*M*=102, *SE*=18.10). Both main effects are shown in Figure 33(a). No other main and interaction effects reached statistical significance.

Second, analysis of number of fixated unique words revealed a statistically significant main effect of subtitle position, F (1, 33) = 45.97, $p < 0.001$, $\eta2 = 0.421$. When reading head-locked subtitles, participants fixated, on average, more unique words (M = 29.40, SE =1.28) than when reading fixed subtitles (M = 15.80, SE =1.55). Analysis showed a marginally significant difference in fixated unique words between colour and monochrome subtitles F (1, 33) =3.65, p = 0.065, $\eta 2$ =0.050. Colour subtitles drew more fixations of unique words (M=24.40, SE=1.64) than monochrome subtitles (M = 20.80, SE = 1.10). Both effects are shown in Figure 33(b). No other effects (main nor interaction) were statistically significant.

Figure 33

Fixation count on subtitle AOIs for viewers depending on subtitle position and subtitle colour



(a) Total fixation count on subtitle AOIs. (b) Fixation count on unique words in subtitle AOIs.

Note. Points represent the estimated means and whiskers represent ±1*SE*

1.1.3. Discussion and conclusions

The observed discrepancy in fixation counts reveals a higher percentage of gaze directed towards head-locked subtitles as opposed to fixed subtitles. This distinction can be explained as head-locked subtitles are more readily fixated because they are "always there". In contrast, fixed subtitles receive fewer fixations due to the initial effort viewers must exert lo locate them, making their processing more demanding. An analysis of gaze data suggests that a smaller number of fixations land on unique words within fixed-position subtitles, indicating that viewers may have limited time to engage with subtitles that are not always available. When it comes to colour, it becomes evident that coloured subtitles are preferred and enhance readability, as evidenced by a higher proportion of fixations compared to monochrome subtitles.

These findings in gaze data appear to be in stark contrast to participants' self-reported impressions of their subtitle reading performance. Participants indicated that they believed they read a greater percentage of the fixed subtitles compared to the head-locked position. However, gaze analysis presents a contrasting picture: a greater proportion of fixations were directed towards head-locked subtitles. This finding proves that there is disconnection between viewer performance and their self-perceived performance, at least in term of the objective evidence provided by eye-tracked data. This discrepancy suggests that self-reported assessments, including those based on questionnaires or focus group, may not accurately reflect objective measurements. This could be attributed to the self-serving bias, which can introduce disparities between self-reported and objectively measured performances.

ANNEX 2

Published articles within this dissertation

2.1. Article 1: Brescia-Zapata, M. (in press). Accessibilities in immersive media environments. In M. J. Varela-Salinas, & C. Plaza Lara (Eds.), *Aproximaciones teóricas y prácticas a la accesibilidad desde la traducción y la interpretación.* Comares.

2.2. Article 2: Brescia-Zapata, M. (2021). The present and the future of accessibility services in VR360 players. *inTRAlinea, 24.*

2.3. Article 3: Brescia-Zapata, M., Krejtz, K., Orero, P., Duchowski, A., & Hughes, C. (2022). VR 360º subtitles: Designing a test suite with eye-tracking technology. *Journal of Audiovisual Translation, 5*(2), 233–258. https://doi.org/10.47476/jat.v5i2.2022.184

2.4. Article 4: Brescia-Zapata, M., Krejtz, K., Orero, P., Duchowski, A., & Hughes, C. (in press). Subtitles in VR 360º video. Results from an eye-tracking experiment. *Perspectives: Studies in Translation Theory and Practice.* https://doi.org/10.1080/0907676X.2023.2268122

2.1. Article 1: Brescia-Zapata, M. (in press). Accessibilities in immersive media environments. In M. J. Varela-Salinas, & C. Plaza Lara (Eds.), *Aproximaciones teóricas y prácticas a la accesibilidad desde la traducción y la interpretación.* Comares.

Accessibilities in immersive media environments

Marta Brescia-Zapata
Departamento de Traducción, Interpretación y Estudios de Asia Oriental
Universitat Autònoma de Barcelona
marta.brescia@uab.cat

Abstract

Research on immersive media content has been intense in the last few years. In this context, 360º videos have become very popular as they are a cheap and effective way to provide immersive experiences. While it is possible to find plenty of accessibility solutions and guidelines for traditional media services, there is a lack of standards regarding accessibility in immersive media. During recent years, both audiovisual translation (AVT) and media accessibility (MA) studies have spotlighted these technologies, discussing how to integrate accessibility services into immersive environments. Moreover, these technologies are gaining popularity due to the COVID-19 crisis; as they enable interactive, hyper-personalised and engaging experiences anytime and anywhere. Thinking about how to implement accessibility services is the next logical step to guarantee universal access to the new virtual world.

This chapter will discuss the main challenges that arise when integrating accessibility services into immersive environments. Minimum requirements for creating accessible, immersive content will also be presented, as well as examples of the latest experiments on how these services could be integrated within 360º videos. The main goal of this work is to contribute to standardise the integration of accessibility services in immersive media.

Keywords: media accessibility, immersive media, virtual reality, 360º videos

Resumen

Durante los últimos años, los entornos inmersivos han sido un campo de investigación muy prolífico. En este contexto, los vídeos 360º son una manera barata y efectiva de crear experiencias inmersivas. Si bien es posible encontrar muchas guías y estándares de servicios de accesibilidad en medios tradicionales, en los entornos inmersivos no sucede lo mismo. El mundo de la traducción audiovisual (TAV) y de la accesibilidad a los medios (MA) también se ha centrado en estas tecnologías con el objetivo de investigar cómo incorporar los servicios de accesibilidad en los entornos inmersivos.

Una de las ventajas de las tecnologías inmersivas, que se ha magnificado durante la crisis derivada de la COVID-19, es que permiten disfrutar de experiencias interactivas y personalizables en cualquier momento y en cualquier lugar. Implementar los servicios de accesibilidad en este tipo de entornos es el paso lógico para garantizar el acceso universal a este nuevo mundo virtual.

En este capítulo se discutirán los retos y desafíos que se plantean a la hora de integrar los servicios de accesibilidad en entornos inmersivos. También se presentarán los requisitos mínimos para crear contenidos inmersivos accesibles para todos, así como ejemplos de los últimos experimentos sobre la integración de estos servicios en videos 360º. Este trabajo pretende contribuir a estandarizar la integración de los servicios de accesibilidad en los entornos inmersivos.

Palabras clave: accesibilidad, medios inmersivos, realidad virtual, videos 360º

1. Introduction

Immersive media have the potential to drastically transform our lives: from the way we work and communicate to the way we understand and experience the world (Creed *et al.,* 2023). Immersive technology is already a reality, with only a few pieces left to fall into place before it grows to become something that encompasses lives and cultures all over the world. We live surrounded by digital media, and more importantly, we must interact with digital media for many basic daily activities such as work, training, shopping, and communication. The COVID-19 pandemic further exposed and exacerbated digital inequalities (Beaunoyer *et al*., 2020), also known as the digital divide. The gap produced by the digital divide —evidenced by related concepts like digital inclusion, digital participation, digital skills, media literacy and media accessibility— needs to be reduced because it is one of the most challenging problems facing the information society (Aissaoui, 2021).

To get the most out of the information and communication technology and now that immersive environments are becoming mainstream, the next logical step is thinking about accessibility. The availability of affordable technology to both produce and consume media content was the challenge when 3D was considered. Early concerns about the individual nature of this format were quickly dispelled by the possibility of projecting immersive media, as seen in the Klimt art exhibitions (2021)[1] or Die Soldaten opera (2018)[2]. In the past, the only way to consume eXtended reality (XR) was with some

[1] For further information about the Klimt exposition, please consult https://www.expo-klimt.be/en/ [retrieved December 20, 2021]

[2] For further information about Die Soldaten opera, please consult https://lafura.com/en/works/die-soldaten/ [retrieved December 20, 2021]

special glasses (head-mounted display) which proved challenging in terms of the amount of time a person could stand to wear them, due to the possible side effects on their health such as nausea and dizziness. More recently, the possibility of a shared end-to-end system for the production and delivery of photorealistic, social, immersive experiences has become a reality (Gunkel *et al.,* 2018).

In the last decade, the UN decided to challenge the effects of the digital divide with the most vulnerable groups and people with disabilities. The Convention on the Rights of Persons with Disabilities (CRPD) was signed by almost all countries worldwide and it had an impact in Europe with the development of three EU Directives: the Audiovisual Media Service Directive (AVMSD), the Web Accessibility Directive (WAD), and the European Accessibility Act (EAA). A commonality across the different directives is the recommendation to thinking of accessibility in the design phase of any product or system. This means that Europe believes in the successful development of accessibility services in the development phase of a new technology. This has been true for many EU funded projects such as DTV4ALL (https://dea.brunel.ac.uk/dtv4all/), when Europe went from analogue to digital; and in Hbb4ALL (http://www.hbb4all.eu/), when Internet technology converged with broadcasting. Europe has continued funding research in that direction with the project called ImAc (https://www.imacproject.eu), which explored how accessibility services could be efficiently integrated with immersive media. In particular, ImAc aimed to ensure that immersive media experiences are inclusive across different languages, addressing the needs not only of those with hearing and low vision problems, but also of people with cognitive or learning difficulties, newcomers, people with low literacy, and the aged, in an interactive and personalised manner. The next challenge is to extrapolate all the ImAc learnings to digital immersive storytelling platforms while developing accessibility further.

This paper will start with a brief historical overview of the transition from traditional to immersive media experiences. It will also introduce the main concepts of XR, as well as a discussion on the main differences between these two environments. Section 3 will focus on the integration of accessibility with immersive media, describing four key areas where XR accessibility should be considered: accessibility of the XR contents, media players, user interfaces, and devices. The last section is a brief reflection on what has been analysed in the chapter, and an open door to a near future in which all traditional and immersive media are born accessible.

2. Traditional media vs new media environments

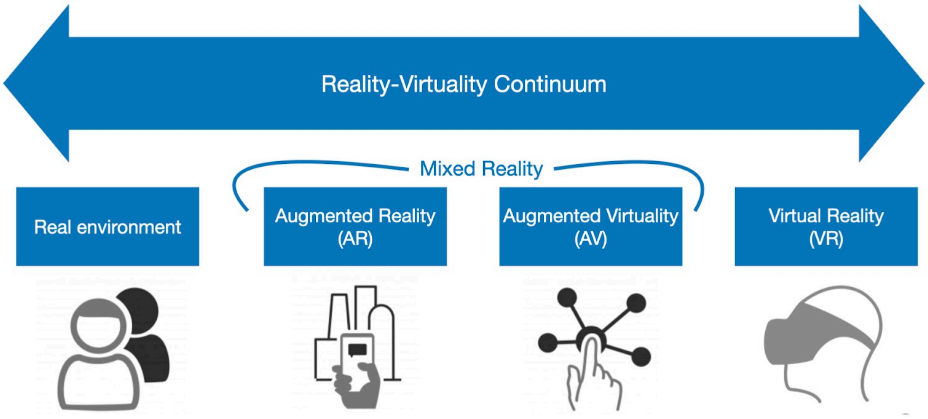
The idea of Virtual Reality (VR) began to take shape long before the appearance in 1956 of the first 3D device, the Sensorama (Puerta Domínguez, 2016). Until then, the concept was limited to science fiction, and all the media content was generated as 2D or flat images —also known as traditional media. Cinema was born at the end of the 19th century, and movie productions were black and white and mostly silent (Crafton, 1999). At the time, technology was not developed enough to allow for music or dialogue in these movies, however, they were accompanied by other elements, such as external music (piano or orchestra) and intertitles (slides that appeared between scenes or in the middle of a scene to provide contextual information for the spectators). With time, technology became more advanced, and sound began to be included in movies: from simple soundtracks or sound effects to dialogues. The expansion into cinema in colour took a little longer, but thanks to Technicolor, its popularity increased, and in the 1950s, black and white movies almost completely disappeared (Molina-Siles *et al.,* 2013).

Technology has continued to develop since then, and still seeks the same goal: to reproduce what we see and hear in our environment with the same quality as a human is capable of. The quest to develop a real digital environment that simulates reality led to the failed attempt to adopt 3D imaging as the mainstream display for audiovisual products in the 1960s. According to Agulló and Matamala (2019, pp. 219), this failure "may have opened a door for VR and 360º content as a new attempt to create engaging immersive experiences."

XR is a term referring to all the technologies that combine real and virtual environments and human-machine interactions generated by computer technology. XR includes the entire Reality-Virtuality Continuum (see Figure 1), also known as "Milgram's Continuum", a concept that was first introduced in 1994 by Paul Milgram to describe all possible combinations of reality and computer-generated images. On the left side of the continuum there is the Real environment, unmodified images just as our eyes see them. On the opposite side, there is the Virtual Reality or VR, where the user is immersed in a computer-generated experience and cannot see the real world. In this case, the content can be accessed using VR devices (Head-Mounted Displays like Oculus Rift, HTC Vive, Samsung Gear, etc.) and can also be consumed as a CAVE (Cave Automatic Virtual Environment), which uses high-resolution projection screens to deliver 360º visual experiences. The area between the real environment and the virtual environment represents Mixed Reality (MR), with various degrees of overlap between reality and virtuality.

Figure 1

Reality-Virtuality Continuum



Source: Author's own elaboration based on Milgram & Kishino (1994)

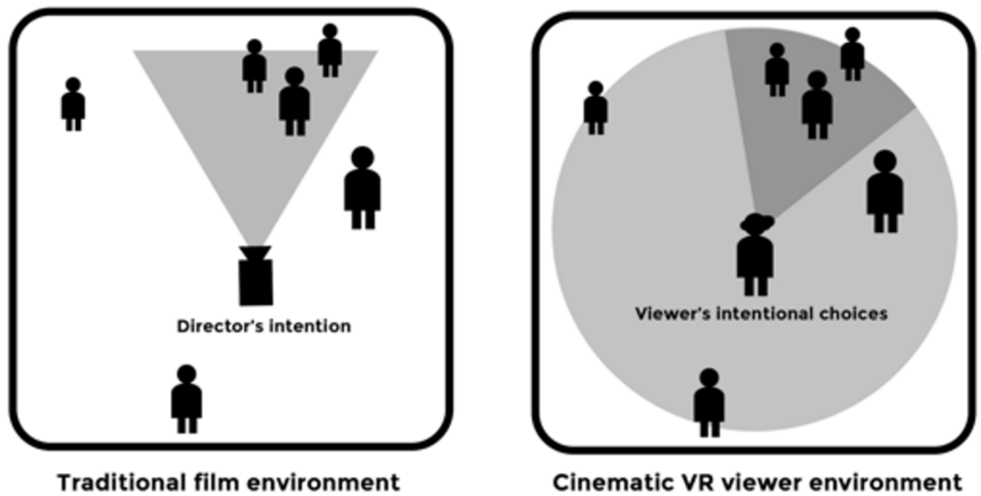
Immersive technologies are mainly designed to elicit presence, the experience of 'being' or 'acting', when physically situated in another place (Slater & Wilbur, 1997). This feeling refers to tricking at least three of the five senses of your brain into believing that you have been taken to another place. Still, immersive videos are limited in nature as they can be unrealistic or make people motion or simulator sick leading to a suboptimal experience that turns people off the immersive experience. According to Slater and Wilbur (1997), the process of achieving spatial presence is divided into three phases: place illusion ("I am here"), plausibility ("this is happening"), and body ownership ("it is my body").

Traditional media (static 2D imagery and 2D video) is the best media format for storytelling, while immersive media (XR experiences) invite users to interact with the content, and thus to modify it. In XR the narrative can be developed using space and movement over the temporal period. This is called "spatial storytelling" and consists of engaging the viewers inside a spatial environment that contains non-linear narratives that can be discovered by exploring the space provided.

Along with XR's new narratives, the main challenge for immersive experiences is that the field of view (FoV) is limited, and the interactor cannot pay attention to the entire visual scene in the given time (Aylett & Louchart, 2003). Traditionally, filmmakers have relied on four core tools to tell stories: cinematography, sound, *mise-en-scène* (the arrangement of the scenery), and editing (Skult & Smed, 2020). These techniques that largely define the current movie-making industry are not suitable for XR experiences. The interactor is free to explore the scene by looking wherever they want, which is why all these core approaches need to be reconsidered. As shown in Figure 2, in XR the viewer does not have a window to a world, he or she is directly present in the world instead. Orientation plays a crucial role especially in cuts since every cut requires the viewer to re-orientate themselves in the new environment. The pacing and playing should be more like it is in traditional theatre with a focus on cinematography.

Figure 2

Traditional film environment in comparison to an XR environment



Source: Ko *et al.* (2018)

Another key feature is that immersive media (as any novel technology) can engage the audience at a higher level than traditional media, but this effect may subside once the novelty wears off. This is called “the novelty effect”, and a study carried out by Vertebrae (2019) —a 3D and Augmented Reality commerce and media platform— showed that participants with less frequent exposure to XR had a stronger emotional response to the medium, validating the novelty expectation. This effect must be considered as a relevant variable when testing immersive technology and/or media with final users.

To sum up, immersive technologies are at a crossroad in their development, and despite the hurdles, it seems likely that the XR industry will steadily continue to improve these technologies, weaving them into more aspects of our personal and professional lives. Understanding the differences between traditional and new immersive media is essential for AVT and MA professionals and researchers in order to take the wheel and integrate these innovative technologies in their work and experiments.

3. Implementing Accessibility in New Media Environments

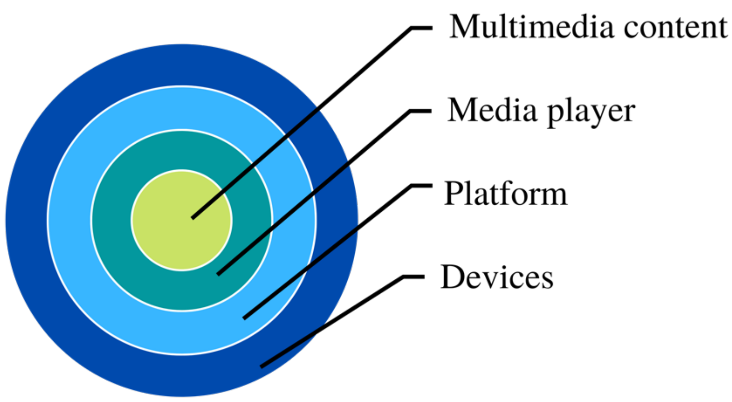
The most popular media accessibility service is translation, which is in fact language accessibility (Orero, 2022). This service shows the universal nature of media accessibility beyond people with disabilities. Immersive media need to be accessible beyond different languages, addressing not only consumers with hearing and visual impairments, but also those with cognitive difficulties, the elderly, tourists, or for example in Europe migrants and refugees who are literate in their language, but can't read the Latin alphabet (also known as “new Europeans”). The way to cater for the needs of the users who cannot access the multimedia content is by integrating different types of accessibilities within these new formats.

XR is often touted as a democratising way to access new worlds and experiences, but in terms of accessibility, there is still a long way to go. An online survey carried out in 2017 by Alice Wong (Brennan, 2017), a disability activist and founder of the Disability Visibility Project in partnership with Lucasfilm, showed that people with a variety of disabilities —from blindness to cerebral palsy or autism— enjoy using VR and believe it be beneficial, but also experience major accessibility issues. The two biggest issues raised by survey responders were the inability to customise the experience and the necessity to move certain parts of the body.

To achieve fully immersive media experiences, media creators have to consider four components: the content, the media player, the platform, and the device(s). As depicted in Figure 3 below, the four components can be organised in a concentric circle: the outermost layer correspond to the device(s), the second layer to the “platform”, the third layer to the “media player” and the innermost layer to the “multimedia content”. This circle interrelates the 4 accessibilities, and it can be interpreted as that the only way to access the centre —the accessible content— is by ensuring that the components of the outer layers also take accessibility into account.

Figure 3

The four accessibilities



Source: Author’s own elaboration.

3.1. Content accessibility

During the last few decades, MA and AVT studies have been a prolific area of research. Best practices and standards for 2D traditional media have been established by both public and private entities (Matamala & Orero, 2018). However, studies on access services integrated within immersive content are still in their infancy; and currently, VR lacks agreed-upon guidelines and/or methods for making content accessible. Moreover, studies involving content accessibility and XR need to adopt a multidisciplinary approach as they borrow concepts such as usability and presence, which belong to other disciplines, mainly Psychology or Computer Science. This fact, linked to the lack of standardized solutions and guidelines, has led to the development of non-unified solutions, meeting only specific requirements (Hughes & Montagud, 2021).

3.1.1. Subtitling for the deaf and hard of hearing

Still, subtitling is the most popular (and also the cheapest) media access service, since there is no need to record voices as in audio description or dubbing. There are also standardised practices regarding many aspects: number of characters, number of lines, speed, position, etc. Traditionally, subtitles for the deaf and hard of hearing (SDH) are referred to as intralingual subtitles, which provide additional details about ambient noises and speaker identification in addition to a written rendition of the script (ISO20071:23:2019). Research in subtitles for XR is a work in progress, and some results have been published already (Agulló *et al.*, 2018; Brown *et al.*, 2018; Montagud *et al.*, 2019; Rothe *et al.*, 2018). Also, different key aspects have been identified so far (Montagud *et al.*, 2021): comfortable viewing fields, presentation modes, guiding methods, and (re-)presentation of non-speech information.

The main challenge when creating SDH in immersive environments is the need to orient the viewer towards the sound's source (i.e., sound effects, or character speaking or not speaking). To facilitate this requirement, location within the 3D space information may be added to each subtitle (Agulló & Matamala, 2019). However, the drawback is that the location is only specified once per caption, and if a person is moving dynamically during this period, the guide could be wrong (Hughes *et al.*, 2019). The ImAc project designed and developed several guiding mechanisms (Agulló *et al.*, 2019), and test results showed two preferred methods: an arrow positioned to the left or right that directs the user to the target, and a radar circle shown in the user view that identifies both the position of the caption and the relative viewing angle of the user.

All the studies carried out so far regarding SDH within immersive media followed a user-centric methodology and chose people with hearing loss for testing. The British Broadcasting Corporation (BBC) was one of the first research organisations to design subtitles in XR (Brown & Patterson, 2017). They first identified the main challenges when developing subtitles for immersive content (Brown *et al.*, 2017) and then designed four solutions for subtitle rendering (Brown, 2017). These rendering modes were tested with several clips (Brown, *et al.,* 2018), and users reported that they preferred subtitles which are always presented in front of the viewer, and that as they turn their head, they follow the movement, staying always in the same location in the headset display (head-locked). This trend is also observed in later studies, such as the one carried out by Rothe *et al.* (2018) and the most recent performed under the umbrella of the ImAc project (Hughes *et al*., 2019).

In the area of standardisation, a W3C community group[3] is focusing on determining and publishing best practices for access, activation, and display settings for captions within different types of immersive media, including AR, VR, and games. They have recently conducted a community survey to gather opinions, but no tests were performed. A small group of users with different hearing levels (Deaf, Hard of Hearing, and Hearing) were asked to evaluate different approaches for captions within immersive environments. Head-locked was identified as the preferred choice, however, it was noted that the most likely reason for this was that it replicated the experience that users were familiar with. It was also acknowledged that it was difficult for users to properly and theoretically evaluate new methods without the opportunity and content to enable them to be experienced.

3.1.2. Audio description

As defined in ADLAB guidelines (Remael, Reviers & Vercauteren, 2015), AD "is a service for the blind and visually impaired that renders Visual Arts and Media accessible to this target group offering a verbal description of the relevant (visual) components of a work of art or media product." In traditional audiovisual content such as films or series, AD is delivered between the dialogues, and it is expected not to interfere with music and other important sound effects (Jankowska, 2015). As well as SDH, many studies on this accessibility service have been carried out to date, however, empirical studies in more immersive environments are almost non-existent (Fidyka & Matamala, 2018).

Given this context, a series of focus groups and experiments on how AD could be integrated into XR was conducted under the umbrella of the ImAc project. The experiments focused on two main topics: how to integrate AD into a 360 environment, and key aspects of the technical realisation and production. Regarding the first topic, two aspects were explored in particular: using different types of script and different sound designs or AD presentation modes (Orero *et al*., 2019).

To assess the quality of experience regarding understanding, enjoyment, and immersion, three approaches were taken to script the AD, including first-person narration, second-person narration, and the traditional standard AD. In a script in first

[3] https://www.w3.org/community/immersive-captions/ [retrieved February 5, 2022]

person, the main character becomes the describer and narrates the story from their own perspective. The style and writing are also similar to how the character sounds and speaks. In a second person script, the describer is sitting next to the viewer describing the scenes, and the style of writing and delivery is casual, informal and friendly. Lastly, a standard AD is a narration in the third person, like any traditional AD, and the describer objectively sets out what is happening in the scene, what the characters look like, etc.

In terms of audio presentation modes, the ImAc project designed three options that where later introduced in the player developed within the project: a) classic, no positioning; b) static: from a fixed point in the scene, and c) dynamic: coming from the direction of the action. This player also offers extended AD tracks for specific scenes, actions, or objects as well as independent volume settings for AD and main audio tracks.

3.1.3. Sign language interpreting

The few efforts devoted to integrating sign language interpretating into immersive content have focused exclusively on representations of 2D avatars. The best example so far is the ImAc player, which allows minimal presentation personalisation in terms of size, position, and language; and also supports non-continuous view streams (showing/hiding the signer window based on the interpreter's activity).

The possibility of introducing a fully customisable 3D avatar with which the user could interact has not been investigated or tested yet. However, according to a statement from the World Federation of the Deaf (WFD) and the World Association of Sign Language Interpreters (WASLI) (2018), animated or digital signing avatars should be avoided as users find them less expressive than a recorded video of humans who can convey the natural quality and skill provided by trained and qualified interpreters. Both organisations insist on using signed avatars only on pre-recordings of real people who are trained and qualified by interpreters and translators.

Integrating sign language interpretating into immersive media is an emerging field, and exploration is encouraged to ensure the future development of quality signing avatars (Saunders *et al.,* 2021). For example, this could be via building a signing avatar that provides a face with fully functioning muscular variables and can successfully parse the nuances of vocal expression and meaning.

3.1.4. Emerging services: easy to read, audio subtitling (AST), and web accessibility

Under-resourced accessibility services are new emerging accessibility services that have never been considered within XR. For example, an innovative test on easy-to-read subtitles carried out by the ImAc project revealed that many users, especially the elderly, enjoyed a simplified version of the subtitles when compared to standard ones. This access service can also be implemented within web accessibility, which allows everyone to perceive, understand, navigate, and interact with the Internet and its contents. It also applies to mobile apps and must consider the developments in technology and trends in recent years.

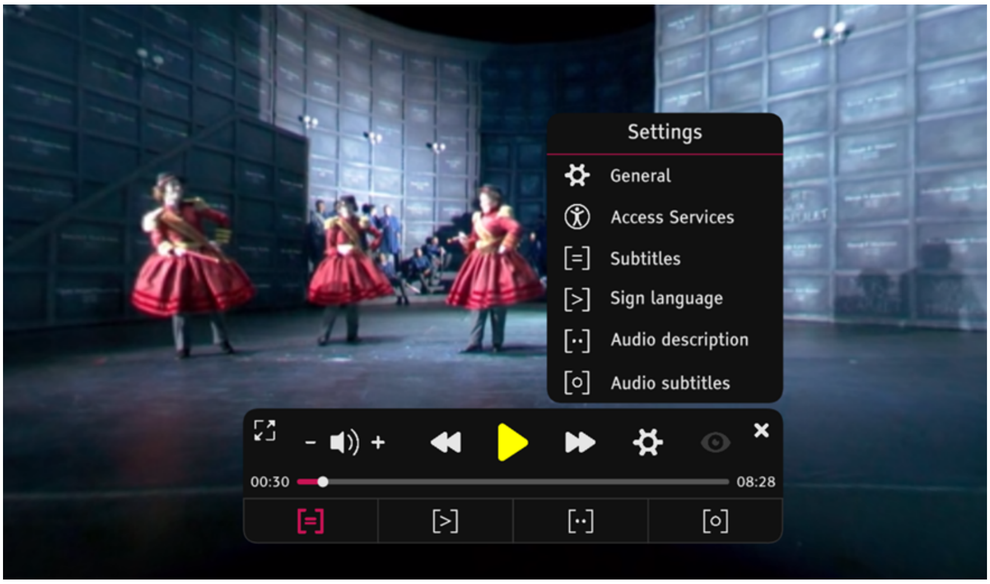
Audio subtitling (AST), also known as spoken subtitles/captions, is defined by ISO/IEC 20071-25 as “captions/subtitles that are read aloud over the audio in a video”; and according to the Access Service Survey published by the European Broadcasting Union (EBU, 2016), it is the least developed access service in Europe. The first attempt to introduce this access service in an immersive environment was for the ImAc project. The ImAc player allows for two audio placement modes: a classic one (no positioning), and a dynamic one (coming from the direction of the speaker). It also has independent volume settings for AST and the main audio tracks, so that the user can customise the volume of each track.

Regarding web accessibility, W3C's Web Content Accessibility Guidelines (WCAG) are the most internationally adopted voluntary web accessibility standard. Even though XR environments are identified as a challenge for accessibility implementation, only a first draft developed as part of the initial findings into potential accessibility-related user needs and requirements for XR is available (see Section 3.3).

3.2. Accessible player

As with traditional 2D media, to create and consume subtitles in XR, a subtitle editor and a subtitle player are needed. Currently, there are commercially available immersive video players offering the ability to play 360º videos, but not many of them support accessible services (Brescia-Zapata, 2022). Moreover, the player has to support different media formats, such as traditional 2D video and 360º video, traditional 2D audio and 3D spatial audio, IMSC subtitle files or support for DASH streaming. This problem has already been addressed by the ImAc project, as one of the main outcomes was the development of an accessible player along with a 360º subtitle editor. The player supports audio description, audio subtitles, and sign language, along with other features (Montagud *et al.,* 2019), as can be seen in Figure 4 below.

Figure 4
ImAc player settings



Source: Imac Player screenshot

One of the challenges for testing immersive subtitles is the difficulty users have in properly evaluating new modalities. The reasons are the cost and time needed to create new prototype subtitle presentations to enable users to experience them. To allow for visualising creative subtitles, an XR subtitle web simulator was developed by Hughes *et al.* (2020). This open-source, web-based simulator was designed for the rapid prototyping of subtitles within a 360º space and can be accessed —along with an editor— from the main project area, where all the imported 360º videos are located. These tools take inspiration from the player and the editor developed by the ImAc project. On the one hand, the ImAc player integrates various accessibility services, although it does not allow much customisation; on the other hand, the player developed by Hughes only allows subtitles, but with a very high degree of customisation. Therefore, the main difference between them is that the ImAc tools are intended to be used by generic audiences (final users), while the tools developed by Hughes are more focused on research and testing.

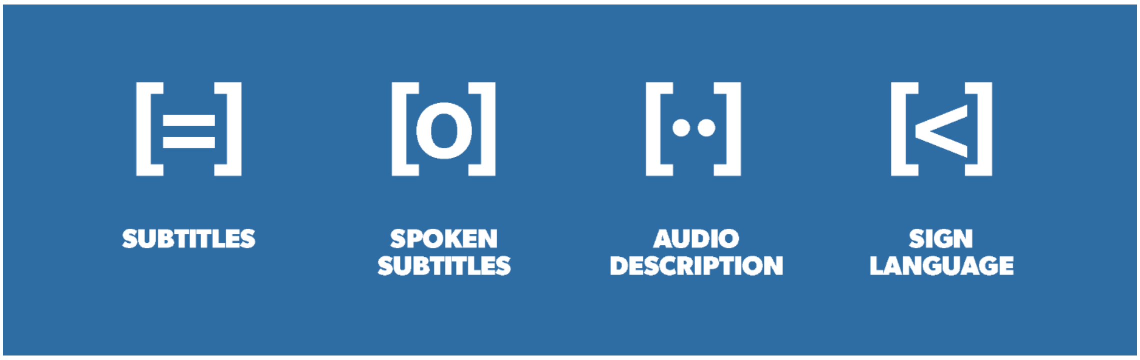
3.3. User interface accessibility

Not only the player itself needs to be accessible, but also the platform or portal where the 360º videos are stored, as well as the user interface. It comes as no surprise that XR interfaces are often less accessible than traditional 2D user interfaces on desktop and mobiles, as the additional spatial component can hamper easy use for users with no accessibility needs. Recently, a W3C working group published a first draft on various accessibility-related user needs for XR. It is mainly focused on gaming, but most insights can be extrapolated to any immersive experience which aims at being accessible. This document outlines personalisation, which involves tailoring aspects of the user experience to meet the needs and preferences of the individual user. In this regard, W3C is exploring areas such as expanding the accessibility information that may be supplied by the author, facilitating preference driven individual personalisation and enabling the author to specify key semantics needed to support users with cognitive impairments. This customisation affects very different aspects, from language selection to colour changes or context magnification. Voice command was also identified as a key feature, as it allows users to navigate, interact and communicate with the environment and also with other users.

Furthermore, platforms and portals must always display their catalogue of videos including the available access services (and languages) for each video. This aspect is especially useful, as it allows users to rapidly identify which content they can or cannot consume. For a better comprehension of the controls, icons or symbols used to identify each access service should be consistent and ideally follow established conventions or standards (showing their more commonly known acronyms, like ST, SL, AST, and AD, in English). Otherwise, the symbol or icons should clearly illustrate the functionality of the represented service. A good example is the icons proposed by the Danish Broadcaster (Eiersø, 2018), also supported by the EBU (see Figure 5 below).

Figure 5

Accessibility services icons for a common European standardisation



Source: © DR Design

3.4. Device accessibility

VR hardware typically makes many assumptions about users' abilities that can lead to accessibility problems. For example, most head-mounted displays (HMDs) are quite heavy and require significant strength to wear them, as well as the ability to execute a large range of movements. They may also be problematic for people who wear glasses or hearing aids. Moreover, the surrounding tracking systems (such as eye-tracking) create an infrastructure that needs to be set up and calibrated before using the HMD. Although the most recent generation of VR headsets aims to reduce the number of hardware components, still cables and chargers are remaining included (as is the case with the HTC VIVE PRO headset) also making the equipment hard to transport.

Flexible and customizable design is becoming more common in VR hardware, which could be a trend for accessibility; for instance, many HMDs allow the user to specify the distance between the eyes or between the eyes and the display in order to optimise the experience (Mott *et al.,* 2019). This flexibility is a great framework that may particularly benefit accessibility, i.e., by supporting novel or customized controllers used by all people regardless of their capabilities. This is the case of AR equipment, which is much lighter than VR headsets, and in the future, they may consider incorporating prescription lenses to benefit end-users with low vision. Other accessibility systems could also be integrated into future devices, such as voice control or easy-to-read instructions to facilitate the set-up of the HMDs.

4. Conclusions

More than ever, technology and new media formats are enabling the empowerment of end-users, both consumers and prosumers. The interaction of users in the current Information Society plays a fundamental role in the social integration and democratic participation of all citizens. However, technology is still designed with accessibility as an afterthought, far from user-centric design. Allowing easy access to content and guiding the user in control of media services are two of the most recent UN and European media regulations. The only way to comply with European regulations, as well as with user expectations, is to find the best way to integrate accessibility within the new immersive media.

The main idea of this chapter is that for XR to be accessible, both the content and the container must also be accessible. Creating immersive, accessible content is an investment that requires a series of technical knowledge, but that is useless if, in the end, users cannot access this content. Both the XR and MA communities should seize this moment as devices and standards continue to evolve to include accessibility in XR from the ground up. That is the only way to create a more usable and inclusive technological future for users of all abilities. In addition, promoting the concepts of accessibility is also a way to guarantee all people equal access to culture.

5. Acknowledgments

The author is member of TransMedia Catalonia, a research group funded by Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya, under the SGR funding scheme (ref. code 2021SGR00077). This article reflects only the author's view, and the funding institutions hold no responsibility for any use that may be made of the information it contains. This article is part of Marta Brescia's PhD in Translation and Intercultural Studies at the Department of Translation, Interpreting and East Asian Studies (Departament de Traducció i d'Interpretació i d'Estudis de l'Àsia Oriental) of Universitat Autònoma de Barcelona.

This research has been partially funded by the H2020 project TRACTION (under Grant Agreement 870610).

6. References

Agulló, Belén, & Matamala, Anna, «Subtitling for the deaf and hard-of-hearing in immersive environments: results from a focus group», *Jostrans. The Journal of Specialised Translation*, vol. 33, 2019, pp. 217-235.

Agulló, Belén, & Orero, Pilar, «3D Movie Subtitling: Searching for the best viewing experience», *Studi di Comunicazione e Mediazione linguistica e culturale COME,* vol. *2*, 2017, nº 1, pp. 91-102.

Agulló, Belén, Matamalana, Anna, & Orero, Pilar, «From disabilities to capabilities: testing subtitles in immersive environments with end-users», *HIKMA*, vol. 17, 2018, pp. 195-220.

Aissaoui, Najeh, «The digital divide: a literature review and some directions for future research in light of COVID-19», *Global Knowledge, Memory and Communication,* 2021*.* DOI: https://doi.org/10.1108/GKMC-06-2020-0075

Aylett, Ruth, & Louchart, Sandy, «Towards narrative theory of virtual reality», *Virtual Reality*, vol. 7, 2–9, 2003. DOI: https://doi.org/10.1007/s10055-003-0114-9

Beaunoyer, Elisabeth, Dupéré, Sophie, & Guitton, Matthieu J., «COVID-19 and digital inequalities: reciprocal impacts and migration strategies», *Computer in Human Behavior,* vol. 111, 2020, nº 6.

Bernabé Caro, Rocío, & Orero, Pilar, «Easy to read as multimode accessibility services». *Hermeneus. Revista de Traducción e Interpretación*, vol. 21, 2019, pp. 53-74.

Brennan, Dominic, (2017, December 6). *New report from ILMxLab & Disability Visibility Project shares insights on VR accessibility design.* Retrieved from: https://www.roadtovr.com/disability-visibility-project-and-ilmxlab-share-lessons-learned-about-vr-accessibility-design/

Brescia-Zapata, Marta, «The present and future of accessibility services in VR360 players», *inTRAlinea*, vol. 24, 2022.

Brown, Andy (2017, October 26). User testing subtitles for 360º content. Retrieved from: https://www.bbc.co.uk/rd/blog/2017-10-subtitles-360-video-virtual-reality-vr

Brown, Andy, Turner, Jayson, Patterson, Jake, Schmitz, Anastasia, Armstrong, Michael & Glancy, Maxine (2018, April 30). *Exploring subtitle behavior for 360º video.* Retrieved from: https://www.bbc.co.uk/rd/publications/whitepaper330

Brown, Andy, & Patterson, Jake, (2017, October 26). *Designing subtitles for 360º content.* Retrieved from: https://www.bbc.co.uk/rd/blog/2017-03-subtitles-360-video-virtual-reality

Brown, Andy, Turner, Jayson, Patterson, Jake, Schmitz, Anastasia, Armstrong, Mike, & Glancy, Maxine, «Subtitles in 360-degree video». *Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video*, 2017, pp. 3–8. DOI: https://doi.org/10.1145/3084289.3089915

Crafton, Donald, *The talkies: American cinema's transition to sound, 1926-1931*, vol. 4, University of California Press, 1999.

Creed, Chris, Al-Kalbani, Maadh, Theil, Arthur, Sarcan, Sayan, & Williams, Ian, «Inclusive AR/VR: accessibility barriers for immersive technologies». *Universal Access in the Information Society*, 2023. DOI: https://doi.org/10.1007/s10209-023-00969-0

Eiersø, Caroline, (2018, Mai 15). *Smart icons design — common European standardization.* Retrieved from https://www.dr.dk/om-dr/about-dr/smart-icons-design-common-european-standardization

Fidyka, Anita, & Matamala, Anna, «Audio description in 360º videos: results from focus groups in Barcelona and Kraków». *Translation Spaces*, vol. 7, 2018, nº 2, pp. 285-303. DOI: https://doi.org/10.1075/ts.18018.fid

Gunkel, Simon, Stokking, Hans M., Prins, Martin J., van der Stap, Nanda, ter Haar, Frank B. & Niamut, Omar A., (2018, June 15). *Virtual reality conferencing: Multi-user immersive VR experiences on the web.* DOI: http://doi.org/10.5281/zenodo.3530681

Hughes, Chris J., & Montagud-Climent, Mario, «Accessibility in 360° video players». *Multimedia Tools and Applications*, vol. 80, 2021, pp. 30993-31020. DOI: https://doi.org/10.1007/s11042-020-10088-0

Hughes, Chris J., Brescia-Zapata, Marta, Johnston, Matthew, & Orero, Pilar, «Immersive captioning: Developing a framework for evaluating user needs». *IEEE AIVR 2020: 3rd International Conference on Artificial Intelligence & Virtual Reality 2020*, 2020, Retrieved from: http://usir.salford.ac.uk/id/eprint/58518/

Hughes, Chris J., Montagud-Climent, Mario, & tho Pesch, Peter, «Disruptive approaches for subtitling in immersive environments», *Proceedings of the 2019 ACM*

International Conference on Interactive Experiences for TV and Online Video, 2019, pp. 216–229. DOI: https://doi.org/10.1145/3317697.3325123

Jankowska, Anna, *Translating Audio Description Scripts*, Berlin, Peter Lang Verlag, 2015. DOI: https://doi.org/10.3726/978-3-653-04534-5

Ko Dong-uk, Ryu Hokyoung, & Kim Jieun, «Making new narrative structures with actor's eye-contact in cinematic virtual reality (CVR)», *Interactive Storytelling: 11th International Conference on Interactive Digital Storytelling, ICIDS,* vol. 11318, 2018, pp. 343–347.

Matamala, Anna, & Orero, Pilar, *Listening to Subtitles. Subtitling for the Deaf and Hard-of-Hearing,* Berna, Peter Lang, 2010.

Matamala, Anna, & Orero, Pilar, «Standardising accessibility: Transferring knowledge to society», *Journal of Audiovisual Translation,* vol. 1, 2018, pp. 139–154. DOI: https://doi.org/10.47476/jat.v1i1.49

Milgram, Paul, & Kishino, Fumio, «A taxonomy of mixed reality visual displays», *IEICE TRANSACTIONS on Information and Systems*, vol. *77*, nº12, 1994, pp. 1321-1329.

Molina-Siles, Pedro Javier, Piquer Cases, Juan Carlos, Cortina Maruenda, Francisco Javier, «El color en los comienzos del cine. De la aplicación manual al Technicolor», *Libro de actas del X Congreso Nacional del Color*, 2013.

Montagud-Climent, Mario, Soler-Vilageliu Olga, Fraile Vila, Isaac, & Fernández Langa, Sergi, «VR360 subtitling: Requirements, technology, and user experience», *IEEE Access*, vol. 9, 2021, pp. 819-2838. DOI: https://doi.org/10.1109/ACCESS.2020.3047377.

Montagud-Climent, Mario, Fraile Vila, Isaac, Meyerson, Einar, Genís, María, & Fernández Langa, Sergi (2019, September 19). Imac player: Enabling a personalized consumption of accessible immersive content. DOI: https://doi.org/10.6084/m9.figshare.9879254.v1

Mott, Martez, Cutrell, Edward, Gonzalez Franco, Mar, Holz, Christian, Ofek, Eyal, Stoakley, Richard, & Ringel Morris, Meredith, «Accessible by Design: An Opportunity for Virtual Reality», *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, 2019, pp. 451-454. DOI: https://doi.org/10.1109/ISMAR-Adjunct.2019.00122

Orero, Pilar, Rai, Sonali, & Hughes, Chis, J., «The challenges of audio description in new formats», *Innovation in Audio Description Research*. London, Routledge, 2019.

Orero, Pilar, «Translation, accessibility, and minorities», *The Routledge Handbook of Translation and Media* (pp. 384-399). London, Taylor Francis, 2022.

Puerta Domínguez, Simón, «Del color al sensorama. Cine e industria cultural», *Razón crítica*, 1, 170–205, 2016. DOI: http://dx.doi.org/10.21789/25007807.1141

Remael, Aline, Reviers, Nina, & Vercauteren, Gert (eds.) *Pictures Painted in Words: ADLAB Audio Description Guidelines*, Trieste, EUT. 2015.

Rothe, Sylvia, Tran, Kim, & Hussmann, Heinrich, «Positioning of subtitles in cinematic virtual reality», *International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*, 2018. DOI: https://doi.org/10.2312/egve.20181307

Saunders, Benjamin, Camgoz, Necati Cihan, & Bowden, Richard, (2021, July 23). Mixed SIGNals: Sign language production via mixture of motion primitives. *Proceedings of the International Conference on Computer Vision (ICCV).* Retrieved from https://openresearch.surrey.ac.uk/esploro/outputs/conferencePaper/Mixed-SIGNals-Sign-Language-Production-via/99609266002346

Skult Natasha, & Smed Jouni, «Interactive storytelling in eXtended Reality: Concepts for the design», *Game user experience and player-centered design. International series on computer entertainment and media technology*, Springer, Cham, 2020. DOI: https://doi.org/10.1007/978-3-030-37643-7_21

Slater, Mel, & Wilbur, Sylvia, «A framework for immersive virtual environments five: Speculations on the role of presence in virtual environments», *Presence: Teleoper. Virtual environ.* Vol. *6,* 1997, pp. 603–616. DOI: https://doi.org/10.1162/pres.1997.6.6.603

Vertebrae. (2109). Immersive vs traditional media. How interactivity and realism influence persuasions. Retrieved from: https://cdn.nrf.com/sites/default/files/2020-02/Vertebrae_3d-AR-eCommerce_Whitepaper-2019%20%281%29_0.pdf

World Federation of the Deaf and World association of sign language interpreters. (2018). WFD and WASLI statement on use of signing avatars. Retrieved from: https://docs.google.com/viewerng/viewer?url=https://wfdeaf.org/wp-content/uploads/2017/02/WFD-and-WASLI-Statement-on-Avatar-FINAL-14032018-Updated-14042018-1.pdf&hl=en_US

XR Accessibility User Requirements. Retrieved from: https://www.w3.org/TR/xaur/

International Organisation for Standardisation. (2018). Information technology — User interface component accessibility — Part 23: Visual presentation of audio information (including captions and subtitles). https://www.iso.org/standard/70722.html

2.2. Article 2: Brescia-Zapata, M. (2021). The present and the future of accessibility services in VR360 players. *inTRAlinea, 24*.

The Present and Future of Accessibility Services in VR360 Players

By Marta Brescia-Zapata (Universidad Autònoma de Barcelona, Spain)

Abstract & Keywords

English:

Technology moves fast and making it accessible turns it into an endless game of cat and mouse. Immersive content has become more popular over the years and VR360 videos open new research avenues for immersive audiovisual experiences. Much work has been carried out in the last three years to make 360º videos accessible (Agulló and Orero 2017; Fidyka and Matamala 2018; Agulló and Matamala 2019) and the chase continues. Unfortunately, VR360 videos cannot yet be classified as accessible because subtitles and/or audio description are not always available. This paper focuses on analysing to what extent the main accessibility services are integrated into the most popular and available commercial VR360 players. It also analyses the development of a fully customisable and accessible player which has been created following the EU standard EN17161 and a user centric approach. The first part deals with the relationship between technology and audiovisual translation (AVT) studies, followed by a brief overview of existing eXtended Reality (XR) content. Section 3 provides a list of the VR360 players available commercially and compares them in relation to the main accessibility services: subtitling for the deaf and the hard-of-hearing (SDH), audio description (AD) and sign language (SL). The last section presents a new player developed by the ImAc project, which has accessibility at its heart.

Keywords: accessibility, virtual environment, 360º video, subtitling, audio description, sign language interpreting, personalisation

---

Introduction

Immersive media such as virtual reality (VR) and augmented reality (AG) are technologies which have the potential to transform the way we work, communicate and experience the world. VR is capable of transforming and innovating traditional sectors such as manufacturing industries, construction, and healthcare. It can also revolutionise education, culture, travelling and entertainment. Immersive environments have enjoyed much popularity in video games and other commercial applications such as simulators for kitchen design or surgical training (Fida et al. 2018; Parham et al. 2019). While tools to generate these environments have been available for some time—the most popular being Unity— photographs and videos are growing in number. This is associated with the availability of cameras to record in 360º. In all cases, the aim is to provide an immersive and engaging audiovisual experience for viewers. The two major types of VR content are 360º videos (web-based) and 3D animations (synthetic or Unity-based). This article deals only with the former.

In this context, 360º video (aka VR360 video) has become a simple, cheap, and effective way to provide VR experiences (Montagud et al. 2020b). The potential of VR360 videos has led to the development of a wide variety of players for all platforms and devices (Papachristos et al. 2017); like computers, smartphones and Head Mounted Displays (HMDs). Traditionally, accessibility has been considered by the media as an afterthought, despite many voices asking for it to be included at the design stage of any process. Moreover, the lack of standardisation and guidelines in this novel medium has resulted in non-unified solutions, focusing only on specific requirements.

This situation has served as the motivation for exploring to what extent have accessibility services been integrated in the available most popular commercial VR360 players. This study is a necessary first step towards identifying the advantages and challenges in this novel field. The second contribution of this paper is the presentation of a fully accessible player, developed under the umbrella of the EU H2020 funded project ImAc**[1]**.

The structure of this paper will now be presented. Section 1 deals with the relationship between new technologies and audiovisual translation (AVT) studies by giving a general overview, with a focus on the media accessibility (MA) field. To do so, this section is divided into two parts: the first outlines the state-of-the-art by summarising the main accessibility guidelines; the second focuses on new technologies as a catalyst, making the interaction between the final users and the main accessibility services (namely AD, SDH and SL) possible. Section 2 offers a general overview of XR content; focusing on VR, AR and 360º video. Section 3 analyses the degree of accessibility in the main commercial VR360 players that are available. After presenting the solutions offered by existing VR360 players, a fully accessible player (the ImAc player) is presented in Section 4. Finally, Section 5 offers a discussion on existing limitations, improvements and solutions provided by the ImAc player, as well as some ideas and avenues for future work.

1. Immersive media and AVT: an overview

The unceasing advances in Information and Communication Technologies (ICT) open the door to new fascinating opportunities within MA studies, despite also posing a series of challenges that this discipline has never before faced. This section deals with how the development of new technologies has affected the field of AVT. Firstly, by revisiting the existing accessibility guidelines and recommendations; and secondly, by analysing the importance of researching the tools and technologies needed to generate accessible content.

1.1. Accessible guidelines and recommendations

It is not yet possible to talk about immersive content "for all" because audio description (AD), subtitling for the deaf and hard-of-hearing (SDH) and sign language (SL) interpreting —among others— are not always available for 2D content, let alone 3D or XR.

Regarding legislation, in 2008 the United Nations (UN) issued the Convention on the Rights of Persons with Disabilities (CRPD), currently ratified by 181 countries. In Article 30 (section 1 b), the convention states that "State Parties […] shall take all appropriate measures to ensure that persons with disabilities: […] (b) Enjoy access to television programmes, films, theatre and other cultural activities, in accessible formats." This convention has helped to promote the proliferation of accessibility services in media content. AVT, and more specifically MA (Remael et al. (eds) 2014; Greco 2016), is the field in which research on access to audiovisual content has been carried out in the last few years, generally focusing on access services such as audio description (AD), subtitling for the deaf and hard-of-hearing (SDH) or sign language (SL) interpreting, among others (Matamala and Orero 2010; Arnáiz-Uzquiza 2012; Romero-Fresco 2015; Fidyka and Matamala 2018).

The CRPD (UN 2006), together with the Convention on the Protection and Promotion of the Diversity of Cultural Expressions (UNESCO 2005) created legal repercussions within the scope of the accessibility of cultural goods in the European Union (EU). These two conventions resulted in two Directives and one Act that demand accessibility services not only on websites, services and spaces; but also for content and information offered at all cultural venues and events (Montagud et al. 2020a). The three pieces of legislation are:

1. The EU Directive on the Accessibility of Websites and Mobile Applications transposed into the law of each EU member state by September 2018. It is based on Web Content Accessibility Guidelines (WCAG) 2.0 guidelines, and references EN301549 as the standard which will enable websites and apps to comply with the law.
2. The Audiovisual Media Services Directive (AVMSD) approved in 2018 gave member states 21 months to transpose it into national legislation. It addresses key issues, for example: rules to shape technological developments that preserve cultural diversity and protect children and consumers whilst safeguarding media pluralism.
3. The European Accessibility Act which takes the form of a legally binding Directive for all member states. It is a law that aims at making many EU products and services (smartphones, computers, TV programs, e-books, websites, mobile apps etc.) more accessible for persons with disabilities.

These three pieces of EU legislation demand accessibility services for information and content and at all cultural venues and events; as opposed to only websites, services and spaces. As is reviewed in the following sections, all audiovisual products need to be accessible, including those providing XR content for their audience.

1.2. The impact of new technologies in AVT and MA studies

Technical advances have not only modified media applications, but accessibility services and profiles also. Even end users have been affected by the advent of new ways to access media content. Traditionally, AVT and MA research has focused on the analysis of translated audiovisual texts and their many transformations or *transadaptations* (Gambier 2003). Some years ago, the translated audiovisual text was static in the sense that one format was distributed and consumed by all. Audiences are now allowed to make decisions about the media they consume, beyond the two traditional values: level of sound and image contrast. These days, technology allows for media personalisation (Orero forthcoming; Orero et al. forthcoming; Oncins and Orero 2020). When watching media content on TV or YouTube, subtitles can be read in different sizes, positions, and colours (Mas Manchón and Orero 2018). The speed of the content reproduction can also be altered, and the audiovisual text is increasingly changing according to individual needs. Media accessibility has shifted from a one-size-fits-all approach to customisation. Technology has allowed for this change. The Internet of Things lets customers decide the way in which each media object is set-up. Artificial Intelligence is now integrated and understands individual choices, needs and preferences.

On the one hand, technology is the basis of the tools used for creating or adapting content; on the other, it is the basis for consuming content (Matamala 2017). Regarding content creation and adaptation, there is a wide range of professional and amateur translation software for both subtitling (e.g., Aegisub, Subtitle Workshop, VisualSubSync, WinCaps, SOftwel, Swift Create, Spot, EZTitles, etc.) and AD (e.g., Softel Swift ADePT, Fingertext, MAGpie2, Livedescribe or YouDescribe). It is not that easy to find studies regarding preferences or comparative analysis of the previously mentioned tools. Concerning subtitling, Aulavuori (2008) analyses the effects of subtitling software on the process. Regarding AD, Vela Valido (2007) compares the existing software in Spain and the USA. Oncins et al. (2012) present an overview of existing subtitling software used in theatres and opera houses and propose a universal solution for live media access which would include subtitling, AD and audio subtitling, among other features.

The tool used by the users to consume audiovisual media is the media player. The choice of device (mobile phone, TV, PC, smart watch, tablet), type of content and the available accessibility services determine the level of personalisation (Gerber-Morón et al. 2020). Subtitles, for example, are displayed differently according to screen size, type of screen, and media format. The choice of the subtitle is made through the media player settings, making understanding its capabilities a basic departure point when analysing translated immersive media content.

The existence of such a wide range of technical solutions opens the door to many research questions. It remains to be seen how new developments will affect the integration of accessible services into existing tools to adapt to the needs of end users: everyone has the right to access the information provided by media services (including immersive experiences). This paper offers a valuable resource to content providers seeking to improve their products, users with accessibility needs in choosing the player that best suits their requirements, and to researchers setting out to identify the challenges and possibilities that this new field poses to AVT and MA.

2. The rise of VR and AR: VR360 video becomes mature

In this section, a general overview of VR, AR and VR360 videos will be given, highlighting the impact that these new technologies may have on our society at different levels.

Although immersive content production is still at an early stage and is generally used in professional environments such as hospitals and universities; in the future it might be used in the daily lives of ordinary users. As with all new technologies, VR will make our lives easier: from going to the supermarket to online shopping (Lee and Chung 2008). Immersive environments are the new entertainment experiences of the 21st century, from museums (Carrozzino and Bergamasco 2010) and theatres to music events such as opera (Gómez Suárez and Charron 2017). They allow users to feel as if they are being physically transported to a different location. Though the most popular applications are cultural representations, it is a great tool for larger audiences and functionalities such as leisure, sport (Mikami et al. 2018), tourism (Guttentag 2010) and health (Rizzo et al. 2008).

There are various solutions that can provide such an experience, such as stereoscopic 3D technology which has re-emerged in films during the last ten years (Mendiburu 2009). Nevertheless, this format is nothing new. It has been available since the 1950s, but the technology has not been ready to deliver quality 3D (González-Zúñiga et al. 2013). Both the quality of the immersive experience and the sense of depth depend on the display designs, which for 3D content are diverse and lacking in standards (Holliman et al. 2011). However, stereoscopy did not become the main display for AV products; perhaps due to the lack of standardisation, the intrusive nature of 3D, and uncomfortable side effects such as headaches or eyestrain (Belton 2012). According to Belén Agulló and Anna Matamala (2019), "the failure to adopt 3D imaging as mainstream display for AV products may have opened a door for VR and 360º content, as a new attempt to create engaging immersive experiences''. VR stands for 'virtual reality' and it takes on several different forms, 360º video being one of them. However, VR and 360º videos are two different mediums (see Table 1). In 360º video, multi-camera rigs (often static) are used to record live action in 360º, giving the consumer a contained perspective of a location and its subjects. VR renders a world in which, essentially, the consumer operates as a natural extension of the creator's environment, moving beyond 360º video by enabling the viewer to explore and/or manipulate a malleable space. In 360º video, the consumer is a passenger in the storyteller's world; in VR, the consumer takes the wheel. The storyteller directs the viewer's gaze through this situational content by using elemental cues such as light, sound and stage movement. The traditional notion of the fourth wall has been eliminated.

	VIRTUAL REALITY	**360º VIDEO**
Photography	Digital environment	Live action
Mobility	Immersive world that you can walk around in.	360º view from camera's perspective. Limited to filmmaker's camera movements.
Video timeline	Video can progress through a series of events. Experiences can be held in an existing world to be explored by the user (6 degrees of freedom).	Video progresses on a timeline created by the filmmaker's camera movements (3 degrees of freedom).
Platforms	A full experience requires an HMD.	Available on 360º compatible players (desktop and mobile).
Story	The filmmaker does not control the physical location of the viewer in the built environment and must capture attention and motivate the user to travel in the direction of the events of the story.	The filmmaker controls the physical location of the camera but must capture the attention of viewers to direct the story.

Table 1: Differences between VR video and 360º video (Based on Sarah Ullman)

Other less commercial immersive technologies are mixed and augmented reality. Paul Milgram and Fumio Kishino (1994: 1321) define those terms as:

> Mixed Reality (MR) visual displays [...] involve the merging of real and virtual worlds somewhere along the 'virtuality continuum' which connects completely real environments to completely virtual ones. Probably the best known of these is Augmented Reality (AR), which refers to all cases in which the display of an otherwise real environment is augmented by means of virtual (computer graphic) objects.

Julie Carmigniani and Borko Furht (2001: 3) define AR as "a real-time direct or indirect view of a physical real-world environment that has been enhanced by adding virtual computer-generated information to it." The properties of AR are that it "combines real and virtual objects in a real environment; runs interactively, and in real time; and registers (aligns) real and virtual objects with each other" (Azuma et al. 2001: 34).

It is not easy to outline a state-of-the-art when talking about immersive environments as the development in VR technology is happening at an unprecedented speed. Moreover, Covid-19 has helped to accelerate virtual experiences. While it is early to evaluate, with the writing of this paper taking place during lockdown, what is now considered "the new normal" will have strong non-presential and virtualization elements. Systems and applications in this domain are presented on a daily or weekly basis. AR/VR technology makes use of sensory devices to either virtually modify a user's environment or completely immerse them in a simulated environment. Specially designed headsets and glasses can be used for visual immersion, while handhelds and wearables offer tactile immersion. Optical devices such as Facebook's Oculus Rift and Sony's PlayStation Virtual have shipped millions of units as consumers look to explore the possibilities offered by virtual environments. Nevertheless, the adoption rate for AR/VR devices is relatively low when compared to other consumer electronics, though many of the world's biggest technological companies see the promise of AR/VR technology and have begun to allocate significant budgets to develop it.

Stable growth of the VR and AR markets is expected both in Europe and around the world, as can be seen in graphic 1. According to a report by Ecorys, the total production value of the European VR & AR industry was expected to increase to between €15 billion and €34 billion by 2020 and to directly or indirectly account for 225,000 to 480,000 jobs. Also, wider supply chain impacts are expected to indirectly increase the production value to between €5.5 billion and €12.5 billion and generate an additional 85,000-180,000 jobs. Due to the strong growth of content-related VR activities, the share of Europe in the global market is expected to increase.

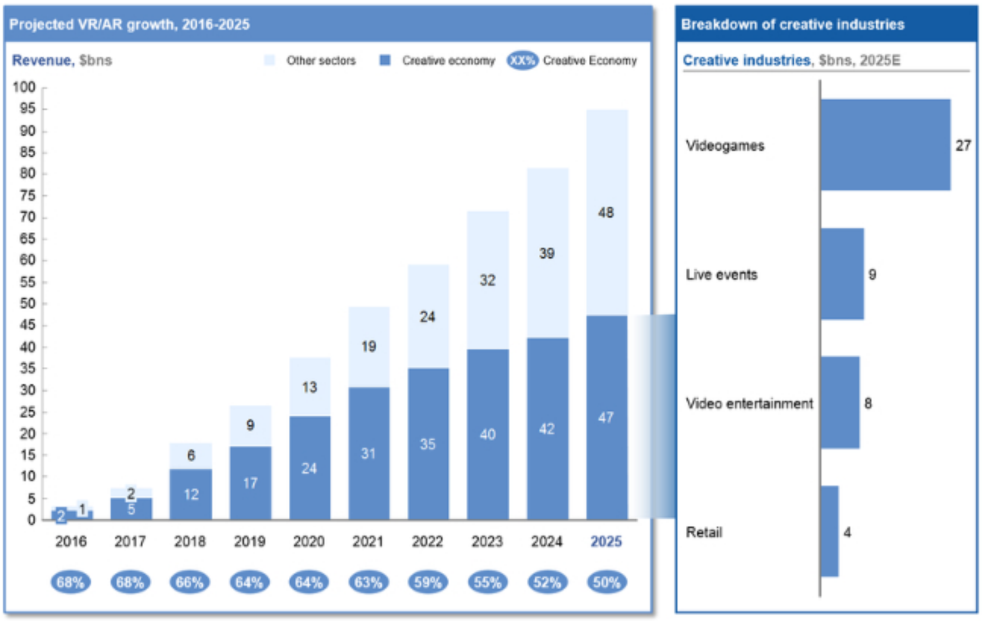


Figure 1. AR and VR global market growth, 2016-2025 (World Economic Forum 2017)

The 360º video has been around for several years with differing levels of sophistication and polish. However, with the advancement of camera technology combined with the development of software for handling the images, 360º video is being used in an increasing number of ways.

2018 was considered the Year of 360º video. Indeed, the trend of watching 360º videos on web browsers, tablets or mobiles has increased. Even if PCs, tablets or mobiles are currently still the main devices for watching 360º video, the use of VR Headsets is starting to grow as well. There are three main factors that could explain why the 360º video market is considered to be already developed: 360º video capture devices are more sophisticated and affordable, the increasing number of web and mobile players (YouTube, Facebook, Twitter and the use of mobile phones as HMDs), and the slightly decreasing price of VR headsets (HDM).

VR360 video is now being used by all kinds of people and organisations for sharing immersive stories and extreme experiences in stunning locations. Estate agents, airlines and the hospitality industry are embracing VR360 video to show off their goods; while broadcasters, educators and social media platforms are also experimenting with the format. Many videographers are now learning about VR360 video creation for various purposes and platforms, for example:

Immersive journalism. The New York Times has started publishing The Daily 360º, a short 360 news report – often in 4K resolution – with multiple cuts between static shots, and a reporter acting as the narrator. Done well, 360º adds a unique eyewitness feel to storytelling and reporting.

360º time-lapses. Many 360º cameras allow the capture of time-lapses, which are sequences of 360º images recorded at set intervals to record changes that take place slowly over time. Speed up the frames and the effect can be exceptional for something as delicate as the Milky Way emerging at night.

Live 360º video cameras can now capture and live-stream spherical imagery and most popular online and social media platforms have recently been updated to support 360º content. Viewers can now tune-in to live 360º broadcasts and download the video afterwards.

3. VR360 players: web and mobile apps

In this section, a list of the main web VR360 players will be presented. Each player will be analysed based on how, and to what extent, they integrate the main accessibility services.

There is an increasing number of VR360 videos available over the internet and the majority of browsers support them, meaning we can enjoy watching VR content online with no need for a VR device. However, it is still necessary to download a 360º video player which can support them. Regular video players such as Windows Media Player do not support them, although they probably will in the future. Nevertheless, there are some players aside from regular video formats that also support VR videos. Most of them are available on the web, but the majority fail in terms of accessibility. The table in the Appendix shows a comparative analysis of accessibility services offered by the most popular executable players.

The selection has been made for two different approaches and from a descriptive perspective. Firstly, following a top-down approach, specialized media such as magazines and papers were taken as a reference to elaborate on the first draft of the list. Secondly, users' opinions and comments in different online forums were taken into account to complete the final list.

GOM is a South Korean product from the Gretech Corporation, one of the best-known video players, mainly used for playing 'regular' videos but also supports 360º video. It is a multilingual player, as you can select the language of the player. It can play 360º videos downloaded to your computer and also directly from YouTube. It is possible to search and upload subtitles and to adjust several features, such as style and position. This player offers the ability to load two separate subtitle files: one displayed at the top of the screen and the other at the bottom.

Codeplex Vr Player is an experimental open-source VR Media Player for Head-Mounted Display devices like Oculus Rift. Not only does it provide the function of playing VR videos, but it also lets users watch 2D and even 3D videos. Its user interface is designed to be intuitive which makes it very easy to use. It provides a free version and a professional paid version.

Total Cinema 360 Oculus Player is a VR App for Oculus Rift developed by Total Cinema 360. It is specifically designed to capture fully interactive, live action spaces in high quality 360º video. It comes equipped with four demonstration videos that allow you to pause, zoom and adjust eye distance. You can also upload your own 360º video content for use with Total Cinema 360.

RiftMax was developed in Ireland by the VR software developer Mike Armstrong. It is not only a VR video player, as it allows you to interact with other people in scenarios such as parties or film screenings. It also enhances video with good effects that come out of the screen. It does not support subtitles and is only available in English.

SKYBOX was developed in the UK by Source Technology Inc and supports all stereo modes (2D and 3D, or 180º and 360º). You can choose multiple VR theatres when you are watching 2D or regular 3D videos; including Movie Theater, Space Station and Void. It supports all VR platforms: Oculus, Vive, Gear VR and Daydream. SKYBOX supports external text-based subtitles (.srt/.ssa/.ass/.smi/.txt) when you are watching a non-VR video. If a video has multiple audio tracks, you can click on the "Track" tab and select the corresponding audio track.

VR Player is a Canadian product developed by Vimersiv Inc. It is specially designed for playing virtual reality videos and is a popular program among Oculus Rift users. It plays not only VR, but also 2D and 3D videos. It opens media from multiple resources, such as YouTube URLs or cloud-based services like Dropbox. It allows "floating subtitles" for watching foreign immersive videos. So far, this player is only available in the English version.

Magix VR-X is a German product from MAGIX. The player supports Android, iOS and Windows with Oculus Rift, HTC Vive and Microsoft Mixed Reality. There are six languages available: English, Spanish, French, German, Italian and Dutch. Captions are only available in the Premium version (Photostory Premium VR).

Simple VR was developed in Los Angeles. It provides users with the simplest functions and can serve as a typical media player for users. You can play, stop and pause VR video through simple controls. In addition, it has a super enhancement mode which can improve the fidelity, contrast and detail of VR videos. It also allows something called "splitters" which take multi-track video files (such as .mkv) and feed the video decoder with specific video/audio tracks and subtitles that you can configure.

 After a quick analysis of the players which allow subtitle files to be loaded, it can be observed that these files are rendered as a 2D overlay onto the video window. As there is still no specific subtitle file for VR360 videos, there is no information about where in the 360 scene the subtitle relates to. Regarding AD, some of the players provide support for selecting alternative audio tracks which can be used for playing the AD track. However, there is no mechanism for mixing the AD over the existing audio track in the player. Concerning SL, none of the players provide any mechanism for adding this access service. They also do not offer the possibility to overlay an additional video stream, which could be used for the SL service.

4. ImAc project and player

This section will present and analyse the VR360 video player created under the umbrella of the H2020 funded ImAc project following the Universal Design approach and the "Born Accessible" concept. The design departed from user specifications and took accessibility requirements into account in the development.

ImAc was a European project funded by the European Commission that aimed to research how access services (subtitling, AD, audio subtitles, SL) could be integrated in immersive media. The project aimed to move away from the constraints of existing technologies into an environment where consumers could fully customise their experience (Agulló, 2020). The key action in ImAc is to ensure that immersive experiences address the needs of different kinds of users. One of the main features of the ImAc project was the user-centred methodological approach (Matamala et al. 2018), meaning that the design and development of the system and tools were driven by real user needs, continuously involving users in every step. The player was developed after gathering user requirements from people with disabilities in three EU countries: Germany, Spain and UK. User input was gathered in two iterations through focus groups and pre-pilot actions.

The first step in the user centric methodology was to define the profile of the end users. Two different profiles were created: professional user and advanced home user. Professional users were considered to be those who would use the tools at work: IT engineers, graphic designers, subtitlers, audio describers and sign language interpreters (signers). On the other hand, the advanced home users were people with disabilities who consumed the media content: the deaf, hard-of-hearing, blind, low vision users, and the elderly. To successfully profile home users, a number of considerations were taken into account beyond disability, such as level of technological knowledge and VR environments. This was decided in order to engage home users in an open conversation regarding their expectations and match them accordingly with the innovation. Only users with knowledge or experience in either functional diversity or technology were consulted. Other profiling features of the home users were oral/written languages (Catalan, German, Spanish and English) and three visual-gestural languages (Catalan Sign Language, German Sign Language and Spanish Sign Language). Other significant profiling factors were level of expertise in the service that the participant was testing (audio description, audio subtitling, sign language, subtitling), sensorial functionality (deaf, hard-of-hearing, blind, low vision) and age. As some degree of hearing or vision loss can often be linked to age, the elderly were included in the home users' category.

Once the two groups of end users (advanced home and professionals) had been defined, they formulated two user requirements: home and professional requirements. The former described the functions exposed by the ImAc services towards consuming media, and the latter described the functions from a working perspective. Three versions or iterations of the requirements were carried out. The first version was based on focus groups in which the user scenarios created by the ImAc partners were evaluated by both professional and home users. User scenario refers to what the already identified user would be experiencing and how (e.g. how the interface deals with AD depending on angle of visualisation). The second was conducted after the pre-pilot tests, where prototypes of accessible immersive media content were presented to the target group of home users (i.e., for the visual access services, the tests focused on the preferred size of the area to display the services and the preferred ways of guiding the users to the speaker). In this way, more extensive feedback on specific issues could be gathered and the home user requirements were subsequently fine-tuned. The final iteration took place after the demonstration pilots which involved both professional and home users. The resulting list of final requirements provides the basis for the further development and quality assurance of the ImAc platform.

After compiling all the information regarding the end user profile and requirements, the user interface (UI) was designed. The aim was to have a concept that was flexible enough to extend the settings later, based on the results of the user testing. The main challenge was to integrate four services with a large number of settings, while avoiding a long and complex menu. The UI design to access accessibility services in the ImAc player was based on existing players (legacy players from catch up TV services, web players for video-on-demand and streaming services and VR players); taking these as a starting point, a design for a "traditional UI" was developed. An "enhanced accessibility UI" was developed in parallel, the two were combined, and the resulting ImAc player UI offers aspects of both. The implementation of the UI allows access to the accessibility services in the ImAc portal and the player reflects their status after feedback from the user tests.

The success of the ImAc player and the reason why it has been presented as the main example of an accessible player is because it follows both the "Universal design" and "Born accessible" concepts. The first term comes from the European Standard EN 17161 (2019) 'Design for All - Accessibility following a Design for All approach in products, goods and services - Extending the range of users.' It specifies the requirements in design, development and provision of products, goods and services that can be accessed, understood, and used by the widest range of users, including persons with disabilities. Along with the European Standard EN17161, there is an EU standard for accessible technologies: the EN301549 (version 3.1.1.: 2019). These two standards secure the concepts of Universal Design and Born Accessible. According to Pilar Orero (2020: 4): "the concept of Born

Accessible closely follows EU legislation and has been proven to be successfully integrated in R&D activities and developments."

5. Conclusions

More than ever, technology is enabling the empowerment of end users both as consumers and prosumers. However, technology is still designed with accessibility as an afterthought: away from user centric design. User interaction in today's Information Society plays a key role in the full social integration and democratic participation of all its citizens. Enabling easy access to content and guiding the user in controlling media services are two of the most recent UN and European media regulations. At the same time, UIs should also be accessible, as the demand for guidance is especially high for accessibility services. Some groups of users need to activate an accessibility service such as subtitles before they can consume the media content. In general, the default setting for a media service is to have all accessibility services switched off. Therefore, it is very important that activating and controlling the accessibility services is made as easy as possible. This article has focused on the accessibility of the media players which are currently available commercially to show the way towards full accessibility in VR, at a time when VR content production is beginning. The article would like to raise awareness of the real possibility of generating accessible VR content from the point of production.

As we have been able to verify by analyzing the most widespread players available commercially, none of them provide access to the full set of accessibility services. VR players do not focus on accessibility services at all and they have not been created departing from user needs. Following the Born Accessible principle to avoid this basic problem, accessibility (and multilingualism) must be considered at the design stage of any process. Most of the players that have been analysed are only available in English, leaving out the users with other linguistic realities. The documented media players that support access to accessibility services do not use aligned conventions for their icons/representation. Using a universal set to represent accessibility services is desirable, and for that reason ImAc uses a set of icons proposed by the Danish Radio for all users in all countries.

To finish, we are living a global change of the consumer landscape due to Covid-19. Confinement measures have changed user demands and moved them further towards the use of online services. The so-called "new normal" will bring a significant increase in remote online activities and virtual experiences will be essential in the near future. Marketing, simulation, leisure, training and communication will need to adapt to new needs, as well as to associate with them. The promotion of the Universal Design and Born Accessible concepts can have a significant role in achieving these goals and supporting the full democratic participation of all people, while protecting their social rights.

Appendix: VR360 players facing accessibility

	SUBTITLE (Deaf and hard-of-hearing)	**AUDIO DESCRIPTION** (Blind and visually-impaired)	**SIGN LANGUAGE AVATAR** (Deaf)	**MULTILINGUAL** (Linguistically-impaired)
GOM Player[2]	Yes	No* (select audio track)	No	Yes
Codeplex VR Player[3]	No	No	No	No
Total Cinema 360 Oculus Player[4]	No	No	No	No
RiftMax VR Player[5]	No	No	No	No
SKYBox VR Video Player[6]	Yes (only in non-VR video)	No (select audio track)	No	No
VR Player[7]	Yes	No	No	No
Magix[8]	No (only in the premium version)	No	No	Yes
Simple VR[9]	Yes	No	No	No

Acknowledgements

The author is member of TransMedia Catalonia, a research group funded by Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya, under the SGR funding scheme (ref. code 2017SGR113). This article reflects only the authors' views and the funding institutions hold no responsibility for any use that may be made of the information it contains). This article is part of Marta Brescia's PhD in Translation and Intercultural Studies at the Department of Translation, Interpreting and East Asian Studies (Departament de Traducció i d'Interpretació i d'Estudis de l'Àsia Oriental) of Universitat Autònoma de Barcelona.

References

Agulló, Belén (2020) “Technology for subtitling: a 360-degree turn”, *Hermeneus*, no. 22.

Agulló, Belén, and Anna Matamala (2019) “The challenge of subtitling for the deaf and hard-of-hearing in immersive environments: results from a focus group”, *The Journal of Specialised Translation*, no. 32, 217–235.

Agulló, Belén, and Pilar Orero (2017) “3D Movie Subtitling: Searching for the best viewing experience”, *CoMe - Studi di Comunicazione e Mediazione linguistica e culturale*, Vol. 2, 91–101, URL: **https://ddd.uab.cat/record/183656** (accessed 8 August 2020).

Arnáiz Uzquiza, Verónica (2012) “Los parámetros que identifican el subtitulado para sordos: análisis y clasificación” *MonTi: Monografías de Traducción e Interpretación*, no. 4, 103–32. URL: https://www.raco.cat/index.php/MonTI/article/view/301235 (accessed 1 August 2020).

Aulavuori, Katja (2008) *Effects of the digital subtitling software on the subtitling process: a survey among Finnish professional television subtitlers*, MA thesis, University of Helsinki.

Azuma, Ronald, Yohan Baillot, Reinhold Behringer, Steven K. Feiner, Simon Julier, and Blair MacIntyre (2001) “Recent advances in augmented reality”, *IEEE Computer Graphics and Applications, 21*(6), 34–47.

Belton, John (2012) “Digital 3D cinema: Digital cinema’s missing novelty phase”, *Film History: An International Journal*, 24(2), 187–195.

Carmigniani, Julie, and Borko Furht (2011) “Augmented reality: An overview” in *Handbook of Augmented Reality*, Borko Furht (ed), Florida, Springer: 3–46.

Carrozzino, Marcello, and Massimo Bergamasco (2010) “Beyond virtual museums: Experiencing immersive virtual reality in real museums” *Journal of cultural heritage,* no. 11, 4:452–458.

EN 17161 (2019) “Design for All - Accessibility following a Design for All approach in products, goods and services - Extending the range of users”, URL: **https://standards.cen.eu/dyn/www/f?p=204:110:0::::FSP_PROJECT,FSP_ORG_ID:62323,2301962&cs=1D28CFDC66E7CEF3CE441294CAA9FEAB** (accessed 15 August 2020).

Fida, Benish, Fabrizio Cutolo, Gregorio di Franco, Mauro Ferrari, and Vicenzo Ferrari (2018) “Augmented reality in open surgery”, *Updates in Surgery, no.* 70, 389–400.

Fidyka, Anita, and Anna Matamala (2018) “Audio description in 360º videos. Results from focus groups in Barcelona and Kraków”, *Translation Spaces*, Vol. 7, no. 2, 285–303.

Gambier, Yves (2003) “Introduction. Screen transadaptation: perception and reception”, *The Translator*, no. 9(2), 171–189

Gerber-Morón, Olivia, Olga Soler-Vilageliu, and Judith Castellà (2020) “The effects of screen size on subtitle layout preferences and comprehension across devices”, *Hermēneus. Revista de Traducción e Interpretación*, Vol. 22, 157–182.

Gómez Suárez, Mónica, and Jean-Philippe Charron (2017) “Exploring concert-goers’ intention to attend virtual music performances”, *AIMAC 14 International Conference on Arts and Cultural Management*, Beiging.

González-Zúñiga, Diekus, Jordi Carrabina, and Pilar Orero (2013) “Evaluation of Depth Cues in 3D Subtitling”, *Online Journal of Art and Design*, no.1, 3:16–29.

Greco, Gian Maria (2016) “On Accessibility as a Human Right, with an Application to Media Accessibility” in *Researching Audio Description. New Approaches*, Anna Matamala and Pilar Orero (eds), UK, Palgrave Macmillan: 11–33.

Guttentag, Daniel (2010) “Virtual reality: Applications and implications for tourism” *Tourism Management*, Vol. 31 (5), 637–651.

Holliman, Nick, Neil A. Dodgson, Gregg E. Favalora, and Lachlan Pockett (2011) “Three-dimensional displays: A review and applications analysis”, *IEEE Trans. Broadcast*, *57*(2), 362–371.

International Organization for Standardization (2015) *ISO/IEC TS 20071-20 Information Technology — User interface component accessibility — Part 20: Software for playing media content,* ISO.

Lee, Kun, and Namho Chung (2008) “Empirical analysis of consumer reaction to the virtual reality shopping mall”, *Computers in Human Behavior*, no. 24, 88–104.

Mas Manchón, Lluis and Pilar Orero (2018) “New Subtitling Possibilities: Testing Subtitle Usability in HbbTV”, *Translation Spaces*, 7(2), 263–284.

Matamala, Anna, and Pilar Orero (eds) (2010) *Listening to Subtitles. Subtitles for the Deaf and Hard of Hearing,* Peter Lang, Berna.

Matamala, Anna (2017) *“Mapping audiovisual translation investigations: research approaches and the role of technology”* in Audiovisual translation-Research and use, Mikolaj Deckert (ed), Frankfurt, Peter Lang: 11-28.

Matamala, Anna, Pilar Orero, Sara Rovira-Esteva, Helena Casas-Tost, Fernando Morales Morante, Olga Soler Vilageliu, Belén Agulló, Anita Fidyka, Daniel Segura Giménez, and Irene Tor-Carroggio (2018). "User-centric approaches in Access Services Evaluation: Profiling the End User", *Proceedings of the Eleventh International Conference on Language Resources Evaluation,* 1–7.

Mendiburu, Bernard (2009) *3D movie making: Stereoscopic digital cinema from script to screen*, Taylor and Francis.

Mikami, Dan, Kosuke Takahashi, Naoki Saijo, Mariko Isogawa, Toshitaka Kimura, and Hideaki Kimata (2018) "Virtual Reality-based Sports Training System and Its Application to Baseball", *NTT Technical Review*, no. 16(3).

Milgram, Paul, and Fumio Kishino (1994) "A taxonomy of mixed reality visual displays", *IEICE Transactions on Information* Systems, no. 77, 1321–1329.

Montagud, Mario, Pilar Orero, and Anna Matamala (2020a) "Culture 4 all: accessibility-enabled cultural experiences through immersive VR360 content", *Personal and Ubiquitous Computing*, URL: https://link.springer.com/article/10.1007/s00779-019-01357-3#citeas (accessed 5 August 2020).

Montagud, Mario, Pilar Orero, and Sergi Fernández (2020b) "Immersive media and accessibility: Hand in hand to the future", *ITU Journal*, URL: **https://www.itu.int/pub/S-JOURNAL-ICTS.V3I1-2020-2** (accessed 11 February 2021).

Oncins, Estel·la, Oscar Lopes, Pilar Orero, and Javier Serrano (2013) "All together now: a multi-language and multisystem mobile application to make live performing arts accessible", *Journal of Specialised Translation*, no.20, 147-164.

Oncins, Este·la, and Pilar Orero (2020) "No audience left behind, one App fits all: an integrated approach to accessibility services", *JosTrans*, no. 34.

Orero, Pilar (2020) *Born Accessible: Beyond raising awareness*, URL: **https://imac-project.eu/wp-content/uploads/2020/05/Policy-Paper_final.pdf** (accessed 5 August 2020).

Orero, Pilar (forthcoming) *"Audio Description Personalisation"* in The Routledge Handbook of Audio Description, Christopher Taylor and Elisa Perergo (eds), London, Taylor Francis.

Orero, Pilar, Rai Sonali, and Chris Hughes (forthcoming) *"The challenges of audio description in new formats"* in Innovation in Audio Description Research, Sabine Braun (ed.), London, Routledge.

Papachristos, Nikiforos M., Ioannis Vrellis, and Tassos A. Mikropoulos (2017) "A Comparison between Oculus Rift and a Low-Cost Smartphone VR Headset: Immersive User Experience and Learning", *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*, 477-481, doi: 10.1109/ICALT.2017.145.

Parham Groesbeck, Eric G. Bing, Anthony Cuevas, Boris Fisher, Jonathan Skinner, Richard Sullivan, and Mulindi H. Mwanahamuntu (2019) "Creating a low-cost virtual reality surgical simulation to increase surgical oncology capacity and capability", *Ecancermedicalscience,* no. 13.

Remael, Aline, Pilar Orero, and Mary Carroll (eds) (2014) *Audiovisual Translation and Media Accessibility at the Crossroads: Media for all 3*, Amsterdam, Rodopi.

Rizzo, Albert, Maria Schultheis, Kimberly A. Kerns, and Catherine Mateer (2004) "Analysis of assets for virtual reality applications in neuropsychology", *Neuropsychological Rehabilitation*, no. 14:1-2, 207–239.

Romero-Fresco, Pablo (2015) *The reception of Subtitles for the Deaf and Hard of Hearing in Europe*, Berlin, Peter Lang.

UN (2006) *Convention on the Rights of People with Disabilities*, URL: **https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities.html** (accessed 15 August 2020).

UNE-EN 301549 (2019) *Accessibility requirements for ICT products and services.*

UNESCO (2005) *Convention on the Protection and Promotion of the Diversity of Cultural Expressions*, URL: **https://en.unesco.org/creativity/convention/texts** (accessed 15 August 2020).

Vela Valido, Jennifer (2007) *Software de audiodescripción: análisis comparativo de los programas existentes en España y EE.UU*, MA thesis, Universitat Autònoma de Barcelona.

Notes

[1] See **https://www.imac-project.org.uk/** [retrieved 11/02/2021]

[2] See **https://www.gomlab.com/** [retrieved 05/08/2020]

[3] See **https://archive.codeplex.com/?p=vrplayer** [retrieved 05/08/2020]

[4] See **http://www.totalcinema360.com/** [retrieved 05/08/2020]

[5] See **https://www.wearvr.com/apps/riftmax-theater** [retrieved 05/08/2020]

[6] See **https://skybox.xyz/en/** [retrieved 05/08/2020]

[7] See **http://www.vrplayer.com/** [retrieved 05/08/2020]

[8] See **https://www.magix.com/us/apps/vrx-player/** [retrieved 05/08/2020]

[9] See **http://simplevr.pro/** [retrieved 05/08/2020]

"The Present and Future of Accessibility Services in VR360 Players", *inTRAlinea* Vol. 24.
Stable URL: https://www.intralinea.org/archive/article/2585

2.3. Article 3: Brescia-Zapata, M., Krejtz, K., Orero, P., Duchowski, A., & Hughes, C. (2022). VR 360º subtitles: Designing a test suite with eye-tracking technology. *Journal of Audiovisual Translation, 5*(2), 233–258. https://doi.org/10.47476/jat.v5i2.2022.184

Journal of
Audiovisual Translation

VR 360º Subtitles: Designing a Test Suite with Eye-Tracking Technology

Marta Brescia-Zapata ✉
Universitat Autònoma de Barcelona

Krzysztof Krejtz ✉
University of Social Sciences and Humanities SWPS

Pilar Orero ✉
Universitat Autònoma de Barcelona

Andrew T. Duchowski ✉
Clemson University

Chris J. Hughes ✉
University of Salford

Abstract

Subtitle production is an increasingly creative accessibility service. New technologies mean subtitles can be placed at any location on the screen in a variety of formats, shapes, typography, font size, and colour. The screen now allows for accessible creativity, with subtitles able to provide novel experiences beyond those offered by traditional language translation. Immersive environments multiply 2D subtitle features to produce new creative viewing modalities. Testing subtitles in eXtended Reality (XR) has expanded existing methods to address user needs and enjoyment of audiovisual content in 360º viewing displays. After an overview of existing subtitle features in XR, the article describes the challenges of generating subtitle stimuli to test meaningful user viewing behaviours, based on eye-tracking technology. The approach for the first experimental setup for implementing creative subtitles in XR using eye-tracking is outlined in line with novel research questions. The choices made regarding sound, duration and storyboard are described.

Citation: Brescia Zapata, M., Krejtz, K., Orero, P., Duchowski, A.T. & Hughes C.J. (2022). VR 360º Subtitles: Designing a Test Suite with Eye-Tracking Technology. *Journal of Audiovisual Translation,* *5*(2), 233–258. https://doi.org/10.47476/jat.v5i2.2022.184

Editor(s): M. Carroll & A. Remael

Received: November 16, 2021

Accepted: July 6, 2022

Published: December 21, 2022

✉ marta.brescia@uab.cat, https://orcid.org/0000-0002-2465-0126
✉ kkrejtz@swps.edu.pl, https://orcid.org/0000-0002-9558-3039
✉ pilar.orero@uab.cat, https://orcid.org/0000-0003-0269-1936
✉ aduchow@g.clemson.edu, https://orcid.org/0000-0003-1681-7878
✉ c.j.hughes@salford.ac.uk, https://orcid.org/0000-0002-4468-6660

Acknowledgements: This study is part of the article-based PhD thesis of Marta Brescia-Zapata in the Department of Translation and Interpreting at Universitat Autònoma de Barcelona (UAB) within the PhD program in Translation and Intercultural Studies. This research has been partially funded by the H2020 projects TRACTION (under Grant Agreement 870610) and MEDIAVERSE (under Gran Agreement 957252). The Commission's support for this publication does not constitute an endorsement of the contents, which reflects the views of the authors only, and the Commission cannot be held responsible for any use which may be made of the information contained herein. Marta Brescia-Zapata and Pilar Orero are members of TransMedia Catalonia, an SGR research group funded by "Secretaria d'Universitats I Recerca del Departament d'Empresa. Coneixement de la Generalitat de Catalunya" (2021 SGR 00077).

Conclusions show that testing subtitles in immersive media environments is both a linguistic and an artistic endeavour, which requires an agile framework fostering contrast and comparison of different functionalities. Results of the present, preliminary study shed light on the possibilities for future experimental setups with eye-tracking.

Key words: subtitles, immersive environments, 360º videos, testing, eye-tracking.

1. Introduction

Immersive media technologies such as Virtual Reality (VR) and 360º videos are increasingly prevalent in society. Their potential has placed them in the spotlight of the scientific community for research and education. Industry has also adopted them not only in the entertainment sector, but also for communication, arts, and culture, which has attracted more and mixed audiences (Montagud et al., 2020). At present, these technologies are gaining popularity very fast due to the COVID-19 crisis as they enable interactive, hyper-personalised, and engaging experiences anytime and anywhere. Moreover, 360º videos, also known as immersive or VR360 videos, are a cheap and effective way to provide VR experiences. Specialised multi-camera equipment that can capture a 360º or 180º Field of View (FoV) instead of the limited viewpoint of a standard video recording, is used to produce content. VR360 videos can be enjoyed both via traditional devices (PC, laptops, smartphones) or VR devices (Head-Mounted Displays). They can also be consumed as a CAVE (Cave Automatic Virtual Environment), which uses high-resolution projection screens to deliver 360◦ visual experiences.

Immersive environments (in eXtended Reality, or XR) are generally used as an umbrella term referring to hardware, software, methods, and experience in Augmented Reality (AR) or VR or in general Mixed Reality (MR). The main goal of any immersive content is to make people believe that they are physically present (Slater & Wilbur, 1997). According to Rupp et al. (2016, p. 2108), VR360 videos can allow for “highly immersive experiences that activate a sense of presence that engages the user and allows them to focus on the video’s content by making the user feel as if he or she is physically a part of the environment”. Immersive videos, however, can also produce negative effects such as motion or simulator sickness, possibly turning people away from VR as a medium (Smith, 2015).

Like any other type of media content, 360º media experiences should be accessible. In almost all media assets, accessibility is added as an afterthought during the postproduction phase, despite many voices asking for accessibility in the creation process (Mével, 2020; Romero-Fresco, 2013). For this research we focus on subtitling, where standardised practices have emerged (Matamala & Orero, 2018), rather than covering different accessibility services. In 2D subtitles, the main aspects to consider are position, character identification, speed, number of lines, and number of characters (Bartoll, 2004; Díaz-Cintas & Remael, 2007; Gottlieb, 1995). Nevertheless, some Audiovisual Translation (AVT) studies have challenged traditional subtitling practices, encouraging more creative and integrated subtitles (Foerster, 2010; Fox, 2018; McClarty, 2012, 2014). The production of creative subtitles requires technology since such subtitles may change any of their paratextual features like the font or size or colour, but also where they are positioned, and more so in immersive environments where 2D features do not apply (Hughes et al., 2015; Lee et al., 2007). The integration of subtitles in XR is yet to be defined, and multiple challenges have emerged. Subtitles should be generated “in an immersive, engaging, emotive and aesthetically pleasing way” (Brown et al., 2017, p. 1), always considering accessibility and usability.

Beyond the challenge of subtitle text creation, XR requires direction to the sound source, as it may be outside the current audience viewpoint. Guiding and readability require the subtitler to preview and tweak formal aspects (Hughes & Montagud, 2020; Orero et al., 2020). This has led to the design of a new, web-based, prototyped framework that generates subtitles in 360º videos. The present article aims to identify how to display such subtitles for an optimal viewing experience. The framework allows for methods used in existing solutions (Brown & Patterson, 2017; Montagud et al., 2019; Rothe et al., 2018) to be easily contrasted and compared, as well as for the quick implementation of new ideas for user testing. After an overview on subtitle features in XR, the article describes the challenges of generating subtitle stimuli to test meaningful user viewing behaviours, based on eye-tracking technology. The approach for the first experimental set up for implementing creative subtitles in XR using eye-tracking is presented, in line with the stated research questions.

2. An Overview of Subtitles in Immersive Environments

Even though XR media was first introduced in the world of videogames, thanks to the development of 360º recording equipment these technologies are now expanding to videos (Hughes et al., 2020a). There are a few significant differences between content created within 2D and 3D environments. 2D means that the content is rendered in two dimensions (flat), while 3D content has depth and volume which allows a rich visual experience. According to Skult and Smed (2020, p. 451), "the key challenge for XR is that the FoV is limited, and the interactor cannot pay attention to the entire virtual scenery at once." The immersive experience, as in real life, moves from passive to active with the user becoming the centre of the story "creating a greater emotional nexus" (Cantero de Julián et al., 2020, p. 418). In a play or opera, the action takes place on the proscenium. However, another activity somewhere in the theatre may distract from that narrative, such as the noise of a lady unwrapping sweets two rows away. The audience in VR has freedom of movement. They can also determine the time spent in any area of interest or field of vision and decide on where to focus their attention. This freedom affects subtitle reading since the development of the narrative may be random, decided by the viewer. Similarly, in VR, the aim is for immersiveness and the concepts of presence and engagement are central, with the ultimate goal of being a witness to the narrative from a first-person viewpoint. This breaks with the concept of a passive audience that reads subtitles, following what Jenkins et al. (2015) define as *spreadable* reading, meaning that the audience spreads its attention across the image as the linear narrative is displayed. In VR images and sound surround, there is no linear narrative and passive viewing moves towards interaction or *drillable* viewing as in video games or transmedia products. In this context, Mittell (2009) explains that "spreadable media encourages horizontal ripples, accumulating eyeballs without necessarily encouraging more long-term engagement. Drillable media typically engage far fewer people but occupy more of their time and energies in a vertical descent into a text's complexities." These features are theoretical principles that have yet to be tested.

The value of VR lies in its potential to tamper with both time and space; hence the experience relies on the viewer. This has a direct effect on the way subtitles are consumed. A person may be watching one part of the scene while there is a person speaking away from the viewing field. A hearing person may be able to locate the sound source but someone with hearing loss will need to be guided.

Another feature that is different from 2D resides in the way media is accessed. There are two options: using CAVE (the naked eye) or a device such as a Head-Mounted Display (HMD). This is a type of display device or monitor that is worn over the head and allows the user to be immersed in whatever experience the display is meant for. The 360º environment accessed when wearing the HMD may be an animation (such as a video game or an animated movie) or live action (such as a movie or a documentary). Depending on the type of media, the content will be installed on a PC or on the HMD itself or stored on the cloud. As Internet speed improves, media content streamed from the web is becoming increasingly popular.

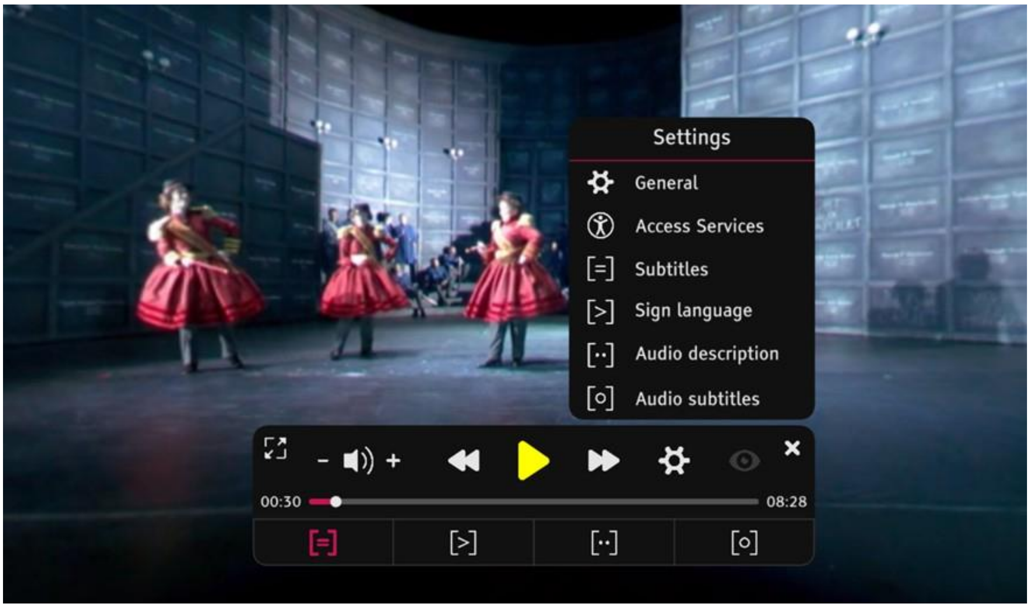
As in traditional 2D media, to create and consume subtitles in XR, a subtitle editor and a subtitle player are needed. Although immersive video players offering the ability to play VR360 video are commercially available, not many of them support accessible services (Brescia-Zapata, 2022). The player needs to be accessible, and the user has to activate the display accessibility. The interface or menu also needs to display the choice of accessibility services available, and finally, the interaction with the terminal or device also needs to be accessible. All these features show the complex ecosystem required for a true XR accessible experience. This, linked to the lack of standardised solutions and guidelines, has led to the development of non-unified solutions, meeting only specific requirements (Hughes & Montagud, 2020). The majority of players seem to have inherited features from the traditional 2D world, instead of addressing the specific features of 360º environments.

This situation served as an inspiration for initiatives like the European H2020 funded Immersive Accessibility (ImAc) project[1] that explored how accessibility services and assistive technologies can be efficiently integrated with immersive media, focusing on VR360 video and spatial audio. Under the umbrella of this project, both an accessible player and a subtitle editor were developed. The accessibility-enabled 360º ImAc player supports audio description, audio subtitles, and sign language along with other features (Montagud et al., 2019) as can be seen in Figure 1.

[1] http://www.imac-project.eu

Figure 1

ImAc Player Settings

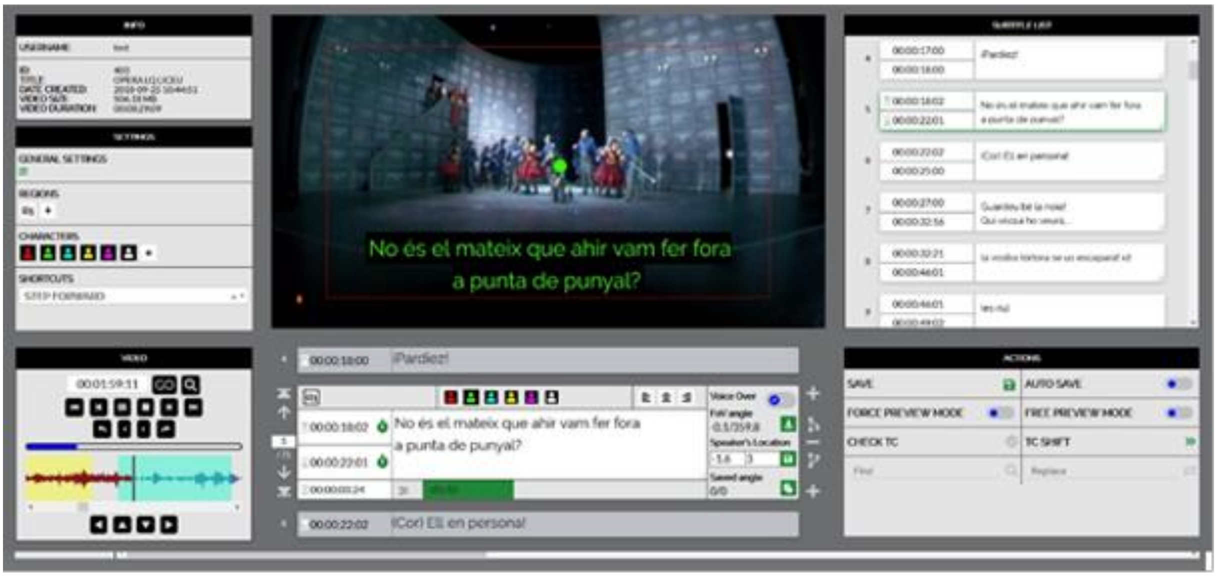


Source: ImAc player screenshot.

In contrast, the ImAc subtitle editor is a commercial web-based editor, and its interface is similar to that of any traditional subtitle editor, as can be seen in Figure 2. The main innovations are related to the FoV in VR360 video, i.e., the extent of observable environment the user is able to see at any given moment. It includes navigation buttons for FoV in spherical space to move up, down, left, and right. There is also a button which moves the FoV to the angle where the speaker of the current subtitle is located. The editor also allows the FoV angle to be changed using the navigation buttons in the video control area or moving the mouse with the left button over the video. By default, the video initially has the current angle as longitude: 0.00º and latitude: 0.00º. In addition, the voiceover option can be marked when there is no speaker in the 360º scene.

Figure 2

Immersive Subtitle Editor Developed in ImAc



Source: ImAc subtitle editor screenshot.

The basic tools to create and consume accessible VR content are now commercially available, e.g., a VR subtitle editor and a VR subtitle player. What is evident is that unless different display modes can be produced, they cannot be tested, and this is one of the shortcomings of the ImAc project which was concluded recently and focused on traditional subtitles projected on immersive environments (Hughes et al., 2020b).

2.1. Related Work

Excluding works that have added (sub)titles at post-editing stages, only three recent studies have focused on investigating subtitles in immersive environments. All the studies followed a user-centric methodology and chose people with hearing loss for testing. Reading skill was not considered within the demographic data.

The British Broadcasting Corporation (BBC) was one of the first research organisations to design subtitles in XR (Brown & Patterson, 2017). The BBC research team first identified the main challenges when developing subtitles for immersive content and, based on these, the following four solutions for subtitle rendering were developed by Brown et al. (2017):

- Evenly spaced: subtitles are placed into the scene in three fixed positions, equally spaced by 120º around the video and slightly below the eye line;

- Follow head immediately: the subtitles are presented as a “head-up display” always in front of you, and slightly below straight ahead. As you turn your head, the subtitle moves with you, always at the same location in the headset display;
- Follow head with lag: the subtitles follow head direction, but only for larger head movements: if you look slightly left or right the subtitle stays in place, but a head movement of a greater amplitude will cause the subtitle to catch up with your head orientation;
- Appear in front, then fixed: each subtitle is placed in the scene in the direction you are looking at the time when it appears and remains fixed in that location in the scene until it disappears.

These four rendering modes were tested with several clips (Brown, 2017), and users reported that while it was easy to locate the evenly spaced captions, they preferred the head-locked options (see Table 1). Head-locked subtitles resemble most traditional ecstatic 2D subtitles, always visible at the bottom of the screen. These results come as no surprise since for years now subtitle testing in Europe has shown that people like what they are used to, even if the data demonstrate that the solution is not ideal, as proved with eye-tracking tests (Mas Manchón & Orero, 2018; Romero-Fresco, 2015).

Table 1

Numbers of People (and Percentages) Who Selected Each Behaviour as Their Favourite or Least Favourite Behaviour. Least Favourite Was Not Specifically Requested, so Was Not Available for All Participants

Behaviour	Favourite	Least favourite
Evenly spaced	1 (4%)	5 (38%)
Follow head immediately	10.5 (44%)	3 (23%)
Follow with lag	7 (29%)	2 (15%)
Appear in front, then fixed	5.5 (23%)	3 (23%)

Source: Authors' own elaboration based on Brown (2017).

The second study (Rothe et al., 2018) compared the two presentation modes: fixed and head-locked subtitles. Although no conclusive results were found, in terms of comfort (i.e., presence, VR sickness, and task load) fixed subtitles led to slightly better results even though fixed captions in general mean that users may not always be able to see the caption as it may be outside their FoV.

The third study, performed under the umbrella of the H2020-funded ImAc project (Hughes et al., 2019), revealed the need to guide users to the sound source of the subtitle (i.e., a sound effect or a character speaking or not speaking). To facilitate this requirement, location within the 3D space information was added to each subtitle (Agulló & Matamala, 2019). This allowed for different modes to be developed which could guide the user to where the speaker was located (Agulló et al., 2019). However, this had the drawback that the location was only specified once per caption, and if a person was moving dynamically, this could affect the exactness of the guiding feature (Hughes et al., 2019).

Nevertheless, the ImAc project designed and developed several guiding mechanisms, and test results showed two preferred methods:

- ImAc Arrow: an arrow positioned left or right directs the user to the target;
- ImAc Radar: a radar circle is shown in the user's view. This identifies both the position of the caption and the relative viewing angle of the user.

In the area of standardisation, a W3C Community Group[2] is focusing on developing new standards for immersive subtitles. They have recently conducted a community survey to gather opinions, but no tests were performed. A small group of users with different hearing levels (Deaf, Hard of Hearing, and Hearing) were asked to evaluate each of the identified approaches for subtitles within immersive environments. Head-locked was clearly identified as the preferred choice. However, it was noted that this was a likely outcome since it replicated the experience that users were familiar with, as indicated above. It was also acknowledged that it was difficult for users to evaluate new methods theoretically without the opportunity to experience them while accessing content. Although all agreed that the head-locked option should be set as default, the respondents maintained that other choices should be made available. Other suggestions included changing the font size and colour and the number of lines (two lines being the default). Consequently, the need to develop a framework enabling delivery of the full experience of each captioning mode, in an environment where an extensive user study could be conducted, would be a priority prior to testing.

3. Methodology for a Pilot Study

Conducting a pilot study before launching a full spectrum study is always desirable. The goal of such a pilot study is not only to try to ensure that the survey questions operate well, but also that the research procedures and measures are adequate and reliable (Bryman, 2004). Especially when research aims to substantiate the validity of a new framework and/or involve the use of novel technology (such as eye-tracking in VR), a pilot study is crucial to ensure that both methodology and the study design are accurate and reliable. The preparation stage for this pilot study involved four main steps: user profile definition, selection of the testing material, implementing the material within the new framework, and design of the test procedure itself.

The procedure followed by the current study consisted of four stages: an introduction, a questionnaire on demographic information, an eye-tracking test using 360º immersive videos, and a focus group. The main aims were (1) to test a new framework for subtitle presentation in 360º videos, (2) to obtain feedback regarding expectations, recommendations, and preferences from users when consuming subtitles, and (3) to explore the visual attention distributions between subtitles and movie scenes while watching videos in VR. To do so, three different subtitle modes were implemented:

---

[2] https://www.w3.org/community/immersive-captions/

mode 1 (following ImAc results), mode 2 (following Fox's (2018) studies), and mode 3 (fully customised).

Before starting the pilot study and taking the previous work in the field as a reference, the following hypothesis was formulated: Fixed, near to the mouth subtitles will allow viewers to spend more time exploring the image instead of reading the subtitles.

3.1. The Live Web Testing Framework

One of the challenges for testing immersive subtitles is the difficulty of having users evaluate new modalities properly because of the cost and time needed to create new prototype subtitle presentations so that users can experience them. To this end, an XR subtitle web simulator was developed by Hughes et al. (2020b). This web-based simulator was designed for rapid prototyping of subtitles within a 360º space, as can be seen in Figure 3 below.

Figure 3

Open-Source Demo Player Developed as Part of This Study

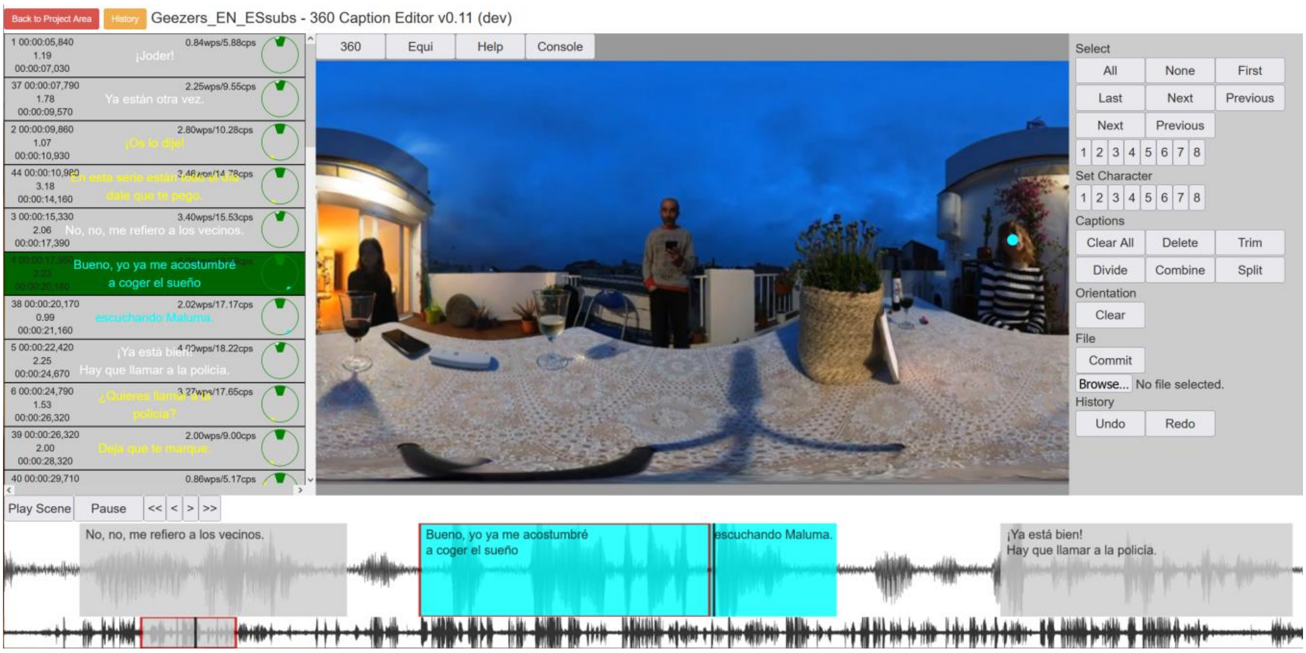
Source: Chris Hughes' demo player screenshot.

This new framework allows for instant immersive subtitle production in up to nine different modes: four of them are fixed, where the subtitle is rendered relative to a fixed location in the world, generally at the position of the character speaking, and five are head-locked, where the subtitle is rendered relative to the user's viewpoint. The main idea behind this demo player is to allow as much personalisation as possible (i.e., subtitle display, placement, timing, render mode, guiding mechanism, etc.); this way, any feature may be activated to define and test subtitles within 360º videos.

Along with this XR subtitle simulator, a web-based editor was also developed (see Figure 4), which allows previously created subtitles to be imported in .srt format or subtitles to be created from scratch. On the one hand, each subtitle can be associated with a character ("Set Character" button), and, on the other hand, each subtitle must have an associated position (FoV), i.e., the place in the 360º scene where it should appear.

Figure 4

Open-Source Editor Developed as Part of This Study



Source: Chris Hughes' subtitle editor screenshot.

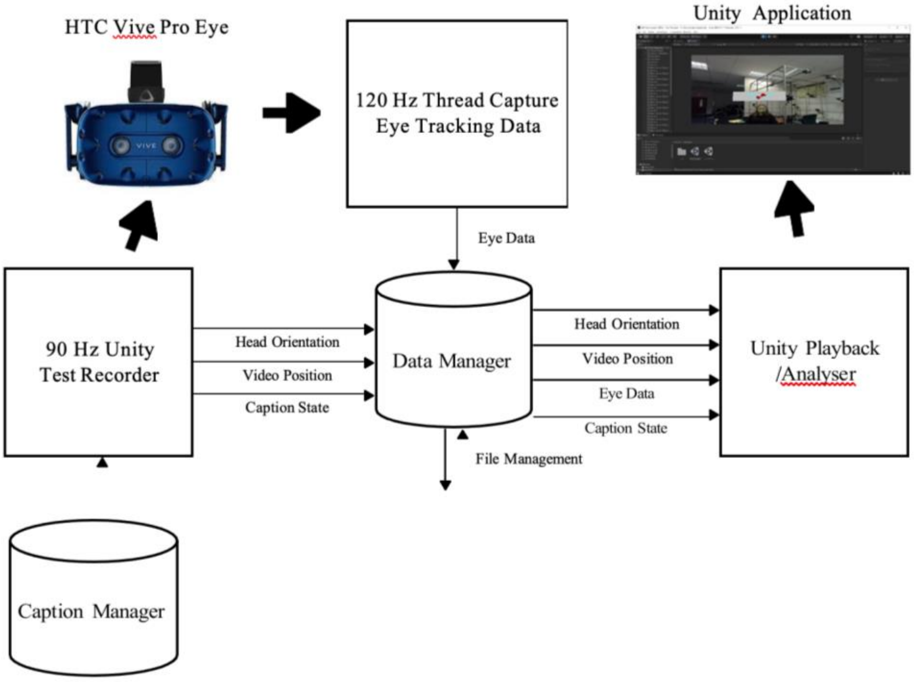
Both the demo player and the editor are open-source and can be accessed from a main project area where all the imported 360º videos are located. These tools take their inspiration from the player and the editor developed in the ImAc project. The main difference between them is that ImAc tools are intended to be used by generic audiences (final users), while the tools used in this study are more focused on research and testing.

3.2. System Architecture

To enable recording of gaze within 360º video, the live web testing framework developed by Hughes et al. (2020a, 2020b) was ported to Unity 3D to allow the display of 360º video content and to capture data from the eye tracker built into the VR device. A new system architecture emerged, as depicted by the schematic in Figure 5.

Figure 5

Eye-Tracking VR System Architecture



Source: Authors' own elaboration.

The system architecture was developed to utilise the HTC Vive Pro Eye, which contains an eye tracker from Tobii built into the display. The application uses two Unity assets, one specifically optimised for recording and the other for playback. At the centre of the architecture is a Data Manager, designed to store all test data. It also handles file management and can generate the output data in a variety of formats as required.

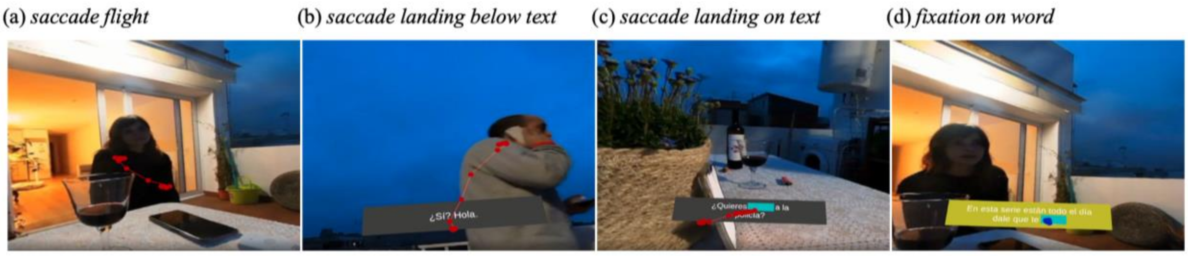
The recording application allows for a specified 360º video to be played with the captions fixed in the scene. During the test, each event and data are logged into the data manager as they become available and timestamped. In order to be able to replay a user viewing session, the system needs to record head orientation, video (frame) position, gaze data (raw and analysed, see below) as well as the subtitle caption state, i.e., which caption from the accompanying subrip format (.srt) file was being displayed and where.

The playback application allows for the data to be retrieved from the Data Manager and the entire test to be replayed. This offers the opportunity to change the analysis process or to include additional Areas of Interest (AOIs) and makes it possible to repeat the analysis. It also permits visual analysis by overlaying the eye data onto the video following capture.

One technical difficulty that had to be overcome was synchronisation of gaze data with video and subtitle data. Gaze data is sampled at 120 Hz while the Unity display refresh rate is 90 Hz. Thus, on average 1.3 gaze samples are expected on any given frame. To enable synchronisation from data streams of different rates, a separate eye-tracking data thread was created to collect gaze data captured at 120 Hz, ensuring no loss of eye movement samples. System playback can be set to either the speed of video or eye tracker, with gaze data drawn atop the projected video, as shown in Figure 6.

Figure 6

Gaze Recording in VR Showing Varying Elements of Gaze to Subtitle: (a) Saccade in Mid-Flight, (b) Saccade Landing Site With Slight Undershoot, (c) Saccade to Midpoint of Subtitle, (d) Fixation Within a Subtitle



Source: Authors' own elaboration.

3.3. Participants

The size of the group was determined in accordance with the pilot nature of the study (Bryman, 2004, p. 507). In the beginning, 7 participants were expected, but due to complications related to the Covid-19 pandemic, only five appeared (2 male and 3 female). All participants were professionals from the Arts, Sciences, or Humanities fields, staying for a few weeks at the residence Faberllull in Olot. The average age was 40 (*SD* = 8.37) and all of them had completed a postgraduate university degree. All were active professionally (1 social worker, 1 cultural manager, 1 music therapist, 1 pre-doctoral researcher, 1 project manager). All participants spoke Spanish and at least one other language.

All participants were familiar with using computers and mobile devices. Two participants reported having previous experience with VR. Most of the participants declared watching different TV content with subtitles at least occasionally (only one of them claimed that she/he never used subtitles).

3.4. Study Materials

One of the main concerns of the study was to find appropriate material for testing. Due to the difficulty in finding royalty-free material that met the needs of the study, a homemade 360º video was recorded using an Insta360 One X2 camera. The duration was 3 minutes and 45 seconds. The camera was settled in the centre of the action and three characters were positioned around the camera so that the action took place throughout the 360º space. The characters followed a script to avoid overlaps because if two characters located at different points in the 360º scene spoke at once, it would be almost impossible for the user to read the subtitles.

There were three types of subtitles yielding three experimental conditions:

- Mode 1: following ImAc results. Same font and colour (b&w) for all the characters, with a grey background and head-locked.
- Mode 2: following Fox's (2018) studies. Same font and colour (b&w) for all the characters, without background and near the mouth.
- Mode 3: fully customised. Different font and colour for each character, with a grey background and near the mouth.

Other conditions, for example subtitles for non-human sound sources, will be presented in future tests.

3.5. Procedure

The study included the following stages. First, participants were welcomed by the facilitator, who briefly explained the aim of the project. The session took place in a meeting room divided into two separate spaces. On one side there was a large TV screen, a computer connected to the screen and chairs for the participants. On the other side, an improvised eye-tracker lab was installed with a computer and a pair of HTC Vive Pro HMD. One researcher took notes and summarised the conclusions in real-time. Second, the aim of the focus group was explained to the participants, and they were asked to sign informed consent forms. The third step consisted of filling in a short questionnaire on demographic information. Finally, the session began. To trigger the discussion, the facilitator gave a short introduction to VR and 360º content and explained how subtitles are integrated within 360º content, showing VR glasses to the participants.

The eye-tracking technology was introduced, as it is integrated within the VR glasses and was one of the data collecting methods in the study. The facilitator explained that 360º content can also be accessed on a flat TV screen using a mouse to move around the 360º scene. Different types of subtitles were presented to give users some idea about how creative subtitling can be implemented in immersive content and to stimulate their imagination.

Then, each participant used the HTC Vive Pro HMD to watch a short video with audio in English and subtitled into Spanish. In total there were three rounds, always using the same video but a different subtitle mode each time. The order of the participants was determined randomly. Immediately after each visualisation, participants filled out a short questionnaire with questions on content understanding, subtitling preferences, and the task load index (NASA-TLX).

After the last round the focus group took place. Together with the stimuli, the facilitator used a list of guiding questions grouped under major topics to generate the participants' reactions. A balance between an open-ended and a structured approach was sought, and the result was a lively discussion in which interesting ideas came up.

4. Pilot Study Results

The data analysis of the study was mainly qualitative accompanied by descriptive statistics of the post-study questionnaire and eye movements captured during the study (see Figure 6).

4.1. Movie Content Understanding

To check the understanding of the stimuli movies, we averaged the accuracy of responses to questions about the content separately for each condition. The highest average accuracy was obtained for the movie with fully customised subtitles (M = 0.64, SD = 0.26). Average accuracy for movies with subtitles in mode 1 (M = 0.52, SD = 0.18) and mode 2 (M = 0.52, SD = 0.36) were the same.

Additionally, when asked about the description of the scenes presented in the movie, participants used, on average, slightly more words after watching the movie in mode 1 (M = 22.2, SD = 12.99) than mode 3 (M = 18, SD = 8.34). The smallest number of words used in the description after watching the movie was in mode 2 (M = 16.20, SD = 9.01).

Qualitative analysis of responses during the focus group interviews showed that some of the participants could not understand the plot until the third visualisation of the clip. This could be related to a learning effect, but also because 3 of the participants had no previous experience with subtitled immersive content. Furthermore, another participant commented that sometimes it was difficult to follow the story because she was distracted exploring the 360º scene. The participant who was the least familiar with new technologies (and the least interested in the immersive format) noted that paying attention to the story stressed her and that she tried to distract herself during the visualisations.

4.2. Subtitle Readability

The participants were asked whether they had been able to read the subtitles after watching each movie. Two responded “yes”, two “no”, and one was “not sure” for mode 1. In mode 2, two responded “yes” and three “no”. The least readable subtitles seemed to be in mode 3. Three participants noted they were not able to read them; only one responded “yes” and one participant was “not sure”. When asked to estimate the percentage of subtitles that they were able to read, the differences were very small: 70% in mode 1, 68% in mode 2, and 67% in mode 3. Both results seem to suggest a slight preference for the subtitles in mode 1 as the most readable ones.

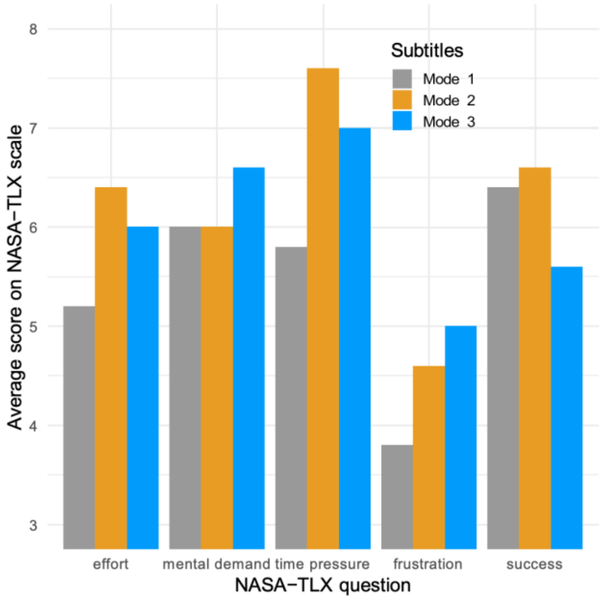
These results comply with the qualitative data extracted from the focus group, since most participants agreed that mode 3 was difficult to read. Only one of the participants noted that she liked the coloured text, and there was a brief discussion about the possibility of customising the subtitles further. Regarding the grey background, there was no consensus: some of the participants found subtitles with no background hard to read, others found them less intrusive. One participant highlighted the reading pace in general, arguing that some captions disappeared "too soon" forcing the user to read faster.

4.3. Self-Reported Task Load

To collect self-reports on the effort elicited by the task of watching stimuli movies in different subtitle modes, the NASA-TLX scale with five questions was analysed (see Figure 7). Subjective evaluation of effort while watching videos in different subtitle modes also suggests a preference for mode 1 (M = 5.20, SD = 2.39). In the participants' opinion, more effort was required to read subtitles in modes 2 (M = 5.80, SD = 2.17) and 3 (M = 6.00, SD = 1.87). However, evaluation of mental load shows a different pattern, namely that modes 1 (M = 6.00, SD = 2.35) and 2 (M = 6.00, SD = 2.35) were equally less demanding than mode 3 (M = 6.60, SD = 2.30). Participants also evaluated reading modes 1 (M = 6.40, SD = 1.52) and 2 (M = 6.60, SD = 2.30) with greater perceived success than mode 3 (M = 5.6, SD = 2.19). These results are not surprising considering the average responses regarding how time-pressured participants felt. Subtitles in mode 2 caused the highest experience of time pressure (M = 7.60, SD = 0.89); this was lower in mode 3 (M = 7.00, SD = 0.71) and lowest in mode 1 (M = 5.80, SD = 2.17). The perceived level of frustration/stress was lowest when watching the video in mode 1 (M = 3.80, SD = 2.17), greater in mode 2 (M = 4.60, SD = 3.21), and greatest in mode 3 (M = 5.00, SD = 2.35).

Figure 7

Task Load Self-Reports With NASA-TLX Scale While Watching Videos in Different Subtitle Modes



Source: Authors' own elaboration.

4.4. Attention Distribution and Cognitive Effort While Reading Captions and Scene Viewing

Gaze was captured as it traversed subtitles when reading the text displayed within the quadrilaterals that contained them. The analysis of the eye movement signal relies on fixation detection, which in turn depends on saccade detection. Fixations are detected within the raw eye movement signal following Nyström and Holmqvist (2010) and by using the Savitzky-Golay filter for velocity-based (I-VT (Salvucci & Goldberg, 2000)) event detection Savitzky and Golay (1964).

The current system architecture allows for detection of fixations falling within arbitrarily defined Areas of Interest (AOIs), including polygons defined over actors and more importantly over quadrilaterals (quads) used to display subtitles as well as quads defined over individual words, see Figure 8 below.

Figure 8

Gaze Recording Showing Fixations Over Areas of Interest: (a) Actor Body, (b) Subtitle Quad, and (c) Individual Word

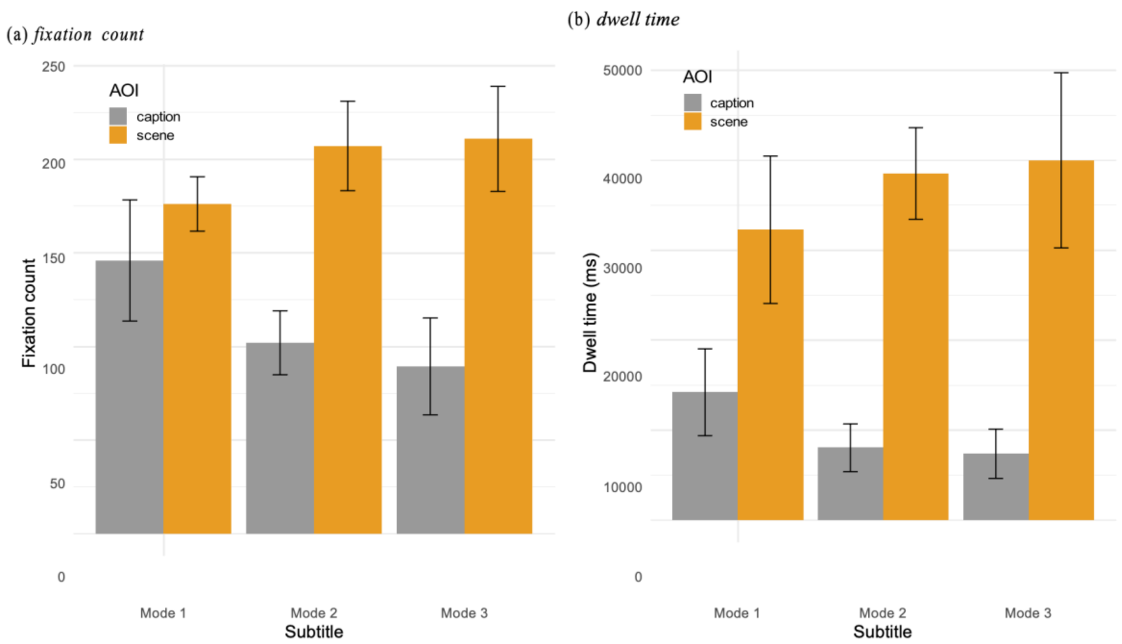
Source: Authors' own elaboration.

The eye movement analysis aimed first at capturing differences in attention to captions and visual scenes in terms of fixation count and dwell time as dependent variables. Descriptive statistics show that in all conditions most fixations were on the visual scene rather than on subtitles. However, the difference is less important for the video in mode 1.

Participants exhibited more fixations on captions (M = 145.8, SD = 32.37) than on the visual scene (M = 176.0, SD = 14.54) in mode 1 compared to modes 2 (for caption M = 101.80, SD = 17.08; for scene M = 207.00, SD = 23.92) and 3 (for caption M = 89.20, SD = 25.90; for scene M = 210.80, SD = 28.05) see Figure 9(a). A similar pattern is observed when analysing dwell time. On average, participants dwelled more on captions than on the visual scene in mode 1 than in modes 2 or 3, see Figure 9(b). Participants appeared to allocate more attention to captions when viewing subtitles in mode 1 than in modes 2 or 3.

Figure 9.

Visual Attention Distribution Over Captions and Visual Scenes While Watching Video With Different Types of Subtitles



Note: Attention distribution is depicted by two metrics: (a) shows fixation counts over captions and visual scene, (b) shows dwell time of captions and visual scene fixating.

Bars height represents mean values and whiskers represent ±1 SD.

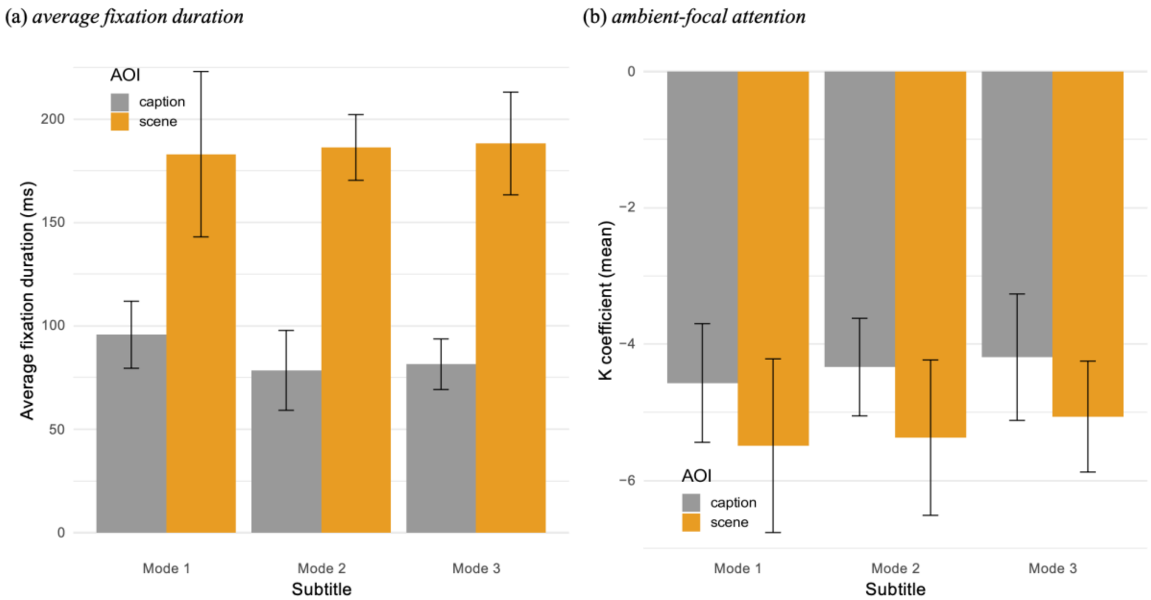
Source: Authors' own elaboration.

We also examined cognitive effort while processing information from captions or visual scene based on average fixation duration (following the *eye-mind assumption* (Just & Carpenter, 1976)) and focus of attention with coefficient *K* (Krejtz et al., 2016), which captures the temporal relation between fixation duration and subsequent saccade amplitude. $K > 0$ indicates focal viewing while $K < 0$ suggests ambient viewing. Focal attention is usually related to higher cognitive effort when processing complex visual or text stimuli (Duchowski et al., 2020; Krejtz et al., 2017; Krejtz et al., 2018). Analysis of descriptive statistics on average fixation duration showed that the visual scene triggered longer average fixation durations than captions in all modes. However, the difference in average fixation durations between the visual scene and the caption is smallest in mode 1.

Moreover, fixation duration on subtitles in mode 1 (M = 95.73, SD = 16.22) is much longer than on subtitles in either mode 2 (M = 78.47, SD = 19.35) or mode 3, (M = 81.43, SD = 12.24), see Figure 10(a). Coefficient *K* showed that viewers were not as focused when reading captions in mode 1 (M = −4.57, SD = 0.87) compared to mode 2 (M = −4.34, SD = 0.72) or 3 (M = 4.19, SD = 0.92), see Figure 10(b). Both fixation duration and coefficient *K* suggest the highest cognitive effort along with decreased focal processing when processing subtitles in mode 1.

Figure 10

Cognitive Processing of Textual and Visual Information From Captions and Visual Scenes While Watching Video With Different Types of Subtitles



Note: (a) shows average fixation duration as a metric for cognitive effort, (b) shows K coefficient as a metric ambient-focal attention.

Bar heights represent mean values and whiskers represent ±1 SD.

Source: Authors' own elaboration.

4.5. Focus Group Insights

A qualitative analysis was carried out on the notes taken during the focus group (the last part of the session, after the participants had watched the video with the different subtitle options). The notes were thoroughly revised and tagged using Atlas.ti. This procedure allowed us to identify three areas that can be associated with the quantitative analysis (subtitle readability, task load, and understanding of the movie content). The analysis also allowed us to define user preferences and identify aspects on which there was consensus among users and issues on which opinions diverged.

In terms of preference, most of the participants agreed that mode 2 was the easiest to read with one participant suggesting adding a colour code to mode 2 (like mode 3). The second preferred option was mode 1 (selected by 2 participants). The main problem in mode 1 seems to have been the difficulty in identifying the character speaking at each given moment.

Regarding creative subtitles, all the participants agreed it would be a great idea to dramatise what is said and that this could add much visual beauty to the content. One of the participants noted that, in some cases, so much creativity distracted her from the content itself.

As in previous studies reviewed earlier, participants highlighted the lack of direction to guide people to the source of the sound (guiding mechanisms). Some of them mentioned that they missed human interaction when watching the immersive content, and that they felt isolated when wearing the HMD for the first time.

5. Discussion

The first hypothesis we wanted to validate was whether fixed, near to the mouth subtitles allow viewers to spend more time exploring the image instead of reading the subtitles when compared to head-locked subtitles. Although the present pilot study cannot yield conclusive evidence, eye movement data together with focus group insights seem to support this hypothesis. Interestingly, eye movement data appear to be consistent with qualitative insights from the focus group, suggesting that participants tend to prefer fixed subtitles located near the mouth of the speaking character (mode 2). These results differ from those obtained in previous studies, in which participants opted for head-locked subtitles.

The results of self-reported cognitive load during movie watching with different subtitle modes suggest a slight preference (less perceived mental effort and higher perceived success in reading captions) for mode 1 (b&w font for all characters, grey background, and head-locked) over modes 2 and 3. However, results carry a large statistical variance and cannot be interpreted decisively. The results may also be biased by a lack of randomisation in the order of presentation and learning effect of the questions during the experimental procedure. Future studies must employ tighter experimental control over stimulus presentation order (e.g., via randomisation or counterbalancing).

Eye movement analysis sheds light on attention allocation (captions vs. scene) and perception. Identification of fixations showed that participants allocated more attention to captions and less to the visual scene when viewing subtitles in mode 1 than in modes 2 or 3. Process measures (average fixation and ambient-focal coefficient) suggest higher cognitive effort paired with less focal processing of subtitles in mode 1.

Subtitles in modes 2 and 3 appear to outperform mode 1 as they may be less distracting from scenes in the movie, but they also seem to require less cognitive effort when focused on reading. We do not know, however, whether mode 2 or 3 is easier to read and less distracting when movie watching. This issue needs to be addressed in a study with more experimental control and a larger sample.

The visualisation of the eye movements analysed, specifically saccades, drawn in red in Figure 11, expose the inadequacy of the velocity-based filtering approach. The I-VT method, while computationally efficient and generally applicable to traditional desktop displays, tends to ignore head-induced gaze movement when captured in the VR HMD. It is likely that a better model of eye and head coupling is required (Guitton et al., 1990), e.g., a fixation detection algorithm suitable for immersive environments (Llanes-Jurado et al., 2020).

Figure 11

The View and Scanpath for Each Participant

Note: rows: participant 1–5, columns: mode 1-3, 00:22. General observations can be drawn, such as that participant 1, although finding the captions in modes 1 and 3, was lost in mode 2 and can be observed saccading between the mouths of the wrong characters, trying to identify the character speaking. The second participant can be seen fixating on the speaking character’s mouth rather than reading the caption in modes 1 and 2. Also Participants 3, 4 and 5 can be observed reading the captions in modes 1 and 2, but not mode 3.

Source: Authors' own elaboration.

6. Conclusions

Immersive environments simulate reality to heighten the immersive experience. Attention should be paid when drawing on results from previous studies on subtitle reading performance in 2D to VR environments. Prior research on evaluation of subtitles has mainly focused on subtitle style, speed of display, and positioning, and has largely been qualitative. In VR we still need to define the challenges faced by subtitle research. Investigating how subtitles are read is one logical quest but finding subtitles when two people are interacting from different fields of vision is also a candidate for testing. The tests carried out in this study have shown that we need some basic understanding of media presentation in VR and user behaviour and habits when consuming media in VR, where the narrative is no longer linear. VR media environments present us with new variables to consider when testing for optimal subtitle presentation. The objective is to find the least disruptive subtitle reading experience for protecting immersivity and the simulation of reality. To understand the visual presentation of subtitles in VR, a framework was developed along three basic presentation modes for 360º videos, which were all piloted. Our contribution is thus two-fold: on the one hand, there is our presentation of the VR subtitle framework and, on the other hand, a new method for triangulation of psycho-physiological (eye movements) self-reports and qualitative (focus group discussions) analyses. To the best of our knowledge, this is the first attempt to advance these two directions when discussing subtitle presentation in VR 360º videos.

Immersive environments need new subtitle presentation modes and reading patterns. This article described the first pilot study using a comprehensive methodological environment to test subtitles in immersive environments. The novel testbed includes a subtitle editor and a VR system designed specifically to collect eye movement data as visual attention is distributed over 360º videos containing subtitles. Both the framework and the methodology tested in this first pilot study can be used to collect quantitative and qualitative behavioural data when viewing subtitled 360º media. Future studies involving more participants are expected to yield new insights and lead to subtitle standardisation in immersive media environments.

References

Agulló, B., & Matamala, A. (2019). Subtitling for the deaf and hard-of-hearing in immersive environments: Results from a focus group. *Journal of Specialised Translation, 32.*

Agulló, B., Montagud, M., & Fraile, I. (2019). Making interaction with virtual reality accessible: Rendering and guiding methods for subtitles. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 33*(4), 416–428. https://doi.org/10.1017/ S0890060419000362

Bartoll, E. (2004). Parameters for the classification of subtitles. In P. Orero (Ed.), *Topics in Audiovisual Translation* (pp. 53–60). John Benjamins. https://doi. org/10.1075/btl.56.08bar

Brescia-Zapata, M. (2022). The present and future of accessibility services in vr360 players. *inTRAlinea, 24.*

Brown, A. (2017, October 26). *User testing subtitles for 360° content*. BBC. https://www.bbc.co.uk/rd/blog/2017-10-subtitles- 360-video-virtual-reality-vr

Brown, A., & Patterson, J. (2017, October 26). *Designing subtitles for 360º content*. BBC. https://www.bbc.co.uk/rd/blog/2017- 03-subtitles-360-video-virtual-reality

Brown, A., Turner, J., Patterson, J., Schmitz, A., Armstrong, M., & Glancy, M. (2017). Subtitles in 360-degree video. *Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video.* The Netherlands, 3–8. https://doi.org/10.1145/3084289.3089915

Bryman, A. (2004). *Social research methods*. Oxford University Press.

Cantero de Julián, J. I., Calvo Rubio, L. M., & Benedicto Solsona, M. Á. (2020). La tenue apuesta por los vídeos en 360º en las estrategias transmedia de las televisiones autonómicas españolas. [The weak bet on videos in 360º in the transmedia strategies of Spanish autonomic televisions]. *Revista Latina de Comunicación Social*, (75), 415–433. https://doi.org/https://doi.org/10.4185/RLCS-2020-1433

Díaz-Cintas, J., & Remael, A. (2007). *Audiovisual translation: Subtitling*. St. Jerome.

Duchowski, A. T., Krejtz, K., Zurawska, J., & House, D. H. (2020). Using microsaccades to estimate task difficulty during visual search of layered surfaces. *IEEE Transactions on Visualization and Computer Graphics, 26*(9), 2904–2918. https://doi .org/10.1109/TVCG.2019.2901881

Foerster, A. (2010). Towards a creative approach in subtitling: A case study. In J. Díaz-Cintas, A. Matamala, & J. Neves (Eds.), *New insights into audiovisual translation and media accessibility: Media for all 2* (pp. 81–98). Rodopi.

Fox, W. (2018). *Can integrated titles improve the viewing experience? Investigating the impact of subtitling on the reception and enjoyment of film using eye-tracking and questionnaire data.* Language Science Press. https://doi.org/10.5281/zenodo.1180721

Gottlieb, H. (1995). Establishing a framework for a typology of subtitle reading strategies – viewer reactions to deviations from subtitling standards. *Communication Audiovisuelle et Transferts Linguistiques – FIT Newsletter, 14*(3–4), 388–409.

Guitton, D., Munoz, D. P., & Galiana, H. L. (1990). Gaze control in the cat: Studies and 256odelling of the coupling between orienting eye and head movements in different behavioral tasks. *Journal of Neurophysiology, 64*(2), 509–531. https://doi.org/10.1152/jn.1990.64.2.509

Hughes, C., Armstrong, M., Jones, R., & Crabb, M. (2015). Responsive design for personalised subtitles. *Proceedings of the 12th International Web for All Conference*, Italy, 1–4. https://doi.org/10.1145/2745555.2746650

Hughes, C., Brescia-Zapata, M., Johnston, M., & Orero, P. (2020a). Immersive captioning: Developing a framework for evaluating user needs. *IEEE AIVR 2020: 3rd International Conference on Artificial Intelligence & Virtual Reality 2020*. http://usir.salford.ac.uk/id/eprint/58518/

Hughes, C., & Montagud, M. (2020b). Accessibility in 360° video players. *Multimedia Tools and Applications*, 1–28. https://doi.org/10.1007/s11042-020-10088-0

Hughes, C., Montagud Climent, M., & tho Pesch, P. (2019). Disruptive approaches for subtitling in immersive environments. *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video. United Kingdom*, 216–229. https://doi.org/10.1145/3317697.3325123

Jenkins, H., Ford, S., & Green, J. (2015). *Cultura transmedia. La creación de contenido y valor en una cultura en red [Transmedia culture. Content creation and value in a network culture].* Gedisa.

Just, M. A., & Carpenter, P. A. (1976). The role of eye-fixation research in cognitive psychology. *Behavior Research Methods & Instrumentation, 8*(2), 139–143. https://doi.org/10.3758/BF03201761

Krejtz, K., Çöltekin, A., Duchowski, A., & Niedzielska, A. (2017). Using coefficient K to distinguish ambient/focal visual attention during map viewing. *Journal of Eye Movement Research, 10*(2). https://doi.org/10.16910/jemr.10.2.3

Krejtz, K., Duchowski, A., Krejtz, I., Szarkowska, A., & Kopacz, A. (2016). Discerning ambient/focal aAt- tention with cCoefficient K. *ACM Transactions on Applied Perception, 13*(3), 11:1–11:20. https://doi. org/10.1145/2896452

Krejtz, K., Wisiecka, K., Krejtz, I., Holas, P., Olszanowski, M., & Duchowski, A. T. (2018). Dynamics of emotional facial expression recognition in individuals with social anxiety. *Proceedings of the 2018 ACM Symposium on Eye-tracking Research & Applications. Warsaw, Poland*, 43:1–43:9. https://doi.org/10.1145/3204493.3204533

Lee, D. G., Fels, D. I., & Udo, J. P. (2007). *Emotive captioning. Comput. Entertain*, *5*(2). https://doi.org/10.1145/1279540.1279551

Llanes-Jurado, J., Marín-Morales, J., Guixeres, J., & Alcañiz, M. (2020). Development and calibration of an eye-tracking fixation identification algorithm for immersive virtual reality. *Sensors, 20*(17). https://doi.org/10.3390/s20174956

Mas Manchón, L., & Orero, P. (2018). Usability tests for personalised subtitles. *Translation spaces, 7*(2), 263–284. https://doi.org/10.1075/ts.18016.man

Matamala, A., & Orero, P. (2018). Standardising accessibility: Transferring knowledge to society. *Journal of Audiovisual Translation, 1,* 139–154. https://doi.org/10.47476/jat.v1i1.49

McClarty, R. (2012). Towards a multidisciplinary approach in creative subtitling. *MonTI: Monografías de Traducción e Interpretación*, 133–153. https://doi.org/10.6035/MonTI.2012.4.6

McClarty, R. (2014). In support of creative subtitling: Contemporary context and theoretical framework. *Perspectives, 22*(4), 592–606. https://doi.org/10.1080/0907676X.2013.842258

Mével, P. A. (2020). Accessible paratext: Actively engaging (with) d/deaf audiences. *Punctum. International Journal of Semiotics, 6*(01). http://doi.org/10.18680/hss.2020.0010

Mittell, J. (2009). *Forensic fandom and the drillable text.* Spreadable media. https://spreadablemedia.org/essays/mittell/index.html#.YgzhV-5_o4g

Montagud, M., Fraile, I., Meyerson, E., Genís, M., & Fernández, S. (2019). Imac player: Enabling a personalized consumption of accessible immersive contents. https://doi.org/10.6084/m9.figshare.9879254.v1

Montagud, M., Orero, P., & Matamala, A. (2020). Culture 4 all: Accessibility-enabled cultural experiences through immersive vr360 content. *Personal and Ubiquitous Computing, 24*(6), 887–905. https://doi.org/10.1007/s00779-019-01357-3

Nyström, M., & Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eye-tracking data. *Behaviour Research Methods, 42*(1), 188–204. https://doi.org/10.3758/BRM.42.1.188

Orero, P., Hughes, C. J., & Brescia-Zapata, M. (2020). Evaluating subtitle readability in media immersive environments. *DSAI 2020: 9th International Conference on Software development and Technologies for Enhancing Accessibility and Fighting Info-exclusion. Online Portugal*, 51–54. https://doi.org/10.1145/3439231.3440602

Romero-Fresco, P. (2013). Accessible filmmaking: Joining the dots between audiovisual translation, accessibility and filmmaking. *The Journal of Specialised Translation, 20*, 201–223.

Romero-Fresco, P. (2015). *The reception of subtitles for the deaf and hard of hearing in Europe.* Peter Lang. https://www.peterlang.com/view/title/36324

Rothe, S., Tran, K., & Hussmann, H. (2018). Positioning of subtitles in cinematic virtual reality. In Bruder, G., Yoshimoto, S., and Cobb, S., editors, *International Conference on Artificial Reality and Telexistence and Eurographics Sysmposium on Virtual Environments*, ICAT-EGVE 2018. The Eurographics Association.

Rupp, M., Kozachuk, J., Michaelis, J., Odette, K., Smither, J., & McConnell, D. (2016). The effects of immersiveness and future vr expectations on subjective – experiences during an educational 360 video. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 60,* 2108–2112. https://doi.org/10.1177/1541931213601477

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the 2000 Symposium on Eye-Tracking Research & Applications.* USA, 71–78. https://doi.org/10.1145/355017.355028

Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry, 36*(8), 1627–1639. http://pubs.acs.org/doi/abs/10.1021/ac60214a047

Skult, N., & Smed, J. (2020). Interactive storytelling in extended reality: Concepts for the design. In B. Bostan (Ed.), *Game user experience and player-centered design* (pp. 449–467). Springer International Publishing. https://doi.org/10.1007/978-3-030-37643-7_21

Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators & Virtual Environments, 6,* London. United Kingdom, 603–616. https://doi.org/10.1162/pres.1997.6.6.603

Smith, W. (2015, November 16). *Stop calling google cardboard's 360-degree video 'vr'*. Wired. https://www.wired.com/_2015/11/360-video-isnt-virtual-reality/

2.4. Article 4: Brescia-Zapata, M., Krejtz, K., Orero, P., Duchowski, A., & Hughes, C. (in press). Subtitles in VR 360º video. Results from an eye-tracking experiment. *Perspectives: Studies in Translation Theory and Practice.* https://doi.org/ 10.1080/0907676X.2023.2268122

Subtitles in VR 360º video. Results from an eye-tracking experiment

***Marta Brescia-Zapata**
https://orcid.org/0000-0002-2465-0126
Universitat Autònoma de Barcelona

Krzysztof Krejtz
http://orcid.org/0000-0002-9558-3039
SWPS University of Social Sciences and Humanities

Andrew T. Duchowski
http://orcid.org/0000-0003-1681-7878
Clemson University

Chris J. Hughes
http://orcid.org/0000-0002-4468-6660
Salford University

Pilar Orero
http://orcid.org/0000-0003-0269-1936
Universitat Autònoma de Barcelona

Abstract
Virtual and Augmented Reality, collectively known as eXtended Reality, are key technologies for the next generation of human-computer-human interaction. In this context, 360º videos are becoming ubiquitous and especially suitable for providing immersive experiences thanks to the proliferation of affordable devices. This new medium has an untapped potential for the inclusion of modern subtitles to foster media content accessibility (Gejrot et al., 2021), e.g., for the deaf or hard-of-hearing people, and to also promote cultural inclusivity via language translation (Orero, 2022). Prior research on the presentation of subtitles in 360º videos relied on subjective methods and involved a small number of participants (Brown et al., 2018; Agulló, 2019; Oncins, 2020), leading to inconclusive results.
The aim of this paper is to compare two conditions of subtitles in 360º videos: position (head-locked vs fixed) and colour (monochrome vs colour). Empirical analysis relies on novel triangulation of data from three complementary methods: psycho-physiological attentional process measures (eye movements), performance measures (media content comprehension), and subjective task-load and preferences (self-report measures). Results show that head-locked coloured subtitles are the preferred option.

Keywords: immersive environments, 360º video, media accessibility, subtitles, captions, eye-tracking

1 Introduction

Immersive content is meant to give the user the illusion of being “physically present” (Slater & Wilbur, 1997), and can provide benefits in a variety of sectors such as entertainment, communication, learning, arts, and culture (Montagud-Climent et al., 2020; Liberatore & Wagner, 2021). 360º videos—also known as immersive or VR360 videos—offer great potential in providing engaging media experiences. Already in 2017, 49 % of the public broadcasters who responded to the European Broadcasting Union (EBU, 2017) report on the use of VR declared offering 360º content. The most popular devices to access this immersive content are head-mounted displays (HMD). According to the newsletter XR today (Greener, 2022), “in 2020, roughly 57.4 people owned a VR headset in the US, although in 2022, this figure increased by 37.7 million.” The demand for VR headsets is expected to increase due to the adoption of VR technology in enterprise, industry, and education sectors.

These new immersive media environments must be accessible for all to fulfil existing accessibility legislation in most world regions. Standards such as the EN301459 recommend Universal Design when developing any system or product. This requirement is becoming mainstream with advice from the United Nations Convention on the Rights of Persons with Disabilities (CRPD, 2006), and now it is also an issue of political will and moral obligation thanks to the Audiovisual Media Services Directive (AVMSD) and the European Accessibility Act (EAA). These pieces of legislation adopted a user-centric approach, a method that is at the heart of Human Rights towards full democratic participation in society by all—which in the 21st century depends on access to media. Following the aforementioned legislation and academic research (Udo & Fels, 2010; Romero-Fresco, 2013) any system or process should be designed with accessibility in mind from the outset, leading to a born accessible system that avoids expensive and complex afterthought solutions.

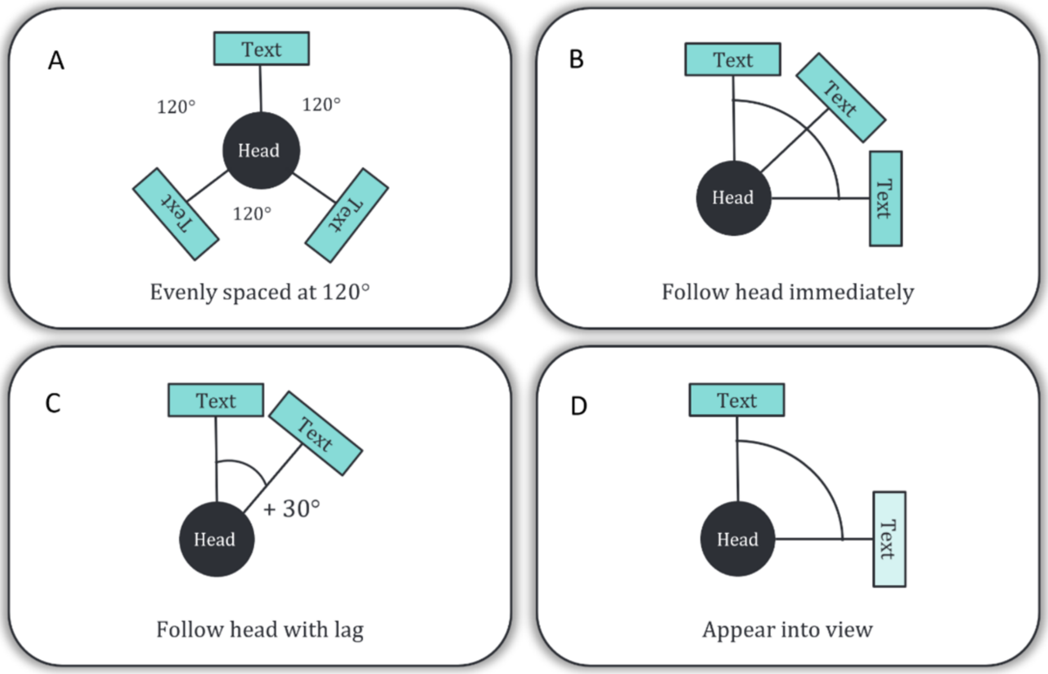
This paper focuses on subtitling, where standardised practices have been adopted in the context of 2D non-immersive media (ISO 20071:23; Matamala & Orero, 2018). Although the trend towards mixed methods in translation accessibility studies is becoming more popular, publications still fail to discuss the mixed-method nature of the study in depth (Hermosa-Ramírez, 2022). Prior reception studies on the evaluation of subtitles in 360º have largely been based on subjective measures, using questionnaires, interviews, and focus groups (Agulló & Orero, 2017; Brown et al., 2018; Roth et al., 2018; Fidyka & Matamala, 2018; Agulló & Matamala, 2020). Objective, psychophysiological measures (such as eye movements, heart rate, and electrodermal activity) have been largely adopted in the context of tests for 2D non-immersive subtitles (Kruger, 2016; Krejtz, Szarkowska & Łogińska, 2016; Szarkowska & Gerber-Morón, 2019; Liao et al., 2021) with only a recent contribution by Ibourk and Al-Aldwan (2019), who included for the first time eye-tracking technology in a small study on immersive subtitling. The present study adopts a mixed-method design, with psycho-physiological process metrics (eye movements), performance metrics (scene comprehension), and subjective self-reports (task-load and preferences) followed by result triangulation. In this article, a critical review of previous studies related to subtitling in immersive media is presented, followed by the characteristics of the study and the methodology. To finish, the results, a discussion, and conclusions are presented.

2 Background

Studies on subtitles in VR are relatively recent. Understanding the displaying mode for subtitles started at the British Broadcasting Corporation (BBC), which first identified the main challenges when developing subtitles for immersive content (Brown et al., 2017). As a result, four solutions for subtitle rendering were designed (Brown, 2017) and tested with 24 hearing participants (Brown et al., 2018). These four conditions were (a) fixed positioned, evenly spaced, (b) follow head immediately—also called Always Visible—, (c) follow with lag and (d) appear in front, then fixed position (see Figure 1). Users reported a preference for subtitles which are always presented in front of the viewer, following head movements—Always Visible subtitles.

Figure 1

Four static subtitle solutions designed and tested by Brown et al. (2018)



Source: Author own elaboration based on Brown et al. (2018)

Rothe et al. (2018) compared Brown et al. (2018) Always Visible subtitles with a new type called dynamic subtitles—placed near the speaker. They go a step further by measuring simulator sickness and task workload. They tested with 34 hearing participants and received feedback from one deaf participant (although this data was not included in their analysis). This study obtained similar results to those of its predecessor and pointed out that "additional usage of an eye tracker could lead to more detailed results in the analysis of the viewing direction" (Rothe et al., 2018, p. 214).

The H2020 European-funded project ImAc[1] based on the results of Brown et al. (2018) and Rothe et al. (2018) added location information within the 3D space to each subtitle (Agulló & Matamala, 2019). Results from Agulló and Matamala (2019) helped to draft user requirements, with a priority for designing a guiding mechanism to help deaf and hard-of-hearing users to locate the sound source related to the subtitle. The project designed and tested several guiding mechanisms with 6 hearing and 2 deaf participants (Agulló, Montagud, & Fraile, 2019). Results showed two preferred guiding methods: an arrow positioned to the left or right of the subtitle directing the user to the sound source, and a radar circle shown in the user field of view that identifies both the position of the sound source and the relative viewing angle of the user. Guiding mechanisms can also be found in videogames, such as *The Last of Us: Part II* (Myers, 2020), which "includes a guide arrow to direct the user to the location of the character speaking" (Hughes et al., 2020). It is worth noting that the replication of Agulló at al. (2019) study with a larger number of participants (*N* = 40: 27 hearing, 20 with hearing loss, 6 hearing impaired and 7 deaf) showed that always-visible subtitles with arrows were the preferred option (Agulló & Matamala, 2020).

On the one hand, all studies to this point conclude that the preferred visualisation mode is head-locked, centred, bottom subtitle, a trend that replicates conventions established for 2D, non-immersive content as defined in ISO/IEC 20071-23:2018 or UNE-153010:2012 among others. On the other hand, all studies focus on two subtitle features: sound source and subtitle position. All previous studies have four limitations: i) demography samples are small in number, ii) the methodology is based solely on subjective opinions and answers (in-depth interviews, focus groups, and/or questionnaires), iii) unstandardised stimuli videos in terms of length of video, language of dialogue in video, language of users, etc.

The most recent work in this field has adopted a new web-based prototyping framework (Hughes et al., 2020). The framework allows for real time subtitle editing and visualisation in 360º videos. It takes into consideration all the previously reported work and is based on two mechanisms for subtitle rendering: (1) head-locked, where the subtitle is rendered relative to the user's viewpoint, and (2) fixed, where the subtitle is rendered relative to a fixed location in the world, usually close to the speaking character. This framework offers the opportunity of instant evaluation which was tested by Brescia-Zapata et al. (2022) in a pilot study leading to a full experiment—whose results are presented in the present paper. The experiment received ethics clearance from the authors' home institutions, according to the ethics and privacy regulations of the H2020 EU funded TRACTION project[2].

The present study was conceived to further clarify which is the best visualisation mode for subtitles in immersive environments for all kinds of users. The aim is two-fold: 1) to detail the design of controlled experiments testing influence of subtitles on users' attention allocation, and immersive content comprehension, and 2) to gather feedback regarding preferences of

[1] https://www.imacproject.eu/
[2] https://www.traction-project.eu/

two characteristics of subtitles in immersive content: position and colour. Special attention has been put on the following methodological aspects i) diverse sample in terms of demography and subtitle usage habits ii) a triangulation of psychophysiological, qualitative and questionnaire methods, and iii) controlled stimuli. It focuses on two hypotheses: the first is related to subtitle positioning and the second is related to subtitle colour.

H1: Head-locked subtitles are easier to follow but more intrusive than fixed subtitles.

H2: Coloured subtitles are helpful to identify a speaking character in a 360º narrative.

3 Method

3.1 Experimental Design and Independent Variables

To test the research hypotheses the experimental study was conducted with 2 x 2 mixed design with two independent variables (IV): subtitle position (head-locked vs fixed) and subtitle colour (monochrome vs. colour). The first one was treated as a between-subject IV and the later one as a within-subjects IV. The order of the experimental conditions was counterbalanced.

The two conditions for subtitle position were the following:

- (a) head-locked subtitles which are always visible and are displayed in front of the viewer (see Figure 2), following head movements. Head-locked subtitles are equivalent to static-follow in Brown et al. (2017), to static subtitles in Rothe et al. (2018), and to always-visible in Agulló and Matamala (2020).

Figure 2

Head-locked subtitles attached to the FoV in Vacations video stimuli

Source: Author own elaboration

- (b) fixed subtitles which appear near to the speaking characters and remain fixed to the scene (Figure 3). Fixed subtitles are equivalent to 120 degrees in Brown et al. (2017), and to dynamic subtitles in Rothe et al. (2018).

Figure 3

Fixed subtitles attached to one position in the sphere in Vacations video stimuli

Source: Author own elaboration

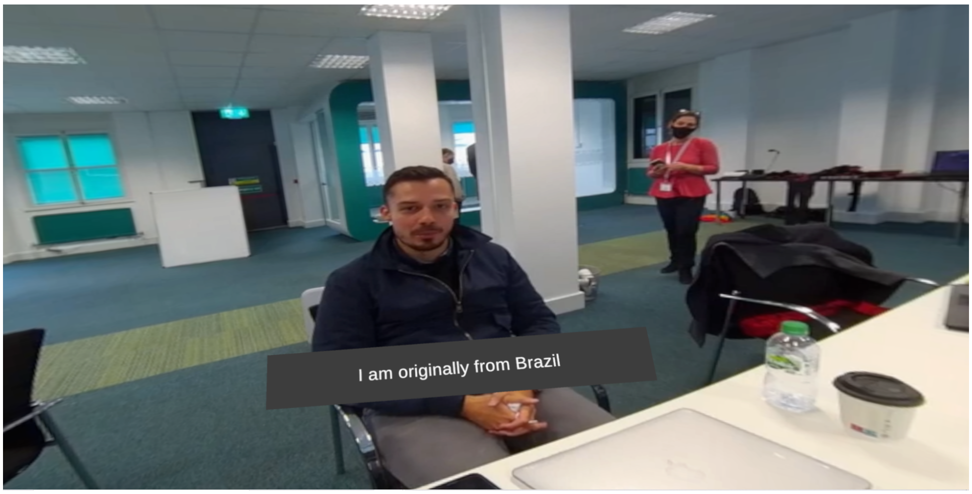
The main reason for testing these two conditions is that results from previous studies (Brown et al., 2018; Rothe et al., 2018; Agulló et al., 2018) only included qualitative data obtained through questionnaires and focus groups.

The two conditions for subtitle colour were the following:

- (a) monochrome (see Figure 4). All the subtitles for each character were in the same colour white on grey background.

Figure 4

Monochrome subtitles in TRACTION video stimuli

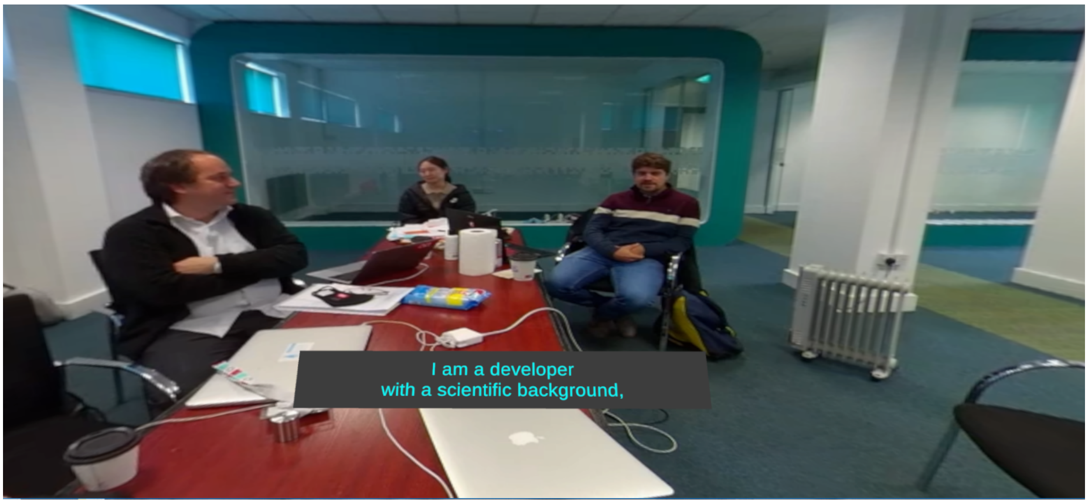


Source: Author own elaboration

- (b) colour (see Figure 5). Four colours have been used to identify the different characters: white, yellow, cyan, and green. The colour coding follows the ISO 20071-23:2018 standard.

Figure 5

Coloured subtitles in TRACTION video stimuli



Source: Author own elaboration

The objective of examining these two conditions is to evaluate the effectiveness of colour coding, which is suggested by certain guidelines such as BBC (2022) as a means of distinguishing between speakers and has demonstrated an enhancement in the immersive experience of multilingual 2D content (Szarkowska & Boczkowska, 2022). However, despite its potential benefits, this feature has not been treated as a variable in previous studies on immersive content (Brown et al., 2018; Rothe et al., 2018; Agulló et al., 2018).

3.2 Participants

In total, 73 volunteers took part in the experiment at three locations: $N = 24$ in Universitat Autònoma de Barcelona (Spain), $N = 24$ in Salford University (UK), and $N = 25$ in SWPS University of Social Sciences and Humanities (Poland). Participants received no incentives for the study. There were no significant differences between the countries in age of the participants $F(2,69) = 2.26$, $p > 0.1$ (see Table 1).

Table 1. Participants' demographic characteristics, attitude, and usage of VR technology in each location

	SPAIN	**UK**	**POLAND**
Number (*N*)	24	24	25
Age	(*M*=33.96, *SD*=11.18)	(*M*=29.41, *SD*=10.34)	(*M*=27.40, *SD*=6.68)
Gender	17 female	4 female	19 female
Age difference between genders	*t*(21)=0.52, *p*=0.61	*t*(22)=0.70, p=0.49	*t*(23)=0.94, *p*=0.36
Occupation	Students (N=11); Researchers/lecturers (N=11); Practitioners (N=2)	Students (N=18), Researchers/lecturers (N=3), Data Analysts/engineers (N=4)	Students (N=16), Researchers/lecturers (N=4), Journalist (N=2), Analyst (N=1)
Vision	Corrected (*N*=12), Uncorrected (N=12)	Corrected (*N*=12), Uncorrected (N=15)	Corrected (N=7), Uncorrected (N=18)
Handedness	Right (*N*=24)	Right (*N*=21)	Right (N=25)
VR interest	(*M*=4.13, *SD*=0.76)	(*M* = 3.96, *SD* = 0.65)	(*M* = 3.00, *SD* = 1.22)
VR experience	(*M*=0.87, *SD*=1.18)	(*M*=0.81, *SD*=1.24)	(*M*=1.24, *SD*=2.01)
Digital device daily usage	(*M*=3, *SD*=1.38)	(*M*=3.59, *SD*=1.28)	(*M*=2.91, *SD*=0.91)
Attitude to subtitles (answers to "I always turn subtitles on")	*(M*=3.97, *SD*=1.16)	*(M*=3.88, *SD*=1.32)	*(M*=4.01, *SD*=0.97)

Source: Author own elaboration

The three locations of the study were chosen in order to ensure the robustness of the obtained results, as each country has a different language translation tradition. Spain has traditionally been considered a dubbing country (Ballester Casado, 1998; Chaume, 2012; Gil Ariza, 2004), although the presence of subtitles has increased in recent years (Matamala et al., 2017). Poland is generally considered a stronghold of voice over (Gottlieb, 1998; Valdeón, 2022), coexisting with other audiovisual translation modes, such as dubbing, subtitling, audio description and subtitling for the deaf and the hard-of-hearing (Szarkowska, 2009). The UK belongs to a large anglophone audiovisual market and is neither a classical "subtitling" nor "dubbing" country (Luyken, 1991), since most audiovisual content is produced in English.

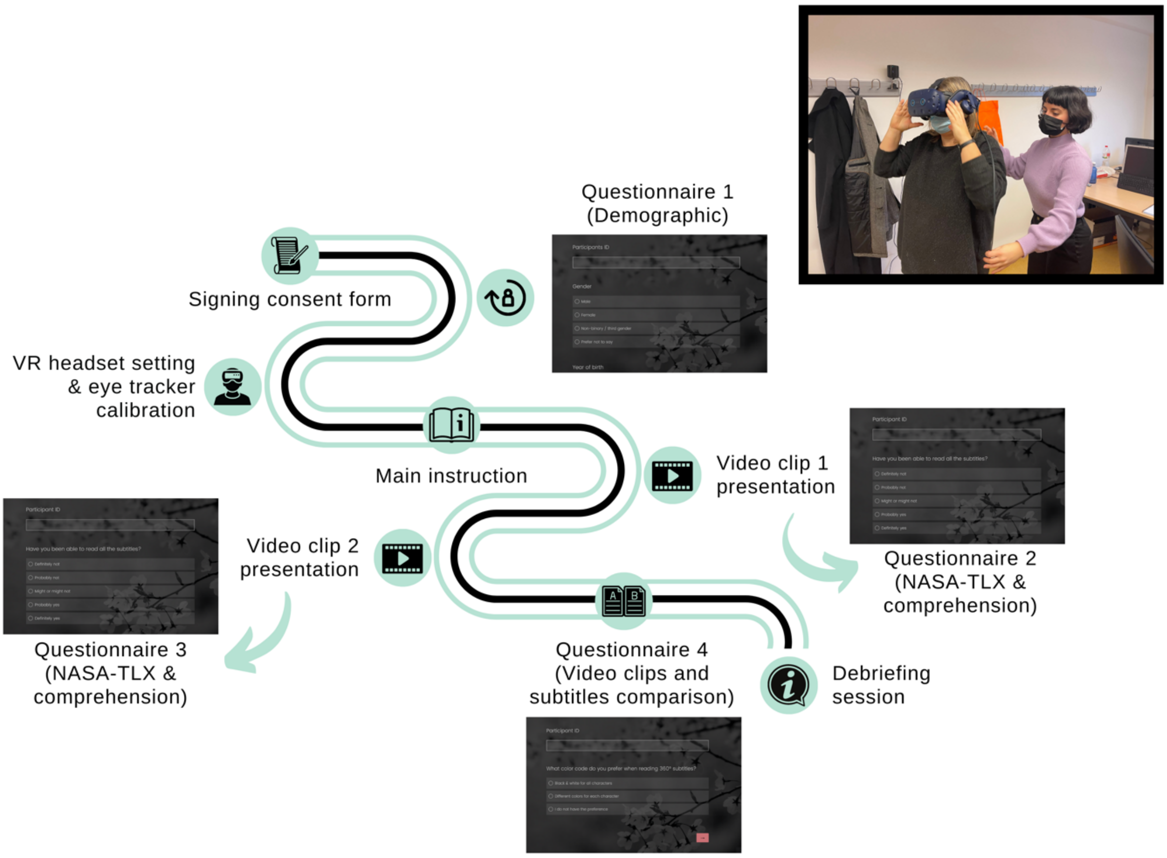
All participants in the sample had above average reading skills. We can also consider them as digital media savvy based on their declared daily usage of digital media and devices although most participants had never experienced VR content before (see Table 1).

3.3 Experimental Procedure

The experimental procedure was the same in the three locations of the present study (see Figure 6).

Figure 6

Experimental procedure schematic timeline and settings (top-right)



Source: Author own elaboration

The experimental procedure was executed individually in the laboratory facilities.
After signing the consent form participants started with the demographic and usage & attitudes questionnaire. Next, they were familiarised with the VR headset. The HMD built-in eye tracker was calibrated. The following main part of the experiment consisted of three steps:

1. watching first video stimuli followed by a comprehension and NASA-TLX questionnaire
2. watching second video stimuli followed by comprehension and NASA-TLX questionnaire
3. a questionnaire comparing subtitles in both viewed videos.

After the experiment participants were thanked and debriefed.
A facilitator was present during the experiment providing information as required, helping the participants to correctly wear the HMD, and assisting with the completion of the

questionnaires (see Figure 6, top-right). All the questionnaires were formulated using Qualtrics system and completed using a desktop computer from the laboratory.
The order of video stimuli presentation was counterbalanced; thus participants were presented with the two videos in random order, one with monochrome and the other with coloured subtitles, but same subtitle position in both cases (either head-locked or fixed). During video stimuli presentation the participants' eye movements were recorded.

3.4 Experimental Materials

3.4.1 Stimuli video

For the present study two custom 360º stimuli videos were recorded using an Insta360 One X2 camera. The first is called *Vacations* and was of a family, speaking Arabic, discussing their vacation plans. In the second stimuli video, called *Traction*, a group of researchers introduce themselves, each speaking in their own language (Spanish, Korean, Catalan, Portuguese, and English). *Vacations* is one minute and 18 seconds long and *TRACTION* is 2 minutes and 32 seconds long. Both videos followed the same features:

1) Speaking characters were placed all around the scene, exploiting 360º movement and immersion,
2) there were no overlaps between speaking characters. If two characters located at different points in the 360º scene spoke at once, it would be almost impossible for the user to read the subtitles,
3) language spoken was different from that of participants. This feature was added after the pilot study (Brescia-Zapata et al., 2022), as during the data collection, several participants claimed that they did not need to read the subtitles because they understood the original audio.

Both videos were subtitled in English, Spanish, and Polish to meet the experimental design (two colour schemes, two positioning schemes). The number of subtitles is the same in all languages: in *Vacations* there are 27 subtitles and in *TRACTION* 39. Participants at UAB were presented with video stimuli subtitles in Spanish, participants at Salford University were presented with video stimuli subtitled in English and participants at Warsaw University were presented with video stimuli in Polish. Subtitles followed the ISO 20071-23:2018 standard and the font type selected was Liberation Sans. Another feature that differentiates the pilot study from this experiment is the grey background box used in both videos. Although some previous studies proved that it obstructs important parts of the image (Brown et al., 2018), results from the pilot test revealed that "some of the participants found subtitles with no background hard to read" (Brescia-Zapata et al., 2022).

3.4.2 Questionnaires

Four questionnaires were designed to gather information about demographics, task-load, comprehension, and preference. The first questionnaire had 20 questions to gather demographic data (sex, age, handedness, occupation) and data related to usage and attitudes towards digital media (e.g., “Which of these devices do you use on a daily basis?”), VR (e.g., “How often do you watch virtual reality (VR) content for instance, 360º videos?”), and subtitles (e.g., “When subtitles are available, do you turn them on?”, “How many hours a day do you watch subtitled content?”).

After each stimuli video, participants filled out a questionnaire with six multiple choice questions on content comprehension. These questions were customised for each stimuli video. After comprehension questions participants completed a task-load index (NASA-TLX) (Hart & Staveland, 1988) consisting of six questions on personal evaluation of task load (e.g., mental load, physical demands, time pressure, etc).

After viewing two stimuli videos and the subsequent questionnaires participants answered three questions on subtitles preferences (“What colour code do you prefer when reading 360º subtitles?”, “Explain in your own words why you prefer the option selected in the previous question”, and “Do you think that the subtitles obstruct important parts of the image?”)

3.5 Experimental Equipment

The experimental procedure was prepared with a framework developed by Hughes et al. (2020) and ported to Unity 3D. We used the HTC Vive Pro Eye headset, which contains a Tobii eye tracker built-in with 120 *Hz* sampling rate. Vive’s eye tracking accuracy estimation is 0.5º—1.1º (Sipatchin et al., 2020). The HMD has two AMOLED screens, with a resolution of 2,880 x 1,600 pixels in total, a display refresh rate of 90 Hz and a FoV of 110º.

4. Results

Hypotheses testing was based on a series of Multilevel Linear Model (MLM) with random effects for subjects and for stimuli videos. Each test we started with the Null model with random effects in it only, then consecutive models added fixed effects for country, subtitle position and colour. When also needed the interaction of position and colour were added as fixed factor. Each consecutive model was tested against the previous one with the chi-square test to check if a new fixed term improves the fit of the model. All the statistical analyses were performed in R language for statistical computing (R Core Team, 2020).

H1 was “head-locked subtitles are easier to follow but more intrusive than fixed subtitles”. To test this hypothesis, we evaluated general comprehension of the content, subjective effort and frustration, and attention allocation over subtitles and scene Areas-of-Interest (AOI). In detail, we expected that head-locked positioning will cause better comprehension of the content, lower frustration, and effort. We also expected that head-locked positioning will cause relatively more fixations on subtitles AOI than fixed positioning.

H2 was “coloured subtitles are helpful to identify a speaking character in a 360º narrative”. To test this hypothesis, we evaluated comprehension of the content. We expected that coloured subtitles allow participants to identify speaking characters more easily than monochrome subtitles.

4.1 Influence of subtitle positioning and colour code on content comprehension and self-reported effort

To test the prediction that head-locked positioning will lead to better content comprehension we have run MLM models on the general comprehension score from the comprehension questionnaires (see Figure 7a and Appendix Table A1 for model detailed tables and Appendix Table A2 for estimated means for the interaction between subtitle position and colour).

Figure 7

Content comprehension (Fig. 7a) and self-evaluation of perceived effort (Fig. 7b) depending on subtitle position (fixed or head-locked) and colour.

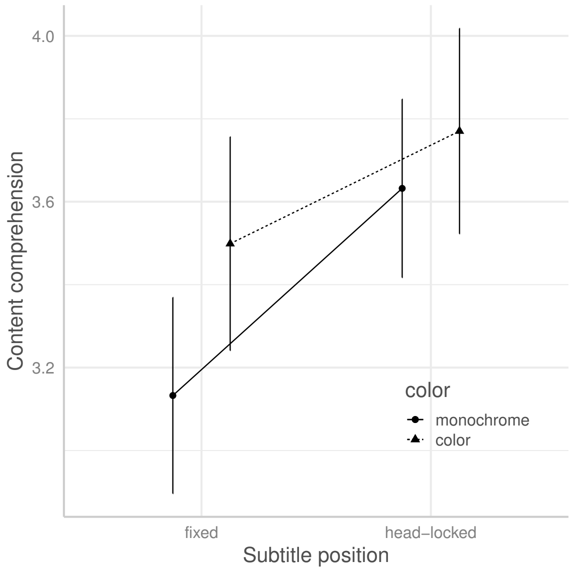


Fig. 7a. Content comprehension depending on subtitle position (fixed or head-locked) and colour.

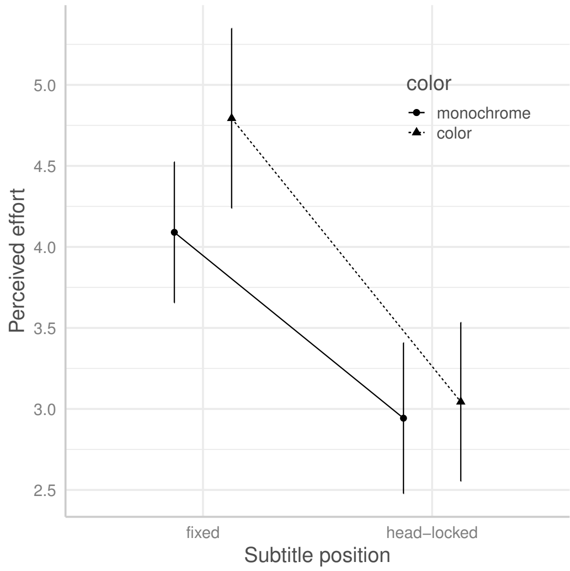


Fig. 7b. Self-evaluation of perceived effort depending on subtitle position (fixed or head-locked) and colour.

Note: bar height represents the estimated means and whiskers ±1*SE.*

The null model with random effects for subject and video showed *SD* = 0.10 for subject intercepts and *SD* = 0.32 for video intercept, and *SD* = 0.99 for residuals. The intercept of the model was significantly higher than zero (b_0 = 3.52, $t(3.12)$ = 13.70, p < 0.001).

The country fixed effect did not influence the model fit (AIC = 1775.1, $\chi^2(2) = 0.74$, $\mathrm{p} = 0.69$). Contrary to expectations subtitle position did not improve significantly the model fit (AIC = 1778.5, $\chi^2(1) = 2.38$, $\mathrm{p} = 0.12$). However, in line with expectations, subtitles colour significantly improve the model fit ($\chi^2(1) = 8.64$, $\mathrm{p} = 0.003$). The full model includes also the interaction of subtitles colour and position as a fixed effect, but it does not influence the fit ($\chi^2(1) = 1.91$, p = 0.17).

The full model fixed effects parameters showed that subtitles colour significantly improved comprehension of the movie ($b = 0.37$, $t(497.99) = 3.05$, $p < .001$). Estimated mean for colour subtitles was ($M = 3.63$, $SE = 0.31$) and for monochrome subtitles ($M = 3.38$, $SE = 0.31$).

Head-locked subtitle position improved comprehension but only marginally significant, (b=0.50, t(88.23) = 1.93, $p = 0.06$), while the interaction of subtitle colour and position did not predict significantly the movie comprehension, (b= -0.23, $t(498.68) = -1.38$, $p = 0.17$).

To test the prediction that head-locked positioning will cause lower effort a similar analysis on comprehension MLM approach was used (see Figure 7b and also Appendix Table A3 for model detailed tables and Appendix Table A4 for estimated means for the interaction between subtitle position and colour). The null model showed *SD* = 2.20 for subject intercepts and SD = 1.01 for video intercept, and *SD* = 1.64 for residuals. The intercept of the model was statistically significant (b_0 = 3.68, $t(2.55) = 4.80$, p = 0.02).

Comparison of models with fixed factors added showed that country did not improve the model (*AIC* = 2247.5, $\chi^2(2) = 2.84$, $p = 0.24$) but both the subtitles position (*AIC* = 2241.9, $\chi^2(1) = 7.58$, $p = 0.005$) and subtitles colour did (*AIC* = 2236.9, $\chi^2(1) = 7.05$, $p = 0.008$) as well as an interaction term between the colour and position of the subtitles (*AIC* = 2234.4, $\chi^2(1) = 4.43$, $p = 0.04$).

The analysis of the full model fixed parameters showed that colour subtitles in general add significant effort to the task ($\beta = 0.70$, t(471.01) = 3.37, $p < 0.001$). Head-locked subtitles lower significantly perceived effort (β = -1.15, t(79.09) = -2.21, p = 0.03). Significant interaction term shows that head-locked subtitles lower the effort mainly when they are presented in colour (β = - 0.60, t(469.41) = -2.11, $p = 0.04$).

4.2 Influence of subtitle positioning on visual attention directed toward subtitles

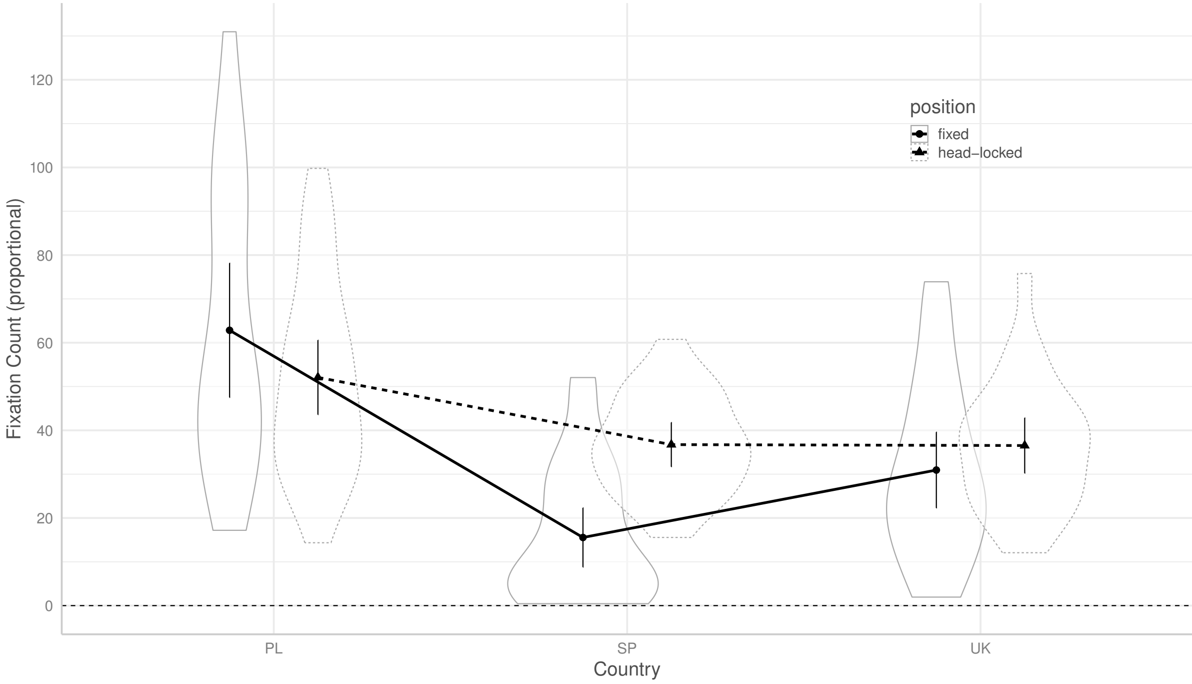
We hypothesised that fixed positioning causes more fixations and longer fixation time over subtitles. To test these predictions, we applied MLM analyses.

4.2.1 Fixation count

The first analysis used proportional fixation count (the number of fixations divided by the number of words in each subtitle) as a dependent variable. The MLM started with the null model with random effects of participants and stimuli video in it. In next steps the model was updated with the fixed effects of country, subtitle position, subtitle colour, and interaction fixed effects of position and colour, position, and country. The models were tested against each other with chi-squared statistics to see if the fixed effect is significantly improving the model fit (see Figure 8 and also Appendix Table A5 for model detailed tables and Appendix Table A6 for estimated means for the interaction between subtitle position and country).

Figure 8

Proportional fixation count on subtitles AOI depending on subtitle position (fixed or head-locked) and country (Poland, Spain or UK).



Note: bar height represents the estimated means and whiskers ±1*SE.* Violin shapes represent the distribution of the data points in each experimental condition in different countries.

The null model showed *SD* = 16.20 for subject intercepts and *SD* = 19.30 for video intercept, and *SD* = 13.32 for residuals. The intercept of the null model was not statistically significant (β_0 = 39.06, *t*(1.04) = 2.83, *p* = 0.21).

Models' comparison shows the significant effect on the model fit of a country as a fixed effect, $\chi^2(2) = 47.03$, $p < 0.001$, $AIC = 1217.10$, $BIC = 1234.90$. Also, the interaction effect of country and position significantly improved the model fit, $\chi^2(2) = 19.53$, $p < 0.001$, $AIC = 1165.28$, $BIC = 1197.87$. The final models' total psycho-$R^2 = 0.78$.

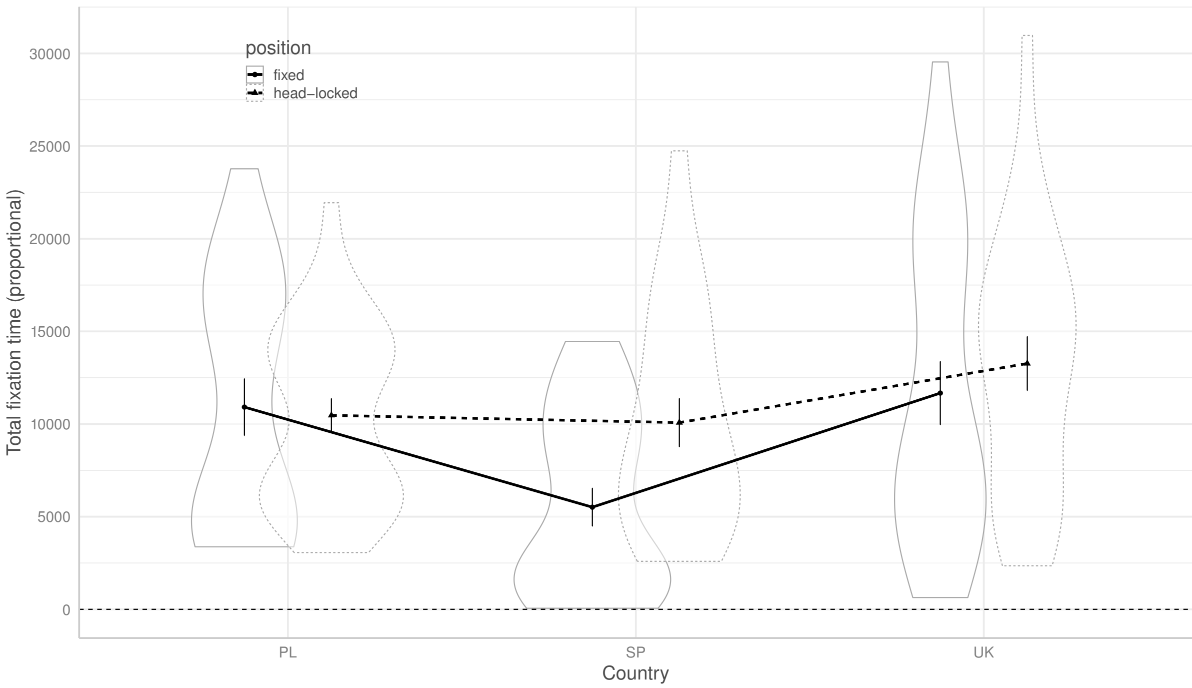
The full model fixed parameters showed that fixed subtitles significantly lower the proportional fixation count comparing to head-locked ($\beta = -12.84$, t(91.72) = -2.39, $p = 0.02$). This effect was quantified by country. The pairwise comparisons of means showed that the differences between fixed and head-locked subtitles were significant in Spain (t(66.6) = -4.22, p < 0.001) and in Poland (t(65.0) = 2.20, p = 0.03) but not in UK (t(65.3) = -1.15, p = 0.26). Note that the difference in fixation count between head-locked and fixed subtitles in Polish sample was in favour of head-locked, unlike in the Spanish sample.

4.2.2 Total Fixation Time

The second MLM analysis used proportional total fixation time (total fixation time divided by visibility time of subtitle) as a dependent variable. The procedure of testing and modelling random and fixed effect structure was identical to the analysis of fixation count described in subsection 5.2.1 (see Figure 9 and also Appendix Table A7 for model detailed tables and Appendix Table A8 for estimated means for the interaction between subtitle position and country).

Figure 9

Proportional total fixation time on subtitles AOI depending on subtitle position (fixed or head-locked) and country (Poland, Spain, or UK).



Note: Bar height represents the estimated means and whiskers ±1*SE.* Violin shapes represent the distribution of the data points in each experimental condition in different countries. The interaction effect is not statistically significant in the full model.

The null model (with random effects only) for total fixation time showed *SD* = 3537.14 for subject intercepts, *SD* = 6696.95 for video intercept, and *SD* = 3238.18 for residuals. The intercept of the null model was not statistically significant (β_0 = 10322.92, *t*(1.02) = 2.17, *p* = 0.27).

The fixed effect of the country significantly improved the model fit ($\chi^2(2) = 16.96$, $p < 0.001$, $AIC = 2808.60$, $BIC = 2826.40$) as well as the fixed effect of subtitle position ($\chi^2(1) =$

4.50, $p = 0.03$, $AIC = 2806.10$, $BIC = 2826.90$). Additionally, the interaction term between country and subtitle position was marginally significant ($\chi^2(2) = 5.83$, $p = 0.054$, $AIC =$ 2808.10, $BIC = 2840.70$). Because of only marginal model fit improvement, the interaction between country and subtitle positioning is hard to interpret.

However, the parameters of the model with the country and position fixed terms (with pseudo-R^2 (total) = 0.85) clearly show that the fixed subtitles gained more total fixation time than head-locked subtitles ($\beta = 1861.30$, $t(58.07) = 2.11$, $p = 0.04$). Significant effect of country shows also that participants in Spain spent significantly less time reading subtitles than participants in Poland ($\beta = 12718.09$, $t(68.28) = -2.51$, $p = 0.01$). See Figure 9 for the estimated means of proportional total fixation time for both effects.

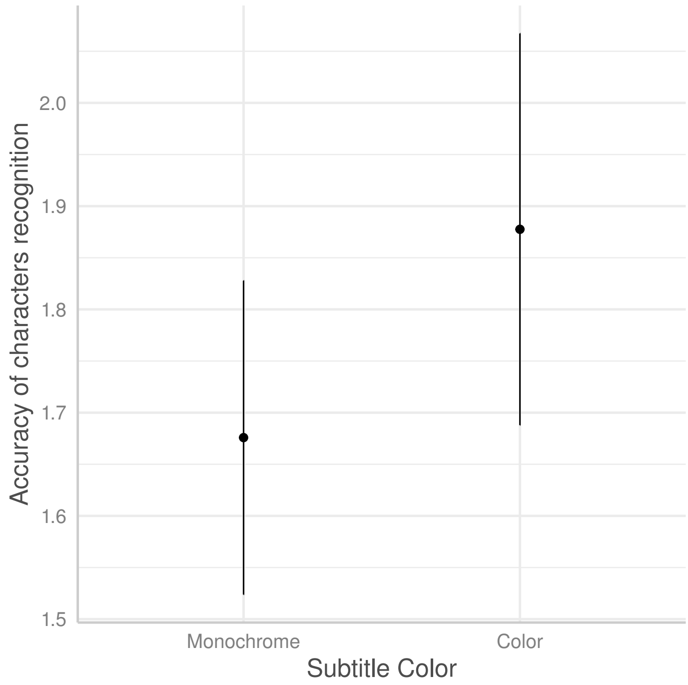
4.3 Influence of subtitle colour on character recognition

Null multilevel model on character recognition, with random effects only for participants and stimuli video, showed *SD* = 0.67 for subject random effect and *SD* = 1.60 for video, and *SD* = 0.69 for residuals. The intercept of the null model was not statistically significant ($\beta_0 = 1.78$, $t(1.01) = 1.57$, $p = 0.36$).

The models fitting with the fixed effects for country, subtitle position, subtitle colour as well as interactions between colour and position, and position and country showed that subtitle colour code significantly improve the model fit ($\chi^2(1) = 12.26$, $p < 0.001$, $AIC = 1378.32$, $BIC = 1413.14$, pseudo-R^2(total) = 0.86). The model parameters showed that colour subtitles code significantly improves the recognition of characters ($\beta = 0.20$, $t(3.51) =$ 3.51, $p < 0.001$). The estimated average recognition of characters with monochrome subtitles was significantly lower ($M = 1.68$, $SE = 1.12$) than with colour subtitles ($M =$ 1.88, $SE = 1.12$), see Figure 10 and Appendix Table A9 for detailed parameters of the model.

Figure 10

Estimated means of fixed effect of subtitle colour on accuracy of characters' recognition



Answers to the question "what colour code do you prefer when reading 360º subtitles?" showed no significant preference between colour and monochrome subtitles ($p > 0.5$). Regarding coloured subtitles, one participant stated that "colours distracted me from what they [the characters] were actually saying", but on the contrary, another participant stated that "it's easier to know which one is talking, especially because you aren't looking to all of them at the same time".

These general results are in line with those obtained in Poland, but not with those obtained in Spain and the United Kingdom (see Table 2).

Table 2

Subtitles colour preferences between countries

		Conditions	
	Monochrome	Coloured	No preference
Poland	42,9 %	40,8 %	16,3 %
Spain	39,1 %	56,5 %	4,3 %
UK	58,3 %	25 %	16,7 %
Total	46,8 %	40,5 %	12,6 %

Source: Author own elaboration

The differences in subtitle colour preference between countries were statistically significant ($p = 0.02$). Spain participants stated a preference for coloured subtitles, while UK participants stated a preference for monochrome subtitles.

5 Discussion

The present study examined 360º video although immersive subtitles can extend to any media asset. It focused on two hypotheses: the first related to subtitle positioning and the second related to subtitle colour. Regarding subtitle positioning, head-locked subtitles were subjectively evaluated as causing less cognitive effort in comparison to fixed subtitles. However, eye tracking data showed that head-locked subtitles engaged more attention from the viewer than fixed subtitles. Head-locked subtitles gained significantly higher fixation counts and total fixation time, both of which are proportional to the number of words in the subtitle and visibility time of the subtitle, respectively. These results suggest that head-locked subtitles are more cognitively engaging, despite their subjective evaluation. The latter might be interpreted via the familiarity heuristic (Ashcraft and Radvansky, 2014) which states that people usually prefer things or situations that they are familiar with. However, it is important to note that this conclusion is specific to the Spanish sample. In the Polish, the difference in fixation count between head-locked and fixed subtitles favoured head-locked subtitles, which contrasts with the findings in the Spanish sample.

Subtitle positioning does not significantly affect video content comprehension. There was only a marginally significant effect of head-locked subtitles improving comprehension. The factor which influenced content comprehension significantly was the subtitles' colour scheme. Use of the different subtitle colours for each character video character improved character recognition and general comprehension, however, colour subtitles were evaluated by participants as causing more effort. Nevertheless, this is the first experiment testing subtitle colour in 360º videos, and it would be interesting to carry out further research on this topic to confirm our results.

The present article points out interesting evidence of discrepancies between self-report measures and more objective, psychophysiological, and comprehension metrics. Thus, we postulate to incorporate psychophysiological and performance methods and indicators in future studies on immersive media accessibility. It also points out the need for test environments or tools similar to what was proposed by Hughes et al. (2020) and that was used in this empirical user study.

An additional contribution of this study is the comparison of the different audiences from the sample. Replicating the experiment in three countries with different language translation traditions in audiovisual media allowed us to work with a heterogeneous sample and report differences between the participants. We found some significant differences both in perceived effort/frustration and AOI fixations depending on the country. Although it is not possible to define tendencies with the number of participants tested, it might be interesting to consider the preferences of people from different countries and cultures when testing subtitled immersive media.

Even though it was not the aim of the study, some participants mentioned the problem of speaker or audio source localisation when answering the questionnaires. This is in line with earlier work (Hughes et al. 2019, Rother et al., 2018; Agulló et al. 2019) calling for user guiding mechanisms such as arrows, radar, or other means. Although arrows have been shown to be the option preferred by users (Agulló, 2020), eye-tracked evaluation of such means holds potential for future work in the field.

6 Conclusions

The present paper presented a review of the main studies regarding subtitles in immersive content and the main challenges posed by this new medium. The first aim of the study was to detail the design of controlled experiments involving immersive environments and subtitles. The methodology followed in the three locations where the experiment took place has been described in detail, including the design and IV, the participants, the procedure, the materials, and the equipment. This allows other researchers to replicate the experiment in other locations and/or using different video stimuli.

The second aim was to gather feedback regarding two characteristics of subtitles in immersive content: position and colour. To confirm the hypotheses, feedback on preferences and task-load was gathered, together with eye movement data. The results have shown that head-locked subtitles, which are always visible and positioned in front of the viewer to align with their head movements, attracted more attention from viewers. However, it should be noted that this finding was observed specifically within the Spanish sample. Additionally, due to the observed trend in the Polish sample pointing in the opposite direction. In all samples, it was observed that participants consistently showed a preference for coloured subtitles as opposed to monochrome subtitles. This preference can be attributed to the enhanced ability to identify the speaking character facilitated by the use of colour. Although colour subtitles were perceived as requiring more effort, particularly when presented in a fixed position, they yielded improved comprehension of the content, especially in the fixed position, and better recall of the characters. Importantly, the impact of colour on comprehension and recall was consistent across the different country samples, indicating that it is not influenced by cultural factors.

The present study had some limitations regarding the type of content. To obtain meaningful results, two videos were recorded in a non-professional environment with non-professional actors. This might have had an impact on the results and a replication of this study using professional content (e.g., a short movie, a documentary, a news program, or interviews) is encouraged.

Head-locked, coloured subtitles could be considered as good practice guidelines when subtitling in 360º videos. On the one hand, results about the position confirm previous studies where results were obtained through Igroup Presence Questionnaires (Agulló and Matamala, 2020). On the other hand, results about subtitle colour are not conclusive and further research in this regard is encouraged.

The study also highlighted the importance of the materials used for testing to ensure the ecological validity of the experiment. Special attention must be paid to the three languages involved in each experiment: the language of the film, the language of the subtitles, and the language of the participants. This problem has already been identified in other studies involving testing of different subtitle implementations, and various solutions have been adopted: the audio was muted (Kurzhals et al., 2017; Agulló and Matamala, 2020) or manipulated (Rothe et al., 2018). When testing with deaf and/or hard-of-hearing participants one more language needs to be added depending on whether they are oralists or if they use sign language (signers).

To conclude, this is a first-of-its-kind analysis of eye-tracked visual attention on subtitles processing in 360º media content with a focus on subtitle position and subtitle colour. Detailed examination of eye movements over subtitles combined with self-reports and performance measures has allowed an analysis of visual attention distribution and scene comprehension. We believe that the present work contributes to the novel frontiers on the implementation of accessibility services in immersive media studies.

Fundings

This study is part of the article-based PhD thesis of Marta Brescia-Zapata in the Department of Translation and Interpreting at Universitat Autònoma de Barcelona (UAB) within the PhD program in Translation and Intercultural Studies. This research has been partially funded by the H2020 projects TRACTION (under Grant Agreement 870610) and MEDIAVERSE (under Gran Agreement 957252); and it is based upon work from COST Action LEAD ME, CA19142, supported by COST (European Cooperation in Science and Technology). The Commission's support for this publication does not constitute an endorsement of the contents, which reflects the views of the authors only, and the Commission cannot be held responsible for any use which maybe made of the information contained herein. Marta Brescia-Zapata and Pilar Orero are members of TransMedia Catalonia, an SGR research group funded by "Secretaria d'Universitats I Recerca del Departament d'Empresa.Coneixement de la Generalitat de Catalunya" (2021 SGR 00077).

References

Agulló, B. & Matamala, A. (2019). Subtitling for the deaf and hard-of-hearing in immersive environments: Results from a focus group. *The Journal of Specialised Translation*, *32*, 217-235. Retrieved from http://www.jostrans.org/issue32/art_agullo.pdf

Agulló, B. & Matamala, A. (2020). Subtitles in virtual reality: Guidelines for the integration of subtitles in 360º content. *Íkala, Revista de Lenguaje y Cultura*, *25*(3), 643-661. https://doi.org/10.17533/udea.ikala.v25n03a03

Agulló, B., Matamala, A., & Orero, P. (2018). From disabilities to capabilities: Testing subtitles in immersive environments with end users. *HIKMA 17*: 195-220.

Agulló, B., Montagud, M., & Fraile, I. (2019). Making interaction with virtual reality accessible: Rendering and guiding methods for subtitles. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, *33*(4), 416-428. https://doi.org/10.1017/S0890060419000362

Agulló, B. & Orero, P. (2017). 3D Movie subtitling: Searching for the best viewing experience. *CoMe - Studi di Comunicazione e Mediazione linguistica e culturale, 2,* 91-101. Retrieved from https://ddd.uab.cat/record/183656

Ashcraft, M. H. & Radvansky, G. A. (2014). *Cognition*. Pearson Education, Boston, sixth edition.

Asociación Española de Normalización y Certificación (AENOR). (2003). *Norma UNE 153010: Subtitulado para personas sordas y personas con discapacidad auditiva. Subtitulado a través del teletexto.* https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma/?Tipo=N&c=Q21 N0029761

Ballester Casado, A. (1998). *Traducción y nacionalismo. La recepción del cine americano en España a través del doblaje, 1920-1948.* [Translation and Nationalism. The Reception of American Cinemain SPain through Dubbing, 1920-1948]. Comares.

BBC. (2022). *Subtitle guidelines*. Version 1.2.1. https://www.bbc.co.uk/accessibility/forproducts/guides/subtitles/. Accessed 25 April 2023.

Brescia-Zapata, M., Krejtz, K., Orero, P., Duchowski, A. T., & Hughes, C. J. (2022). VR 360◦ subtitles: Designing a test suite with eye-tracking technology. *Journal of Audiovisual Translation, 5*(2). (In press.)

Brown, A. (2017, October 26). *User testing for 360º content.* BBC. https://www.bbc.co.uk/rd/blog/2017-10-subtitles-360-video-virtual-reality-vr

Brown, A., Turner, J., Patterson, J., Schmitz, A., Armstrong, M., & Glancy, M. (2017). Subtitles in 360-degree video. *Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video.* The Netherlands, 3–8. https://doi.org/10.1145/3084289.3089915

Brown, A., Turner, J., Patterson, J., Schmitz, A., Armstrong, M., & Glancy, M. (2018). *Exploring subtitle behaviour for 360° video* (White Paper WHP 330) https://www.bbc.co.uk/rd/publications/whitepaper330

Chaume, F. (2012). *Audiovisual translation: Dubbing.* Manchester: St Jerome Publishing.

European Broadcasting Union (EBU). (2017). Virtual reality: How are public broadcasters using it? https://www.ebu.ch/publications/virtual-reality-how-are-public-broadcasters-using-it

Fidyka, A., & Matamala, A. (2018). Audio description in 360º videos: Results from Focus Groups in Barcelona and Kraków. *Translation Spaces, 7*(2), pp. 285–303. http://doi.org/10.1075/ts.18018.fid

Gejrot, E., Kjellstrand, S., & Laurin, S. (2021). New realities unlocking the potential of XR for persons with disabilities. Tech. rep., Funka Nu AB, Stockholm, Sweden.

Gil Ariza, M. C. (2004). A case study: Spain as a dubbing country. *Translation journal*, 29. http://translationjournal.net/journal/29movies.html

Gottlieb, H. (1998). Subtitling. In M. Baker (Ed.) *Routledge encyclopaedia of translation studies.* London and New York: Routledge, 244-8.

Greener, R. (2022, December 19). *Virtual reality statistics to know in 2023.* XR today. https://www.xrtoday.com/virtual-reality/virtual-reality-statistics-to-know-in-2023/#:~:text=In%202020%2C%20roughly%2057.4%20people,%2C%20industry%2C%20and%20education%20adoption.

Hart, S. G. (2006). Nasa-task load index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50*(9), 904-908. https://doi.org/10.1177/154193120605000909

Hermosa-Ramírez, I. (2022). Physiological instruments meet mixed methods in media accessibility. *Translation Spaces, 11*(1), 38-59. https://doi.org/10.1075/ts.21020.her

Hughes, C., Brescia-Zapata, M., Johnston, M., & Orero, P. (2020). *Immersive captioning: Developing a framework for evaluating user needs.* IEEE AIVR: 3rd International Conference on Artificial Intelligence & Virtual Reality, IEEE.

Hughes, C., Montagud, M., & tho Pesch, P. (2019). Disruptive approaches for subtitling in immersive environments. In *Proceedings of the 2019ACM International Conference on Interactive Experiences for TV and Online Video – TVX '19*. http://doi.org/10.1145/3317697.3325123

Ibourk, E. M., & Al-Adwan, A. (2019). Subtitling virtual reality into arabic: Eye tracking 360-degree video for exploring viewing experience. *Lebende Sprachen, 64*(2), 286-308.

International Organization for Standardisation. (2018). *Information technology — User interface component accessibility* (ISO/IEC 20071) https://www.iso.org/standard/70722.html

Krejtz, I., Szarkowska, A., & Łogińska, M. (2016). Reading function and content words in subtitled videos. *The Journal of Deaf Studies and Deaf Education, 21*(2), 222–232. https://doi.org/10.1093/deafed/env061

Kruger, J. (2016). Psycholinguistics and audiovisual translation. *Target. International Journal of Translation Studies, 28*(2), 276–287. https://doi.org/10.1075/target.28.2.08kru

Kurzhals, K. Cetinkaya, E., Hu, Y., Wang, W., & Weiskopf, D. (2017). Close to the Action: Eye-Tracking Evaluation of Speaker-Following Subtitles, *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*, 6559–6568. https://doi.org/10.1145/3025453.3025772

Liao, S., Yu, L., Reichle, E. D., & Kruger, J. L. (2021). Using eye movements to study the reading of subtitles in video. *Scientific Studies of Reading, 25*(5), 417–435.

Liberatore, M. J., & Wagner, W. P. (2021). Virtual, mixed, and augmented reality: A systematic review for immersive systems research. *Virtual Reality, 25*(3), 773–799. https://doi.org/10.1007/s10055-020-00492-0

Luyken, G. M. (1991). *Overcoming language barriers in television: Dubbing and subtitling for the European audience.* European Institute for the Media.

Matamala, A., & Orero, P. (2018). Standardising accessibility: Transferring knowledge to society. *Journal of Audiovisual Translation, 1*(1), 139–154. https://doi.org/10.47476/jat.v1i1.49

Matamala, A., Perego, E., & Bottiroli, S. (2017). Dubbing versus subtitling yet again? An empirical study on user comprehension and preferences in Spain. *Babel. Revue internationale de la tradution/International Journal of Translation, 63*(3), 423–41. https://doi.org/10.1075/babel.63.3.07mat

Montagud-Climent, M., Orero, P., & Matamala, A. (2020). Culture 4 all: Accessibility-enabled cultural experiences through immersive VR360 content. *Personal and Ubiquitous Computing, 24*(6), 887-905.

Myers, Maddy. (June 12, 2020). "The Last of Us Part 2 review: We're better than this". Polygon. Vox Media. Archived from the original on June 12, 2020. Retrieved July 6, 2022.

Oncins, E., Bernabé, R., Montagud, M., & Arnáiz Uzquiza, V. (2020). Accessible scenic arts and virtual reality: A pilot study with aged people about user preferences when reading subtitles in immersive environments, *MonTI. Monografías de Traducción e Iterpretación 12*(12), 214–41. https://doi.org/10.6035/MonTI.2020.12.07

Orero, P. (2022). Translation, accessibility and minorities. In E. Bielsa (Ed.), *The Routledge handbook of translation and media* (pp. 384-399). Taylor Francis.

Orero, P., Doherty, S., Kruger, J., Matamala, A., Pedersen, J., Perego, E., Romero-Fresco, P., Rovira-Esteva, S., Soler-Vilageliu, O., & Szarkowska, A. (2018). Conducting experimental research in audiovisual translation (AVT): A position paper. *JoSTrans, The Journal of Specialised Translation,* (30), 105–126. https://www.jostrans.org/issue30/art_orero_et_al.pdf

Orero, P., Sonali R., & Hughes, C. (2019). The challenges of audio description in new formats. In S. Braun (Ed.), *Innovation in audio description research*. Routledge.

R Core Team. (2020). R: A language and environment for statistical computing. R foundation for statistical computing. Vienna, Austria. http://www.R-project.org

Romero-Fresco, P. (2013). Accessible filmmaking: Joining the dots between audiovisual translation, accessibility and filmmaking. *The Journal of Specialised Translation, 20*, 201–223.

Rothe, S., Tran, K., & Hussmann, H. (2018). Positioning of subtitles in cinematic virtual reality. In G. Bruder, S. Yoshimoto, & S. Cobb (Eds.). *International Conference on Artificial Reality and Telexistence and Eurographics Sysmposium on Virtual Environments*, *ICAT-EGVE 2018.* The Eurographics Association.

Sipatchin, A., Wahl, S., & Rifai, K. (2020). Eye-tracking for low vision with virtual reality (VR): testing status quo usability of the HTC Vive Pro Eye. *bioRxiv.*

Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and Virtual Environments, 6*(6), 603–616. https://doi.org/10.1162/pres.1997.6.6.603

Szarkowska, A. (2009). The audiovisual landscape in Poland at the dawn of the 21st century. In: A. Goldstein, & B. Golubović (Eds.), *Foreign language movies – dubbing vs. subtitling* (pp. 185–201). Verlag Dr. Kovač.

Szarkowska, A., & Boczkowska, J. (2022). Colour coding subtitles in multigoal films – a reception study. *Perspectives: Studies in Translatology, 30*(3), 520–536. https://doi.org/10.1080/0907676X.2020.1853186

Szarkowska, A., & Gerber-Morón, O. (2019). Two or three lines: A mixed-methods study on subtitle processing and preferences. *Perspectives: Studies in Translatology, 27*(1), 144–164. https://doi.org/10.1080/0907676X.2018.1520267

Udo, J.-P., & Fels, D. I. (2010). Universal design on stage: Live audio description for theatrical performances. *Perspectives: Studies in Translatology, 18*(3), 189–203. http://doi.org/10.1080/0907676X.2010.485683

Valdeón, R. A. (2022). Latest trends in audiovisual translation. *Perspectives: Studies in Translatology*, 30(3), 369–381. https://doi.org/10.1080/0907676X.2022.2069226

Appendices

1.1 Tables with detailed statistics for the model on content general comprehension

Table A1. Model parameters for content general comprehension

	estimate (β)	***SE***	***t***	***df***	***p***
Intercept	3.25	0.33	9.70	11.10	0.00
Country Spain	-0.11	0.30	-0.37	72.02	0.71
Country UK	-0.22	0.30	-0.76	71.94	0.45
Position head-locked	0.50	0.26	1.93	88.23	0.06
Colour colourful	0.37	0.12	3.05	497.99	0.00
Position head-locked X Colour colourful	-0.23	0.16	-1.38	498.68	0.17

Table A2. Estimated means from the model for content general comprehension for interaction effect of subtitle position and colour

position	**colour**	**est. mean**	***SE***
fixed	monochrome	3.13	0.34
head-locked	monochrome	3.63	0.33
fixed	colourful	3.50	0.34
head-locked	colourful	3.77	0.33

Table A3. Model parameters for self-perceived effort

	estimate (β)	***SE***	***t***	***df***	***p***
Intercept	4.57	0.90	5.07	4.52	0.01

Country Spain	-0.31	0.62	-0.50	68.96	0.62
Country UK	-1.12	0.61	-1.83	68.45	0.07
Position head-locked	-1.15	0.52	-2.21	79.09	0.03
Colour colourful	0.70	0.21	3.37	471.01	0.00
Position head-locked X Colour colourful	-0.60	0.29	-2.11	469.41	0.04

Table A4. Estimated means from the model for self-perceived effort for interaction effect of subtitle position and colour

position	**colour**	**est. mean**	***SE***
fixed	monochrome	4.09	1.08
head-locked	monochrome	2.94	1.07
fixed	colourful	4.79	1.08
head-locked	colourful	3.04	1.07

Table A5. Model parameters for proportional fixation number

	estimate (β)	***SE***	***t***	***df***	***p***
Intercept	64.93	14.23	4.56	1.17	0.11
Country Spain	-47.29	5.15	-9.18	66.55	0.00
Country UK	-31.90	5.11	-6.24	65.29	0.00
Position head-locked	-12.84	5.36	-2.39	91.72	0.02
Colour monochrome	-4.20	3.29	-1.28	69.19	0.21
Position head-locked	4.18	4.49	0.93	69.46	0.35

X Colour monochrome					
country Spain: Position head-locked	31.95	7.00	4.56	65.80	0.00
country UK: Position head-locked	16.34	6.90	2.37	65.11	0.02

Table A6. Estimated means from the model for proportional fixation number for interaction effect of subtitle position and country

position	**country**	**est. mean**	***SE***
fixed	Poland	62.8	14.1
head-locked	Poland	52.1	14.0
fixed	Spain	15.5	14.1
head-locked	Spain	36.7	14.1
fixed	UK	30.9	14.0
head-locked	UK	36.5	14.0

Table A7. Model parameters for proportional total fixation time

	estimate (β)	***SE***	***t***	***df***	***p***
Intercept	10756.03	4916.41	2.19	1.13	0.25
Country Spain	-5401.21	1556.41	-3.47	66.92	0.00
Country UK	756.23	1547.46	0.49	65.83	0.63
Position head-locked	-317.28	1575.22	-0.20	83.65	0.84
Colour monochrome	316.28	806.71	0.39	69.40	0.70
Position head-locked :	-257.78	1100.44	-0.23	69.58	0.82

Colour monochrome					
country Spain : Position head-locked	5010.05	2118.56	2.36	66.45	0.02
country UK : Position head-locked	2041.71	2088.48	0.98	65.86	0.33

Table A8. Estimated means from the model for proportional total fixation time for interaction effect of subtitle position and country

position	**country**	**est. mean**	***SE***
fixed	Poland	10914	4900
head-locked	Poland	10468	4856
fixed	Spain	5513	4880
head-locked	Spain	10077	4888
fixed	UK	11670	4878
head-locked	UK	13266	4878

Table A9. Model parameters for accuracy of character' recognition

	estimate (β)	***SE***	***t***	***df***	***p***
Intercept	1.53	1.13	1.35	1.05	0.40
Country Spain	0.14	0.21	0.68	68.19	0.50
Country UK	0.12	0.21	0.60	68.10	0.55
Position head-locked	0.12	0.17	0.70	68.11	0.48
Colour colourful	0.20	0.06	3.51	500.42	0.00

ANNEX 3

Ethical Committee Documentation

3.1. Written informed consent forms and information sheets for participants in the experiments (Spanish)

3.2. Written informed consent forms and information sheets for participants in the experiments (English)

3.1. Written informed consent forms and information sheets for participants in the experiments (Spanish)

TRACTION Opera co-creation for a social transformation H2020-SC6-TRANSFORMATIONS- 870610	Traction Opera co-creation for a social transformation
Hoja informativa	

Proyecto: **TRACTION «Opera co-creation for a social transformation».**

Investigadora principal: Mikel Zorrilla
Asesora ética: Pilar Orero

Nos gustaría invitarle a participar en una prueba para un estudio de investigación. El objetivo del estudio es el desarrollo de una herramienta tecnológica que ayude a generar subtítulos en contenidos 360.

El objetivo de la prueba es estudiar el uso de subtítulos en escenas inmersivas. La actividad constará de tres partes. En la primera, se realizará una explicación de qué son los contenidos inmersivos y el subtitulado creativo. La segunda parte son unas pruebas individuales donde se le pedirá que visualice unos contenidos inmersivos con un casco de realidad virtual. Este dispositivo recogerá datos mediante *eye-tracking* para analizar su comportamiento visual ante los contenidos y los subtítulos. La tercera y última parte es la participación en un grupo de discusión sobre la experiencia, formatos y posibilidades de los subtítulos en escenas inmersivas.

Sus respuestas y las de otros participantes se agruparán y se utilizarán para orientar las siguientes fases del proyecto. Los datos que recojamos durante las pruebas contribuirán a la consolidación del sistema y al despliegue de la versión completa del sistema desarrollado en TRACTION hacia finales de 2022.

Puede parar el cuestionario sin justificación alguna.

Por favor, lea ahora el formulario de consentimiento.

TRACTION Opera co-creation for a social transformation H2020-SC6-TRANSFORMATIONS- 870610	Traction Opera co-creation for a social transformation

Consentimiento informado

El objetivo del estudio es el desarrollo de una herramienta tecnológica que permita generar y consumir subtítulos en contenidos 360. Los resultados servirán para definir nuevos requisitos desde una perspectiva centrada en el usuario.

Se trata de una actividad individual durante la cual se recogerán datos mediante eye-tracking, una tecnología de seguimiento ocular que permite saber los comportamientos visuales del usuario, integrada en este caso en un casco de realidad virtual. Para conseguir estos datos, se visualizarán dos videos en 360 utilizando el dispositivo de realidad virtual y siguiendo siempre las indicaciones del investigador. Después de cada visualización, se rellenará un cuestionario breve para obtener datos sobre la compresión del contenido y las preferencias de cada usuario.

La información que proporcione durante la actividad será anónima y se usará exclusivamente para el proyecto. Toda la información que proporcione será tratada con la más estricta confidencialidad y solo se usará en este estudio. No se le pedirán más datos personales que le identifiquen después de firmar el formulario de consentimiento.

Los investigadores principales de la Universitat Autònoma de Barcelona (UAB), encargados de la participación de personas en el proyecto, guardarán este formulario en un lugar seguro y lo destruirán cinco años después de la finalización del proyecto. No se compartirá con nadie fuera de la UAB. Los datos analizados y la base de datos estarán disponible para otros investigadores interesados en repositorios de datos abiertos. La información disponible será anónima.

Su participación en el taller es absolutamente voluntaria y no hay compensación económica. Puede decidir no participar. Puede interrumpir su participación en el estudio en cualquier momento sin justificación previa. Esto no tendrá repercusiones ni consecuencias negativas de ningún tipo.
TRACTION es un proyecto europeo liderado por VICOMTECH (España). La asesora ética responsable de los procedimientos éticos es la Dra. Pilar Orero. Puede ponerse en contacto con ella para pedir más información sobre el proyecto y sus resultados. Además, puede ejercer sus derechos reconocidos por la Regulación Europea sobre Protección de datos. Para ello, póngase en contacto con Dra. Orero, con su petición (rellenando y enviando el formulario 'ARCO' en la siguiente página de este documento, junto con una fotocopia de su DNI, NIE o pasaporte:

This Project has received funding from the EU programme H2020

Dra. Pilar Orero (pilar.orero@uab.cat), Dept. de Traducció i Interpretació i Estudis de l'Àsia Oriental. Campus UAB, Plaça del Coneixement, MRA/126. 08193 Bellaterra (Cerdanyola del Vallès)

Autorizo:

El análisis cualitativo y cuantitativo de mis respuestas
El uso de mis respuestas para la investigación y diseminación científica

Si está dispuesto a participar, confirme las siguientes declaraciones firmando al final de este documento.

• He leído y comprendido la información dada para esta investigación,

• He tenido la oportunidad de hacer preguntas sobre la investigación,

• Doy mi consentimiento para participar en las sesiones de investigación.

Nombre y apellidos del entrevistado	Fecha	Firma
Nombre y apellidos del investigador	Fecha	Firma

FORMULARIO DE ACCESO, RECTIFICACIÓN, CANCELACIÓN, OPOSICIÓN Y PORTABILIDAD DE DATOS PERSONALES (ARCO)

Está participando en un proyecto de investigación denominado TRACTION «Co-creación de Opera para la transformación social». De conformidad con el Reglamento 2016/679 del Parlamento Europeo y el Consejo de la UE, de 27 de abril de 2016 sobre protección de datos (GDPR), y la Ley orgánica 15/1999, de 13 de diciembre, sobre la protección de datos personales y las regulaciones que administran a continuación, le informamos que los datos personales recopilados en el marco de este estudio formarán parte de un archivo de datos personales, del cual TransMedia Cataluña (Universidad Autónoma de Barcelona) es responsable, con el único propósito de llevar a cabo la investigación mencionada anteriormente.

Si en el futuro desea ejercer sus derechos de acceso, rectificación, cancelación, oposición y portabilidad de sus datos personales ante la persona responsable del archivo, complete este formulario y envíelo a cualquiera de los investigadores responsables del estudio junto con una fotocopia de su documento de identidad o cualquier otro medio de identificación.

DATOS DE LA PERSONA RESPONSABLE DEL ARCHIVO.
Dra. Pilar Orero (pilar.orero@uab.cat), Dpto. De Traducción y Interpretación y Estudios de Asia Oriental. Campus de la UAB, Plaça del Coneixement, Edificio MRA, Oficina 126. 08193 Bellaterra (Cerdanyola del Vallès).

SUS DATOS / O SU REPRESENTANTE LEGAL
Yo, __________________, mayor de edad y residente en ____________________________ Localidad _________ Provincia ______________ Código Postal __________ con DNI / NIE / OTRO _____________________, del cual adjunto una copia, mediante este escrito deseo ejercer mis derechos para proteger los datos personales con referencia a los datos recolectados en el marco del estudio « TRACTION Co-creación de Opera para la transformación social », de conformidad con la Ley Orgánica 15/1999, del 13 de diciembre, sobre la protección de datos personales y las regulaciones que la desarrollan y por consiguiente,

SOLICITO (seleccione la opción que corresponda)

This Project has received funding from the EU programme H2020

- ❑ Ejercer mi derecho de acceso a mis datos personales, y que en un plazo máximo de un mes a partir de la recopilación de esta solicitud, se me proporcione la información que se almacena sobre mí en relación con el estudio mencionado anteriormente.
- ❑ Que mis datos personales se rectifiquen, dentro de los días hábiles posteriores a la recepción de esta solicitud, y que se me notifique por escrito del resultado de la rectificación realizada.
- ❑ Que se continúe con la cancelación de mis datos personales, dentro de los días hábiles posteriores a la recepción de esta solicitud, y que se me notifique por escrito del resultado de la cancelación practicada.
- ❑ Me opongo a que mis datos personales se almacenen en dicho registro dentro de los días hábiles posteriores a la recepción de esta solicitud, y que se me notifique por escrito el resultado de la rectificación practicada.
- ❑ Solicito que mis datos personales no se compartan con terceros, siempre y cuando no estén debidamente anonimizados de forma completa e irreversible.

Firma

Fecha

This Project has received funding from the EU programme H2020

3.2. Written informed consent forms and information sheets for participants in the experiments (English)

TRACTION Opera co-creation for a social transformation H2020-SC6- TRANSFORMATIONS- 870610	Traction Opera co-creation for a social transformation
CONSENT FORM	

The main goal of this study is the development of a technological tool that allows generating and consuming subtitles within 360 contents. The results will help towards drafting new requirements from a user-centred approach.

It is an individual activity, and the data will be collected using eye-tracking technology, a process of measuring eye movements to determine where a person is looking, what they are looking at, and how long their gaze is in a particular spot. To collect this data, each participant —always guided by the researcher's instructions— will watch two 360º videos wearing a virtual reality head-mounted display. After each visualisation, participants will fill out a brief questionnaire to obtain data about comprehension, task-load, and preferences.

The information you provide will be used in the project, but it will remain anonymous. All information that you give will be treated in the strictest confidence and it will only be used for the purposes of this study. No personal data will be collected that permits your identification. This consent will be kept in a safe place by the principal investigators and will be destroyed five years after the investigation has finished. Once the study has been completed and the data analysed, the entire database will be available to other interested researchers after five years after the project has been completed.

Your participation in this study is absolutely voluntary, and there is not economic compensation. There is no penalty for not participating and there are no risks of any kind in your participation. You can discontinue your involvement in the study at any time without prior justification. This shall have no repercussions or negative consequences of any sort.

TRACTION is a European project led by VICOMTECH (Spain). The ethical adviser responsible of ethical procedures is Pilar Orero. You can contact Pilar Orero at Pilar.Orero@uab.cat and ask for more information about the project and the project results. The researcher administering the test is Marta Brescia Zapata. As well you can exercise your rights under the European General Data Protection Regulation by making a request to the UAB Data Protection Office (https://www.uab.cat/web/coneix-la-uab/itineraris/proteccio-de-dades/drets-de-les-persones-interessades-1345764799916.html).
You may also file a claim before the Catalan Data Protection Authority (https://apdcat.gencat.cat/ca/contacte).

I authorise:

TRACTION Opera co-creation for a social transformation
H2020-SC6-TRANSFORMATIONS- 870610
This Project has received funding from the EU programme H2020

- The qualitative and quantitative analysis of my responses
- The use of my answers for scientific research and dissemination

If you are willing to participate, please confirm the following statements by signing at the end of this document.

- I have read and understood the information given for this research or have had the information read to me,
- I have had the opportunity to ask questions about the research.
- I consent to take part in the research sessions.

____

Name of the participant Date Signature

____

Name of the researcher Date Signature

This Project has received funding from the EU programme H2020

ANNEX 4

Questionnaires

4.1. Demographic questionnaire – Article 3

4.2. Postquestionnaire – Article 3

4.3. Demographic questionnaire – Article 4 and Annex 1

4.4. Post stimuli questionnaire for the video called “TRACTION” – Article 4 and Annex 1

4.5. Post stimuli questionnaire for the video called “Vacations” – Article 4 and Annex 1

4.6. Postquestionnaire on preferences – Article 4 and Annex 1

4.1. Demographic questionnaire – Article 3

360 FOR ALL

Nuevas formas de visualizar subtítulos en VR360.

* Indica que la pregunta es obligatoria

1. Número de identificación *

____

2. Género *

Marca solo un óvalo.

- ⬭ Masculino
- ⬭ Femenino
- ⬭ Otro: ______________________________

3. Año de nacimiento *

____

4. Visión *

Marca solo un óvalo.

- ⬭ Sin corregir
- ⬭ Corregida con gafas
- ⬭ Corregida con lentillas
- ⬭ Otro: ______________________________

5. Lateralidad *

Marca solo un óvalo.

- Diestro
- Zurdo
- Otro: ______________________

6. Nivel de estudios acabados *

Marca solo un óvalo.

- Educación primaria
- Educación secundaria
- Licenciatura/Grado
- Máster
- Doctorado o superior

7. Profesión *

____

8. Lengua(s) materna(s) *

____

9. ¿Qué idiomas dominas y a qué nivel? *

Marca solo un óvalo por fila.

	No lo domino	Principiante	Intermedio	Avanzado	Fluido
Catalán	◯	◯	◯	◯	◯
Castellano	◯	◯	◯	◯	◯
Inglés	◯	◯	◯	◯	◯
Francés	◯	◯	◯	◯	◯
Alemán	◯	◯	◯	◯	◯
Otro	◯	◯	◯	◯	◯

10. ¿Cuál de estos dispositivos usas a diario? Puedes elegir más de una opción *

Selecciona todos los que correspondan.

- [ ] Televisión
- [ ] Ordenador de sobremesa
- [ ] Ordenador portátil
- [ ] Teléfono móvil
- [ ] Tablet
- [ ] Gafas de realidad virtual
- [ ] Consola
- [ ] Otro: ______________________________

11. ¿Con qué frecuencia ves contenido de realidad virtual (por ejemplo, vídeos 360º)? *

Marca solo un óvalo por fila.

	Nunca	Ocasionalmente	Al menos una vez al mes	Al menos una vez a la semana	Todos los días
En un teléfono móvil					
En una tablet					
En un ordenador					
En un teléfono móvil con gafas de realidad virtual					
Con gafas de realidad virtual conectadas a un ordenador					

12. Si nunca has visto contenido en realidad virtual como vídeos 360º o solo lo has hecho en alguna ocasión puntual, indica el motivo. Puedes elegir más de una opción. *

Selecciona todos los que correspondan.

- [ ] Porque no me interesa.
- [ ] Porque no es accesible.
- [ ] Porque no he tenido la oportunidad de usarlo.
- [ ] Otro: ______________________

13. Indica tu nivel de acuerdo con la siguiente afirmación: «Me interesa el contenido en realidad virtual (por ejemplo, vídeos 360º)». *

Marca solo un óvalo.

- Muy de acuerdo
- De acuerdo
- Ni de acuerdo ni en desacuerdo
- En desacuerdo
- Muy en desacuerdo

14. ¿Tienes algún dispositivo para acceder a contenido de realidad virtual? *

Marca solo un óvalo.

- Sí
- No
- No lo sé o no quiero responder

15. Si la respuesta a la pregunta anterior fue «sí», por favor especifica qué dispositivos.

16. ¿Te gusta ver los siguientes tipos de contenido en televisión o en línea? *

Marca solo un óvalo por fila.

	Me gusta mucho	Me gusta	Ni me gusta ni me disgusta	No me gusta	No me gusta nada
Noticias					
Ficción (series, películas)					
Programas de entrevistas					
Documentales					
Deportes					
Dibujos animados					

17. Cuando hay subtítulos disponibles, ¿los activas para los siguientes tipos de contenido? *

Marca solo un óvalo por fila.

	Siempre	A veces	Alguna vez	Nunca
Noticias				
Ficción (series, películas)				
Programas de entrevistas				
Documentales				
Deportes				
Dibujos animados				

18. Si los subtítulos están disponibles y no los activas, indica los motivos. *

Marca solo un óvalo.

- Porque la interfaz no es accesible.
- Porque no quiero subtítulos en todos los contenidos, solo en algunos contenidos.
- Otro: ____________

19. ¿Cuántas horas al día ves contenido subtitulado? *

Marca solo un óvalo.

- Ninguna
- Menos de una hora
- De una a dos horas
- De dos a tres horas
- De tres a cuatro horas
- Cuatro horas o más

20. ¿Para qué utilizas los subtítulos? Puedes elegir más de una opción. *

Selecciona todos los que correspondan.

- [ ] Me ayudan a entender el contenido.
- [ ] Son mi única manera de acceder al diálogo.
- [ ] Los uso para aprender idiomas.
- [ ] Otro: ____________

4.2. Postquestionnaire – Article 3

After-movie questionnaire

* Indica que la pregunta es obligatoria

1. Número de identificación *

2. ¿Qué versión del corto subtitulado acabas de ver? *

   Marca solo un óvalo.

   - Version 1
   - Version 2
   - Version 3

3. ¿Qué le molesta al personaje masculino en el corto? *

4. ¿Qué le dice el personaje femenino a la policía por teléfono? *

5. ¿Cuál es la excusa del personaje masculino para subir a la fiesta? *

6. ¿Quién utiliza por primera vez el término «puretas» en el corto? *

Marca solo un óvalo.

- Agus
- María (la chica morena)
- Anna (la chica rubia)
- No lo sé

7. ¿Quién de los tres es el «amigo borracho»? *

Marca solo un óvalo.

- Agus
- María (la chica morena)
- Anna (la chica rubia)
- No lo sé

8. ¿Cómo es el escenario en el que tiene lugar el corto? Anota los detalles que recuerdes. *

9. ¿Has podido leer todos los subtítulos? *

Marca solo un óvalo.

- Sí
- No
- No lo sé

10. Estima el porcentaje de subtítulos que has podido leer. *

11. ¿Cuánto esfuerzo mental has necesitado para leer los subtítulos? *

Marca solo un óvalo.

	Muy poco
1	⬭
2	⬭
3	⬭
4	⬭
5	⬭
6	⬭
7	⬭
8	⬭
9	⬭
10	⬭
	Mucho

12. ¿Cómo de rápido ha sido el ritmo de lectura de los subtítulos? *

Marca solo un óvalo.

Muy bajo

1 ⬭

2 ⬭

3 ⬭

4 ⬭

5 ⬭

6 ⬭

7 ⬭

8 ⬭

9 ⬭

10 ⬭

Muy alto

13. Respecto a la tarea de leer los subtítulos ¿cuál es tu grado de satisfacción? *

Marca solo un óvalo.

Muy bajo

1 ⬭

2 ⬭

3 ⬭

4 ⬭

5 ⬭

6 ⬭

7 ⬭

8 ⬭

9 ⬭

10 ⬭

Muy alto

14. ¿Cómo de difícil era leer todos los subtítulos? *

Marca solo un óvalo.

Muy poco

1 ⬭

2 ⬭

3 ⬭

4 ⬭

5 ⬭

6 ⬭

7 ⬭

8 ⬭

9 ⬭

10 ⬭

Mucho

15. Durante la visualización del corto, ¿en qué medida te has sentido insegurx, desalentadx, irratdx o tensx? *

Marca solo un óvalo.

Muy bajo

1 ◯

2 ◯

3 ◯

4 ◯

5 ◯

6 ◯

7 ◯

8 ◯

9 ◯

10 ◯

Muy alto

4.3. Demographic questionnaire – Article 4 and Annex 1

Demographics

Q1 Participants ID

____

Q2 Gender

○ Male (1)

○ Female (2)

○ Non-binary / third gender (3)

○ Prefer not to say (4)

Q3 Year of birth

____

Q4 Vision

○ Uncorrected (1)

○ Corrected with glasses (2)

○ Corrected with contact lenses (3)

○ other (4) ____

Q5 5. Handedness

- ○ Right (1)
- ○ Left (2)

Q6 Highest level of education obtained

- ○ Primary school (1)
- ○ High school (2)
- ○ University Bachelor (3)
- ○ University Master (4)
- ○ University PhD and higher (5)

Q7 Occupation

____

Q8 Mother language(s)

____

Q9 Which languages do you speak and on what level?

	Not at all (1)	Beginning (2)	Intermediate (3)	Advanced (4)	Fluent (5)
English (1)	○	○	○	○	○
Spanish (2)	○	○	○	○	○
Catalan (3)	○	○	○	○	○
Portuguese (4)	○	○	○	○	○
Korean (5)	○	○	○	○	○
Arab (6)	○	○	○	○	○

Q10 Which of these devices do you use on a daily basis?
Multiple answers are possible.

- ▢ Television (1)
- ▢ Desktop computer (2)
- ▢ Laptop (3)
- ▢ Mobile phone (4)
- ▢ Tablet (5)
- ▢ Virtual reality glasses (6)
- ▢ Game console (7)
- ▢ Others (8) ____

Q11 How often do you watch virtual reality (VR) content for instance, 360º videos?

	Never (1)	Occasionally (2)	At least once a month (3)	At least once a week (4)	Every day (5)
On smartphone (1)	○	○	○	○	○
On a tablet (2)	○	○	○	○	○
On a computer or laptop (3)	○	○	○	○	○
On smartphone plugged to VR glasses (4)	○	○	○	○	○
With VR glasses connected to a computer (5)	○	○	○	○	○

Mostrar esta pregunta:

If Q11 [Never] (Cuenta) = 7

Q12 If you have never used VR content such as 360º videos or only occasionally, please indicate why. Multiple answers are possible.

○ Because I am not interested (1)

○ Because it is not accessible (2)

○ Because I haven't had the chance to use it (3)

○ Other (4) ____

Q13 Please state your level of agreement with the following statement: “I am interested in virtual reality content (such as 360º videos).”

- ◯ I strongly agree (1)
- ◯ I agree (2)
- ◯ Neither agree nor disagree (3)
- ◯ Disagree (4)
- ◯ Strongly disagree (5)

Q14 Do you own any device to access virtual reality content?

- ◯ Yes (1)
- ◯ No (2)
- ◯ I don't know (3)
- ◯ I don’t want to reply (4)

Mostrar esta pregunta:

If Q14 = Yes

Q15 Please specify which device(s)

____

Q16 Do you like to watch the following types of content on TV or online?

	I like it very much (1)	I like it (2)	I neither like or dislike it (3)	I don't like it (4)	I don't like it at all (5)
News (1)	○	○	○	○	○
Fiction (series, movies) (2)	○	○	○	○	○
Talk shows (3)	○	○	○	○	○
Documentaries (4)	○	○	○	○	○
Sports (5)	○	○	○	○	○
Cartoon (6)	○	○	○	○	○

Q17 When subtitles are available, do you turn them on for the following types of content?

	Always (1)	Sometimes (2)	Occasionally (3)	Never (4)
News (1)	○	○	○	○
Fiction (series, movies) (2)	○	○	○	○
Talk shows (3)	○	○	○	○
Documentaries (4)	○	○	○	○
Sports (5)	○	○	○	○
Cartoon (6)	○	○	○	○

Q18 If subtitles are available and you don't turn them on, state why

- ○ Because the interface is not accessible (1)
- ○ Because I don't want subtitles in all the contents, only in some contents (2)
- ○ Other (3) ____

Q19 How many hours a day do you watch subtitled content?

- ○ None (1)
- ○ Less than one hour (2)
- ○ One to two hours (3)
- ○ Two to three hours (4)
- ○ Three to four hours (5)
- ○ Fours hours of more (6)

Mostrar esta pregunta:
If Q17 != Never

Q20 What do you use subtitles for?

- ○ They help me understand the content (1)
- ○ They are my only way to access the dialogue (2)
- ○ I use them to learn languages (3)
- ○ Other (4) ____

4.4. Post stimuli questionnaire for the video called "TRACTION" – Article 4 and Annex 1

Questionnaire “TRACTION”

Q00 Participant ID

____

Q1 Have you been able to read all the subtitles?

- ○ Definitely not (1)
- ○ Probably not (2)
- ○ Might or might not (3)
- ○ Probably yes (4)
- ○ Definitely yes (5)

Q2 Please estimate the percentage of subtitles you were able to read.

	0 10 20 30 40 50 60 70 80 90 100
Move the slider to set a number between 0 and 100 percent ()	

Q3 Answer the questions below indicating your effort with sliders.
0 means very low and 10 means very high

	0 10
How mentally demanding was it to follow the subtitles? ()	
How hurried or rushed was the pace of reading subtitles? ()	
How successful were you in reading all the subtitles? ()	
How hard was it to read all the subtitles? ()	
How discouraged, irritated, stressed, and annoyed were you with the subtitles in the movie? ()	

Q4 What is the name of the project this team is working on?

- ○ Traction (1)
- ○ e-balance (2)
- ○ iMac (3)
- ○ I do not know (4)

Q5 In which city in South Korea was Sueyoon Lee born?

- ○ Pohang (1)
- ○ Ulsan (2)
- ○ Busan (3)
- ○ I do not know (4)

Q6 How old is Pablo?

- ◯ 45 (1)
- ◯ 55 (2)
- ◯ 65 (3)
- ◯ I do not know (4)

Q7 What is Anderson's field of work?

- ◯ Computer graphics (1)
- ◯ Graphic designer (2)
- ◯ Computer networks (3)
- ◯ I do not know (4)

Q8 Where does James work?

- ◯ Royal Opera House (1)
- ◯ Irish National Opera (2)
- ◯ English National Opera (3)
- ◯ I do not know (4)

Q9 What city does Nacho work in?

◯ Amsterdam (1)

◯ Rotterdam (2)

◯ Antwerp (3)

◯ I do not remember (4)

Q11 Rate your experience in VR
1 means disagree and 10 means agree

	1	2	3	4	5	6	6	7	8	9	10
I had a sense of being there ()											
I felt like I was just watching a movie ()											

4.5. Post stimuli questionnaire for the video called "Vacations" – Article 4 and Annex 1

Questionnaire “Vacations”

Q00 Participant ID

____

Q1 Have you been able to read all the subtitles?

- ◯ Definitely not (1)
- ◯ Probably not (2)
- ◯ Might or might not (3)
- ◯ Probably yes (4)
- ◯ Definitely yes (5)

Q2 Please estimate the percentage of subtitles you were able to read.

	0 10 20 30 40 50 60 70 80 90 100
Move the slider to set a number between 0 and 100 percent ()	

Q3 Answer the questions below indicating your effort with sliders.
0 means very low and 10 means very high

	0 10
How mentally demanding was it to follow the subtitles? ()	
How hurried or rushed was the pace of reading subtitles? ()	
How successful were you in reading all the subtitles? ()	
How hard was it to read all the subtitles? ()	
How discouraged, irritated, stressed, and annoyed were you with the subtitles in the movie? ()	

Q4 In what month of the year does the story take place?

- ○ September (1)
- ○ November (2)
- ○ December (3)
- ○ I do not know (4)

Q5 What have the girls done during the summer?

- ○ Go to the beach (1)
- ○ Visiting friends (2)
- ○ Visiting parents (3)
- ○ I do not know (4)

Q6 What city did Bassima want to travel to in the summer?

- ◯ Paris (1)
- ◯ Barcelona (2)
- ◯ Madrid (3)
- ◯ I do not know (4)

Q7 What does the mother offer the girls when she comes home?

- ◯ Alcoholic drinks (1)
- ◯ Soft drinks (2)
- ◯ Tea (3)
- ◯ I do not know (4)

Q8 Who went to get the drinks?

- ◯ Maryam (1)
- ◯ Bassima (2)
- ◯ Mother (3)
- ◯ I do not know (4)

Q9 Why is Bassima upset?

- ◯ Because she hasn't enjoyed the Summer holidays. (1)
- ◯ Because she hasn't enjoyed the Winter holidays. (2)
- ◯ Because she does not like the drink. (3)
- ◯ I do not remember (4)

Q15 Rate your experience in VR
1 means disagree and 10 means agree

	1 — 10
I had a sense of being there ()	
I felt like I was just watching a movie ()	

4.6. Postquestionnaire on preferences – Article 4 and Annex 1

Preferences

Q1 Participant ID

____

Q2 What color code do you prefer when reading 360º subtitles?

- ◯ Black & white for all characters (1)
- ◯ Different colors for each character (2)
- ◯ I do not have the preference (3)

Mostrar esta pregunta:

If What color code do you prefer when reading 360º subtitles? != I do not have the preference

Q3 Explain in your own words why you prefer the option selected in the previous question

____

Q4 Do you think that the subtitles obstruct important parts of the image?

- ◯ Definitely not (1)
- ◯ Probably not (2)
- ◯ Might or might not (3)
- ◯ Probably yes (4)
- ◯ Definitely yes (5)

ANNEX 5

Digital and multimedia resources

5.1. Stimuli used in the reception studies

Available in the following web server (login needed): https://www.chxr.org/360/index.php